HAL
open science

L'AIS : une donnée pour l'analyse des activités en mer

Damien Le Guyader, David Brosset

To cite this version:

Damien Le Guyader, David Brosset. L'AIS : une donnée pour l'analyse des activités en mer. Conférence Internationale de Géomatique et d'Analyse spatiale (Spatial Analysis and GEOmatics) - SAGEO 2013, Sep 2013, Brest, France. pp.53. hal-00935417

HAL Id: hal-00935417

https://hal.univ-brest.fr/hal-00935417v1

Submitted on 23 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sageo 13 Brest

Conférence internationale de Géomatique et d'analyse spatiale

ACTES de la Conférence

IUEM • Brest

23 - 26 Septembre 2013

Comité scientifique / Scientific committee

Responsables scientifiques :	
Rémy Thibaud (IRENav, Brest)	Françoise Gourmelon (CNRS, LETG-Brest Géomer)

Comité scientifique :	
Anne Ruas (IFSTTAR, Paris)	Françoise Gourmelon (CNRS, LETG-Brest Géomer)
Jérôme Gensel (Univ. Grenoble)	Christophe Claramunt (IRENav, Brest)
Thierry Joliveau (EVS, St Etienne)	Rémy Thibaud (IRENav, Brest)

Comité d'organisation / Steering committee

IRENav :	*LETG-Brest Géomer :*
David Brosset	Françoise Gourmelon
Christophe Claramunt	Iwan Le Berre
Cyril Ray	Adeline Maulpoix
Eric Saux	Mathias Rouan
Rémy Thibaud	Cyril Tissot

Comité de lecture / Reviewers

Banos Arnaud, CNRS, Paris

Batton-Hubert Mireille, Ecole Nationale Sup. des Mines, St-Etienne

Billen Roland, Univ. Liège (Belgique)

Boulier Joël, Univ. Paris 1

Brosset David, Ecole Navale, Lanvéoc-Poulmic

Bucher Benedicte, IGN, St-Mandé

Chardonnel Sonia, CNRS, Grenoble

Claramunt Christophe, Ecole Navale, Lanvéoc Poulmic

Collet Claude, University of Fribourg (Suisse)

Davoine Paule-Annick, Univ. Grenoble

De Ruffray Sophie, Univ. Rouen

De Sède-Marceau Marie-Hélène, Univ. Besançon

Debaine Françoise, Univ. Nantes

Del Mondo Géraldine, INSA, Rouen

Desjardin Eric, Univ. Reims

Devillers Rodolphe, Memorial University, Terre-Neuve (Canada)

Devogele Thomas, Univ. Tours

Donnay Jean-Paul, Univ. Liège (Belgique)

Douvinet Johnny, Univ. d'Avignon

Feyt Grégoire, Univ. Grenoble

Fournier Sébastien, Univ. Marseille

Gaio Mauro, Univ. Pau

Gançarski Pierre, Univ. Strasbourg

Gensel Jérôme, Univ. Grenoble

Gourmelon Françoise, CNRS, Brest

Guilbert Eric, Hong Kong Polytechnic University (Chine)

Guth Peter, US Naval Academy, Annapolis (Etats-Unis)

Houet Thomas, CNRS, Toulouse

Joliveau Thierry, Univ. St-Etienne

Josselin Didier, CNRS, Avignon

Lardon Sylvie, AGROPARISTECH, Aubière

Le Ber Florence, ENGEES, Strasbourg

Le Berre Iwan, Univ. Brest

Leibovici Didier, Univ. of Nottingham (Angleterre)

Libourel Therese, Univ. Montpellier

Mainguenaud Michel, INSA, Rouen

Mathian Hélène, CNRS, Paris

Moreau Guillaume, Ecole Centrale, Nantes

Mustiere Sébastien, IGN, St-Mandé

Noucher Matthieu, CNRS, Bordeaux

Passouant Michel, CIRAD, Montpellier

Pierkot Christelle, Exia, Pau

Pinet François, IRSTEA, Clermont-Ferrand

Puissant Anne, Univ. Strasbourg

Quénol Hervé, CNRS, Rennes

Ray Cyril, Ecole Navale, Lanvéoc Poulmic

Robin Marc, Univ. Nantes

Rouan Mathias, CNRS, Brest

Ruas Anne, IFSTTAR, Marne-la-Vallée

Sallaberry Christian, Univ. Pau

Sanders Lena, CNRS, Paris

Saux Eric, Ecole Navale, Lanvéoc-Poulmic

Servigne Sylvie, INSA, Lyon

Sheeren David, ENSAT, Toulouse

Sidonie Christophe, IGN, St-Mandé

Sèdes Florence, Univ. Toulouse

Tannier Cécile, CNRS, Besançon

Thibaud Rémy, Ecole Navale, Lanvéoc-Poulmic

Tissot Cyril, CNRS, Brest

Zeitouni Karine, Univ. Versailles St Quentin en Yvelines

Avant-Propos

La Conférence Internationale de Géomatique et d'Analyse spatiale SAGEO (Spatial Analysis and GEOmatics) est un évènement majeur de la communauté de recherche en géomatique.

Soutenue par le GDR CNRS MAGIS (Méthodes et Applications pour la géomatique et l'information spatiale), SAGEO 2013 a été co-organisée par l'UMR 6554 CNRS LETG-Brest, Géomer (Université de Bretagne Occidentale) et par l'Institut de recherche de l'Ecole navale (EA 3634). Elle s'est déroulée à l'Institut Universitaire Européen de la Mer (IUEM), du 24 au 26 Septembre 2013.

S'adressant aux chercheurs en informatique, géographie et autres domaines dont l'objet d'étude est l'information géographique, les thématiques abordées traitent des modèles, des outils et des méthodes permettant l'analyse des milieux et des activités qui s'y développent. Les articles rassemblés dans les actes de la Conférence SAGEO 2013 traduisent parfaitement la pluralité des problématiques et des approches scientifiques concernées. Répartis dans six chapitres reprenant l'organisation des sessions proposées lors de la conférence, ils valorisent des travaux aboutis présentés sous forme de contributions orales ou des travaux en cours présentés par des posters.

Ces actes incluent également les résumés des présentations des trois conférenciers invités qui nous ont fait l'honneur d'accepter de participer à SAGEO 2013 : Rodolphe Devillers (Université Memorial de Terre-Neuve), Martin Raubal (Institut Fédéral Suisse de Technologie de Zurich) et Marius Thériault (Université Laval, Québec).

L'organisation de cet événement a été soutenue par plusieurs institutions et collectivités locales et régionales que nous remercions chaleureusement. Nous remercions également les membres du comité de lecture et du comité scientifique pour leur expertise dans la sélection des articles et leur accompagnement dans la mise en place de la manifestation. Enfin, un grand merci à tous nos collègues qui se sont impliqués dans la préparation de l'évènement. Sans leur aide, leur efficacité et leur dévouement, l'organisation de SAGEO 2013 à Brest et la publication de ses actes n'auraient pas été possible.

Rémy Thibaud
IRENav
Institut de Recherche de l'Ecole Navale

Françoise Gourmelon
CNRS LETG-Brest, Géomer
IUEM, Université de Bretagne Occidentale

Orateurs invités / *Guest speakers*

Rodolphe Devillers est professeur agrégé au département de géographie de l'université Memorial de Terre-Neuve au Canada où il dirige depuis 8 ans un groupe de recherche en géomatique marine. Ses travaux de recherche utilisent principalement les systèmes d'information géographiques, l'analyse spatiale et la cartographie acoustique sous-marine.

Son expertise porte par exemple sur la cartographie des habitats marins, les méthodes géographiques utilisées pour la protection de la biodiversité marine, l'utilisation de statistiques spatiales pour l'étude des écosystèmes marins et la géovisualisation en support à l'analyse de trajectoires maritimes.

Rodolphe Devillers est également l'éditeur associé du journal *Geomatica* en charge de la géomatique marine et membre du comité de rédaction du journal *Le géographe canadien*. Il a publié une centaine d'articles dans des revues à comités de lecture et conférences portant sur la géomatique fondamentale et appliquée. Il est aussi actif dans le domaine de la qualité de l'information géographique, thème sur lequel il a publié trois ouvrages, incluant un traité IGAT.

Martin Raubal est professeur de géo-ingénierie à l'Institut Fédéral Suisse de Technologie de Zurich. Ses recherches se situent dans les domaines des SIG mobiles et services géolocalisés (LBS), de l'ingénierie spatiale cognitive, du mobile eye-tracking et des SIG pour l'analyse des énergies renouvelables.

Martin Raubal est actuellement membre du conseil d'AGILE (Association des laboratoires d'information géographique en Europe) et a été membre du conseil d'UCGIS (Consortium Université des Sciences d' Information Géographique).

Il est membre des comités de rédaction de *Transactions of GIS, Journal of Location Based Services*, *Journal of Spatial Information Science*, and *Geography Compass*.

Il est auteur et co-auteur de plus de 90 livres et de documents de recherche publiés dans des revues à comité de lecture et actes de conférences.

Marius Thériault est professeur-chercheur au Centre de Recherche en Aménagement et Développement de l'Ecole Supérieure d'Aménagement du Territoire et de Développement Régional à l'Université de Laval (Québec).

Ses recherches portent sur la modélisation des dynamiques urbaines, sur la modélisation des processus dans les SIG et les bases de données spatio-temporelles ainsi que sur les applications des méthodes d'aide à la décision et des SIG en aménagement du territoire. Il s'intéresse également aux statistiques appliquées et à l'analyse spatiale ainsi qu'à la perception de l'environnement et la modélisation des impacts environnementaux.

Il est membre du comité de rédaction de la *Revue internationale de géomatique*.

Il a reçu plusieurs prix et distinctions et a publié plus de 70 articles dans des revues à comité de lecture.

Résumés des conférences

« De la cartographie de notre "planète bleue" à l'aide de cartes papier numériques, à une meilleure connaissance des milieux marins à l'aide d'approches géographiques »

Rodolphe Devillers

Avec plus de 70% de sa surface couverte d'eau, la "Terre" porte souvent le nom plus approprié de "planète bleue". Pour la représenter, nous continuons cependant d'utiliser des méthodes développées dans les années 60 pour numériser des cartes papier du milieu terrestre, découpant les océans en tranches statiques et représentant une réalité complexe à l'aide de quelques primitives géométriques, des approches ne permettant pas la plupart du temps de capturer la complexité requise à la compréhension de ces régions les plus méconnues de notre planète. Les environnements marins continuent de défier les méthodes géomatiques classiques par leur véritable nature 3D, leur complexité et dynamique et par les efforts et coûts requis pour les échantillonner. A travers des exemples de projets de recherche récents allant de la géovisualisation de mouvements de navires à l'utilisation de statistiques spatiales pour analyser les ressources halieutiques différemment, cette présentation plaidera pour un rôle plus important des approches géographiques dans l'étude des environnements marins et lancera un appel pour rediriger encore plus nos efforts de la cartographie des océans à des méthodes pouvant améliorer leur compréhension. Cinq défis que la communauté géomatique devrait relever dans la prochaine décennie seront présentés, pouvant permettre une meilleure représentation et compréhension des environnements marins.

« GIScience & Mobility »

Martin Raubal

In recent decades a majority of our civilization has become a mobile information society. People's increased mobility has impacted various areas, such as travel and tourism, communication, social behavior, and the environment. Due to the dynamic nature of mobility, it has also led to a change in the space-time behavior of society, and created novel spatio-temporal problems and decision situations for people. Geographic Information Science can help answer many of the important questions regarding people's mobility in space and time: by providing data collection and analysis methods for the investigation of mobility patterns; by conceiving novel ways of visualizing such patterns; or by applying novel technologies for exploring people's spatio-temporal behavior and decision-making.

In this talk I will first demonstrate how human mobility patterns can be derived from mobile phone data and what kind of knowledge can be gained from the analysis of such patterns. I will then introduce location-aware mobile eye-tracking as a novel method for explaining people's spatio-temporal behavior. An improved understanding of these issues will help us to gain a better picture of people's mobility and provide input for the design of Location Based Services that facilitate mobile human-computer interaction and provide high-quality spatio-temporal decision-making support to their users.

« Modélisation et simulation des dynamiques sociales dans l'espace géographique »

Marius Thériault

Une combinaison judicieuse des procédés d'analyse statistique et de simulation informatique permet d'analyser les processus de diffusion spatiale et de construire des outils prévisionnels pour plusieurs domaines d'intervention.

Basé sur quelques principes théoriques de modélisation des processus de diffusion, cet exposé présente des stratégies de modélisation/simulation des dynamiques socio-territoriales à travers trois exemples spécifiques:

1) modélisation de la diffusion d'une innovation (adhésion à l'autopartage) à Québec de 1996 à 2004 (modèle de Poisson);
2) modélisation de la diffusion de l'urbanisation sur le pays de Brest de 1968 à 2008 (modèle logistique);
3) simulation de la diffusion d'une épidémie avec un modèle à compartiments doublé d'une enquête de mobilité quotidienne (simulation spatiale hybride de type multi-agents et automates cellulaires).

La présentation se termine par un retour sur les principes de base qui permettent de développer des modèles spatio-temporels à partir des données disponibles, tout en visant des résultats utiles pour améliorer la prise de décision lors des interventions publiques.

Table des articles par session

S.1 *Mer et littoral*

Modélisation ontologique pour l'analyse de comportements de navires à risques
Arnaud Vandecasteele, Aldo Napoli 13

Modélisation des activités de pêche à la drague en Rade de Brest
Cyril Tissot, Mathias Rouan, David Brosset, Damien le Guyader 25

Simulation cartographique au service de l'innovation collective
Géoprospective territoriale à l'île de La Réunion
Xavier Augusseau, Félicie Aulanier, Agnès Bégué, Jeremy Bourgoin, Daniel David, Gilbert David, Pascal Degenne, Mireille Fargette, Erwann Lagabrielle, Gilles Lajoie, Jean-Baptiste Laurent, Valentine Lebourgeois, Guillaume Lestrelin, Danny Lo Seen, Maud Loireau, Emmanuel Tillard, Pierre Todoroff, Jean-Philippe Tonneau, François Vitry 37

L'AIS : une donnée pour l'analyse des activités en mer
Damien Le Guyader, David Brosset 53

Aide à la détection de routes maritimes avec l'algorithme Mean Shift
Romain Gallen 57

S.2 *Usage de l'information géographique et de ses représentations*

Vers une première évaluation de la contribution des Infrastructures de Données Géographiques à la gestion des territoires
Jade Georis-Creuseveau 71

Processus cognitifs experts vs étudiants lors de la réalisation de cartes de susceptibilité
Protocole expérimental pour une analyse comparée
Raffaella Balzarini, Paule-Annick Davoine, Muriel Ney 87

Usages des modèles spatiaux pour la prospective
Thomas Houet 101

Itinéraire Méthodologique de Diagnostic Partagé des enjeux environnementaux. Comment accompagner le changement dans l'ingénierie territoriale d'Etat ?
Sylvie Lardon, Aurélie Dumont, Cécile Lacroix, Marie-Neige Lebourg 113

Analyse des patrimoines de données géographiques nationaux
Comparaison de trois infrastructures nationales de données géographiques (France, Brésil, Bolivie)
Julie Pierson, Matthieu Noucher, Pierre Gautreau, Louca Lerch, Olivier Pissoat, Aurély Jautard, Sylvain Lesage 127

Capitalisation de l'information géo¬historique dans le cadre de la Directive Inondation
Principes et développements de la Base de Données Historiques sur les Inondations (BDHI)
Marlène Villanova-Oliver, Anthony Hombiat, Cécile Saint-Marc, Paule-Annick Davoine, Jérôme Gensel 133

S.3 *Analyse spatiale*

Cartographie historique : outil clé pour caractériser la continuité temporelle
Application à des travaux en écologie forestière
Sylvie Ladet, Judicael Lopez, Audrey Grel, Lionel Valladares, Antoine Brin 139

Caractérisation des trajectoires d'évolution de l'occupation et de l'usage des sols
Approche multi-scalaire appliquée aux Pyrénées
Laure Vacquié, Thomas Houet, Christelle Vigneau 155

Appariement hiérarchique de réseaux hydrographiques imparfaits
Benoît Costes 169

Visibilité et prévisibilité : cartographie de la part cachée d'un paysage urbain
Etude comparée de quatre indicateurs qualifiant le champ visuel en milieu urbain
Thomas Leduc, Guillaume Chauvat 185

Reconstruction 3D de bâtiments à partir de cartes postales anciennes
Vers un SIG 4D collaboratif
Lara Younes, Barbara Romaniuk, Eric Desjardin, Eric Bittar 197

S.4 *Mobilités*

CTRD : une base de données spatiotemporelle pour l'analyse de la mobilité des chercheurs dans les firmes multinationales
Mahfoud Boudis, Virginie Jacquier-Roux, Claude Paraponaris 211

Dispositif mobile et géoweb pour la collecte de données d'inventaire sismique du bâti en milieu urbain
L'application Isibat
Paule-Annick Davoine, Laurent Poulenard, Philippe Gueguen, Jérôme Gensel, Jose Bringel-Filho 227

Proposition d'une méthode de génération de matrices de flux temporelles issues de l'activité de téléphones mobiles
Ana-Maria Olteanu-Raimond, Françoise Bahoken, Thomas Couronné, Zbigniew Smoreda 241

Simulation des effets de structure des réseaux sur les polluants émis par l'automobile personnelle
Julie Prud'homme, Cyrille Genre-Grandpierre, Didier Josselin, Rémi Prud'homme 259

Analyse de la mobilité d'un champ de Barkhanes (Mauritanie) à partir des images SAR
Kamel Hachemi, Yves-François Thomas 273

Utiliser les Smartphones et les outils de géolocalisation pour améliorer les connaissances et les prévisions des inondations rapides : enjeux, contraintes et perspectives
Jules Sekeouia Kouadio, Johnny Douvinet, Loic Grasland 287

ARCAMA-3D : Une application mobile à base de réalité augmentée connectée au Web des Données
Betül Aydin, Jérôme Gensel, Philippe Genoud, Sylvie Calabretto, Bruno Tellez 293

S.5 *Modélisation et gestion des données spatiales*

Optimisation des performances des opérations d'agrégation au sein des entrepôts de grilles spatialisées
Myoung-Ah Kang, Mehdi Zaamoune, François Pinet, Sandro Bimonte, Philippe Beaune 307

Interfaçage de bases de données photographiques et géographiques par appariement de lignes
Youssef Attia, Thierry Joliveau, †Eric Favier 319

S2PViewer : un prototype de visualisation de motifs spatio-temporels
Hugo Alatrista-Salas, Sandra Bringay, Frédéric Flouvat, Nazha Selmaoui-Folcher, Maguelonne Teisseire 335

Exploitation du vague spatial dans le SOLAP : vers une approche de conception prenant en compte les risques d'usage
Elodie Edoh-Alove, Sandro Bimonte, François Pinet, Yvan Bédard 347

WPS Bassin Versant
Un outil de modélisation hydrologique intégré à une Infrastructure de Données Spatiales
Roderic Béra, Hervé Squividant, Pierre Aurousseau 361

SPATIODATA, une approche alternative pour la gestion des données spatialisées relatives aux bâtiments
Cyril Carré, Vincent Delfosse, Roland Billen, Aurélie Jeunejean, Christelle Boulanger, Anne Vandervreken, Pierre-Henri Lefèbvre, Benoît Michel, Pierre Leclercq 373

Une combinaison de graphes spatiaux pour la modélisation de la structure spatiale des territoires bâtis
Mathieu Domingo, Rémy Thibaud, Christophe Claramunt 385

Une méthode de construction de données spatio-temporelles appliquée aux rues anciennes de Paris
Bertrand Dumenieu, Julien Perret, Anne Ruas 399

s.6 *Modélisation et fouille de données*

Managing flash flood hazards susceptibility coupling GIS and Cellular Automata Experimentations in Upper-Normandy (France)
Johnny Douvinet, Florian Mallet, Aurélie Escudier, Aurélien Christol, Daniel Delahaye, Patrice Langlois 419

OPITER : Fouille de données d'opinion pour les territoires
Eric Kergosien, Pierre Maurel, Mathieu Roche, Maguelonne Teisseire 435

Usage de modèles dynamiques et spatialement explicites pour la spatialisation de scénarios prospectifs normatifs
Application à la dynamique des plantations forestières industrielles dans le sud chilien
Nicolas Maestripieri, Thomas Houet, Martin Paegelow 453

Etude de faisabilité d'une cartographie multi niveaux des végétations du Nord-Ouest de la France
Vanessa Sellin 467

Modélisation spatiale de la température dans un site viticole des Coteaux du Layon
Cyril Bonnefoy, Malika Madelin, Hervé Quénol 471

Fouille d'images géoréférencées avec RapidMiner
Thomas Guyet, Hervé Nicolas, Boussad Ghedamsi, Élise Athané 475

Visualisation de données relationnelles
Intégration des graphes spatiaux dans QGis
Thomas Guyet..... 487

s.7 *Traitements d'images*

Change Detection in Floodable Areas of the Danube Delta using Radar Images
Simona Niculescu, Cédric Lardeux, Grégoire Mercier, Laurence David 499

Vectorisation automatique des forêts dans les minutes de la carte d'Etat-Major du 19ème siècle
Pierre-Alexis Herrault, David Sheeren, Mathieu Fauvel, Martin Paegelow 519

Radar hyperfréquence embarqué appliqué à la cartographie du corridor fluvial
M. Jaud, R. Rouveure, P. Faure, M-O. Monod 533

SESSION 1 :

Mer et littoral

Modélisation ontologique pour l'analyse de comportements de navires à risques

A. Vandecasteele[1], A. Napoli[2]

1. Memorial University of Newfoundland, Department of Geography
St. John's (NL), A1B 3X9, Canada
a.vandecasteele@mun.ca

2. MINES ParisTech – Centre de recherche sur les Risques et les Crises
Rue Claude Daunesse – CS 10207
F-06904 Sophia Antipolis, France
aldo.napoli@mines-paristech.fr

RÉSUMÉ. Cet article se propose d'étudier les potentialités des ontologies spatiales à la fois comme objet de modélisation, de partage et d'inférence afin d'améliorer l'analyse de comportements des navires à risque. Pour cela, un système de surveillance maritime composée de trois principales ontologies a été développé. Puis, les règles d'inférences nécessaires à la détection des alertes ont été définies par les experts du domaine maritime. Enfin, un moteur de Raisonnement à Partir de Cas permet de déterminer automatiquement les scénarios potentiels issues de l'interprétation d'une situation à risque. Au final, l'objectif est de fournir aux experts du domaine un environnement adapté permettant la modélisation des connaissances spatiales. L'approche adoptée a été mise en application au sein du prototype FishEye.

ABSTRACT. This paper explores the modeling, sharing and inferring capacities of spatial ontologies to analyze abnormal maritime behavior. To achieve this, a maritime surveillance system composed of three ontologies has been developed. Then, the inference rules needed for the detection of the alerts have been defined by the maritime experts. Finally, a Case Based Reasoning engine that analyzes a potential risky situation is used to automatically determine the potential scenarios. The purpose of this system is to provide to the maritime experts a suitable environment where spatio-temporal knowledge can be modeled. This approach has been implemented into the FishEye prototype.

MOTS-CLÉS : ONTOLOGIE, MODELISATION DES RISQUES, SURVEILLANCE MARITIME
KEYWORDS: ONTOLOGY, RISK MODELLING, MARITIME SURVEILLANCE

1. Introduction

Chaque jour, plusieurs dizaines de milliers de navires transitent par les océans. Dans ce flot continu et en raison des limites des systèmes de surveillance actuels, certains d'entre eux profitent de cette situation pour exercer des activités illégales (trafic de drogues ou de clandestins, pêche illégale, dégazage, etc.). A titre d'exemple, il est estimé que 70% de la cocaïne est transportée par voie maritime et 82% des réserves halieutiques en Méditerranée sont surexploitées (UNCD, 2011).

Face à ces différentes menaces, les opérationnels en charge de la surveillance maritime du domaine disposent d'outils spécifiques à la gestion et à l'analyse du trafic comme SPATIONAV pour la France, SIVE pour l'Espagne ou encore MEVAT pour la Finlande. Si ces outils apportent aux opérationnels une aide certaine, plusieurs limites ont néanmoins été observées (Riveiro and Falkman, 2010). L'une des principales portent sur la représentation des connaissances au sein de ces systèmes. En effet, pour définir un comportement suspect, ces experts utilisent des heuristiques acquis pour la plupart de par leur expérience. Ces heuristiques s'appuient sur des faits comme « *le navire est proche d'une côte* ». Mais, si pour l'expert en charge de la surveillance maritime ce fait est déduit naturellement, sa transcription dans un système informatique est problématique. En effet, de quel type de navires s'agit-il ? Ce fait est-il spécifique à tous les navires ou à une seule catégorie ? Qu'entend-t-on par proche ? Répondre à ces questions au sein d'un système d'aide à la décision nécessite de disposer de formalismes de représentation riches permettant de traduire la complexité de ces situations. Pour cela, nous proposons dans cet article d'explorer l'utilisation des ontologies à la fois comme réceptacle de connaissances mais aussi comme moteur d'inférence. La méthode proposée a ensuite été mise en application au sein d'un système de surveillance pour l'analyse de comportements de navires à risques.

Cet article est composé de quatre principales sections. La première présente le domaine des ontologies et plus particulièrement celui des ontologies spatiales. La seconde section décrit la méthode adoptée ainsi que l'ontologie développée. Dans la dernière section, différents cas basés sur des données réelles illustre l'approche proposée. Une analyse de cette approche ontologique ainsi que les futures évolutions potentielles sont abordées en conclusion.

2. Formalisation de la connaissance spatiale au sein des ontologies

La représentation de la connaissance dans un système informatique est intrinsèquement liée au formalisme de représentation utilisé. A l'heure actuelle, c'est le domaine des ontologies qui propose les structures de représentation les plus riches.

2.1. Des ontologies aux ontologies spatiales

Bien qu'originellement rattachée au domaine de la philosophie, la notion d'ontologie est considérée dans cet article sous l'angle de l'Intelligence Artificielle (IA). Studer *et al.* (1998) définissent l'ontologie comme : « une spécification formelle et explicite d'une conceptualisation partagée ». D'une manière plus formelle, une ontologie est constituée d'un vocabulaire spécifique permettant une description du domaine sous la forme d'un graphe constitué de concepts et de relations. Celui-ci peut alors être décrit de la manière suivante : 0:{*C, R, HC, rel, A0*} où *C* et *R* sont un ensemble disjoints de concepts et de relations, *HC* une hiérarchie (taxonomie) entre ces concepts, *rel* une fonction spécifiant la relation (non taxonomique) de *R* et enfin *A0* un ensemble d'axiomes exprimé dans un langage ontologique (Baglioni et al., 2011).

Tout comme dans le domaine de l'Intelligence Artificielle (IA), on retrouve dans l'utilisation des ontologies appliquées aux Systèmes d'Information Géographique (SIG) cette orientation à la fois philosophique et informatique. Bien que Laurini (2007) ait démontré que les ontologies pouvaient être utilisées à la fois pour aboutir à un consensus entre les acteurs d'un domaine et à une interopérabilité entre systèmes d'information, Agarwal (2005) souligne que les différents travaux scientifiques alliant SIG et ontologies se sont essentiellement appuyés sur les potentialités de chacun des domaines sans pour autant former une véritable discipline commune. Par exemple, bien qu'il existe déjà des projets modélisant la dimension spatiale sous une forme ontologique (ex : Ordonnance Survey[1], W3C[2] etc.) ainsi que des extensions du langage OWL (Cullot et al., 2003, Stocker and Sirin, 2009), il n'existe actuellement aucun standard permettant la représentation de données spatiales dans les ontologies ni de procédures de raisonnement spatial dans les moteurs d'inférence (Vandecasteele, 2012).

2.2. Représentation ontologique des objets mobiles

Dans la littérature, plusieurs modèles pour la représentation des trajectoires ont été proposés. Ces derniers varient tant au niveau des thématiques auxquels ils s'appliquent qu'aux besoins auxquels ils répondent. Par exemple, le modèle proposé par Mouza et Rigaux (2005) permet uniquement l'identification de motifs de trajectoires dont les déplacements seraient déjà connus. Le modèle de Brakatsoulas *et al.* (2004) définit un modèle sémantique ainsi que les relations potentielles mais uniquement pour le domaine routier. Plus générique, le modèle proposé par Spaccapietra *et al.* (2008) introduit un ensemble de concepts permettant de relier des informations à des événements spécifiques (*stops*, *moves*, *etc.*) et ainsi d'obtenir une trajectoire enrichie sémantiquement. Du fait de sa modularité, l'ontologie présentée dans le paragraphe suivant s'appuie en partie sur le modèle de Spaccapietra.

---

1 http://www.ordnancesurvey.co.uk/oswebsite/ontology/

2 http://www.w3.org/2005/Incubator/geo/

3. Une approche ontologique pour l'analyse des comportements de navires

Modéliser les comportements anormaux de navires nécessite au préalable de formaliser la connaissance du domaine. L'utilisation d'un langage de formalisation comme les ontologies permet d'obtenir une description enrichie tout en permettant d'effectuer des raisonnements. Pour cela, en s'appuyant sur le modèle proposé par Yan (2011) une infrastructure composée de trois ontologies a été développée. L'union de ces différentes ontologies permet alors de disposer des notions nécessaires à la description du comportement d'un navire. Cette description a été réalisée à l'aide de la connaissance des experts puis intégrée au sein de différentes ontologies qui forment le cœur de la plate-forme développée dans ces travaux de recherche. Bien que le fonctionnement général du système ainsi que l'utilisation des règles d'inférence aient déjà été présentés (Vandecasteele et Napoli, 2012; 2012a), nous nous attachons ici à décrire les concepts ontologiques sur lesquels reposent ces travaux.

3.1. Modélisation du contexte sémantique

Afin de comprendre le comportement d'un navire, sa position seule ne suffit pas. En effet, ce comportement s'analyse au travers d'une multitude d'informations (condition météorologique, navires à proximité, etc.) qu'il est nécessaire de prendre en compte. C'est là toute l'idée du concept de comportement sémantique. Ainsi, la position, le segment ou le trajet d'un navire sont enrichies par l'ajout d'informations contextuelles qui seront ensuite utilisées lors du processus d'inférence. Pour cela, en s'appuyant sur le modèle initialement développé par Yan (2011) trois principales ontologies ont été créées :

• Une ontologie des trajectoires contenant les différents concepts spatio-temporels nécessaires à la description géométrique des trajectoires. Ces concepts géométriques peuvent être donnés de manière absolue sous la forme d'objets géographiques (points, lignes, etc.) ou relative comme une adresse;

• Une ontologie géographique contenant les concepts spécifiques à la description du territoire (routes, ports, baies, etc.). Cette ontologie est reliée à la fois à l'ontologie des trajectoires par les concepts géométriques qu'elle utilise et à l'ontologie du domaine pour les thématiques spécifiques qui peuvent être ajoutées en fonction des besoins ;

• Une ontologie du domaine qui, comme son nom l'indique, est relative au domaine étudié. Celle-ci contient les concepts permettant d'enrichir les informations initialement contenues et d'apporter ainsi une connaissance spécifique au sein de l'application.

Dans le cadre de ces travaux, le modèle proposé par Yan (2011) a été étendue afin non pas de modéliser uniquement les trajectoires, mais aussi d'en déduire des événements (alertes) et surtout d'identifier automatiquement les scénarios potentiels. Bien que l'ontologie pour l'analyse des comportements anormaux s'appuie sur trois

sous-ontologies, l'essentiel des travaux réalisés a porté sur celle concernant la modélisation du domaine maritime. En effet, conformément au précepte ontologique de réutilisabilité, celles sur les trajectoires et la géographie ont été créées en s'appuyant sur des ontologies déjà existantes. Par exemple, les éléments géographiques et temporels ont été définies en utilisant les ontologies *GeoRSS Simple*[3] et *OWL Time*[4]. Pour les trajectoires, l'approche présentée par Spaccapietra a été implementée. De ce fait, seule l'ontologie du domaine sera décrite dans cet article.

3.2. Ontologie du domaine : l'analyse de comportements anormaux de navires

En accord avec la typologie donnée par Guarino (1998), l'ontologie du domaine presentée dans ce document doit être considérée comme une ontologie d'application. La modélisation du domaine a été réalisée lors de séances de travail avec les experts maritimes (garde-côtes, gendarmerie maritime, etc.) mais aussi en s'appuyant sur la littérature scientifique existante (Nilsson et al., 2008; Van Laere and Nilsson, 2009; Roy and Davenport, 2010). Les éditeurs d'ontologies actuels ne permettant pas une utilisation intuitive sans connaissance préalable, il n'a pas été possible pour les experts de modéliser directement les notions souhaitées. Cette étape a donc été réalisée dans un second temps par le modélisateur. Puis le résultat, sous la forme d'une ontologie, a été présenté aux experts pour validation. De plus, il est nécessaire de préciser que même si en situation à risque les manœuvres entreprises ne sont pas toujours respectueuses du règlement maritime (Le Pors, 2010), seules les règles formelles ont été implementées.

3.2.1 Principales classes de l'ontologie du domaine

Quatre classes principales structurent l'ontologie du domaine permettant l'analyse de comportements des navires :

• La classe *Alert* : cette classe contient les différentes alertes relatives à un navire. Ces dernières ont été définies en collaboration avec les experts du domaine. Différentes sous-classes ont été créées en fonction de la thématique des alertes. Sont par exemple répertoriées, celles relatives à l'état de la mer, à la localisation ou à la cinématique du navire. Certaines de ces alertes nécessitent la prise en compte de la dimension spatiale. Dans ce cas, il est nécessaire d'ajouter aux individus la propriété de donnée *DP_Analysis_DataPath* indiquant la localisation des données géographiques à utiliser. Les alertes peuvent être reliées aux autres concepts de l'ontologie grâce à la propriété d'objet *has_alert*.

• La classe *Vessel* : cette classe contient toutes les notions relatives aux navires. Elles permettent de décrire un navire en précisant par exemple son type ou encore son pavillon d'origine.

---

[3] http://www.w3.org/2005/Incubator/geo/
[4] http://www.w3.org/2006/time

• La classe *Scenario* : dans cette classe les experts du domaine définissent le comportement d'un navire. Un scénario est composé de notions provenant des trois classes précédentes. Les experts construisent alors un scénario en fonction des alertes, des éléments du contexte et des caractéristiques du navire ou des navires.

3.2.2. Modélisation des alertes et des scénarios

Dans l'ontologie du domaine, la classe *Alert* permet de spécifier un événement inhabituel ou suspect. Deux grands types d'alertes peuvent être distingués. Les alertes à caractère spatial et celles à caractère non spatial. Les concepts définis dans cette classe seront ensuite utilisés pour la description des scénarios et l'enrichissement de la base de connaissance. Les alertes spatiales s'appuient directement sur les éléments géographiques. Il s'agit par exemple de signaler l'entrée d'un navire dans un espace soumis à une réglementation spécifique, d'identifier une trajectoire particulière ou encore de détecter une zone où les conditions météorologiques sont mauvaises (Figure 1). Ces alertes spatiales s'appliquent aussi bien à des événements ponctuels (mouillage d'un navire) que linéaires (trajectoire anormale).

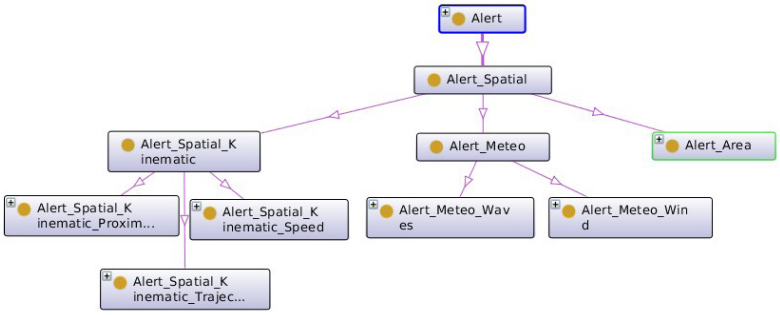


Figure 1 : Sous-classes de la classe Alert_Spatial

La notion de scénario, telle que définie dans l'ontologie, est composée d'une suite d'événements qui peut être enrichie par des éléments contextuels. Les scénarios potentiels ont été renseignés par les experts du domaine. Tout comme les alertes, les scénarios potentiels ont été divisés par thématique. A l'heure actuelle, quatre sous-classes ont été définies (Figure 2) :

• *Scenario_Attack :* il décrit les cas d'attaques potentiels. Il s'agit par exemple d'une attaque terroriste ou encore d'une attaque à l'aide d'un tanker.

• *Scenario_Fishing :* il s'agit dans ce cas de scénarios liés à la pêche. Sont notamment définis ceux en rapport au transbordement ou encore la pêche illégale.

• *Scenario_Traffic :* ce scénario est relatif à tous les types de trafic potentiel. Il peut s'agir d'un trafic de marchandises, d'armes ou de drogues.

• *Scenario_Trajectory :* il porte sur les scénarios prenant en compte la trajectoire du navire. Cela peut être par exemple nécessaire pour détecter une future collision.

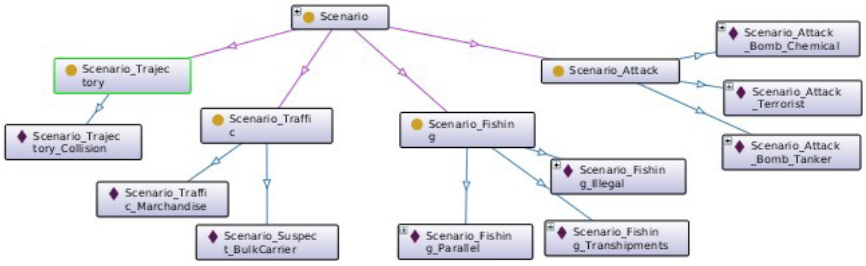


Figure 2 : Sous-classes et instances pour la description des scénarios

La modélisation des scénarios utilise les différentes notions de l'ontologie provenant de la classe Alert, Context ou Vessel. Ces scénarios sont par la suite utilisés par le moteur de Raisonnement À Partir de Cas (RÀPC) pour identifier les comportements anormaux.

3.2.4. Modélisation du navire

Au cœur de l'analyse de comportement, il y a bien évidemment l'objet à étudier, le navire lui-même. Celui est représenté dans l'ontologie par la classe Vessel qui regroupe l'ensemble des notions importantes à sa caractérisation (Figure 3). Différentes sous-classes ont été définies et permettent par exemple d'annoter son comportement, son historique ou encore son type. Les navires identifiés comme suspects sont stockés dans l'ontologie sous la forme d'instance de la classe VOI (*Vessel Of Interest*).

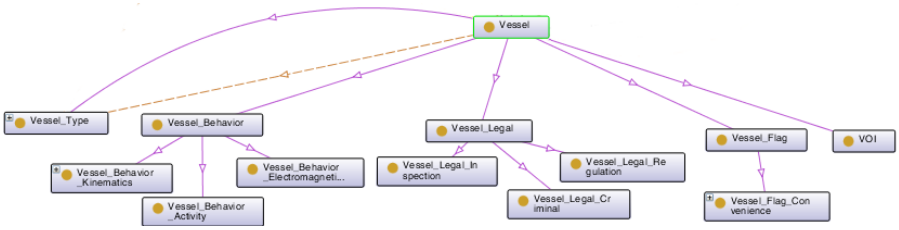


Figure 3 : Sous-classes de la classe Vessel

3.3. Modélisation des règles d'inférence et enrichissement du SWRL

Une fois le domaine modélisé, il est alors possible d'ajouter des règles permettant l'analyse du comportement des navires. Celles-ci peuvent être vues comme des déclencheurs qui vont modifier l'ontologie en fonction de la satisfaction ou non de la règle. La représentation et l'intégration des règles dans l'ontologie ont été réalisées à l'aide du langage Semantic Web Rule Language (SWRL). Celui-ci enrichit la sémantique d'une ontologie en combinant le langage OWL-DL et RuleML

(Horrocks et al., 2004). Basée sur les logiques de description et les règles de Horn, la structure d'une règle SWRL est de la forme antécédent(s) -> conséquence(s) qui signifie : « si l'antécédent est vrai, alors la conséquence l'est aussi ». En plus des prédicats OWL, SWRL possède également des fonctions supplémentaires (mathématiques, booléennes, etc.) nommées 'built-in' (Grau et al., 2008). Ces derniers étendent les fonctions initialement proposées par le langage OWL en permettant par exemple d'effectuer des comparaisons et des calculs. De la même manière, cette approche sous la forme de '*built-in*' a été adoptée afin d'intégrer des fonctionnalités spatiales aux SWRL. Ainsi, tout comme les opérateurs mathématiques ou ceux portant sur les chaînes de caractères, l'intégration de la dimension spatiale se fait au moyen de mots clé. Ces mots-clés, que l'on retrouve également dans le domaine des bases de données spatiales, sont par exemple : *touches*, *intersects*, etc. Cette approche permet ainsi de bénéficier de toute la richesse sémantique proposée par les ontologies tout en intégrant la composante géographique.

Par exemple, pour spécifier que l'entrée d'un navire dans une zone réglementée doit déclencher une alerte, la règle SWRL ci-dessous a été définie. Celle-ci prend en paramètres, le navire (*vessel*) ainsi que sa position (*position*), la couche de données sur laquelle portera l'analyse (*Alert_Area_Restricted*) et enfin la fonction spatiale à réaliser (*intersects*). Si cette analyse est positive alors une alerte est automatiquement ajoutée au navire correspondant (*Alert_Area_Restricted*) :

Règle SWRL 1. Entrée d'un navire dans une zone réglementée

1: *Vessel(?vessel),*
2: *DP_Analysis_DataPath(Alert_Area_Restricted, ?geoData),*
3: *DP_Vessel_hasPosition(?vessel, ?position),*
4: *intersects(?position, ?geoData)* →
5: *OP_Alert_hasAlert(?vessel, Alert_Area_Restricted)*

Les règles spatiales sont directement prises en compte et calculées par le système. Pour cela, le moteur d'inférence Pellet[5] a été étendu par l'ajout d'un module géographique. Ce module ainsi que les fonctions spatiales ont été développés dans le cadre de ces travaux de recherche.

4. Expérimentations et évaluations

L'approche proposée a ensuite été mise en application au sein du système de surveillance maritime développé par le Centre de recherche sur les Risques et les Crises : FishEye. Celui-lui offre aux experts du domaine un ensemble d'éléments pour la modélisation et l'analyse de comportements anormaux de navires. Afin d'illustrer le fonctionnement de cette plate-forme, le cas d'un navire à risque entrant dans une zone réglementée est présenté cette section.

[5] http://clarkparsia.com/

4.1. Scénario d'un navire dangereux dans une zone réglementée

Certaines zones sont, pour des questions de sécurité, interdites à la navigation. L'entrée d'un navire dans l'une d'entre elles doit donc générer une alerte afin d'informer l'opérateur en charge de la surveillance. Ce type d'analyse fait appel à une fonction spatiale de type intersection (*intersects*) et aux données du SHOM[6] dans lesquelles sont décrites ces zones. Dans le cas présenté, il s'agit plus précisément d'identifier les navires ayant été potentiellement détournés afin de servir comme arme potentielle.

4.2. Modélisation du scénario

Ce scénario nécessite de disposer d'informations concernant notamment le type du navire, les conditions de navigation ainsi que les zones traversées. Pour cela, deux requêtes spatiales sont nécessaires. La première qui va permettre de savoir si un navire s'est rendu dans le port d'un pays identifié comme étant à risque et la seconde qui signale si ce même navire entre dans une zone réglementée (Règle SWRL 1). Ces deux requêtes font appel à la même fonction spatiale *intersects* mais avec des jeux de données différents.

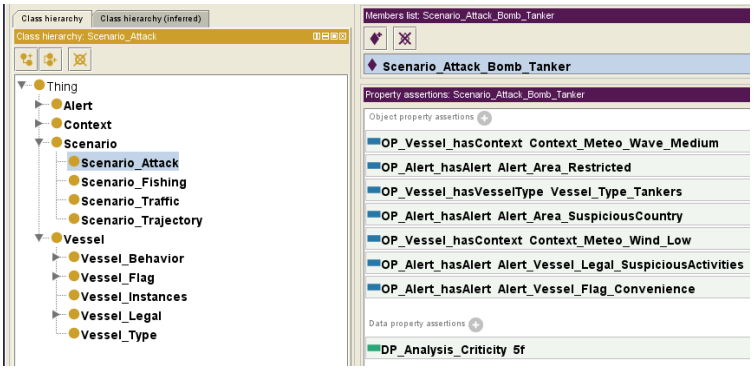


Figure 4 : Description du scénario de type attaque de tanker

En fonction du type du navire, trois scénarios, constitués de sept propriétés d'objets (contexte dans lequel évolue le navire, son type, les zones, etc.), ont été créés. L'illustration ci-dessus (Figure 4), présente les paramètres nécessaires à l'identification du scénario par l'attaque d'un tanker. Les différentes alertes (*Alert_Area_Suspicious*, *Alert_Vessel_Flag_Convenience*, etc.) inhérentes à ce scénario ainsi que les éléments contextuels (*Context_Meteo_Wind_Low*, *Context_Meteo_Wave_Medium*, etc.) ont été définis. Enfin, pour chaque scénario un niveau de criticité est spécifié (*DP_Analysis_Criticity*). Celui-ci sera utilisé pour définir le niveau de dangerosité du navire et le mettre en évidence lors de l'affichage sur l'interface cartographique FishEye.

[6] Service Hydrographique et Océanographique de la Marine (http://www.shom.fr/)

4.3. Analyse du comportement et visualisation des résultats dans FishEye

Une fois les scénarios modélisés, il est alors possible de créer dans l'ontologie les individus et de vérifier la concordance des règles et des scénarios. Pour cela deux étapes sont nécessaires. Dans un premier temps, le moteur d'inférence (spatial et non spatial) va détecter les alertes potentielles (*Alert_Area_Restricted,* etc.) en fonction des caractéristiques du navire, de sa position et de sa trajectoire. Puis, dans un second temps, le module de Raisonnement à Partir de Cas va mettre en correspondance les scénarios préalablement définis et la situation observée. Les informations automatiquement calculées par le moteur d'inférence peuvent ensuite être analysées par le module de RàPC. Celui-ci va alors automatiquement définir les scénarios ayant la plus forte occurrence et définir le niveau de criticité en comparant les alertes et les éléments contextuels du navire avec ceux définis pour chaque scenario.

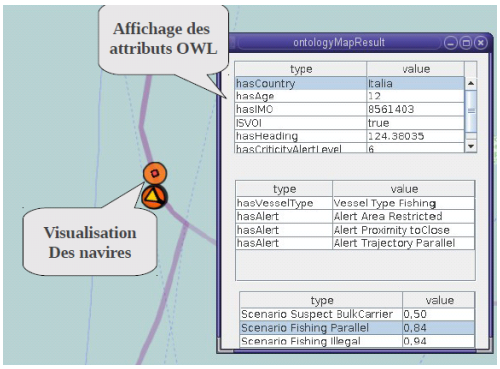


Figure 6: Détection de comportements anormaux dans FishEye

Enfin, l'affichage du résultat des analyses effectuées se fait à partir du système de surveillance maritime FishEye (Figure 6). Cette visualisation offre un accès direct et intuitif aux informations sémantiques. Néanmoins, pour des questions de performances et malgré une étape préalable d'optimisation des données, le système proposé ne peut, pour le moment, être utilisé en temps réel. En effet, l'ensemble du processus se fait directement à partir du fichier OWL. A terme, une base de données sémantique (triplestore) sera utilisée. Cela permettra d'optimiser le processus et ainsi envisager l'utilisation du système en temps réel.

5. Conclusion

La modélisation des navires à risque sous la forme d'une ontologie proposée dans ce document offre un cadre formel tout en permettant de disposer de procédures de raisonnement. À partir de cette ontologie une plateforme de modélisation a été conçue permettant de spécifier les alertes et les scénarios maritimes à détecter. Un risque maritime étant spatio-temporel, le langage SWRL a été enrichi afin de

manipuler des objets spatiaux ainsi que des fonctions spatiales comme l'inclusion ou l'intersection. Enfin, la correspondance entre les scénarios préalablement définis et une situation donnée se fait grâce à un moteur de Raisonnement à Partir de Cas.

Bien que fonctionnelle, plusieurs améliorations de la plateforme Fisheye ont été envisagées. Tout d'abord, une extension intégrant le langage de requête SPARQL et son extension spatiale GeoSPARQL serait pertinente. Celle-ci permettrait alors aux utilisateurs d'explorer les résultats obtenus afin de confirmer ou affiner les scénarios détectés. De plus, bien que les langages ontologiques offrent les éléments nécessaires à la modélisation des connaissances, la formalisation de celles-ci restent difficiles pour des personnes extérieures au domaine. Même si certains éditeurs d'ontologies comme Protégé offre une interface conviviale, ceux-ci s'adressent néanmoins à des utilisateurs connaissant déjà les concepts inhérents aux ontologies. Or, dans notre ontologie la définition des règles permettant de spécifier les alertes ainsi que la création des scénarios se font en utilisant directement le langage OWL et SWRL. Il est bien évidemment difficile de demander à un expert maritime (CROSS, gendarmerie maritime, etc.) ne disposant pas de ce type de compétences de réaliser cette implémentation. C'est pourquoi la création d'interfaces spécialement adaptées au domaine étudié pourrait être une solution. Enfin, le stockage des individus (navires, scénarios, etc.) se fait actuellement directement dans le fichier OWL. Si cette solution est techniquement plus simple, elle entraîne également des performances moindres. De ce fait, une évolution du système intégrant une base de données sémantique est envisagée.

Bibliographie

Agarwal, P., (2005). Ontological considerations in GIScience. International Journal of Geographical Information Science 19, 501–536.

Baglioni, M., Masserotti, M.V., Renso, C., Spinsanti, L., (2011). Improving Geodatabase Semantic Querying Exploiting Ontologies, in: GeoSpatial Semantics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 16–33.

Brakatsoulas, S., Pfoser, D., Tryfona, N., 2004. Modeling, storing and mining moving object databases. Database Engineering and Applications Symposium, 2004. IDEAS '04, pp. 68 – 77.

Cullot, N., Parent, C., Spaccapietra, S., Vangenot, C., 2003. Des ontologies pour données géographiques. Revue internationale de géomatique, Les SIG sur le web 13, 285–306.

Grau, B.C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U., (2008). OWL 2: The next step for OWL. Web Semantics: Science, Services and Agents on the World Wide Web 6, 309–322.

Guarino, N., (1998). Formal Ontology and Information Systems, in: Proceedings of FOIS'98. IOS Press, Trento, Italy, pp. 3–15.

Horrocks, I., Patel-Schneider, P., Boley, H., Tabet, S., Grosof, B., Dean, M., 2004. SWRL: A Semantic Web Rule Language Combining OWL and RuleML.

Kuhn, W., (2001). Ontologies in support of activities in geographical space. International Journal of Geographical Information Science 15, 613–631.

Laurini, R., (2007). Pre-consensus Ontologies and Urban Databases., in: Ontologies for Urban Development, Studies in Computational Intelligence. Springer, pp. 27–36.

Le Pors, T., (2010) Simulation cognitive de la prise de décision d'experts. Application au trafic maritime. Thèse de Doctorat, IRENav - LESTICC, Univ. de Bretagne Sud.

Mouza, C. du, Rigaux, P., 2005. Mobility Patterns. Geoinformatica 9, 297–319.

Nilsson, M., van Laere, J., Ziemke, T., Edlund, J., (2008). Extracting rules from expert operators to support situation awareness in maritime surveillance, in: 11th International Conference on Information Fusion. Presented at the 2008 11th International Conference on Information Fusion, IEEE, pp. 1–8.

Riveiro, M., Falkman, G., (2010). Supporting the Analytical Reasoning Process in Maritime Anomaly Detection: Evaluation and Experimental Design, in: Information Visualisation (IV), 2010 14th International Conference. Presented at the Information Visualisation (IV), 2010 14th International Conference, pp. 170 –178.

Spaccapietra, S., Parent, C., Damiani, M.L., de Macedo, J.A., Porto, F., Vangenot, C., 2008. A conceptual view on trajectories. Data & Knowledge Engineering 65, 126–146.

Roy, J. et Davenport, M., (2010). Exploitation of maritime domain ontologies for anomaly detection and threat analysis, in: Waterside Security Conference (WSS), 2010 International. Presented at the Waterside Security Conference (WSS), 2010 International, pp. 1–8.

Stocker, M., Sirin, E., 2009. PelletSpatial: A Hybrid RCC-8 and RDF / OWL Reasoning and Query Engine. Language 529.

Studer, R., Benjamins, V.R., Fensel, D., (1998). Knowledge engineering: Principles and methods. Data & Knowledge Engineering 25, 161–197.

UNCD, (2011). Review of Maritime Transport 2011 (No. UNCTAD/RMT/2011). United Nations Conference on Trade and Development.

Vandecasteele A., et Napoli A. (2012). Using Spatial Ontologies for Detecting Abnormal Maritime Behaviour ; Conférence internationale IEEE OCEANS 2012 ; Yeosu, Korea.

Vandecasteele A. et Napoli A., (2012a). Enhancement of Ontology with Spatial Reasoning Capabilities to Support Maritime Anomaly Detection ; Conférence internationale IEEE SOSE 2012 (7th International Conference on System Of Systems Engineering), Gênes, Italie, 16-19 juillet 2012.

Van Laere, J., Nilsson, M., (2009). Evaluation of a workshop to capture knowledge from subject matter experts in maritime surveillance, in: 12th International Conference on Information Fusion, 2009. FUSION ’09. IEEE, pp. 171–178.

Yan, Z., (2011). Semantic Trajectories: Computing and Understanding Mobility Data (Phd).

Modélisation des activités de pêche à la drague en Rade de Brest

Cyril Tissot*, Mathias Rouan*, David Brosset, Damien le Guyader***

*LETG/Brest Géomer, Institut Universitaire Européen de la Mer, place Nicolas Copernic 29280 Plouzané
** IRENAV, Ecole Navale Lanvéoc-Poulmic, 29240 BREST CEDEX 9, France

RÉSUMÉ. *Cette contribution présente une démarche de modélisation dédiée à la simulation d'activités de pêche à la drague en Rade de Brest. Fondée sur le couplage d'un modèle multi-agents avec une base d'information spatio-temporelle, l'approche proposée vise à restituer le déroulement de la pêche sous un angle rétrospectif (évaluation de la variabilité inter et intra-annuelle de l'activité) en intégrant l'évolution des contraintes de pratique. En établissant une relation spatio-temporelle formelle entre les activités et l'évolution des conditions environnementales du milieu dans lesquels elles évoluent, le modèle présenté permet d'envisager des développements prospectifs à base de scénarios.*

ABSTRACT. *This paper presents a modeling approach dedicated to the simulation of dredge fishing in the Bay of Brest. This approach is based on the coupling of a multi-agent model with a spatio-temporal database. Results are fitted fishery operation under a retrospective angle (assessment of variability) by integrating the evolution of practical constraints. By establishing a formal spatio-temporal relationship between activities and environmental condition the model allows to explore forecasting scenarios.*

MOTS-CLÉS : *modèle multi-agents, base de données spatio-temporelles, couplage, activités marines, simulation rétrospective.*

KEYWORDS: *agent-based model, spatiotemporal databases, marine activities, coupling simulation.*

1. Introduction

L'examen des différentes contributions scientifiques traitant de la modélisation des interactions homme/milieu démontre la difficulté d'intégration des dimensions spatiales et temporelles dans les approches proposées (Gould, 1987 ; Stonebraker, 1990 ; Allen, 1991 ; Muxart, 1992 ; Snodgrass, 1992 ; Cheylan, 1993 ; Claramunt,

1999 ; Parent, 1999 ; Legay, 2000, Pelekis, 2005, Tang 2008). La principale limite est inhérente à la variabilité de ces interactions.

Cette remarque vaut tout particulièrement lorsque l'on s'intéresse au milieu marin, traditionnellement considéré comme « un espace de liberté » et, où il est particulièrement délicat d'appréhender les interactions homme/milieu. L'analyse de ces interactions est donc un des objectifs majeurs de la recherche pour l'aide au développement durable des sociétés littorales (Coelho *et al.*, 2003; Dronkers *et al.*, 1999). L'étude des modes d'utilisation et d'exploitation du milieu par l'Homme dans une double perspective spatiale et temporelle en est une composante essentielle (Holligan 1994; Vallega 1999).

Dans ce contexte, les tentatives de modélisation dédiées aux applications marines sont de plus en plus nombreuses et peuvent concerner, par exemple, la lutte antipollution (Howlett 1997; Pelot *et al.* 2008), la gestion des pêches (Maury *et al.* 1999) ou l'aide à la navigation maritime (Fournier *et al.* 2003; Goralski *et al.* 2007). Néanmoins les études visant à modéliser le déroulement d'activités en mer côtière et d'en étudier les interactions réciproques et avec l'environnement sont plus rares (Le Tixerant *et al.*, 2012).

Les travaux présentés dans cet article se placent dans cette perspective. Ils visent à modéliser le déroulement des activités de pêche à la drague en rade de Brest et à en étudier les interactions avec leur environnement de manière à évaluer les pressions qu'elles exercent sur la ressource et les habitats benthiques associés. Cette démarche se place dans la perspective de proposer une approche rétrospective visant à reconstituer l'évolution des stocks de coquilles Saint-Jacques et de praires en tenant compte des facteurs environnementaux et des changements d'intensité de la pêche à la drague..

2. Modélisation quantitative des activités de pêche à la drague

L'approche développée vise à construire un modèle quantitatif permettant de restituer le déroulement de la pêche à la drague et d'analyser les changements d'intensité de cette activité aux échelles inter et intra-annuelles.
Cet objectif nécessite de produire un modèle intégrant l'ensemble des contraintes techniques, environnementales et réglementaires influençant le déroulement de la pratique. Dans ce contexte un prototype a été implémenté au sein d'une plateforme multi-agents développée au laboratoire LETG/Brest Géomer depuis un dizaine d'année (Tissot *et al.*, 2013). Cette plateforme baptisée Dynamique des Activités HUmaines (DAHU) propose un environnement de modélisation pouvant intégrer des contraintes spatio-temporelles multi-sources et multi-échelles en tant que variables de forçage[i] au sein de modèles agents contraints[ii] (figure 1). Son architecture permet de formaliser de manière explicite les relations spatio-temporelles entre des entités spatiales, des processus environnementaux et des activités simulées.

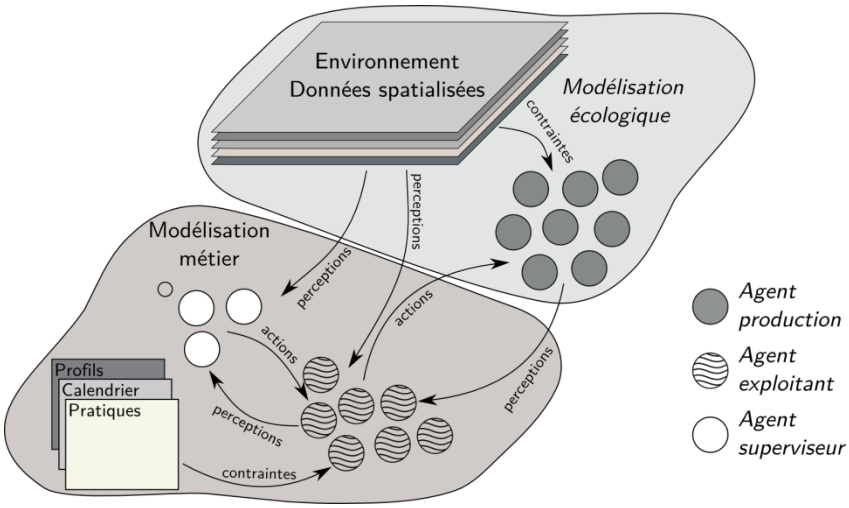


Figure 1 : structuration de la plateforme de modélisation Dynamique des Activités HUmaines (DAHU)

DAHU a été conçue dans l'optique de décrire le fonctionnement et la distribution spatio-temporelle d'activités humaines, modélisés sous la forme d'agents réactifs contraints par des variables exogènes (contraintes bio-physiques, socio-économiques et réglementaires). Il considère l'écosystème étudié comme un potentiel de ressources exploité par les hommes en fonction des techniques dont ils disposent et de leur organisation sociale à une époque donnée. Cette exploitation n'est pas systématiquement associée à un prélèvement, il peut également s'agir d'une affectation de territoire à une utilisation spécifique.
Chaque activité est représentée par un agent autonome capable de réagir à un environnement vraisemblable préalablement modélisé. Ces agents restent à des niveaux de spécification génériques et ne possèdent aucune spécificité liée à l'implantation spatiale des activités modélisées. En revanche ils intègrent des capacités de réaction et d'adaptation à l'évolution de leur environnement, celui-ci résultant d'une combinaison de contraintes naturelles et anthropiques associées à un territoire.

Trois types d'agents permettent de hiérarchiser les relations entre les activités anthropiques et leur environnement :

- les Agents Superviseurs représentent les structures de régulation. Ils orientent et contrôlent le bon déroulement des activités,
- les Agents Exploitants restituent le déroulement des pratiques associées à chaque activité. Ils possèdent des objectifs et élaborent une stratégie,
- les Agents Production ont pour objectif de produire des ressources dans un environnement contraint. Ils représentent l'élément central du simulateur

car ils déterminent la nature et la qualité des ressources disponibles pour les Agents Exploitants. Ils jouent donc un rôle essentiel dans la stratégie des Agents Exploitants et dans l'attitude des Agents Superviseurs qui, en fonction de l'état de cette ressource ou de son environnement, peuvent prendre des mesures restrictives.

Chaque Agent Production est construit, à partir d'un archétype, comme un processus "hors-sol" de manière à appréhender l'ensemble des échelles temporelles en excluant toute perturbation liée au particularisme des contraintes d'environnement. A ce stade, le temps est considéré comme un élément structurant chargé de faire le lien entre des agents en perpétuelle évolution et des états de l'environnement observés à des instants t.
Ce schéma relationnel est rendu possible grâce à la gestion conjointe des dimensions spatiales et temporelles au sein de la plateforme DAHU. Les différents modèles d'activités intègrent des informations provenant d'une base de données spatio-temporelle, administrées sur un serveur postgresl/postgis, comme autant de contraintes au fonctionnement des agents. En retour, chaque agent est doté de capacités réactives qui se manifestent par un comportement adaptatif en fonction de l'évolution de son environnement.

Ces différentes classes d'agents ont été transcrites pour répondre aux besoins spécifiques de la modélisation des activités de pêche à la drague (figure 2). Trois types d'agents ont ainsi été implémentés dans un module dédié "Rad2Brest" :

- les Agents "ComiteDesPeches" dont le rôle est d'encadrer les pratiques de pêche à la drague et d'administrer les campagnes de réensemencement visant au maintien des ressources benthiques ciblées par la pêche. Ils ont également une fonction réglementaire dans la mesure où ils définissent les modalités de déroulement des campagnes (périodes autorisées, zones de réserve à accès interdit ou restreint, quotas éventuels) ;

- les Agents "Pecheur" réalisent un prélèvement sur une espèce cible correspondant à leur armement. Ils possèdent des capacités de réaction à l'évolution de leur environnement (qualité de l'eau, contraintes réglementaires, effort de pêche, variation des prix, …) et peuvent donc adapter leurs stratégies (choix du site, période de pêche) en fonction de ces contraintes ;

- les Agents "ZoneDePeche" sont considérés comme des entités de production des ressources benthiques. Ils possèdent un ensemble de caractéristiques qui déterminent le stock et le niveau de maturité des individus qu'ils abritent.

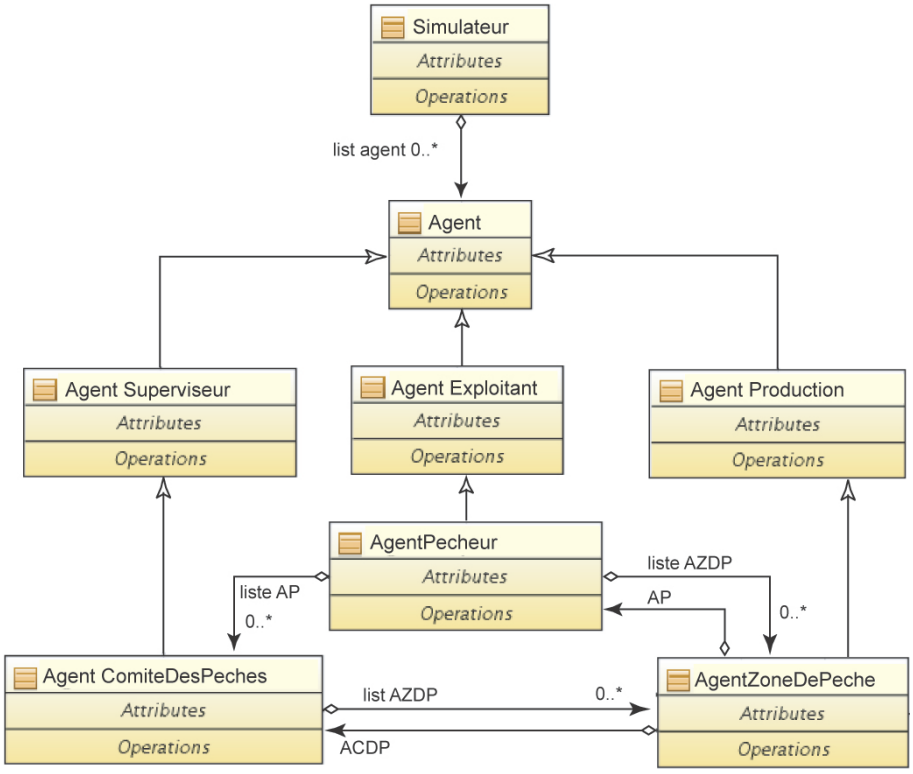


Figure 2 : Diagramme de classes du module Rad2Brest

Chaque agent hérite des attributs et méthodes de la classe supérieure. Sur le plan informatique, un agent est identifié à un thread intégrant des données techniques propres au fonctionnement de l'activité simulée et des contraintes spatio-temporelles inhérentes au domaine de simulation. Chaque agent possède des capacités de réaction génériques en fonction de la classe à laquelle il est intégré (Agent Superviseur, Agent Prélèvement ou Agent Production). La stimulation de ces capacités réactives s'effectue via l'intégration de contraintes propres à chaque type d'agent. Les réponses données à ces stimuli dépendent donc du profil des agents. Un agent "ZoneDePeche" par exemple, évoluera en fonction des contraintes biophysiques du milieu mais également en fonction de la stratégie de l'agent "ComiteDesPeches" qui peut modifier la gestion de la pêcherie (choix des sites de réensemencement, quantité de juvéniles...).

Compte tenu de l'hétérogénéité des données intégrées au sein de la chaîne de calculs, l'étalonnage temporel du modèle est piloté par une classe simulateur qui a pour rôle de synchroniser les différents threads manipulés par le modèle. Dans le cas présent, les agents possèdent des pas de temps journaliers. L'ordonnancement nécessite également de tenir compte de l'inertie temporelle nécessaire aux agents production

pour atteindre un état cyclique. Une période d'équilibrage permettant de constituer une population d'individus de différentes classe d'âge est donc calculée pour l'agent "ZoneDePeche" en fonction du type d'espèce concernée (coquilles Saint-Jacques, Praires...). Lors du lancement d'une simulation certains threads sont lancés à une date antérieure à la date courante (déterminée en fonction de la durée d'équilibrage de l'agent) afin d'éviter un lancement simultané de l'ensemble des activités au premier jour de la simulation.
A partir de cette structuration, l'objectif est d'évaluer la variabilité des ressources benthiques en fonction de l'évolution des activités (nombre de licence, quotas, nombre de jours de pêche autorisés), des conditions environnementales et de la dynamique de repeuplement (réensemencements, mises en réserve).

3. Intégration des contraintes d'environnement

L'ensemble des contraintes d'environnement est administré au sein d'un serveur Postgresql/PostGis couplé au modèle multi-agents. La structuration proposée permet de calculer, les conditions de pratique pour chaque type d'agent à partir d'une combinaison de contraintes influençant le déroulement des activités modélisées.

Dans le cas de la pêche à la drague ces calculs sont réalisés au pas de temps quotidien. Les principales contraintes sont inhérentes à la qualité de l'eau, aux conditions de marées (horaires, coefficients) et à l'abondance des ressources benthiques (taille, poids). Pour la classe AgentComiteDesPeches, l'objectif est de connaître l'état du bassin de production (suivi des stocks) et la qualité des eaux. Pour les agents de la classe AgentPecheur, l'objectif est d'élaborer la meilleure stratégie de pêche, c'est-à-dire d'adapter leurs pratiques à l'évolution des contraintes d'environnement et des captures réalisées au tour précédent. Pour la classe AgentZonedePeche, le contrôle d'un ensemble de variables environnementales (t° de l'eau, salinité,…) permet de simuler la croissance des coquilles Saint-Jacques et des praires au pas de temps quotidien (régression multiple). Cette relation dynamique est construite via des séquences de requêtes spatio-temporelles (en langage SQL) mobilisant dans certains cas des traitements numériques de l'information géographiques (calcul d'intersections, différences symétriques, calcul de zones tampons…). Ces requêtes fournissent aux agents l'ensemble des contraintes de production ou d'exploitation des ressources.

La mise en relation de ces différents éléments pose d'importante question en terme de granularité spatiale. Définir la granularité spatiale nécessaire à la modélisation d'un phénomène est un des points clés et un préalable indispensable à la construction d'un modèle sous contraintes spatio-temporelles (Pereira *et al.*, 2004). Une pratique couramment utilisée et relativement intuitive est de considérer le plus petit dénominateur commun des objets manipulés comme l'unité spatiale de référence. Il en va de même pour le pas de temps de simulation. Si dans de nombreuses applications cette démarche donne de bons résultats, elle a l'inconvénient d'amener potentiellement à manipuler des objets dont la dynamique n'est pas connue ou sans réalité tangible.

Dans le cas des activités de pêche à la drague il est particulièrement complexe de mettre en relation une zone de pêche avec des espèces benthiques et des activités de prélèvement qui mobilise des engins différents en fonction de l'espèce cible.
Dans ce contexte, le modèle développé mobilise des procédures d'agrégation et de désagrégation spatiale permettant de disposer de données en adéquation avec le type d'activité modélisée. Cette problématique pose d'importantes questions méthodologiques concernant la réalisation de transferts d'échelle au sein des modèles. Les contributions scientifiques traitant de ce type de démarche sont rares et mettent en avant la complexité des procédures d'agrégation et de désagrégation spatiale [Gotway, 2002 ; Duboz, 2004]. Les principales limites sont inhérentes à la perte de cohérence de données jugées homogènes à une échelle donnée mais qui peuvent s'avérer hétérogènes à une autre échelle. Il s'avère donc nécessaire de maîtriser les variabilités de petite échelle dans les modèles utilisés, soit de manière explicite (modélisations emboîtées), soit sous forme paramétrisée.

Afin d'être en cohérence avec les spécificités de la pêche à la drague et de limiter les erreurs liées à une perte de cohérence des données intégrées à différents niveaux scalaires, deux niveaux d'échelles cohabitent au sein du modèle :

- un niveau élémentaire, composé de quadrat[iii] dont la surface est déterminée en fonction du type de pêche. Cette échelle précise permet de restituer une dynamique des espèces benthiques cohérente avec les mesures réalisées in situ par les biologistes;
- un niveau agrégé, correspondant à la zone de pêche qui permet de restituer une dynamique globale en relation avec les stratégies de pêche.

Les agents pêcheurs interagissent avec le niveau élémentaire. Compte tenu du nombre d'objets à manipuler par le modèle (plusieurs millions de quadrats à chaque tour) seul le niveau agrégé est représenté spatialement. Malgré cette précaution les temps de calcul restent très chronophages (environ trois semaines pour simuler 10 ans de pêche pour deux espèces cibles).

4. Simulation

A l'initialisation du modèle, chaque zone de pêche est divisée en un ensemble de quadrat dont la dimension est fonction du type d'engin et de l'espèce ciblée. Dans le cas de la coquille Saint-Jacques cette valeur est de 1.80 mètre ce qui correspond à la largeur réglementaire d'une drague.

La phase de simulation se déroule en deux temps :

- dans un premier temps le modèle reconstitue un gisement partir de campagnes d'ensemencements de juvéniles (réalisées par l'AgentComitedesPêches). Afin de disposer d'une ressource benthique hétérogène, en termes de classe d'âge et de taille des individus, la simulation démarre avant la date choisie de manière à

construire un stock représentatif des conditions réelles (8 ans avant dans le cas de coquilles Saint-Jacques. Durant, cette période, déterminée par rapport au cycle de croissance des espèces (Emery, 2008), aucun prélèvement n'est réalisé et seule la mortalité naturelle, la prédation ou les mortalités accidentelles (pollutions) régulent la population des individus.

- une fois cette phase achevée, les AgentsPêcheurs choisissent une zone de pêche en fonction des conditions d'environnement (état de la mer, coefficient de marées) et de l'effort de pêche consenti au tour précédent (rapport prises/temps de pêche). Lorsque plusieurs zones de pêche offrent les mêmes caractéristiques, ce choix s'effectue de manière aléatoire. Une fois la zone pêche déterminée, chaque agent tire au sort un nombre de quadrats calculé en fonction du temps de pêche autorisé et des contraintes techniques de la pratique (vitesse du navire, durée du trait de drague). Cette liste de quadrats détermine un potentiel de capture lié au nombre et à la taille des individus présents dans chaque quadrat. En fonction du temps de pêche autorisé (fixé chaque année par l'AgentComitedesPêche) et du tirage de quadrats réalisé, chaque AgentPêcheur obtient un résultat de pêche très variable. Cette variabilité, rapportée au temps consacré à la capture, donne l'effort de pêche pour chaque journée.

Les résultats obtenus permettent de réaliser des bilans de capture et d'évaluer les changements d'intensité de la pression sur les ressources benthiques. Ils peuvent être agrégés à différents niveaux d'échelles spatiales (quadrat, zone de pêche, rade de Brest) et temporelles (journée, semaine, mois, année). L'examen des données obtenues montre une forte variabilité spatio-temporelle de cette pression à l'intérieur de la rade de Brest (cf. figures 3 et 4). Confrontée à la politique de réensemencement et de mise en réserve (délimitation d'une zone interdite à la pêche pendant un temps déterminé afin de permettre un renouvellement du stock) pratiquée par le comité local des pêches, cette information fournie des éléments d'aide à la décision sur les choix réglementaires en terme de zonages et de périodes d'ouverture des campagnes de pêche. La validation de ces résultats reste cependant difficile car les bilans de capture officiel totalisent les prises débarquées en criée et n'intègrent pas les ventes directes (elles peuvent représenter jusqu'à 30% des prises certaines journées).

Afin de combler ces lacunes et d'identifier des seuils de développement pérennes de l'activité de pêche à la drague à la coquille Saint-Jacques au regard des ressources benthiques disponibles et du coût de la politique de réensemencement qui pèse de manière importante dans le prix de la licence de pêche, une collaboration avec le comité local des pêches est mise en œuvre. Cette démarche se place dans la perspective de proposer une approche rétrospective visant à reconstituer l'évolution des stocks de coquilles Saint-Jacques en tenant compte des facteurs environnementaux et des changements d'intensité de la pêche à la drague. En complément une approche prospective visant à une optimisation de l'activité (obtenir le meilleur rapport entre nombre de licences, quotas, stocks disponibles et rendements de la pêche) est envisagée.

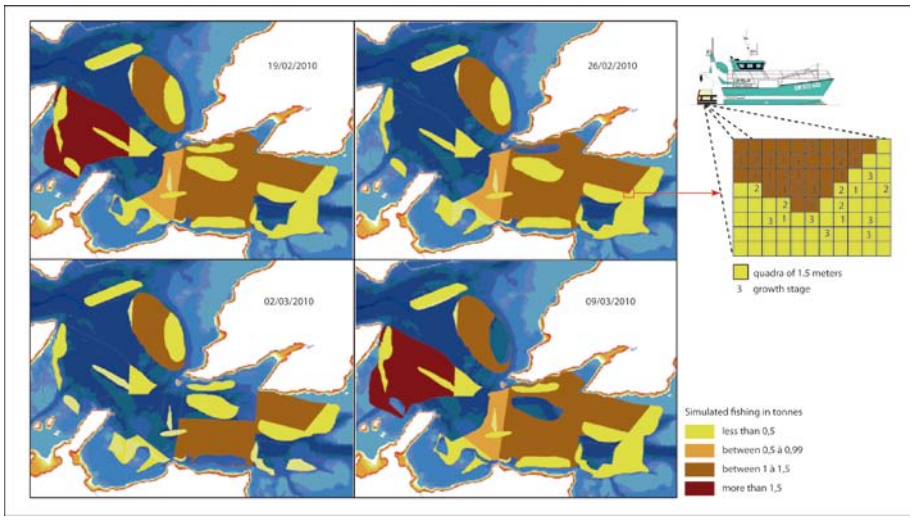


Figure 3 : variabilité de la pêche à la drague à l'échelle journalière en rade de Brest

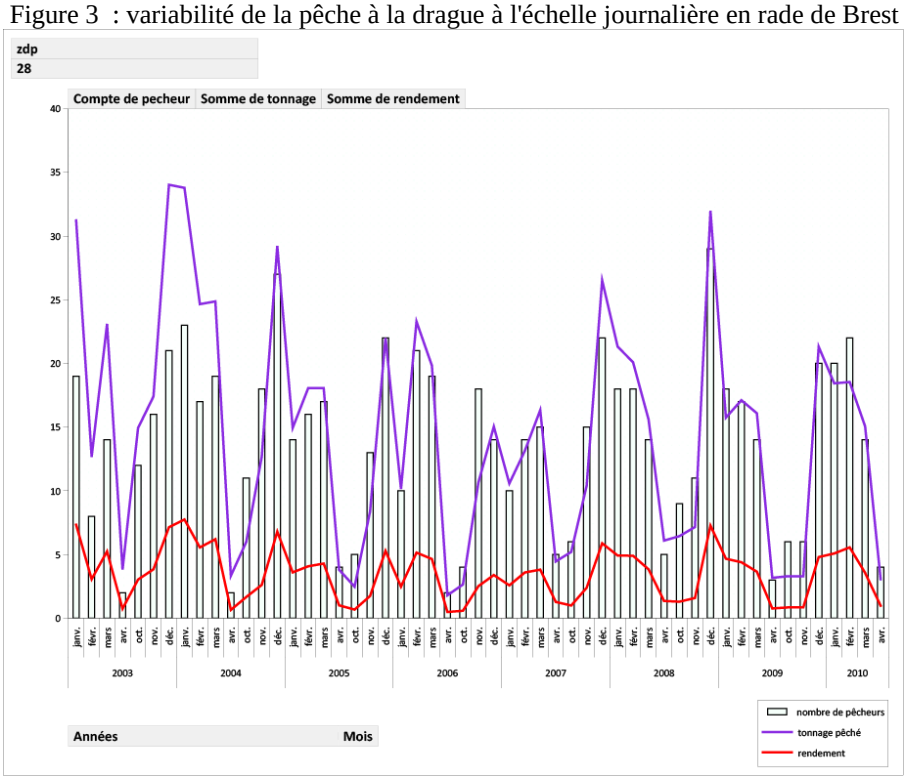


Figure 4 : variabilité inter-annuelle de la pêche à la drague à la coquille Saint-Jacques à l'échelle d'une zone de pêche

5. Conclusion

L'approche présentée dans cet article se place en rupture par rapport aux démarches classiquement mises en œuvre en modélisation multi-agents. L'espace, souvent considéré comme un support de modélisation, est ici traité comme une variable de forçage exogène calculée par combinaison de données spatio-temporelles. L'environnement des agents est construit à chaque pas de temps de simulation. Leur perception de cet environnement est donc évolutive en fonction de leur profil.

Sur le plan méthodologique, le modèle développé formalise les relations spatio-temporelles entre des entités spatiales, des processus environnementaux et des activités anthropiques.
Sur le plan thématique, les développements réalisés en Rade de Brest donnent des résultats encourageant sur la capacité du modèle à restituer le fonctionnement des activités de pêche à la drague en interactions avec des écosystèmes benthiques perturbés. Les changements d'intensité de ces activités s'avèrent particulièrement intéressant à étudier car ils sont traduisent la variabilité de ces interactions.

Les perspectives de développement du modèle sont nombreuses. Elles concernent principalement l'amélioration de la restitution des interactions entre les différentes classes d'agents et l'introduction de boucle de rétroaction dans la chaîne de simulation (variables de sorties d'un pas de temps réintroduites comme variables de forçage à t+1). Cette évolution permettrait de positionner la stratégie de l'agent en fonction de conditions antérieures et de tester les points de rupture de l'activité de pêche. Le contrôle de ces paramètres améliorerait sensiblement la portée du modèle auprès des pêcheurs et favoriserait le développement d'une approche prospective abordant la gestion de la pêcherie sur le long terme.

5. Bibliographie

Allen J.F., « Time and time again : the many ways to represent time ». *International Journal of Intelligent Systems*, vol 6 : pp. 341-355, 1991.

Cheylan J.-P., Lardon S., Mathian H., Sanders L., « Les problématiques liées au temps dans les SIG. *Revue Internationale de Géomatique* », vol 4 : pp. 287-305, 1994.

Claramunt C., Parent C., Spaccapietra S., Thériault M., « Database Modelling for Environmental and Land Use Changes ». In : Openshaw S. Geertman S., Stillwell J., (coord) - *Geographical Information and Planning : European Perspectives*. Springer-Verlag, pp. 173-194, 1999.

Duboz R., 2004. Intégration de modèles hétérogènes pour la modélisation et la simulation de systèmes complexes. Application à la modélisation multi-échelles en écologie marine. Université de Calais, 230 p.

Emmery A., 2008. Modélisation de la croissance et de la reproduction de la coquille Saint Jacques *Pecten Maximus* selon la théorie « Dynamic Energy Budjet » : variabilité environnementale et croissance individuelle. *Rapport de stage de Master II en Sciences Biologiques Marines*, Institut Universitaire Européen de la Mer, Laboratoire des Science et de l'Environnement Marin.

Fournier S., Devogele T., Claramunt C., 2003. "A role-based multi-agent model for concurrent navigation systems". *6th AGILE Conference on Geographic Information Science*. Presse polythechniques et universitaires romandes, p. 623-632.

Goralski R.I., Gold C.M., 2007. "The development of a dynamic GIS for maritime navigation safety". *ISPRS Workshop on Updating Geo-spatial Database with imagery & The 5th ISPRS Workshop on DMGISs*, p. 47-50.

Gotway C.A., Young L.J., 2002. Combining incompatible spatial data, Journal of the American Statistical Association 97, pp. 632-648.

Gould S.J., Time's Arrow,Time's Cycle: Myth and Metaphor in the Discovery of Geological Time. Harvard University Press, Cambridge, 222 p., 1987.

Holligan P.M., 1994. *Land Ocean Interaction In the Coastal Zone (LOICZ): Implementation Plan*. IGBP, 215 p.

Howlett E., 1997. "Environmental and geographical data management tools for oil spill modelling applications". *20th Artic and Marine Oilspill Program (AMOP) technical seminar*, vol. 2. Environment Canada, Vancouver, p. 893-908.

Legay J.-M., « Les temps de l'environnement ». *In :* Barrué-Pastor M. and Bertrand G. (coord) - *Les temps de l'environnement*. Presses Universitaires du Mirail, Toulouse, pp. 19-32, 2000.

Le Tixerant M., Gourmelon F., Tissot C., Brosset D., 2012, Modelling of human activity development in coastal sea areas. Journal of Coastal Conservation Volume 15, Number 4, pp. 407-416

Maury O., Gascuel D., 1999. "SHADYS (Simulateur HAlieutique de Dynamiques Spatiales), a GIS based numerical model of fisheries. Example application : the study of a marine protected area". *Aquatic Living Ressource*, volume 12, p. 77-88.

Muxart T., Blandin P., Friedberg C., « Hétérogénéité du temps et de l'espace : niveaux d'organisation et échelles spatio-temporelles ». *In :* Jollivet M. (coord) - *Sciences de la nature, sciences de la société. Les passeurs de frontières*. CNRS, Paris, pp. 243-258, 1992.

Parent C., Spaccapietra S., Zimanyi E., « Spatio-Temporal Conceptual Models : data structures + space + time ». Actes du Colloque *Advance in GIS*, Kansas City, november 5-6, pp. 26-33, 1999.

Pelot R., Plummer L., 2008. "Spatial analysis of traffic and risks in the coastal zone". *Journal of Coastal Conservation*, vol. 12, p. 201-207.

Pelekis N., Theodoulidis B, Kopanakis I., Theodoridis Y., 2005 - Literature Review of Spatio-Temporal Database Models, The Knowledge Engineering Review Journal, 19(3), 235-274.

Pereira, A.; Reis, L. & Duarte, P. EcoSimNet: A Multi-Agent System for Ecological Simulation and Optimization. *Progress in Artificial Intelligence, Springer,* **2009**, 473-484

Snodgrass R.T., « Temporal Databases ». *In :* Campari I. Frank A., Fromentini O., (coord) - *Theories and methods of spatio-temporal reasoning in geographic space*. Springer-Verlag, pp. 22-64, 1992.

Stonebraker M., Rowe L., Hirohama M., « The implementation of POSTGRES ». *IEEE Transaction of Knoledge and Data Engineering*, vol 2 : pp. 125-142, 1990.

Tang, W., 2008, Simulating complex adaptive geographic systems: A geographically aware intelligent agent approach. *Cartography and Geographic Information Science*, 35(4): 239-263

Tissot C., Brosset D., Rouan M., Le Tixerant M., Le Guyader D., 2013. Modeling Human Activities under Environmental Constraints. *Environemental Modelling and Software*, submitted.

Vallega A., 2003. "From Rio to Johannesburg: the role of coastal GIS". *CoastGIS'03*, Genova (Italia).

[i] Variable d'entrée contraignante pour l'agent

[ii] Un thread est un processus léger et autonome possédant des fonctions prédéfinies comme la synchronisation avec d'autres threads ou encore la possibilité d'utiliser des sections critiques permettant d'éviter les conflits de partage de sources de données communes.

[iii] Le quadrat est un carré ou rectangle de surface déterminée, classiquement utilisé pour échantillonner des données de terrain.

Simulation cartographique au service de l'innovation collective

Géoprospective territoriale à l'île de La Réunion

Augusseau X.[1], Aulanier F.[2], Bégué A.[3], Bourgoin J.[3], David D.[4], David G.[5], Degenne P.[3], Fargette M.[5], Lagabrielle E.[6], Lajoie G.[6], Laurent J.B.[2], Lebourgeois V.[7], Lestrelin G.[1], Lo Seen D.[3], Loireau M.[5], Tillard E.[8], Todoroff P.[2], Tonneau J.P.[3], Vitry F.[2]

1. CIRAD, UMR TETIS
Station de la Bretagne, BP 20, 97438 Saint Denis, La Réunion, France
xavier.augusseau@cirad.fr, guillaume.lestrelin@cirad.fr,
2. CIRAD, UPR SCA
Station de Ligne-Paradis, 7 chemin de l'IRAT, F-97410 Saint-Pierre, La Réunion, France
felicie.aulanier@cirad.fr, jean.baptiste.laurent@cirad.fr, pierre.todoroff@cirad.fr, francois.vitry@cirad.fr
3. CIRAD, UMR TETIS
Maison de la Télédétection, 500 Av. J-F Breton, F-34093 Montpellier, France
agnes.begue@cirad.fr , jeremy.bourgoin@cirad.fr, pascal.degenne@cirad.fr, danny.lo_seen@cirad.fr, jean-philippe.tonneau@cirad.fr
4. IRD, UMR ESPACE-DEV,
Parc Technologique Universitaire, 2 rue Joseph Wetzell, F-97492 Sainte-Clotilde, La Réunion, France
daniel.david@ird.fr
5. IRD, UMR ESPACE-DEV
Maison de la Télédétection, 500 rue Jean-François Breton, F-34093 Montpellier, France
gilbert.david@ird.fr, mireille.fargette@ird.fr, maud.loireau@ird.fr
6. *Université de La Réunion, UMR ESPACE-DEV*
Parc Technologique Universitaire, 2 rue Joseph Wetzell, F-97492 Sainte-Clotilde, La Réunion, France
erwann.lagabrielle@univ-reunion.fr, gilles.lajoie@univ-reunion.fr
7. CIRAD, UMR SELMET
Station de Ligne-Paradis, 7 chemin de l'IRAT, F-97410 Saint-Pierre, La Réunion, France
emmanuel.tillard@cirad.fr

2

8. CIRAD, UMR TETIS
Station de Ligne-Paradis, 7 chemin de l'IRAT, F-97410 Saint-Pierre, La Réunion, France
Valentine.lebourgeois@cirad.fr

RÉSUMÉ.
L'objectif de cet article est de présenter une démarche de construction d'un exercice de prospective territoriale basé sur un outil de modélisation spatiale. Cette démarche est mise en œuvre dans le cadre du projet ANR DESCARTES dont l'objectif est de construire un outil de simulation cartographique pour analyser différents scénarios d'affectation de l'usage des sols à l'Ile de La Réunion, en termes de services environnementaux. La plateforme de simulation cartographique est composée de deux applications complémentaires (i) le langage de modélisation Ocelet et son environnement de simulation de paysages dynamiques, et (ii) la plateforme Margouill@. Le développement puis la présentation, en atelier, d'un premier démonstrateur sur la consommation des terres agricoles par l'urbanisation a permis de tester l'outil cartographique comme support d'un processus d'innovation collective entre les parties prenantes, et d'ouvrir de nouveaux champs de recherche sur l'analyse de la démarche par les apprentissages, la spatialisation et la simulation prospective des services écosystémiques, et la prise en compte du changement d'échelle.

ABSTRACT.
The objective of this paper is to present an approach for experimenting territorial prospective analysis based on spatial modelling. This approach is carried out in the framework of the DESCARTES project which aims at developing a spatial simulation tool to support the design and analysis of different scenarios for land-use allocation in Reunion Island in terms of environmental services. The spatial modelling tool is composed of two complementary applications: (i) the Ocelet modelling language and its land dynamics simulation environment, and (ii) the Margouill@ platform. The first demonstrator, a model of farm land consumption by urbanization, was developed and presented during workshops in order to test the role of the spatial simulation tool in support of a collaborative innovation process among stakeholders, and to foster new research on social learning, spatial simulation of environmental services, and scale change issues.

MOTS-CLES : prospective territoriale, modélisation paysagère, services environnementaux, apprentissage collectif, Ile de la Réunion

KEYWORDS: scenario planning, landscape modeling, ecosystem services, social learning, Reunion Island

1. Introduction

Les processus de périurbanisation de l'agriculture se traduisent par une forte tension sur le foncier agricole et l'émergence d'une demande en agriculture « de services ». De nouvelles fonctions sont attribuées à l'agriculture, comme la gestion des aménités rurales et environnementales (Beuret et Cadoret, 2009). L'espace agricole devient ainsi le support de multiples fonctions (production primaire et services environnementaux) et l'objet de controverses entre acteurs urbains, environnementalistes et agricoles (Bertrand *et al.*, 2006). L'agriculture représente désormais l'enjeu de négociations qui dépassent largement le champ de la production, avec de nouveaux acteurs et instances de gouvernance (Jarrige *et al.*, 2006).

Ces nouvelles situations de gouvernance « partagée » rendent difficiles les choix et la recherche de consensus. Le recours à l'information et à la modélisation spatialisée et « l'usage raisonné des représentations spatiales » semblent être des voies privilégiées pour mettre en avant et débattre de différents points de vue et, potentiellement, dégager des consensus (Lardon et Piveteau, 2005 ; Maurel *et al.*, 2007). Ainsi, la mise en débat des services environnementaux est organisée autour de concepts tels que les « landscape services », censés permettre une meilleure compréhension des acteurs et créer un cadre unificateur pour les discussions entre décideurs, scientifiques et acteurs locaux autour de projets de territoire (De Groot, 2006 ; Termorshuizen and Opdam, 2009).

En lien avec ces concepts, la spatialisation des services environnementaux émerge comme une problématique importante (Kienast *et al.*, 2009). La cartographie des services écosystémiques s'attache ainsi à établir des relations entre unités de paysage (occupation/utilisation du sol), couplée à d'autres indicateurs (densité de population par exemple), et services environnementaux (type, quantité, qualité et distribution spatio-temporelle) (Eigenbrod *et al.*, 2010). Des plateformes de modélisation comme les projets InVEST (Integrated Valuation of Ecosystem Services and Tradeoffs) et MIMES (Multiscale Integrated Models of Ecosystem Services) sont également développées pour spatialiser des scénarios d'aménagement et réaliser des bilans en termes de services environnementaux (Golstein *et al.*, 2010).

Au-delà des concepts et outils de modélisation et d'évaluation, la diversification des questionnements autour de l'agriculture et la prise en compte d'une pluralité d'acteurs et de perspectives dans les négociations et la prise de décision, nécessitent également des innovations organisationnelles. Or, dans la littérature récente sur la modélisation paysagère (Nelson *et al.*, 2009 ; Willemen *et al.*, 2010), la participation des acteurs et le rôle joué par la plateforme de modélisation dans les processus de négociation et de construction de scénario ne sont souvent qu'évoqués. La démarche « participative » n'est pas véritablement construite et le rôle des modèles comme objets intermédiaires (Van Egmond and Zeiss, 2010) n'est pas réellement analysé.

Si l'utilisation de modèles comme objets intermédiaires peut jouer un rôle important dans la médiation et les processus d'apprentissage (Vinck, 2000 ; Antona

et al., 2005), il est donc nécessaire de promouvoir des démarches de co-construction itératives pour assurer une bonne appropriation des modèles et la mise en débat par les acteurs concernés (Van Ittersum *et al.*, 1998 ; Volk *et al.*, 2010).

Nous présentons ici les premiers résultats de la construction d'un dispositif de géoprospective qui mobilise un outil de simulation de dynamiques paysagères comme support d'un exercice participatif de prospective territoriale. Ce dispositif est élaboré dans le cadre du projet DESCARTES (ANR 11-AGRO-002-01).

2. La démarche du projet

2.1. La démarche globale

Dans un contexte insulaire, le projet DESCARTES vise à construire une démarche d'élaboration participative d'un diagnostic sur la place et la contribution de l'agriculture à des projets de territoire sur l'île de la Réunion (quel espace agricole ? pour quelle production ? pour quels services ?). Dans la perspective de développer des apprentissages, cette démarche intègre les partenaires institutionnels et la communauté scientifique sur l'ensemble de la mise en œuvre du projet. Le travail en partenariat s'appuie sur un outil de simulation cartographique visant à tester les « conséquences et cohérences », en termes de services (production et services environnementaux), de différents scénarios d'affectation de l'usage des sols dans un projet de territoire. Cet outil est prévu pour dialoguer avec des bases de données et des modèles, et pour être mobilisé comme le support de processus d'innovation collective entre les parties prenantes.

La mobilisation de modèles comme objets intermédiaires doit s'inscrire dans une démarche organisant les interactions entre scientifiques et acteurs concernés par le projet. Ces interactions sont considérées comme nécessaires à un processus d'apprentissage collectif, gage de l'utilisation des résultats du projet par les acteurs.

La première étape est la mise en place et la consolidation d'une arène qui garantisse la légitimité de l'exercice de prospective et en précise la finalité. Pour cela nous avons mobilisé un premier modèle démonstrateur (Figure 1). Les interactions avec les partenaires se font au cours de groupes de travail, la plupart sous la forme d'ateliers de modélisation, durant lesquels les modèles et les simulations sont discutés et consolidés.

La démarche commence par un exercice de prospective qui vise à définir collectivement des scénarios et les indicateurs pour les évaluer, et s'achève avec la présentation et la mise en débat des simulations des scénarios. Les scénarios consistent à explorer différentes hypothèses de mise en œuvre des différentes politiques publiques sur le territoire étudié. Dans les différents ateliers de modélisation, les différents scénarios sont mobilisés pour produire des modèles « thématiques » et paramétrer des simulations.

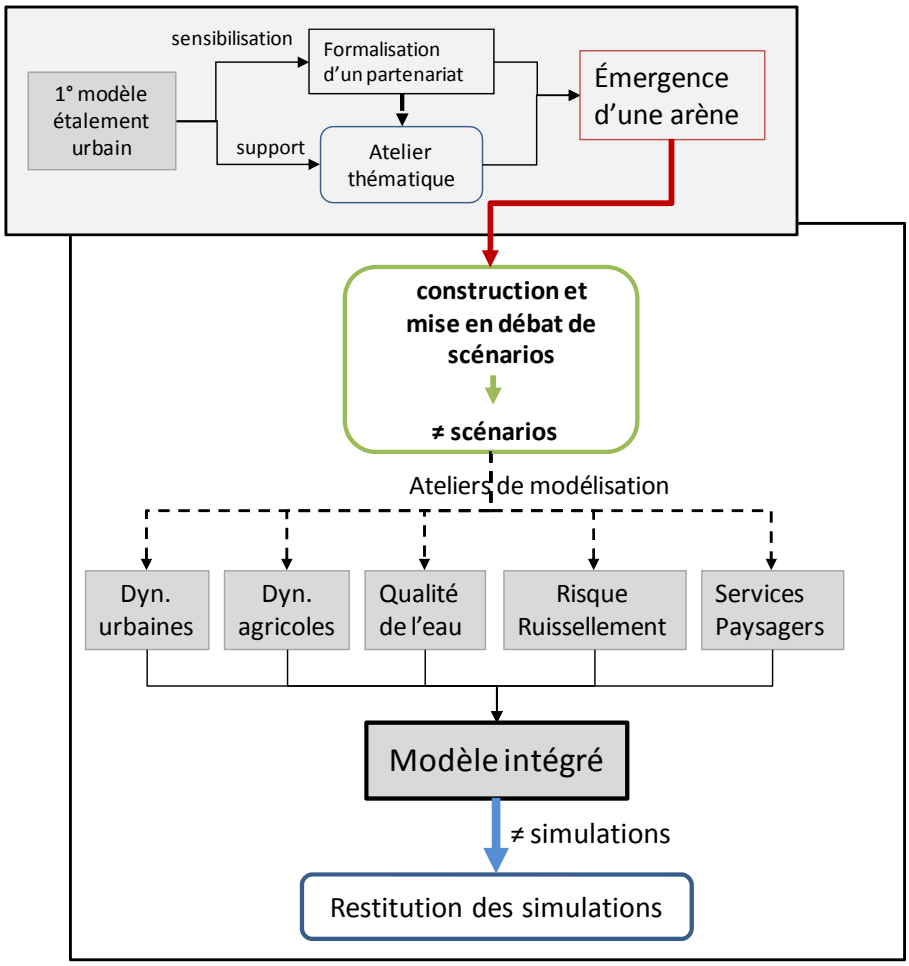


Figure 1 : Schéma général de la mise en place du dispositif partenarial de géoprospective.

2.2. L'approche de modélisation

L'outil de simulation cartographique proposé est basé sur des travaux de recherche récents menés par les équipes du projet, portant sur : (i) le développement d'un langage (Ocelet) et d'un environnement de modélisation de paysages dynamiques (projet ANR Blanc STAMP *Modelling dynamic landscapes with Spatial, Temporal and Multiscale Primitives*) (Degenne *et al.*, 2009 ; 2010 ; Degenne, 2012) et (ii) la constitution d'un système d'information agricole en ligne (plateforme Margouill@ ; www.margouilla.net) (Laurent *et al.*, 2011). Il est

construit pour hériter à la fois de la flexibilité de modélisation de l'approche Ocelet et de la convivialité de la plateforme Margouill@.

Employés comme objets intermédiaires, les modèles de simulation ont l'avantage de pouvoir intégrer un grand nombre de connaissances sur le fonctionnement des composantes du territoire et de montrer rapidement les évolutions possibles en fonction de différents scénarios. Elaborés au cours d'ateliers collectifs, ils permettent de faire émerger des représentations partagées et acceptées par les différents acteurs. En effet, si la construction de modèles et la simulation constituent bien souvent des exercices de réflexion et d'application peu compatibles avec une démarche participative (où le débat, parfois animé, est central), la flexibilité d'un langage de modélisation comme Ocelet permet de combler le décalage entre la nécessité des acteurs d'interagir en temps réel lors des arènes, et le temps nécessaire pour développer les simulations. A cet égard, un ensemble d'interfaces de représentation graphique des indicateurs (de services environnementaux notamment) est en cours de développement afin de faciliter la visualisation et la comparaison des simulations.

2.3. L'intégration des services écosystémiques dans le processus de planification spatiale

Le concept de service écosystémique est apparu à partir des années 1970 en économie de l'environnement, puis s'est imposé à partir des années 1990 comme un concept-outil de légitimation écologique et économique des politiques de conservation face aux pressions anthropiques.

L'intégration des services écosystémiques dans l'aménagement de l'espace est un nouveau défi (De Groot *et al.*, 2010). Ainsi, la spatialisation des services écosystémiques peut aider et faciliter leur appropriation par les décideurs avec l'appui des scientifiques (Fisher and Turner, 2008).

La spatialisation et l'évaluation des services écosystémiques dans le cadre du projet DESCARTES se fait en partenariat avec les gestionnaires, tout en assumant le rôle moteur (i.e. d'innovation) des scientifiques dans cette démarche (Lagabrielle *et al.*, 2010). Nous considérons que le concept de service écosystémique est structurant pour modéliser le fonctionnement biophysique d'un territoire à l'échelle du paysage, sans pour autant chercher à imposer a priori ce concept dans le débat d'aménagement. Aussi le concept « jargonneux » de service écosystémique est introduit dans le projet de manière implicite. L'équipe du projet joue ensuite le rôle d'intégrateur des connaissances et des données dans ce cadre conceptuel.

La démarche consiste (i) à catégoriser les services écosystémiques, (ii) à établir des associations entre des types d'occupation du sol et ces services à l'échelle du territoire, (iii) à identifier des bénéficiaires pour chaque type de service, puis (iv) à évaluer les impacts de différents scénarios prospectifs d'occupation du sol sur ces services.

3. situation étudiée

3.1. Le contexte de l'étude

Le projet est mené à La Réunion, petit espace insulaire de 2512 km^2 situé dans l'Océan Indien, qui se caractérise par une dynamique démographique et une économie qui doit composer avec de fortes contraintes d'espace. La gestion du foncier est source de divergences. Depuis deux décennies, les acteurs du monde agricole tentent de protéger les terres agricoles, progressivement déclassées par le développement urbain, alors que les collectivités territoriales cherchent à assurer les besoins économiques et sociaux d'une population croissante (dans la partie « utile » de l'île, les densités sont proches de 600 hab./ha). La création récente du Parc National de La Réunion (le cœur du parc couvre 40% de l'île) et d'une Réserve Naturelle Nationale Marine sur les récifs coralliens de la Côte Ouest, inscrivent dans l'espace géographique un projet de territoire insulaire orienté vers la conservation et la valorisation du patrimoine naturel.

L'espace agricole de l'île (environ 50 000 ha) est le support de fonctions de production « primaire », mais devient aussi le support d'une production potentielle de services environnementaux. Les différentes filières agricoles doivent ainsi réaliser leurs propres objectifs (production et surfaces), potentiellement concurrentiels, tout en contribuant au développement durable de l'île. Néanmoins les différentes politiques sectorielles se superposent sans qu'il n'y ait vraiment de lieux (forums) et d'outils pour diagnostiquer et débattre des interactions. Ainsi les différents diagnostics territoriaux des projets de planification (PLU[1], SCOT[2], …) présentent généralement une vision segmentée des différents enjeux du territoire, bien loin de l'intégration nécessaire.

3.2. Les sites d'étude

Le TCO a été choisi car il concentre des problématiques de l'île où les enjeux environnementaux sont particulièrement aigus, notamment avec la présence en aval d'un lagon, dont les ressources sont directement impactées par les activités anthropiques en amont et le cœur du Parc National dans les hauts. Le bassin versant choisi, futur bassin d'alimentation de captage Grenelle, se caractérise par des enjeux environnementaux importants et un exutoire qui se situe au cœur de la Réserve Marine dans le lagon.

Deux échelles emboitées sont prévues dans le projet : une communauté de communes, les Territoires de la Cote Ouest (TCO) et un bassin versant (Figure 2). Ces deux échelles sont traitées de façon homogène, avec la même approche basée sur des interactions entre entités spatiales à des échelles compatibles, mais

1 Plan Local d'Urbanisme

2 Schéma de COhérence Territoriale

exprimant des processus différents, pertinents à ces échelles, et avec des acteurs/décideurs qui ont des niveaux d'implication plus ou moins forts selon le territoire de gestion considéré. L'échelle de la communauté de communes est le niveau où l'on teste des scénarios « stratégiques ». Ces scénarios servent de forçage à l'échelle locale, comme lorsque par exemple l'application stricte d'un Instrument d'Action Publique (IAP) induit un changement de pratiques culturales dans certaines parcelles situées en zone d'alimentation de captage d'eau.

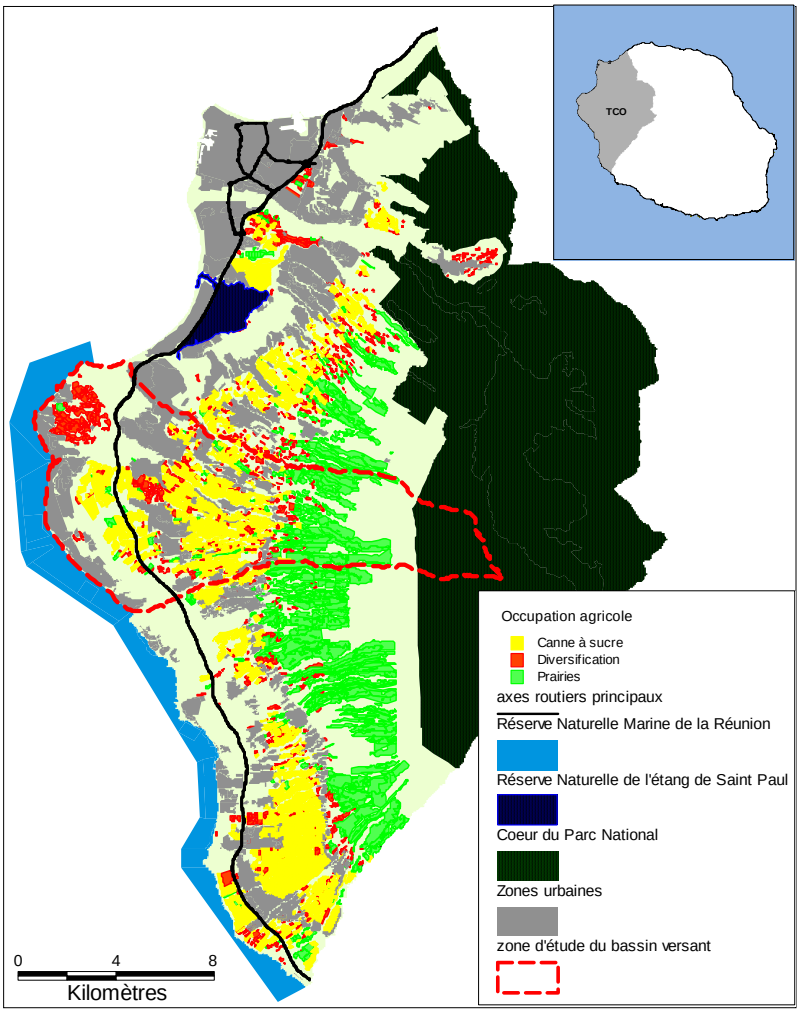


Figure 2 : Carte de localisation du TCO (Territoires de la Cote Ouest) à La Réunion et organisation du territoire.

4. Mise en place de la démarche et premiers résultats

4.1. Développement d'une plateforme de simulation

La plateforme de simulation cartographique se compose de deux applications complémentaires. La première hérite de l'environnement de modélisation Ocelet (Degenne *et al.*, 2009, 2010 ; Degenne, 2012) basé sur la manipulation de graphes (objet mathématique constitué d'un ensemble de nœuds et d'un ensemble d'arcs reliant certains des nœuds) qui sont employés ici de manière innovante. Avec le concept de graphe d'interaction, où les entités du modèle sont les nœuds du graphe et les fonctions d'interaction sont portées par les arcs, il est possible de modéliser avec une grande liberté d'expression une portion de territoire et de simuler son évolution dans le temps. L'outil se présente comme un environnement logiciel avec les fonctions nécessaires pour construire des modèles et des simulations. Cette application est mobilisée au cours d'ateliers pour simuler des scénarios.

La deuxième application est un enrichissement des fonctionnalités de Margouill@ (Laurent *et al.*, 2011) et s'utilise en ligne (sur un serveur web). Connecté à plusieurs bases de données spatiales, le portail Margouill@ met l'accent sur l'ergonomie, la convivialité et la facilité d'utilisation et intègre un outil cartographique en ligne respectant les standards de l'OGC (*Open Geospatial Consortium*).

Un modèle généré par Ocelet est interfacé à Margouill@ à trois niveaux :

- Margouill@ intègre une interface de paramétrage du modèle qui est générée de façon dynamique « à la volée ». L'utilisateur peut ainsi manipuler les paramètres d'entrée du modèle et tester plusieurs simulations sans avoir à intervenir sur la configuration de celui-ci (Figure 3).
- Lors de l'exécution du modèle, celui-ci utilise les données agro-environnementales stockées dans la base de données PostgreSQL/Postgis de Margouill@.
- Les résultats de simulations sont principalement des cartes animées au format KML de Google Earth. Ces cartes sont intégrées au site de Margouill@ à l'aide d'un plugin de visualisation.

L'outil DESCARTES en ligne permet de lancer des simulations avec les modèles construits par la première application, de faire varier les paramètres des modèles (Figure 3), de visualiser les évolutions des différentes simulations et d'évaluer les résultats.

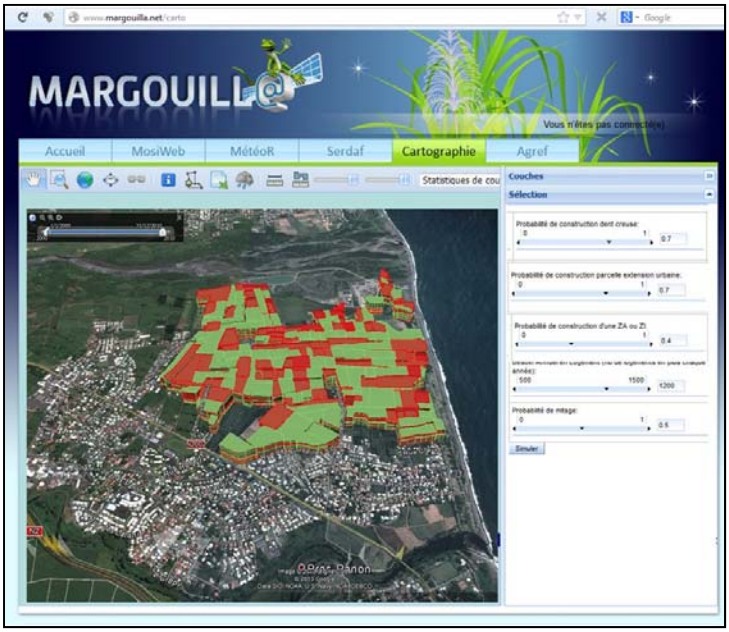


Figure 3 : Prototype de la plateforme Margouill@-Ocelet utilisant le plugin Google Earth pour afficher les résultats de sortie.

4.2. Construction d'un modèle de consommation des terres agricoles par l'urbanisation

En croisant l'expertise des partenaires du projet DESCARTES sur les processus d'urbanisation à La Réunion (Lajoie, 2007 ; David, 2010 ; David *et al.*, 2012) et les connaissances issues des interactions avec les acteurs de l'aménagement sur le TCO, nous avons entrepris la co-construction d'un modèle permettant de simuler l'évolution de l'urbanisation et le processus de consommation des terres agricoles sur ce territoire. La version courante de ce modèle est le fruit d'évolutions successives. Un premier prototype – simplifié et volontairement non calibré – a ainsi été développé entre chercheurs dans le but explicite d'exposer au TCO notre savoir-faire en matière de modélisation et de simulation cartographique sur une zone test (d'échelle communale). Suite à ce prototype, nous avons enclenché avec les équipes du TCO la démarche de modélisation qui a permis d'aboutir à un modèle plus complet dans lequel nous avons conceptualisé, implémenté et couplé deux dynamiques : l'urbanisation planifiée et l'urbanisation non-contrôlée.

Plusieurs entités spatiales multi-échelles sont représentées et manipulées dans ce modèle, en particulier le TCO lui-même, ses pôles urbains, et ses parcelles cadastrales. Chaque année, dans la dynamique d'urbanisation planifiée, une partie des parcelles cadastrales qui se trouvent au sein des zones d'extension urbaine

définies dans les documents de planification (notamment le Plan Local de l'Habitat du SCOT) s'urbanise en fonction des objectifs de nombre et de densité de logements associés à chacun des pôles urbains (à l'horizon 2020). En fonction des contraintes et des scénarios de simulation, la part des besoins en logement qui n'est alors pas satisfaite vient enrichir la dynamique d'urbanisation non-contrôlée. Dans cette dernière, des maisons sont construites sur des parcelles cadastrales se situant en dehors des zones urbaines règlementaires mais qui possèdent un fort potentiel en raison de leur attractivité et de leur proximité au réseau routier, aux pôles urbains et/ou aux zones historiques de mitage des terres agricoles.

Les résultats des simulations issues de cette version du modèle ont été exportés sous forme de fichiers *kml* temporels pour comparer les évolutions spatiales associées à des scénarios d'aménagement tests (Figure 4). Ces premiers résultats ont étés présentés aux élus du TCO ainsi qu'à un panel des acteurs de l'aménagement et les discussions engendrées ont permis d'enrichir notre modèle conceptuel, que ce soit au niveau des critères à prendre en compte dans les processus (par exemple l'ensemble des types de réseaux, qu'ils soient routiers, de distribution en eau, d'assainissement, ...), au niveau de l'identification de certains facteurs clés (la disponibilité des financements, les effets des lois de défiscalisation, ...), ou au niveau même de la caractérisation des dynamiques d'urbanisation (« renouvellement urbain » et « extension urbaine » d'une part, « mitage traditionnel » et « mitage vases-communicants » d'autre part). Le modèle conceptuel obtenu – qui servira de base à la prochaine version du modèle implémenté – est ainsi le fruit d'une démarche progressive de co-construction et d'apprentissage collectif.

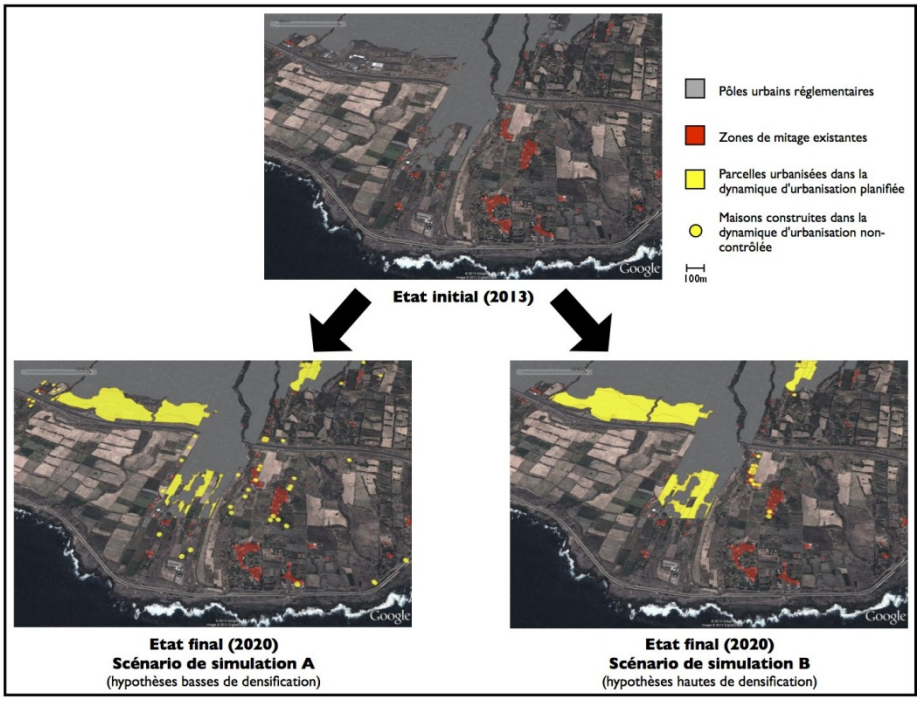


Figure 4 : Représentation des dynamiques urbaines.

4.3. Emergence d'un dispositif de géoprospective

L'émergence du dispositif de géoprospective est structurée par différentes étapes et changements d'état dans les relations entre acteurs du projet. Dans un premier temps, le développement d'un modèle démonstrateur sur l'étalement urbain et la présentation de simulations cartographiques issues de ce modèle ont eu pour effet d'attirer l'attention et de sensibiliser l'équipe technique du TCO sur l'intérêt potentiel des méthodes et outils développés pour la résolution de questions de planification et prospective (Figure 1). Cette étape de sensibilisation a été facilitée par l'actualité interne de l'équipe technique du TCO, en phase stratégique de redéfinition des grandes orientations de son territoire (ex. validation du projet de SCOT et révision du SAGE[3]). Des groupes de travail sur la modélisation des dynamiques d'urbanisation et la restitution de simulations sous *Google Earth* ont notamment permis d'alimenter des discussions entre l'équipe scientifique du projet et l'équipe technique du TCO autour du projet de SCOT et des différentes orientations possibles de l'urbanisation en ciblant localement des incohérences ou des compromis potentiels. Suite à ces échanges, la collaboration a été formalisée par

3 Schéma d'Aménagement et de Gestion des Eaux

une charte de partenariat précisant la finalité du travail de prospective, les modalités d'interaction entre les acteurs du projet, et positionnant l'Intercommunalité du TCO au centre d'un dispositif de réflexion et de débat sur l'évolution du territoire.

Sur ces bases, le dispositif de géoprospective a pu être étendu pour intégrer un plus grand nombre d'acteurs et répondre à des besoins stratégiques. Ainsi, dans le cadre de ses activités de planification et de prospective, l'Intercommunalité souhaite développer une charte rurale et identifier des orientations possibles pour une agriculture en cohérence avec le projet de territoire général. L'exercice est cependant rendu difficile par un dialogue quasiment inexistant avec les institutions agricoles. Pour répondre à ces objectifs et enjeux, l'accent a donc été mis sur la dimension heuristique du travail de modélisation et le potentiel d'intégration et d'intermédiation associé aux modèles et simulations cartographiques. Les interactions entre aménageurs, acteurs du monde agricole et de l'environnement ont pu être multipliées et structurées par une série de groupes de travail et d'un atelier collectif organisés sur le thème des dynamiques historiques, des impacts et de la modélisation du mitage des terres agricoles. Ces activités ont notamment permis de dépasser les visions statistiques pour faire émerger des perspectives plus intégrées et territorialisées, de mettre en évidence des perceptions divergentes, d'analyser et de débattre collectivement du processus et de sa régulation. Elles ont aussi permis de faire évoluer rapidement le modèle d'étalement urbain vers une « meilleure » représentation[4] du mitage et de ses interactions avec d'autres dynamiques et variables territoriales.

5. Conclusion

Cet article présente la démarche de construction d'un exercice de prospective territoriale à la Réunion, dans le cadre du projet ANR DESCARTES. Le projet est en cours et les premiers résultats décrivent la construction des premiers modèles de simulation et leur utilisation pour mettre en place une arène qui vise à organiser les interactions entre l'équipe du projet et ses partenaires.

Une question importante soulevée par le projet DESCARTES concerne les impacts de la démarche de modélisation participative et prospective sur le renforcement des compétences et l'apprentissage. L'apprentissage collectif est notamment considéré comme un moyen d'augmenter les capacités d'adaptation des décideurs et gestionnaires et, partant, de renforcer la résilience des systèmes socio-écologiques (Armitage *et al.*, 2008 ; Reed *et al.*, 2010). Au travers d'enquêtes de diagnostic initial puis de suivi des activités du projet, l'équipe scientifique s'intéresse ainsi à l'émergence de nouvelles pratiques, normes et formes d'interaction sociale au sein du collectif de participants. Une analyse du dispositif de géoprospective comme « acteur-réseau » (Akrich *et al.*, 2006) visera notamment à

4 En accord avec Cash et al. (2003), l'idée de « meilleure » représentation est considérée ici comme intégrant trois dimensions : la crédibilité, la légitimité et la pertinence d'une représentation.

réintégrer la question des rapports de pouvoir au centre du questionnement sur l'apprentissage collectif et l'innovation territoriale (Fox, 2000), à évaluer la capacité de médiation d'objets intermédiaires, tels que les modèles et les scénarios, et, en particulier, à mettre en évidence leur rôle dans la production des savoirs et la reproduction ou la reconfiguration des rapports de pouvoir (Cooren, 2010). A cet égard, nos premières observations suggèrent que la démarche collaborative et les technologies de modélisation et de simulation spatiale employées ont d'ores et déjà contribué à structurer de façon significative les relations et interactions entre acteurs du territoire. En particulier, elles ont contribué à faire émerger un réseau de réflexion et de débat sur la prospective territoriale et à motiver le choix de l'Intercommunalité du TCO de se constituer en « organisation frontière », porteuse de l'arène du projet DESCARTES, à l'interface entre recherche et prise de décision.

Remerciements

Cette recherche est en partie financée par l'Agence Nationale de la Recherche à travers le projet DESCARTES (ANR 11-AGRO-002-01 ; http://pad.teledetection.fr/descartes).

Bibliographie

Akrich M., Callon M., Latour B. (2006). *Sociologie de la traduction. Textes fondateurs*. Presses des Mines de Paris, Paris.

Antona M., D'Aquino P., Aubert S., Barreteau O., Boissau S., Bousquet F., Daré W., Etienne M., Le Page C., Mathevet R., Trébuil G., Weber J. (2005). La modélisation comme outil d'accompagnement, *Natures Sciences Sociétés*, n°13, p. 165-168.

Armitage D., Marschke M., Plummer R. (2008). Adaptive co-management and the paradox of learning. *Global Environmental Change*, vol. 18, n° 1, p. 86-98.

Bertrand N., Souchard N., Rousier N., Martin S., Micheels M.C. (2006). Quelle contribution de l'agriculture périurbaine à la construction de nouveaux territoires: consensus ou tensions ? *Revue d'Économie Régionale et Urbaine*, n°3, p. 329-353.

Beuret J.-E., Cadoret A. (2009). Une gouvernance territoriale endogène de l'environnement : contours et enjeux. *3èmes journées de recherches en sciences sociales*, Montpellier.

Cash D. W., Clark W. C., Alcock F., Dickson N. M., Eckley N., Guston D. H., Jagger J.,et Mitchell R.B. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, n° 14, p. 8086–8091.

Cooren F. (2010). Ventriloquie, performativité et communication. Ou comment fait-on parler les choses ? *Réseaux*, vol. 5, n° 163, p. 33-54.

David D. (2010). *Prospective territoriale par simulation orientée agent.* Thèse de doctorat, Université de la Réunion.

David D., Gangat Y., Payet D., Courdier R. (2012). Reification of emergent urban areas in a land-use simulation model in Reunion Island. *ECAI workshop on Intelligent Agents in Urban Simulations and Smart Cities* (IAUSSC2012), Montpellier.

Degenne P. (2012). *Une approche générique de la modélisation spatiale et temporelle : application à la modélisation des paysages*. Thèse de Doctorat, Université Paris-Est, Marne-la-Vallée.

Degenne P., Ait Lahcen A., Curé O., Forax R., Parigot D., Lo Seen D. (2010). Modelling with behavioural graphs. Do you speak Ocelet? In: *proceedings of International Congress on Environmental Modelling and Software*. Ottawa.

Degenne P., Lo Seen D., Parigot D., Forax R., Tran A., Ait Lahcen A., Curé O., Jeansoulin R. (2009). Design of a domain specific language for modelling processes in landscapes. *Ecological Modelling*, vol. 220, n° 24 p. 3527-3535.

de Groot R.S., Alkemade R., Braat L., Hein L. and Willemein L. (2010). Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making. *Ecological Complexity*, n°7, p.260-272.

de Groot R. (2006). Function-analysis and valuation as a tool to assess land use conflicts in planning for sustainable, multi-functional landscapes. *Landscape and Urban Planning*, n°75, p. 175-186.

Eigenbrod F., Armsworth P.R., Anderson B.J., Heinemeyer A., Gillings S., Roy D.B., Thomas C.D., Gaston K.J. (2010). The impact of proxy-based methods on mapping the distribution of ecosystem services. *Journal of Applied Ecology*, n°47, p. 377-385.

Fisher B., Turner R.K. (2008). Ecosystem services: classification for valuation. *Biological Conservation*, n°141, p. 1167–1169.

Fox S. 2000. Communities of practice, Foucault and Actor-Network Theory. Journal of Management Studies, vol.37, n°6, p. 853-868.

Goldstein J., Caldarone G., Colvin C., Duarte T.K. et al. (2010). *The Natural Capital Project, Kamehameha Schools, and InVEST: Integrating Ecosystem Services into Land-Use Planning in Hawai*. Report for the Natural Capital Project Hawai Case for TEEB D2.

Jarrige F., Thinon P., Nougaredes B. (2006). La prise en compte de l'agriculture dans les nouveaux projets de territoires urbains. Exemple d'une recherche en partenariat avec la Communauté d'Agglomération de Montpellier. *Revue d'Économie Régionale et Urbaine*, n°3, p. 393-414.

Kienast F., Bolliger J., Potschin M., de Groot R.S., Verburg P.H., Heller I.,Wascher D., Haines-Young R. (2009). Assessing Landscape Functions with Broad-Scale Environmental Data: Insights Gained from a Prototype Development for Europe. *Environmental Management*, n° 44, p.1099-1120.

Lagabrielle E., Botta A., Daré W., David D., Aubert A., Fabricius C. (2010). Modelling with stakeholders to integrate biodiversity conservation into land-use planning – Lessons learnt in Réunion Island (Western Indian Ocean), *Environmental Modelling & Software*, vol. 25, n°11, p. 1413-1427.

Lajoie G. (2007) *Recherches en modélisation urbaine*, Rapport d'Habilitation à Diriger des Recherches, Université de la Réunion.

Lardon S., Piveteau V. (2005) Méthodologie de diagnostic pour le projet de territoire : une approche par les modèles spatiaux. *Géocarrefour*, vol. 80, n° 2, p.75-90.

Laurent J-B., Todoroff T., Mézino M., El Biache Y., Degenne P. (2011). Margouill@ : un système d'information géographique en ligne pour le pilotage de la production de canne à sucre. *Congrès STIC & Environnemen*t, St Etienne.

Maurel P., Craps M., Cernesson F., Raymond R., Valkering P., Ferrand N. (2007). Concepts and methods for analysing the role of Information and Communication tools (IC-tools) in Social Learning processes for River Basin Management. *Environmental Modelling & Software*, vol. 22, n°5, p. 630-639.

Nelson E., Mendoza G., Regetz. J., Polaski S., Tallis H., Cameron D.R., Chan K.M., Daily G.C., Goldstein J., Kareiva P.M., Lonsdorf E., Naidoo R., Ricketts T.H., Shaw M.R. (2009). Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the Environment*, vol.7, n°1, p. 4-11.

Reed M., Evely A., Cundill G., Fazey I., Glass J., Laing A., Newig J., Parrish B., Prell C., Raymond C., Stringer L. (2010). What is social learning? *Ecology and Society*, vol. 15, n° 4, http://www.ecologyandsociety.org/vol15/iss4/resp1/

Termorshuizen J.W., Opdam AE.P. (2009). Landscape services as a bridge between landscape ecology and sustainable development. *Landscape Ecology*, n° 24, p. 1037-1052.

Van Egmond C., Zeiss R. (2010). Modeling for Policy: Science-based models as performative boundary objects for Dutch policy making. *Science Studies*, vol. 23, n° 1, p. 58-78.

Van Ittersum M.K.,Rabbinge R., van Latesteijn H.C. (1998). Land Use Studies and Their Role in Strategic Policy Making. *Agricultural Systems*, Vol. 58, n°. 3, p. 309-330.

Vinck D. (2000). *Pratiques de l'interdisciplinarité. Mutation des sciences, de l'industrie et de l'enseignement*, PUG, Grenoble.

Volk M., Lautenbach S., van Delden E., Newham L., Seppelt R. (2010). How Can We Make Progress with Decision Support Systems in Landscape and River Basin Management? Lessons Learned from a Comparative Analysis of Four Different Decision Support Systems. *Environmental Management*, n° 46, p.834–849

Willemen L., Hein L., van Mensvoort M., Verburg P. (2010). Space for people, plants, and livestock? Quantifying interactions among multiple landscape functions in a Dutch rural region. *Ecological Indicators*, n°10, p. 62–73.

L'AIS : une donnée pour l'analyse des activités en mer

Damien Le Guyader[1], David Brosset[2]

1. LETG-Brest, Géomer, IUEM, Université de Bretagne Occidentale
Technopôle Brest-Iroise, Place N. Copernic, 29280 Plouzané, France
damien.leguyader@univ-brest.fr

2. IRENAV, École Navale Lanvéoc-Poulmic, 29240 Brest, France

RESUME. Cette contribution présente des éléments méthodologiques pour la description des activités humaines en mer dans une perspective d'aide à la gestion. Différentes procédures, combinant l'exploitation de bases de données spatio-temporelles issue de données AIS archivées à des analyses spatiales au sein d'un SIG, sont testées afin de caractériser le transport maritime en Mer d'Iroise (Bretagne, France) sur les plans spatiaux, temporels et quantitatifs au cours d'une année.

ABSTRACT. This paper presents methodological elements for the description of human activities in coastal sea for management support purposes. The objective is to identify marine transportation patterns in Iroise Sea (Brittany, France) over a period of one year by their spatial, temporal and quantitative characteristics. Using AIS spatiotemporal databases combined with GIS spatial analysis, different methods are tested.

MOTS-CLES: AIS, bases de données spatio-temporelles, SIG, transport maritime.

KEYWORDS: AIS, spatiotemporal databases, GIS, marine transportation.

1. Introduction

La concentration et la diversité des activités qui se développent en mer côtière exercent une pression croissante sur cet espace et les milieux associés générant des interactions parfois conflictuelles entre activités. La compréhension de ces interactions implique, entre autre, de disposer d'informations pertinentes permettant de caractériser la distribution spatio-temporelle de ces activités. La mise à disposition relativement récente de données issues de système d'identification automatique AIS (*Automatic Identification System*) archivées constitue alors une ressource très utile à ces fins. Destiné à la sécurité et la surveillance du trafic maritime, le système AIS produit un important flux de données en temps réel. Leur archivage et leur analyse imposent la mise en place de plateformes dédiées et le

recours à des protocoles de fouille de données. L'exploitation de ces données archivées, en particulier pour l'extraction de connaissances dans une perspective d'aide à la gestion, pose cependant de nombreuses questions organisationnelles, méthodologiques et éthiques.

Cette contribution s'inscrit dans un travail en cours[1] qui porte sur la documentation de l'acquisition de données AIS jusqu'à leur exploitation dans un cadre d'aide à la gestion. Il ambitionne d'apporter des réponses aux questions suivantes : quelles sont les modalités techniques de réception et de d'archivage des messages AIS ? Les données AIS archivées sont-elles accessibles? Quels sont les méthodes et outils mobilisables pour caractériser le transport maritime à partir de ces données ?

2. Les données AIS

L'Organisation Maritime Internationale impose depuis 2002 à certains navires de disposer d'un Système AIS (les navires d'une jauge brute égale ou supérieure à 300 tonneaux (tjb) qui effectuent des voyages internationaux ; les navires de charge d'une jauge brute égale ou supérieure à 500 (tjb) qui n'effectuent pas de voyages internationaux ; les navires de passagers, quelles que soient leurs dimensions). Cette disposition est complétée par une directive Européenne (Directive 2011/15/UE) concernant les navires de pêche professionnelle de plus de 15 mètres naviguant dans les eaux sous juridiction d'un état membre de l'Union Européenne.

Le système AIS repose sur la transmission automatisée par voie hertzienne (VHF) de messages pré-formatés et numérotés (27 messages primaires) entre navires et entre navires et centres de surveillance maritimes. Afin d'étendre la couverture de réception de l'AIS au-delà des 40 milles, le système de réception terrestre a tendance à être complété par un système satellitaire en orbite terrestre basse. Ces messages fournissent trois catégories d'informations: des informations statiques qui identifient le navire émetteur, des informations dynamiques qui rendent compte de sa position et de son déplacement et des informations spécifiques au trajet effectué. En France, le réseau de stations de bases étatique est principalement composé d'une cinquantaine de stations AIS déployées dans le cadre du projet SPATIONAV. Depuis 2011, le Centre d'Etudes Techniques Maritimes et Fluviales (CETMEF) dispose d'un système nommé ENVISIA (système de mesures ENVironnementales par Système d'Identification Automatique) qui permet l'archivage de données collectées à partir des stations AIS. Actuellement les données stockées concernent les stations déployées dans les CROSS (Centre Régionaux Opérationnels de Surveillance et de Sauvetage) de la Manche (Guichoux *et al*. 2011).

---

1 Ce travail est réalisé dans le cadre d'un projet de recherche sur la CARTographie des Activités HUmaines en mer côtière (CARTAHU) soutenu par la Fondation de France, et appliqué à la Zone Atelier Brest Iroise.

3. Caractérisation du transport maritime à partir de données AIS

3.1. Travaux réalisés

Dans le cadre du projet Rad2Brest (Gourmelon *et al*. 2013), nous avons mobilisé des données AIS fournie par l'IRENaV mises en forme dans une base de données spatio-temporelles pour caractériser le transport maritime de charge et de passagers en rade de Brest au cours d'une année. Le trafic maritime journalier est identifié, quantifié et cartographié suivant une méthode d'analyse en quatre étapes. Elle combine l'exploitation d'une base de données PostgreSQL/PostGIS et l'utilisation de *Geospatial Modelling Environment* (GME) (Beyer *et al.* 2010) permettant le calcul d'une enveloppe spatiale comprenant une part significative des trajectoires de navigation. Suite à l'intégration des données ponctuelles brutes dans une base d'informations (étape 1), les trajectoires quotidiennes de chaque navire sont construites (étape 2) de manière à identifier les couloirs de navigation en tant que zones de pratiques (étape 3). Les calendriers de pratiques (informations temporelles et quantitatives) ont été reconstitués par comptage du nombre de navires par jour pour chaque couloir de navigation (étape 4). La méthode est décrite en détail dans Le Guyader *et al*. (2011).

3.2. Travaux en cours

Les travaux en cours portent sur les étapes d'acquisition, d'archivage et d'analyse de données AIS.

3.2.1. Acquisition et archivage

Une première étape consiste à documenter les méthodes et les modalités techniques de réception et de d'archivage des messages AIS. Une description des méthodes permettant de convertir les données brutes reçues en des données ponctuelles utilisables est également réalisée. La seconde étape comprend la comparaison de deux sources de données AIS archivées : l'une à l'initiative d'université grecque (*Marine Traffic*) et la seconde à l'initiative de l'État français et porté par le CETMEF (ENVISIA). La comparaison porte sur les conditions d'accès (convention, dispositions financières, temps de mise à disposition), puis sur la nature des données mises à disposition (contenu, format, post-traitement avant mise à disposition).

Le délai constaté entre la première demande de fourniture des données couvrant la zone d'étude sur une période d'un an et leur réception s'étend sur 4 mois. Pour l'instant, en raison des conditions financières d'accès aux données, seules celles provenant de *Marine Traffic* ont été acquises (fourniture à titre gratuit pour une utilisation dans un cadre de recherche).

3.2.2. Analyse

Les données AIS acquises auprès de *Marine Traffic* concernant le transport maritime de charge et de passagers de mai 2011 à mai 2012, sont exploitées dans le but d'identifier les zones et des calendriers de pratiques sur le site d'étude. Pour cela différentes méthodes sont testées. La première porte sur la transposition d'une méthode employée en rade de Brest (Le Guyader *et al*. 2011) sur la mer d'Iroise. La seconde consiste à mobiliser un outil récemment mis à disposition par la *National Oceanic and Atmospheric Administration* (NOAA): *AIS data Handler*.

Les premiers résultats montrent que la transposition de la méthode développée en rade de Brest est peu contrainte par la gestion d'un plus gros volume de données. En revanche, elle nécessite d'être modifiée en raison du caractère plus « ouvert » du site d'étude. La sélection semi-automatique des trajets reliant un port à un autre, préalable à l'analyse spatiale permettant de discrétiser les couloirs de navigation, peut être réalisée pour le transport de passagers. Ceci ne convient pas pour une partie du transport de charge en raison de l'absence de port d'escale sur le site et du caractère plus diffus des trajets sur le plan spatial. Pour y remédier, une solution non-supervisée est envisagée.

4. Conclusion

La récente mise à disposition de données AIS archivées ouvre d'intéressantes perspectives pour la caractérisation des activités maritimes sur les plans spatiaux, temporels et quantitatifs. Toutefois, les résultats préliminaires présentés dans cette contribution indiquent que leur exploitation dans une perspective d'aide à la gestion de la zone côtière et du milieu marin, semble pour l'instant limitée en raison de contraintes d'accès, de la nécessité de gérer d'importants volumes de données et de la complexité de l'analyse qu'elles requièrent. L'enjeu consiste alors à identifier ou proposer des méthodes et outils permettant la production d'informations pertinentes dans un cadre de gestion et qui soit transférable hors de la sphère scientifique.

Bibliographie

Beyer H.L., Jenness J., Cushman S.A., (2010). Components of Spatial Information Management in Wildlife Ecology: Software for Statistical and Modeling Analysis. *Spatial Complexity, Informatics, and Wildlife Conservation*. Springer Japan, p. 245‑253.

Gourmelon F., *et al*. (2013). *Modélisation et scénarisation des activités humaines en rade de Brest*. Rapport définitif. MEDDE (Programme LITEAU III).

Guichoux Y., *et al*. (2011). ENVISIA – A scalable archiving system for AIS data storage. *Actes du colloque Global sharing of Maritime Data*, IALA , Paris.

Le Guyader D., Brosset D., Gourmelon F., (2011). Exploitation de données AIS (Automatic Identification System) pour la cartographie du transport maritime. *Mappemonde*, vol. 104, n° 4-2011. http://mappemonde.mgm.fr/num32/articles/art11405.html

Aide à la détection de routes maritimes avec l'algorithme Mean Shift

Romain GALLEN[1]

1. Centre d'Études Maritimes et Fluviales (CETMEF)
155 rue Pierre Bouguer, technopôle Brest-Iroise,BP5, 29280 Plouzané, France
romain.gallen@developpement-durable.gouv.fr

Résumé.
La détection de routes maritimes est un sujet d'intérêt pour l'analyse de risque. L'obligation faite aux navires à passagers, tankers et navires de gros tonnage d'emporter un transpondeur AIS (Automatic Identification System) a facilité ces dernières années l'obtention de grandes bases de données de trafic dans les mers côtières. C'est en se fondant sur l'utilisation de ces bases de données que certaines méthodes d'évaluation quantitatives du risque fondées sur l'analyse du trafic maritime ont été proposées. Elles utilisent en entrée des routes maritimes prédéfinies manuellement par un expert afin d'estimer les risques de collision le long de ces routes. Les routes maritimes et la topologie du réseau peuvent également être utilisées à l'étude des évolutions des déplacements à des fins d'analyse économique ou d'analyse des transports. Enfin, ces routes maritimes peuvent être utilisées pour des systèmes de simulation du trafic qui prennent en compte un réseau réaliste.
Afin d'identifier correctement les routes maritimes, un simple affichage des positions se montre insuffisant dans les zones de haute densité de trafic du fait de l'étalement des positions le long des routes et du nombre potentiellement élevé de points. Notre méthode est centrée sur la simplification de l'affichage de données de positionnement afin d'assister les experts lors de la définition des routes maritimes dans les systèmes d'information géographiques.

Abstract. The detection of maritime routes is a matter of interest for risk analysis. The mandatory requirement for passenger ships, tankers and large ships to carry an AIS (Automatic Identification System) device has simplified the collection of data representative of the maritime traffic in coastal seas. Based on these data, quantitative risk assessment methods linked to the maritime traffic have been proposed. In the domain of maritime safety, quantitative risk analysis methods based on traffic characterization use user-defined maritime routes as input before calculating collision risks. Such AIS datasets can serve as well to study the mobility trends from ships for economical or transportation purposes. In this case, user-defined maritime routes are used as a prerequisite to the study of the network topology and traffic characteristics. Maritime routes can also be needed as a prerequisite for maritime traffic simulation tools that account for a real maritime network

In order to properly identify maritime routes, the simple display of AIS positions proves to be of limited use in high density traffic zones because of the spreading of positions along the

maritime routes and the potential high number of points. Our method is focused on simplifying the display of large datasets of position data in order to assist humans when defining maritime routes in GIS systems.

Mots-clés : analyse de risque, sécurité maritime, routes maritimes, trajectoires, affichage, données ais

Keywords: risk analysis, maritime safety, maritime routes, trajectories, display, AIS data

1. Introduction

Afin d'améliorer la sécurité maritime, la préservation de la santé et de la sécurité des personnes, la protection de l'environnement et des biens, l'Organisation Maritime Internationale (OMI) (*cf.* OMI 2002) et l'Association Internationale de Signalisation Maritime et des Aides à la Navigation (AISM) (*cf.* AISM 2008) ont recommandé depuis des années que l'utilisation de méthodes d'évaluation formelle de la sécurité (Formal Safety Assessment, FSA) devrait être utilisées autant que possible. L'analyse de risque et les analyses coûts bénéfices permettant de réaliser de telles évaluation formelles se sont concentrées dans un premier temps sur les défaillances matérielles, humaines, et les situations d'urgence par l'utilisation d'arbres de défaillances, d'arbres d'événements (*cf.* Winter 1995) ou en incluant des analyses de fiabilité humaines (Human Reliability Analysis, HRA). Les évolutions récentes des FSA appliquées au domaine maritime se sont concentrées sur l'identification des zones présentant des risques de collisions afin d'adapter les mesures de sécurité telles que le balisage, les dispositifs de séparation du trafic (DST) ou toutes autres mesures de réductions des risques telles que le pilotage, les centres de surveillance et de sauvetage (Vessel Traffic Services, VTS), le positionnement des remorqueurs ou des dispositifs antipollution.

Ces processus de FSA appliqués à l'analyse du trafic maritime s'appuyaient principalement sur des analyses qualitatives de risque et le jugement d'experts lorsque les données sur le positionnement des navires étaient rares ou difficiles à collecter. L'obligation pour certains navires (*cf.* OMI 2001) et l'utilisation de plus en plus commune de transpondeurs AIS a permis de collecter de grandes bases de données géolocalisées et enrichies concernant les déplacement des navires dans les eaux côtières, améliorant la connaissance et la sécurité maritime (*cf.* Tsou 2010). Le développement récent de moyen satellitaires de détection de signaux AIS promet une couverture globale des transports maritimes dans le future.

Dans ces approches FSA, la détection des routes maritimes est souvent une tâche difficile pour les experts bien que de nombreux outils ou aides peuvent les aider dans cette tâche lors de l'affichage de données maritimes géolocalisées. Différentes techniques telles que l'affichage à grande échelle, les calculs de densité, l'utilisation d'échelles d'intensité logarithmiques, les calculs de densité par méthodes à noyaux, permettent de donner à l'utilisateur une meilleure perception de la localisation précise des routes (*cf.* BMT Isis 2009, Goerlandt et al. 2012, Marin 2012). Néanmoins, la détermination précise du positionnement des routes maritimes reste une tâche difficile pour les experts dans les zones à haute densité de trafic. Ce n'est qu'une fois que ces routes ont été définies qu'elles peuvent être utilisées pour la caractérisation du trafic à l'aide d'outils automatiques permettant d'estimer la distribution des vitesses, des types de navires, la dispersion des trajectoires autour

des routes ou les risques de collisions lorsque ces routes se croisent ou sont partiellement superposées (*cf.* Fujii et al. 1974, MacDuff 1974 et Montewka et al. 2010).

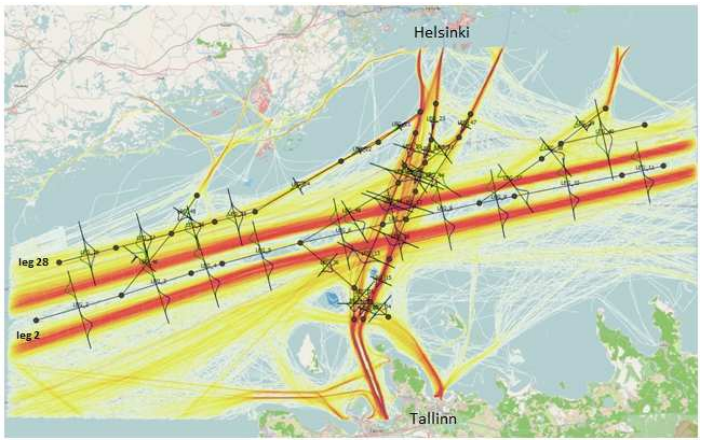


Figure 1. Utilisation de routes maritimes définies manuellement afin d'estimer le risque de collision à l'aide du logiciel de l'AISM, IWRAP MkII (cf. AISM 2009*) .*

Dans cet article, nous proposons d'utiliser une variante de l'algorithme Mean Shift destiné à simplifier l'apparence visuelle des routes maritimes tout en préservant la topologie du réseau de navigation. Mean Shift est un algorithme itératif puissant, adaptatif et non paramétrique qui peut être utilisé à des fins différentes telles que la fouille de données, la classification ou la segmentation de données multidimensionnelles, *etc*. Il a été proposé initialement par Fukunaga et Hostetler (1975) et a été étendu à d'autres domaines et en particulier au traitement d'images. Nous montrons qu'il peut être également appliqué à des données de trajectoires de navires afin d'aider à la détection de routes maritimes par les experts, en particulier dans les zones de haute densité.

Différentes méthodes de classification et de segmentation des routes maritimes ont été proposées par Andrienko et Andrienko (2013), Le Guyader *et al.*(2011) ou Etienne *et al.* (2010). Ces méthodes ne répondent que partiellement aux attentes des experts d'analyse de risque dans la mesure où elles sont semi-supervisées et nécessitent la définition manuelle préalable de zones de passage avant de traiter les trajectoires comme c'est le cas dans les travaux de Andrienko et Andrienko (2013) ou de Etienne *et al.* (2010). Elles ne peuvent donc être facilement être implémentées sur des larges étendues géographiques comme la haute mer où les positions AIS se font plus rares au fur et à mesure de la distance aux stations côtières. De plus sur des données d'emprise régionales, le nombre de zones à définir manuellement peut s'avérer très grand (supérieur à 100). Dans des travaux tels que ceux de Le Guyader *et al.* (2011), l'approche semi-supervisée (filtrage préalable de points statiques, sélection des classes d'intérêt avant affichage) et l'affichage de couloirs de

navigation larges représentatifs de la trace de 90% des trajectoires d'un cluster ne permet pas d'obtenir une définition fine et précise de la position des routes comme attendu dans un logiciel tel qu'IWRAP MkII. Ce manque de méthode non supervisée produisant des affichages fins des routes maritimes nous a amenés à envisager d'utiliser l'algorithme Mean Shift.

Cet article est organisé comme suit : nous présentons dans un premier temps l'algorithme Mean Shift dans son fonctionnement nominal. Ensuite nous présentons les adaptation nécessaires à l'application à des positions de navires, puis l'application à des données réelles et les résultats obtenus sont présentés et discutés. Enfin, nous concluons, présentons nos perspectives et proposons des pistes d'évolution permettant d'améliorer la caractérisation du trafic maritime et des routes maritimes en particulier.

2. L'algorithme Mean Shift

2.1. Fonctionnement nominal de Mean Shift

Mean Shift considère l'espace des caractéristiques comme une fonction de densité de probabilité empirique. Si les données d'entrées sont un jeu de points, celui-ci est vu comme la réalisation d'une densité de probabilité sous-jacente. Si des régions denses (groupes ou clusters de points) sont présentes dans l'espace des caractéristiques, ces régions sont considérées comme étant les modes (maximas locaux) de la densité de probabilité. À l'origine, Mean Shift était destiné à identifier des classes ou groupes associés à ces modes.

Le fonctionnement de Mean Shift est assez simple. Pour chaque point séléctionné, Mean Shift l'associe avec le mode le plus proche de la densité de probabilité du jeu de données. Pour chaque point, Mean Shift définis une fenêtre autour du point considéré et calcule la moyenne des points à l'intérieur de la fenêtre. On déplace alors le centre de la fenêtre vers cette moyenne et on itère la méthode jusqu'à convergence. Après chaque itération, on peut considérer que la fenêtre se déplace vers une région plus dense du jeu de donnés.

Algorithme 1. Processus du Mean Shift

1: ***MeanShift*** (*Portee*)
2: {
3: *PointsProches* $\leftarrow \varnothing$
4: *MoyenneLocale* $\leftarrow \varnothing$
5: $P_i' \leftarrow P_i$
6: **pour tout** $P_i \in$ (*Position AIS navire*) **faire**
7: *PointsProches* $\leftarrow \{P_j, \|\overrightarrow{P_i P_j}\| \leq Portee\}$
8: *MoyenneLocale* $\leftarrow$ *PointsProches*
9: $P_i' \leftarrow$ *MoyenneLocale*
10: **fin boucle**

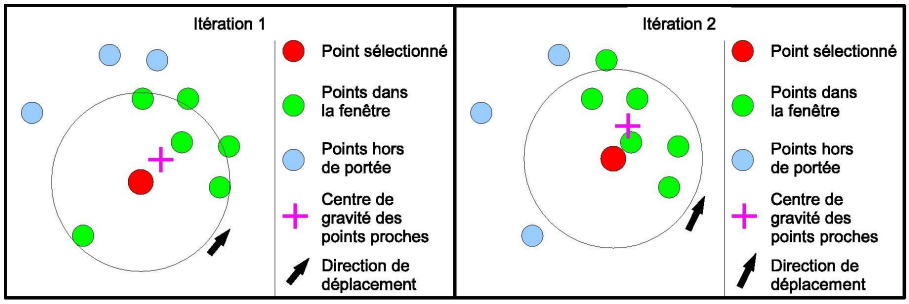


Figure 2. Fonctionnement de l'algorithme Mean Shift

Dans notre configuration, on ne peut considérer les trajectoires comme des réalisations d'un processus stochastique. Ces trajectoires sont le résultat de la volonté de se déplacer d'un port de départ vers un port de destination. Idéalement, tous les navires suivraient la trajectoire optimale en absence de contraintes environnementales (la plus courte en absence de tous facteurs externes tels que les courants ou les vents). Afin de pouvoir appliquer Mean Shift à ces trajectoires, nous faisons l'hypothèse que le positionnement latéral le long des routes maritimes peut être considéré sur une longue période de temps comme un processus aléatoire dépendant de nombreuses contraintes telles que l'environnement, la bathymétrie, les vents et courants, et tous les obstacles potentiels (balisage, autres navires, *etc.*). Notre objectif en appliquant Mean Shift à des trajectoires de navires est de réduire l'étalement latéral des positions le long des routes et ainsi faciliter le travail des experts lors de la définition du réseau de routes maritimes.

2.2. Adaptation de Mean Shift à des trajectoires

Afin de préserver la topologie du réseau, nous ne voulons pas que Mean Shift déplace un point vers le mode le plus proche de manière aveugle. En effet, les points entourant le point sélectionné peuvent souvent appartenir à d'autres routes dans les zones de haute densité de trafic. L'algorithme aurait alors tendance à réunir des points appartenant à des routes différentes sur un simple critère de distance sans aucune prise en compte du contexte de navigation.

Nous adaptons l'algorithme de telle sorte que pour chaque point sélectionné, il ne prenne en compte dans le processus de mesure de la moyenne et de déplacement que des points ayant un cap proche du cap du point considéré. Nous nous appuyons pour cela sur la donnée de Course over Ground (CoG) donnée par l'AIS et calculée d'après la route fond du navire. Ces points partageant un même CoG dans une zone d'étendue limitée ont plus de chances d'appartenir à une même route, il semble donc plus logique de les intégrer dans le processus de Mean Shift.

Algorithme 2. Mean Shift modifié pour des trajectoires

1: ***MeanShift*** (*Portee,DiffCap*)
2: {
3: *PointsProches* $\leftarrow \varnothing$
4: *MoyenneLocale* $\leftarrow \varnothing$
5: $P_i' \leftarrow P_i$
6: **pour tout** $P_i \in$ (*Position AIS navire*) **faire**
7: $PointsProches \leftarrow \{P_j, \|\overrightarrow{P_i P_j}\| \leq Portee, Cap(P_i)\text{-}Cap(P_j) \leq DiffCap\}$
8: $MoyenneLocale \leftarrow PointsProches$
9: $P_i' \leftarrow MoyenneLocale$
10: **fin boucle**

3. Application sur des données AIS réelles

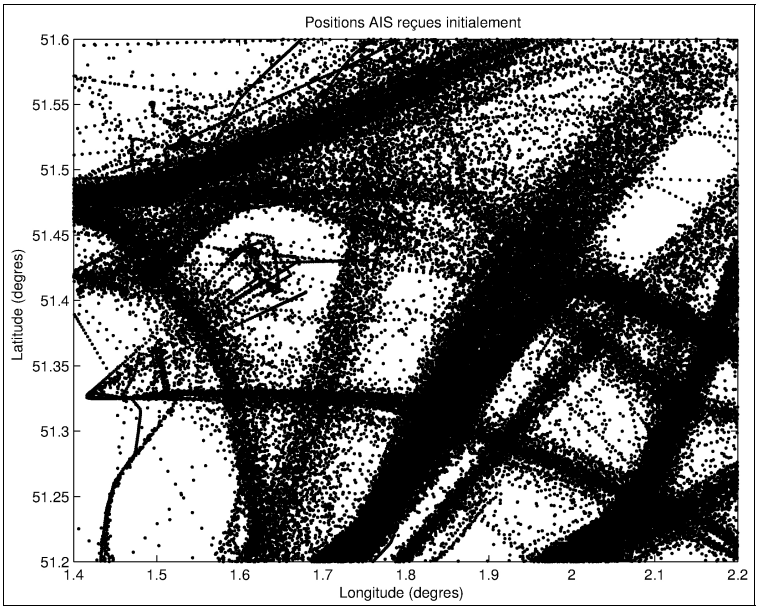


Figure 3. Positions AIS reçues par les stations de base

La figure 3 présente les données AIS originales, cumulées durant une semaine, telles que reçues par les stations de base AIS disposées sur le littoral. Dans cette zone à haute densité de trafic, même en utilisant des méthodes de prétraitement des données afin d'améliorer l'affichage, il est difficile de distinguer les différentes routes maritimes, comme illustré sur la figure 4. L'étalement et le recouvrement des positions rendent impossible une définition précise du réseau de route par un expert. Jouer sur la taille des cellules de calcul de densité ou en modifiant l'échelle d'intensité colorée donne ne donne qu'un faible gain qualitatif.

8

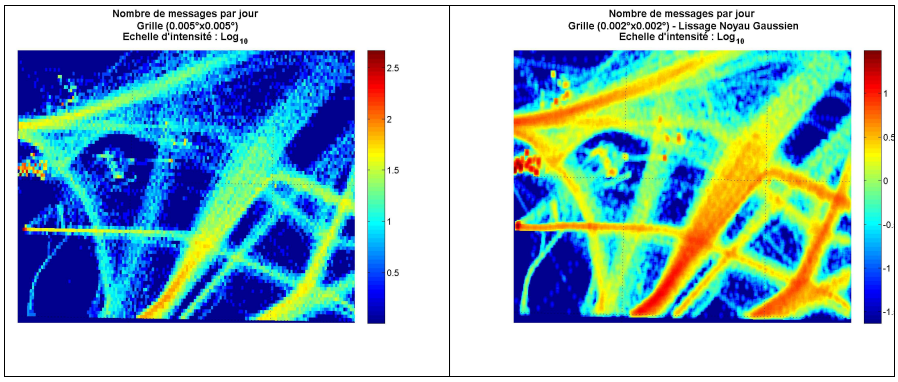


Figure 4. Cartes de densité, échelle d'intensité logarithmique, taille de cellule de 0,005 degré (à gauche), de 0,002 degré avec lissage par noyau gaussien (à droite)

Les résultats présentés sur la figure 5 sont en revanche plus simples à exploiter pour des experts souhaitant dessiner le réseau des routes maritimes. Les résultats présentés ont été obtenus en prenant une fenêtre autour des points de 1 mille nautique de rayon et une tolérance sur les caps voisins de 10 degrés. L'algorithme Mean Shift modifié préserve certaines caractéristiques particulières des routes initialement présentes qui n'apparaissaient pas sur les figures précédentes telles que des routes parallèles en sens opposé se superposant fortement qui ne pouvaient pas être différenciées au sein des larges routes visibles sur les figures précédentes.

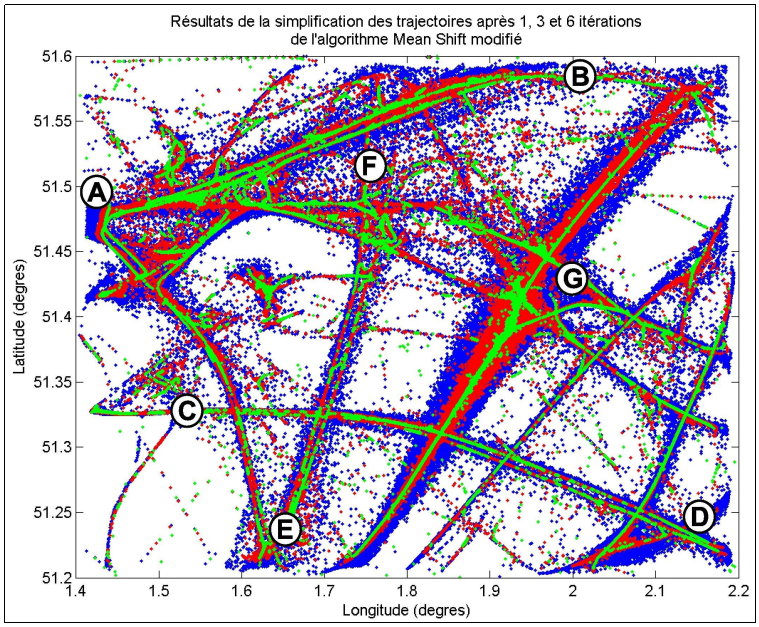


Figure 5. Positions après 1, 3 et 6 itérations de l'algorithme Mean Shift modifié

On notera en particulier que sur les routes AB et CD, on peut aisément distinguer la présence de deux routes parallèles, vraisemblablement de cap opposés. Elles apparaissent plus clairement que sur aucune des images précédentes (brute ou avec prétraitement). C'est le type d'information qui peut avoir une importance majeure lors de l'appréciation des risques de collisions (cas des chocs frontaux tels que définis dans Fujii *et al.* 1974). La zone G qui présente à la fois le cas d'un étalement particulièrement large des positions cumulé a un croisement de routes multiples illustre également très bien les apports de la méthode. D'une part on distingue le lieu de la route principale mais on voit également apparaître avec précision les points de croisement de routes orthogonales.

La route EF en revanche permet de voir deux routes distinctes après trois itérations de l'algorithme (en rouge) alors qu'avec six passes de l'algorithme sur les données, les routes sont assez déstructurées (en vert). Cette route illustre les limites de l'algorithme, celui-ci, après un grand nombre d'itérations aura tendance à regrouper en paquets compacts les points proches, le nombre de paquets et leur répartition dépend de la taille de la fenêtre de voisinage autour des points.

Il n'y a pas de méthode permettant à coup sûr de s'arrêter au bon nombre d'itérations, en particulier dans le cas de l'application de l'algorithme à des données de trajectoire. Fukunaga et Hostetler (1975) puis Comaniciu et Meer (2002) ont montré la convergence de l'algorithme dans des cas de classification pour un nombre de classes connues mais nous ne pouvons pas parler de classes dans cette approche fondée sur l'étude des trajectoires.

La figure 6 illustre néanmoins un moyen de fixer un seuil au nombre d'itérations de l'algorithme. Nous illustrons sur la figure 6 (gauche) le déplacement moyen des points après chaque itération. Celui-ci suit une progression régulière avant de marquer un plateau puis de reprendre sa progression. Nous proposons de s'appuyer sur cette évolution du déplacement moyen et d'arrêter les itérations ou de ne présenter que les n premières itérations en se fondant sur l'analyse de ces courbes.

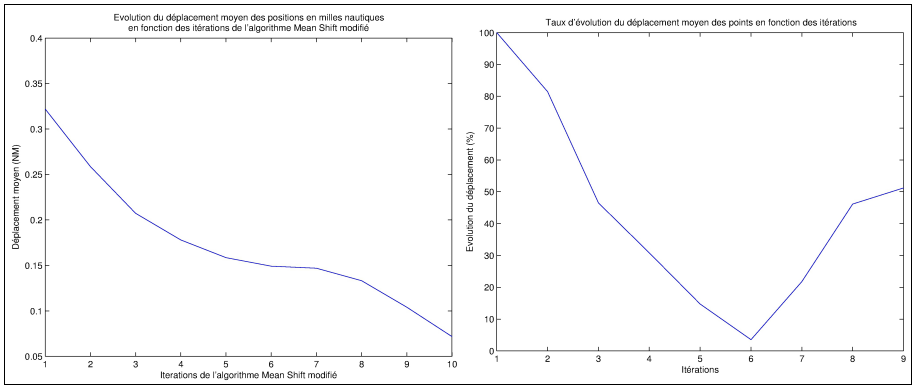


Figure 6. Déplacement moyen des points en fonction du nombre d'itérations (gauche) et taux de déplacement par rapport à la première itération (droite)

L'expérience nous a montré que lors des premières itérations les points au large des routes tendent à se rapprocher rapidement suivant une direction orthogonale à la route. Cela est dû au fait que la route constitue un pôle d'attraction local. Une fois sur la route en revanche, du fait que de nombreux points se situent de part et d'autre le long de la route, la moyenne d'un voisinage n'est plus orientée et le déplacement des points baisse. Enfin, des singularités (trous dans les observations ou densités locales supérieures) vont générer de nouveau des déplacements au loin ou vers ces singularités, provoquant une déstructuration progressive des routes et cet effet de mitage qui peut être observé sur certaines portions des routes vertes de la figure 5. D'après la figure 6, nous pouvons soit fixer le nombre d'itérations souhaité par un expert à la vue de l'évolution du déplacement moyen (Fig. 6 gauche), soit fixer un arrêt automatique de l'algorithme lorsque l'évolution du déplacement moyen à l'itération *n* ne représente plus qu'un déplacement marginal (moins de 10% ou moins de 5%) par rapport au déplacement de la première itération (Fig. 6 droite).

3. Conclusion et perspectives

L'algorithme Mean Shift présente un intérêt pour l'analyse des trajectoires de navires. Il peut être utilisé pour simplifier l'affichage des routes maritimes et ainsi faciliter le travail des experts lors de la définition de la topologie du réseau de route. Il est particulièrement utile dans les zones à haute densité de trafic et est robuste à une forte dispersion des points autour des routes maritimes principales.

Non seulement ses applications peuvent servir dans le domaine de la sécurité maritime mais également dans le domaine des transports et de l'économie des transports par l'analyse des réseaux et des déplacements. Enfin, il peut être utilisé dans le domaine de la simulation de trafic maritime pour l'optimisation des mesures de réduction des risques ou des stratégies de routage des navires.

Il présente néanmoins un défaut principal : sa complexité algorithmique en $O(n^2)$ le rend très lent à utiliser si ce n'est impraticable à très grande échelle. Des travaux tels que ceux de Comaniciu et Meer (2002), de Carreira-Perpinan (2006), de Yang *et al.*(2010) ou de Wang et Carreira-Perpinan (2010) visent à améliorer à la fois sa vitesse et son efficacité pour la segmentation et la classification dans des situations complexes.

Nous envisageons de chercher une implémentation de l'algorithme qui soit robuste à l'application aux données réelles et qui en outre permette de brider les problèmes de mitage et de sur-segmentation (apparition de paquets de positions) qui tend à donner des routes discontinues. Il nous faudra également évaluer la robustesse et la validité de la méthode d'arrêt automatique de l'algorithme fondée sur le déplacement moyen. Enfin, une analyse fine de la taille de fenêtre optimale pour l'application de cet algorithme à des données de trajectoire permettra probablement d'en faire un outil utilisé dans les analyses de trafic à l'avenir.

Bibliographie et références

Andrienko, N et Andrienko G. (2013), Exploring Traffic in Brest Harbor by Trajectory Aggregation, *MOVE-COSTS AIS Workshop*, Brest.

Association Internationale de Signalisation Maritime et des Aides à la Navigation (2008), *Risk Management*, Guideline 1018, Rev. 2.

Association Internationale de Signalisation Maritime et des Aides à la Navigation (2009), http://iala-aism.org/wiki/iwrap/index.php?title=Main_Page

BMT Isis (2009), UK/France *Formal Safety Assessment: Reducing Risk in the English Channel/La Manche Traffic Separation Schemes*, Tech. Rep.

Carreira-Perpinan, M. A.. (2006), Fast Nonparametric Clustering with Gaussian Blurring Mean-Shift, *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA, pp. 153-160.

Comaniciu, D. et Meer, P. (2002), Mean Shift: A Robust Approach Toward Feature Space Analysis, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.24, n°5, pp. 603-619.

Etienne L., et Devogele T. et Bouju A. (2010) Spatio-Temporal Trajectory Analysis of Mobile Objects following the same Itinerary, *Proceedings of the International Symposium on Spatial Data Handling (SDH)*, Hong Kong.

Fujii, Y. et Yamanouchi, H. et Mizuki, N. (1974) *Some factors affecting the frequency of accidents in marine traffic. ii: The probability of stranding, iii: The effect of darkness on the probability of stranding*, Journal of Navigation, vol. 27., n°2, p.239-247.

Fukunaga, K. et Hostetler, L. (1975), The estimation of the gradient of a density function, with applications in pattern recognition, *IEEE Transactions on Information Theory*, vol.21, pp. 32-40.

Goerlandt, F., Montewka, J., Sonne Ravn, E., Hanninen, M. et Mazaheri, A. (2012), *Analysis of the near-collision using ais data for the selected locations in the baltic sea*, EfficienSea, Tech. Rep.

Le Guyader, D. et Brosset, D. et Gourmelon, F. (2011) Exploitation de données AIS pour la cartographie du transport maritime, *Mappemonde*, vol.104, n°4.

Marin (2012), *Safety assessment study for offshore wind farm*, Tech. Rep., Bonn Agreement Project: BeAware.

MacDuff, T. (1974), The Probability of Vessel Collisions, *Ocean Industry*, pp. 144-148.

Montewka, J., Hinz, T., Pentti, K. et Matusiak, J. (2010), Probability modelling of vessel collisions, *Reliability and System Safety*, vol. 95, pp. 573–589.

Organisation Maritime Internationale (2001),*Guidelines for the installation of a shipborne automatic identification system (AIS), IMO Resolution A.917(22).*

Organisation Maritime Internationale (2002), *Formal Safety Assessment*, Guidelines MSC/Circ.1023/MEPC/Circ.392, MSC/Circ.1180-MEPC/Circ.474, MSC-MEPC.2/Circ.5.

Tsou, M.C., (2010), Discovering Knowledge from AIS Database for Application in VTS, *The journal of Navigation*, vol.63, 449-469.

Wang, W. et Carreira-Perpinan, M.A. (2010) Manifold Blurring Mean Shift Algorithms for Manifold Denoising, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1759-1766.

Winter, M.W., (1995), *Software Fault tree analysis of an Automated Control System Device in ADA*, thesis, Naval Postgraduate School, Monterey (CA).

Yang, Z., Zhao, Q., Tao, H. et Liu, W. (2010) Evolving Mean Shift with Adaptive Bandwidth: A Fast and Noise Robust Approach, *Computer Vision – ACCV 2009 Lecture Notes in Computer Science*, vol. 5994, 2010, pp 258-268.

SESSION 2 :

Usage de l'information géographique et de ses représentations

Vers une première évaluation de la contribution des Infrastructures de Données Géographiques à la gestion des territoires

Jade Georis-Creuseveau[1]

1. LETG-Brest (UMR 6554 CNRS), Geomer, Université de Bretagne Occidentale, Institut Universitaire Européen de la Mer
Technopôle Brest-Iroise, 29280 Plouzané cedex, France
jade.georis-creuseveau@univ-brest.fr

RESUME. *Les Infrastructures de Données Géographiques (IDG) peuvent contribuer à la mise en place des politiques publiques concernant les territoires à travers les systèmes d'informations, les politiques et les dispositions institutionnelles qu'elles rassemblent dans le but de faciliter la mise à disposition et l'accès aux données géographiques. Mais leur contribution effective dépend des réponses qu'elles apportent aux acteurs des territoires. Basée sur un questionnaire en ligne diffusé en octobre 2012, notre démarche vise à recueillir les points de vue des utilisateurs ou producteurs d'information géographique de la sphère publique. La mise en œuvre d'analyses statistiques factorielles permet de dresser une typologie des usagers et de proposer une première évaluation de la contribution effective des IDG à la gestion des territoires.*

Abstract. Spatial Data Infrastructure (SDI) can contribute to the implementation of public policies concerning the territories through a combination of technologies, policies and institutional arrangements to improve data access and sharing processes. However, their real contribution depends on the answers they provide to public actors. Based on an online survey released in October 2012, our approach is to gather the various points of view of public users. Multivariate data analysis provides an initial user's typology and a first evaluation of the effective contribution of IDG to management of territories.

MOTS-CLES : *Infrastructures de Données Géographiques (IDG), usage, pratique, questionnaire en ligne, analyse statistique, territoires*

KEYWORDS: *Spatial Data Infrastructure (SDI), usage, practice, web survey, statistical analysis, territories*

2

1. Contexte et objectifs

Le besoin exprimé au niveau international de faciliter l'accès, l'utilisation et le partage des données géographiques détenues par les producteurs a conduit partout dans le monde et depuis les années 1990, au développement d'Infrastructures de Données Géographiques (IDG) nationales (Crompvoets *et al.*, 2004). Les IDG font référence à des solutions fédérées qui rassemblent les données, les réseaux informatiques, les normes et standards, les accords organisationnels et les ressources humaines nécessaires pour faciliter et coordonner le partage, l'accès et la gestion des données géographiques (Rajabifard *et al.*, 2002). A l'image des premières initiatives nationales et de l'infrastructure européenne instituée par la directive européenne INSPIRE[1], des IDG sont mises en œuvre à différents niveaux (international, national, régional, local (Masser, 2010) ou pour répondre aux problématiques de diverses thématiques (EUROGI/eSDI-Net+, 2011).

En France, la mise en œuvre de la directive INSPIRE impose de moderniser et d'harmoniser l'utilisation et la diffusion de l'information géographique dans les services publics. Le Géoportail[2], implémentation française de la directive européenne, est l'outil de référence pour l'accès aux données géographiques publiques produites sur le territoire national. En avril 2013, l'AFIGEO (Association française pour l'information géographique) recensait 55 IDG mises en œuvre du niveau départemental au niveau national. Elles participent avec le Géoportail, à la promotion d'une meilleure diffusion et utilisation de l'information géographique en France. Ce chiffre est en constante progression depuis 2008 (AFIGEO, 2013).

La coopération entre l'ensemble des producteurs et des utilisateurs de données semble indispensable (Masser, 2005; Nedovic-Budic *et al.*, 2011; Vandenbroucke, 2011) pour atteindre les objectifs affichés des IDG, à savoir le développement économique, la stabilité sociale, la bonne gouvernance et la gestion durable de l'environnement (Williamson *et al.*, 2003; Masser, 2010). Cependant, malgré les progrès techniques réalisés ces dernières années, les barrières organisationnelles et la réticence de nombreuses institutions à partager leurs données continuent d'entraver la mise en œuvre des IDG, notamment dans certains domaines (Strain *et al.*, 2006; Canessa *et al.*, 2007; Gourmelon *et al.*, 2010). Notre hypothèse est que les IDG peuvent contribuer à la mise en place des politiques publiques concernant les territoires à condition qu'elles répondent aux besoins et aux attentes de leurs usagers professionnels de la sphère publique (service de l'état, collectivité territoriale, établissement public...). Dans cette perspective, le questionnaire « Information Géographique, Infrastructure de Données Géographiques (IDG) et gestion des

[1] Directive 2007/2/CE du 14 mars 2007 élaborée par la Direction générale de l'environnement de la Commission européenne, qui vise la mise en œuvre d'une IDG européenne en s'appuyant sur des infrastructures interopérables mises en place par les Etats membres. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:01:FR:HTML

[2] http://www.geoportail.gouv.fr/accueil

territoires : quels usages et quelles pratiques ? » a été diffusé fin 2012. Cet article a pour objectif de présenter les résultats des premières analyses du questionnaire.

2. Méthode

2.1. Questionnaire

Dans le cadre de l'adoption des technologies par les individus, Orlikowski (1999) souligne que l'accès et la disponibilité ne suffisent pas, « c'est l'utilisation qui donne sa valeur à l'outil ». Le questionnaire a pour objectif d'évaluer la contribution des IDG à la gestion des territoires à travers (1) l'utilisation des IDG par les acteurs publics des territoires et (2) les retombées de cette utilisation dans le cadre de leurs missions territoriales.

Notre approche se base, pour évaluer ces deux éléments, sur le recueil des points de vue des usagers. Selon Boltanski et Thevenot (1991) et De Singly (1992), l'analyse des opinions apporte un éclairage sur « le sens que les acteurs donnent à leur conduite et les valeurs associées ». Elle permet ainsi de mieux comprendre les « critères de jugements qui servent de support subjectif à l'élaboration de leurs pratiques ». Les termes « usage » et « pratique » font référence à la définition de Jouët (1993) qui place le terme d'usage « dans un contexte plus restrictif qui renvoie à la simple utilisation tandis que la pratique est une notion plus élaborée qui recouvre non seulement l'emploi des techniques (l'usage) mais aussi les comportements, les attitudes et les représentations des individus qui se rapportent directement ou indirectement à l'outil ».

Dans la perspective de recueillir un grand nombre de données numériques uniformes, le choix de mettre en œuvre un questionnaire en ligne est apparu pertinent (Couper, 2000; Dillman, 2007). Le questionnaire s'adresse aux acteurs impliqués dans les IDG aux différents niveaux institutionnels, qu'ils soient coordinateurs d'IDG, animateurs ou membres de pôle métier[3], ayants droit[4], ou simple utilisateur. Toutes les organisations (services de l'Etat, collectivités locales, organismes publics, etc.) sont visées quel que soit leur domaine d'activité en lien avec les territoires, leur protection, aménagement, développement…

2.1.1. La conception du questionnaire

Le questionnaire a été élaboré sur la base d'indicateurs qui permettent d'approcher le concept, objet de la recherche, et de le mesurer à travers des traits caractéristiques sélectionnés par rapport à une réalité par essence multidimensionnelle (De Singly, 1992; Lebaron, 2006). Le concept central du

[3] Lieu d'échange dédié à une thématique et à des outils pour favoriser la mise en cohérence des données et leur partage. Il est composé des représentants de l'ensemble des partenaires travaillant sur le thème.
[4] Signataire du dispositif de partenariat mis en place par l'IDG.

questionnaire est la contribution des IDG à la gestion des territoires. Pour approcher cette notion, 4 indicateurs ont été identifiés et ont été recueillis à l'aide de 12 questions, non seulement au sujet de l'IDG au sein de laquelle les enquêtés sont impliqués, mais aussi au sujet des IDG mises en œuvre par les différents niveaux institutionnels et qu'ils utilisent dans le cadre de leurs missions territoriales. En complément des questions relatives aux indicateurs identifiés, 7 questions concernant les facteurs sociaux influant sur les conduites et les représentations ainsi que 8 questions relatives aux variables contextuelles ont été formulées (De Singly, 1992). Enfin, 3 questions ouvertes ont été proposées dans le but de mieux cerner les principaux avantages et problèmes rencontrés dans le cadre de ces usages (Tab.1). L'ensemble de ces questions a été structuré en 5 parties.

Tableau 1. Présentation synthétique du questionnaire

Concept	Dimension	Indicateur	Exemple de question
Contribution à la gestion des territoires	-IDG au sein de laquelle l'enquêté est impliqué -Autres IDG mises en œuvre par les différents niveaux institutionnels	Contribution à la gestion des territoires	-Selon vous, cette IDG vous aide-t-elle ou aide-t-elle votre organisme aujourd'hui à gérer votre territoire ?
		Usages de l'IDG	-A quelle fréquence utilisez-vous les ressources et fonctionnalités proposées par cette IDG ? -Quelle ressource et fonctionnalité proposée par cette IDG, considérez-vous être utile pour gérer votre territoire ?
		Rôle au sein de l'IDG	-Quel est votre rôle au sein de l'IDG ? -Depuis quand ?
		Bénéfices associés	-Depuis votre adhésion / utilisation / implication dans cette IDG, considérez-vous en retirer des bénéfices ?
Expression libre	N/A	-Avantages et problèmes -Améliorations souhaitées	-D'une manière générale, quelles sont les principales réussites des IDG en lien avec vos missions territoriales ?
Facteurs sociaux	N/A	Variables contextuelles	-Votre organisme est-il impliqué dans une IDG ? -Dans votre organisme, combien de personnes travaillent en géomatique / cartographie ?
		-Désignation biologique -Capital scolaire et professionnel	-Dans le cadre de vos missions territoriales, quel est votre rôle ? -Depuis quand ? -Quel est votre domaine de formation ?

Le questionnaire a été piloté par une interface en ligne gérée à l'aide du logiciel LimeSurvey[5]. Afin de diversifier le recrutement des enquêtés de ce type

[5] http://www.limesurvey.org/

d'échantillon volontaire (Martin, 2007) et de pouvoir appréhender (1) la diversité des individus impliquée dans les IDG et (2) la multiplicité des points de vue relatifs à la contribution des IDG à la gestion des territoires, plusieurs réseaux de diffusion ont été activés. Une collaboration avec l'AFIGEO a permis de solliciter directement les coordinateurs d'IDG afin de diffuser le questionnaire à l'aide (1) d'une invitation transmise directement à tous les ayants droit de l'IDG, (2) d'un lien vers le questionnaire à partir du site éditorial de l'IDG et (3) d'un article concernant le questionnaire dans la lettre d'information de l'IDG. En complément, plusieurs réseaux professionnels, associations, journaux et forums de discussion ont également relayé l'invitation au questionnaire.

2.2. Analyse statistique exploratoire multidimensionnelle

Le questionnaire est principalement exploité par une analyse factorielle, qui occupe une place de choix parmi la statistique exploratoire multidimensionnelle. L'analyse est fondée sur la complémentarité de deux approches : l'analyse des correspondances multiples (ACM) et la classification ascendante hiérarchique (CAH). Ces méthodes « permettent d'explorer les données, de se familiariser avec leur réalité, de trouver progressivement des cohérences et d'identifier les relations qu'elles entretiennent » (Escofier et Pagès, 2008).

2.2.1. L'analyse des correspondances multiples

L'analyse des correspondances multiples est une application particulière de l'analyse factorielle adaptée notamment au traitement des données qualitatives issues de questionnaire. Elle permet de réduire l'information contenue dans de volumineux tableaux de données numériques par la création de variables synthétiques (Lebart *et al.*, 2004). Deux nuages de points représentant respectivement les lignes et les colonnes du tableau étudié sont construits. La proximité géométrique entre deux points traduit la ressemblance statistique entre les objets qu'ils représentent (Escofier et Pagès, 2008). La géométrie des nuages de points et les calculs des proximités se basent sur la distance du χ^2 qui permet de faire jouer un rôle symétrique aux lignes et aux colonnes du tableau de données[6]. En outre, elle assure la robustesse des résultats de l'analyse vis-à-vis de l'arbitrage du découpage en modalités des variables nominales (Lebart *et al.*, 2004). Les deux nuages de points représentant les lignes et les colonnes sont projetés chacun sur une suite d'axes orthogonaux maximisant la variance (également appelée inertie) du nuage de points.

[6] Cette distance tire son nom du fait que « la dispersion des nuages de point calculée avec cette distance est proportionnelle au χ^2 de Karl Pearson utilisée pour tester l'indépendance des données » (Lebart *et al.*, 2004).

2.2.2. La classification ascendante hiérarchique

Compléter l'ACM par une classification permet notamment de prendre en compte la dimension réelle du nuage de points par une analyse dans l'espace tout entier. La compréhension de la structure des données, parfois nombreuses et complexes à interpréter, en est améliorée (Lebart *et al.*, 2004; Escofier et Pagès, 2008). Basées sur des calculs algorithmiques, les méthodes de classification permettent de découper l'échantillon en groupes homogènes d'individus en fonction de « la ressemblance de leurs réponses aux questions » (Martin, 2007). Parmi les stratégies utilisées pour agréger les individus, la classification ascendante hiérarchique (CAH) « consiste à agréger petit à petit les individus en créant des classes d'individus de plus en plus nombreux ». A l'aide d'un dendrogramme, elle présente l'avantage de fournir des critères permettant de déterminer le nombre optimal de classes (Martin, 2007).

La complémentarité entre ACM et CAH peut se traduire par une mise en œuvre conjointe de ces techniques. C'est ce que propose le logiciel R[7]et sa bibliothèque de fonctions Factominer[8], spécialement dédiée aux analyses factorielles (Husson *et al.*, 2009; R Development Core Team, 2009). Factominer propose une démarche adaptée aux variables qualitatives qui se base sur le critère d'agrégation de Ward[9].

3. Résultats

Accessible du 15 octobre 2012 au 18 mars 2013, le questionnaire a été relayé par 21 coordinateurs d'IDG, 10 responsables de réseaux professionnels et associations, 4 journaux et 2 forums de discussion. Le questionnaire en ligne a été consulté 1413 fois. Un premier tri a permis d'identifier un jeu de 729 réponses suffisamment complet dans les 5 parties du questionnaire pour permettre une exploitation statistique de qualité.

3.1. Portrait des participants

L'analyse des réponses au questionnaire permet de dresser un premier portrait des organisations et des individus ayant participé à l'enquête. En termes d'organisations, 38 % des répondants travaillent dans une collectivité territoriale et 20 % dans un service de l'état. Les Technologies de l'Information Géographique (TIG) y sont relativement bien développées. Près de 70 % des enquêtés travaillent dans une organisation impliquée dans la mise en œuvre d'une IDG ou faisant partie des partenaires de la plateforme. Les enquêtés utilisent l'information géographique et la géomatique pour l'aménagement et le développement territorial à 39%, à un échelon variant du régional à l'intercommunal (68%). Les enquêtés se caractérisent

[7] http://www.r-project.org/
[8] http://factominer.free.fr
[9] Il minimise la perte d'inertie calculée sur un tableau de coordonnées factorielles elles-mêmes issues d'un calcul d'inertie (Husson *et al.*, 2013).

par un profil de géomaticien à 45 % et de décideur à seulement 4 %. Ils sont des utilisateurs d'IDG à 53 % (Tab. 2). Au-delà d'une vocation initialement descriptive, ces analyses permettent de définir progressivement les contours d'une population encore mal connue (Martin, 2007).

Tableau 2. Principales caractéristiques des participants et de leur institution

Questions	% de réponses
Vous travaillez dans :	
Un service de l'Etat	20
Une collectivité territoriale	**38**
Une entreprise privée	9
Autre (établissement public, chambre consulaire, association...)	33
Total (N=728)	100
Quelle est la fréquence d'utilisation des Systèmes d'Information Géographique (SIG) et autres outils géomatiques au sein de votre service ?	
Pas d'utilisation	1
Ponctuel	13
Régulier	28
Permanent	**58**
Total (N=722)	100
Votre organisme est-il impliqué dans une IDG ?	
Non	33
Oui	**67**
Total (N=643)	100
Quel(s) est/sont le(s) domaine(s) d'activité pour le(s)quel(s) vous utilisez les données géographiques, les outils géomatiques et les IDG ?	
Aménagement et développement territorial	**39**
Activités en mer	6
Administration et gestion	6
Intervention technique	20
Service à la population	9
Prévention et sécurité	12
Autre	8
Total (N=1469)	100
Vous travaillez principalement au niveau :	
Communal	13
Intercommunal	21
Du Pays (au sens de la loi LOADDT, dite loi Voynet)	2
Départemental	20
Régional	**24**
National	9
Européen	2
Autre	9
Total (N=727)	100
Dans le cadre de vos missions territoriales, vous êtes principalement un :	
Décideur	4

Gestionnaire	21
Thématicien	18
Géomaticien	**45**
Autre	12
Total (N=551)	100
Au sein de l'IDG identifiée, vous êtes :	
Coordinateur	10
Animateur	5
Membre de pôle métier	11
Partenaire	21
Utilisateur	**53**
Total (N=722)	100

Les utilisateurs, partenaires et membres de pôle métier représentent 85 % de l'échantillon enquêté. En se basant sur l'approche centrée sur les usagers proposée notamment par Orlikowski (1999) et Roche (2004) et compte tenu du peu d'attention accordée aux besoins des utilisateurs finaux des IDG (Vandenbroucke, 2011), ces 3 catégories d'enquêtés (617 individus) ont été considérées prioritaires pour étudier les usages et pratiques des IDG et à travers eux, analyser la contribution des plateformes à la gestion des territoires. Les résultats qui suivent concernent uniquement cette sous-population, nommée par commodité « usagers » dans la suite du document.

3.2. Vers une typologie des usagers des IDG

Dans le but de fournir une image synthétique et simplifiée de l'information recueillie concernant les usagers, leurs usages et leurs pratiques, une ACM a été réalisée sur la sous-population statistique constituée par les modalités « utilisateur », « partenaire » et « membre d'un pôle métier » de la variable « Quel est votre rôle dans l'IDG ? ». Après suppression des modalités incertaines quant à leur interprétation (notamment les non-réponses et la modalité « ne se prononce pas »), l'analyse a été conduite sur 246 individus à l'aide de :

- 6 variables actives qui reflètent le mieux les indicateurs présentés au tableau 1 et qui ont été identifiés comme déterminants pour évaluer la contribution des IDG à la gestion des territoires (Tab. 3) ;

- 63 variables explicatives[10]. Les variables explicatives (aussi appelées supplémentaires) ne participent pas à la constitution des axes ni de la classification, mais permettent d'élargir le contexte d'interprétation induit par les éléments actifs (Lebart *et al.*, 2004).

[10] Les 63 variables explicatives sont issues de 12 questions à choix unique et de 5 questions à choix multiples qui ont été « éclatées » en 51 variables correspondant aux modalités proposées.

Tableau 3. Variables et modalités des variables actives de l'ACM

Variables actives	Nb de modalités
- Quel est votre rôle dans l'IDG ?	5
- Depuis quand ?	4
- A quelle fréquence utilisez-vous les données et métadonnées de l'IDG ?	4
- A quelle fréquence participez-vous au pôle métier de l'IDG	4
- Depuis votre adhésion à l'IDG, considérez-vous en retirer des bénéfices ?	2
- Selon vous, cette IDG vous aide-t-elle ou aide-t-elle votre organisme à gérer le territoire ?	4

Les variables explicatives introduites dans l'ACM concernent :

- les caractéristiques du répondant et celles de son organisation ;
- les points de vue sur l'utilité de l'IDG au sein de laquelle le répondant est impliqué, et des IDG qu'il utilise ;
- la contribution des IDG au développement durable des territoires.

L'interprétation conjointe du nuage des modalités, complétée par l'analyse des coordonnées et des contributions met en évidence, le long du premier axe factoriel, un gradient exprimant le lien entre d'une part la fréquence d'utilisation des ressources et fonctionnalités proposées par les IDG, et d'autre part les bénéfices associés et la contribution à la gestion des territoires. Afin de compléter l'interprétation de l'ACM, une classification a été réalisée à l'aide de la fonction HCPC de Factominer. L'allure du dendrogramme et le diagramme des inerties suggèrent une partition en 6 classes et permettent de proposer une première typologie d'usagers des IDG reposant sur certains traits caractéristiques de leurs pratiques (Tab. 4). Cette typologie met en évidence un gradient d'usages et de pratiques dans le cadre de la gestion territoriale.

Tableau 4. Description des classes suggérées par la CAH

Classe	Nb d'individus	Description	Usages et pratiques
1	39	Utilisateurs récents et de manière peu intensive des IDG. Ils déclarent leur méconnaissance des IDG malgré une utilisation régulière des TIG dans le cadre de leur emploi.	---
2	17	Utilisateurs occasionnels des IDG. Ils considèrent les fonctionnalités proposées par celles-ci peu ou pas utiles dans le cadre de leurs missions territoriales. Aucun bénéfice n'est ressenti.	-
3	48	Individus au profil de géomaticien qui travaillent au sein d'une organisation partenaire d'une IDG. Ils ne	+

		retirent pas de bénéfices à utiliser l'IDG au sein de laquelle ils sont membres depuis 1 à 3 ans. Cependant, certaines ressources et fonctionnalités proposées par l'IDG sont considérées utiles pour leurs missions territoriales (notamment données de références et téléchargement) même s'ils déclarent ne pas utiliser souvent les données de l'IDG.	
4	65	Partenaires d'une IDG, ces individus retirent de ce partenariat un bénéfice important. Ils utilisent de manière occasionnelle les IDG mises en œuvre par les différents niveaux institutionnels.	++
5	14	Membres de pôle métier, ces individus sont convaincus de l'utilité de ce type de groupe de travail collaboratif auquel ils participent très souvent. Ils utilisent également très fréquemment les IDG mises en œuvre pas les différents niveaux institutionnels. Ils considèrent importants les bénéfices associés à l'usage des IDG, spécifiquement au niveau territorial.	+++
6	63	Membres de pôles métiers depuis 3 à 5 ans. Les ressources et fonctionnalités proposées par l'IDG leur procurent des bénéfices importants pour gérer leurs territoires (document de cadrage, catalogage, données thématiques, cartothèque, site éditorial, pôle métier…). Ils illustrent cet aspect par plusieurs principes de développement durable.	+++

3.3. Echelon territorial des IDG utilisées

Les usagers ont identifié 31 IDG au sein desquelles ils se sont déclarés être impliqués. Ces IDG sont réparties sur 4 échelons territoriaux, dont l'échelon régional se caractérise par la majorité d'usagers (45 %). En complément de ces IDG, préférentiellement utilisées, les usagers ont déclaré à 27 %, utiliser les IDG nationales (Tab.5). Les plateformes nationales et régionales jouent donc un rôle important dans la diffusion et le partage de données géographiques françaises, tel que souligné par le rapport de la France sur la mise en œuvre de la directive INSPIRE (Point de contact INSPIRE France, 2013) qui considère le niveau régional comme le « niveau-clé de la réussite de la mise en œuvre d'INSPIRE ». Cependant, 49 % des usagers se sont déclarés impliqués auprès de plusieurs IDG mises en œuvre par de multiples niveaux institutionnels (jusqu'à 11 IDG sur 6 niveaux) tendant à démontrer que les usages se diversifient à travers une complémentarité du global au local et du général au thématique.

Tableau 5. Echelon territorial des IDG utilisées

Niveau territorial	Nb d'IDG	% d'individus
IDG au sein de laquelle l'usager s'est déclaré être impliqué		
IDG nationale	4	5
IDG régionale	20	45
IDG départementale	6	0,9
IDG infra-départementale	1	0,1
Plusieurs IDG mises en œuvre par divers niveaux institutionnels	N/A	49
Total	(N=31)	100 (N=600)
IDG utilisée par l'usager dans le cadre de ses missions territoriales en complément de l'IDG au sein de laquelle il est impliqué		
IDG internationale		9
IDG européenne		13
IDG nationale		27
IDG régionale	N/A	18
IDG départementale		13
IDG infra-départementale		20
Total		100 (N=1622)

3.4. Contribution des IDG à la gestion des territoires

Les usagers expriment leur accord à 65 % à la question « Selon vous, l'IDG au sein de laquelle vous êtes impliqués, vous aide-t-elle ou aide-t-elle votre organisme à gérer votre territoire ? » (Fig. 1).

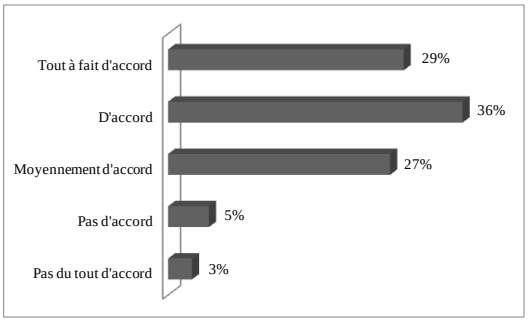


Figure1. Contribution de l'IDG à la gestion des territoires (N=357)

Une analyse combinée d'une nouvelle ACM[11] réalisée sur les variables actives (1) contribution, (2) bénéfices et (3) ressources et fonctionnalités et d'une analyse

[11] L'ACM a été réalisée sur 56 individus et 32 variables actives issues d'une question à choix unique et de 2 questions à choix multiples qui ont été « éclatées » en 23 variables correspondant aux modalités proposées.

thématique des avantages identifiés par les répondants[12] permet de proposer une première évaluation de la contribution des IDG à la gestion des territoires. Cette analyse croisée suggère 4 modalités qui s'expriment par certains bénéfices et mobilisent différentes ressources (Tab. 6). Elle permet également de replacer certains groupes d'usagers identifiés par la typologie présentée au tableau 3.

Tableau 6. Modalités de contribution des IDG à la gestion des territoires

Type de contribution	Bénéfice associé	Principales ressources et fonctionnalités	Points forts identifiés	Catégorie d'usagers
Inventaire	-Individuel -Institutionnel	-Métadonnées -Catalogage	-Economie (temps, coût…) -Qualité de la donnée -Valorisation de la donnée	3
Partage	Inter-organisationnel	-Données de référence -Documents de cadrage	-Accès aux données de référence - encadrement normalisé de la production	4
Technologie	Bénéfices multiples	-Flux -Extraction de données - Exportation de fonctions	Progrès technologiques (rapidité, outils d'analyse…)	-
Collaboration	Bénéfices multiples	-Données thématiques -Pôle métier	Collaboration entre acteurs	5 et 6

La combinaison des analyses multidimensionnelles et thématiques permet de mettre en évidence le rôle joué par les IDG dans les processus de gestion territoriale à travers les différentes composantes qui les caractérisent[13]. Même si les notions de connaissance des territoires et d'aide à la décision ont été (faiblement) formulées en tant que points forts des IDG, ces différentes modalités ne mettent pas clairement en avant la dimension territoriale de la contribution.

Soulignons enfin que pour 35 % des répondants, les IDG n'apportent pas ou peu de support à la gestion des territoires. De nombreuses mesures d'amélioration sont formulées au premier rang desquelles figurent la simplification de la technologie,

[12] L'analyse thématique a été réalisée sur les réponses formulées à la question ouverte « D'une manière générale, quelles sont les principales réussites des IDG en lien avec vos missions territoriales ? ».

[13] Différentes définitions de l'IDG existent. Nous retiendrons celle proposée par Rajabifard *et al.* (2002) qui considèrent 5 composantes : les données, les réseaux informatiques, les normes et standards, les accords organisationnels et les ressources humaines nécessaires pour faciliter et coordonner le partage, l'accès et la gestion des données géographiques.

l'augmentation des moyens (humains, financiers, en temps, compétence), l'amélioration de la qualité des données (mise à jour, échelle, précision, thématique) et de son accès.

4. Conclusion et perspectives

La gestion du territoire associe de multiples acteurs, mobilise de nombreuses données et repose sur divers outils notamment de production et de partage d'informations géographiques. Dans le cadre général d'ouverture des données publiques, les IDG ont pour objectif une meilleure production, gestion et diffusion des données géographiques.

Par notre étude, un cadre d'évaluation de l'opérationnalité des IDG dans le contexte de la gestion des territoires est proposé. Il repose sur une enquête ouverte analysée statistiquement pour saisir la variabilité des opinions et des comportements des individus impliqués à différents niveaux dans les IDG. Une première typologie des usagers ainsi qu'une identification des principaux modes de contribution territoriale sont proposées et tendent à confirmer que les politiques publiques en matière de gestion des territoires s'appuient de manière croissante sur les technologies géomatiques et les informations géographiques qu'elles produisent, gèrent et/ou diffusent.

Au-delà des typologies proposées, de nombreuses nuances sont à prendre en compte dans le contexte dynamique qui caractérise la mise en œuvre des IDG. De plus, les résultats du questionnaire mettent également en évidence les problèmes rencontrés par les usagers professionnels dans le cadre de leurs missions territoriales.

Des analyses complémentaires qui questionnent tant les modes d'interaction entre usagers et IDG (Hanneman et Riddle, 2005) que les pratiques des acteurs en situation (Argyris et Schon, 1978; Roche et Caron, 2004) sont en cours. Elles permettront d'apporter un éclairage complémentaire sur les facteurs qui favorisent ou limitent les usages des IDG dans le cadre de la gestion des territoires. Les usages considérant le territoire comme « un facteur d'intégration des différents phénomènes et un support de faits sociaux » (Joliveau, 2004) à travers une « vision intégrée orienté vers une gestion et une planification environnementale à long terme » (De Sede et Thériault, 1996) seront particulièrement recherchés.

Remerciements

L'auteur remercie toutes les personnes ayant répondu au questionnaire ainsi que l'AFIGEO qui, par son soutien, a permis une large diffusion du questionnaire. Elle remercie aussi Françoise Gourmelon, Christophe Claramunt, Frédérique Chlous-Ducharme, Matthieu Noucher, Adeline Maulpoix et François-Gilles Carpentier pour leurs conseils concernant la réalisation et l'analyse du questionnaire.

Bibliographie

AFIGEO. (2013). *Catalogue des IDG 2013*, http://www.afigeo.asso.fr/documentation/publications.html?download=630%3Acatalogue_idg_integral_2013

Argyris, C., Schon, D. (1978). *Organizational Learning: A theory of action perspective*, Reading, Mass, Addison-Wesley.

Boltanski, L, Thevenot, L (1991). *De la Justification. Les économies de la grandeur», Gallimard*, Paris.

Canessa, R., Butler, M., Leblanc, C., Stewart, C., Howes, D. (2007). Spatial Information Infrastructure for Integrated Coastal and Ocean Management in Canada. *Coastal Management*, vol. 35, n° 1, p. 105-142.

Couper, M. P. (2000). Review: Web Surveys: A Review of Issues and Approaches. *The Public Opinion Quarterly*, vol. 64, n° 4, p. 464-494.

Crompvoets, J., Bregt, A., Rajabifard, A., Williamson, I. (2004). Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science*, vol. 18, n° 7, p. 665-689.

De Sede, MH, Thériault, M (1996). La représentation systémique du territoire: un concept structurant pour les SIRS institutionnels. *Revue internationale de Géomatique*, vol. 6, n° 1, p. 27-50.

De Singly, F. (1992). *L'enquête et ses méthodes : le questionnaire*, Nathan.

Dillman, D.A. (2007). *Mail and internet surveys: The tailored design method*, John Wiley & Sons, New York.

Escofier, B., Pagès, J. (2008). *Analyses factorielles simples et multiples : objectifs, méthodes et interprétation*, Dunod, Paris.

EUROGI/eSDI-Net+. (2011). *SDI Self-Assessment Framework*, http://www.eurogi.org/esdinet-publications/file/5-sdi-selfassessmentframework-pdf.html#10

Point de contact Inspire France (2013). *Rapport de la France sur la mise en oeuvre de la directive INSPIRE pour la période 2010-2012*. CGDD/DRI/MIG.

Gourmelon, F., Robin, M., Maanan, M., Georis-Creuseveau, J., Pennober, G., Simao Da Silva, A., Fattal, P., Zourarah, B., Affian, K., Hauhouaut, C. (2010). Geographic Information System for Integrated Coastal Zone Management in Developing countries: cases studies in Mauritania, Côte d'Ivoire, Guinea-Bissau and Morocco. *Geomatic solutions for Coastal Environments*, New York, Nova Science Publishers, p. 347-359

Hanneman, R. A. , Riddle, M. (2005). *Introduction to social network methods*, http://faculty.ucr.edu/~hanneman/nettext/

Husson, F., Josse, J., Le, S. , Mazet, J. (2013). Package 'FactoMineR'.

Husson, F., Lê, S., Pagès, J. (2009). *Analyse de données avec R*, Presses universitaires de Rennes, Rennes.

Joliveau, T. (2004). *Géomatique et gestion environnementale du territoire: recherche sur un usage géographique des SIG*. Mémoire d'Habilitation à Diriger des Recherches en Sciences Humaines, Université de Rouen.

Jouët, J. (1993). Usages et pratiques des nouveaux outils de communication. *Dictionnaire critique de la communication*, vol. 1, p. 371-376.

Lebaron, F. (2006). *L'enquête quantitative en sciences sociales: recueil et analyse des données*, Dunod.

Lebart, L., Morineau, A., Piron, M. (2004). *Statistique exploratoire multidimensionnelle*, Dunod, Paris.

Martin, O. (2007). *L'enquête et ses méthodes: l'analyse de données quantitatives*, Armand Colin.

Masser I. (2005). *GIS worlds: creating spatial data infrastructures*, ESRI press, Redlands, California.

Masser, I. (2010). *Building European spatial data infrastructures*, ESRI Press, Redlands, California.

Nedovic-Budic, Z., Crompvoets, J., Georgiadou, Y. (2011). *Spatial Data Infrastructures in Context: North and South*, CRC Press.

Orlikowski, W. (1999). *L'utilisation donne sa valeur à la technologie*, http://www.lesechos.fr/formations/manag_info/articles/article_8_5.htm

Rajabifard, A., Feeney, M.-E., Williamson, I.P. (2002). Future directions for SDI development. *International Journal of Applied Earth Observation and Geoinformation*, vol. 4, n° 1, p. 11-22.

Roche, S., Raveleau, B. (2004). Usages sociaux et modèles d'adoption des SIG. *Aspects organisationnels des SIG*, p. 121-145.

Roche, S., Caron, C. (2004). *Aspects organisationnels des SIG* Hermes Sciences Publications, Lavoisier Paris.

Strain, L., Rajabifard, A., Williamson, I. (2006). Marine administration and spatial data infrastructure. *Marine Policy*, vol. 30, n° 4, p. 431-441.

R Development Core Team. (2009). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.

Vandenbroucke, D. (2011). *INSPIRE & NSDI State of Play: D3.2 - Detailed survey on use and usability of INSPIRE & NSDI*, http://inspire.jrc.ec.europa.eu/reports/stateofplay2011/INSPIRE_NSDI_SoP_-_D_3_2_-_Result_Detailed_Survey_Year_2-_v3.1.pdf

Williamson, I.P., Rajabifard, A., Feeney, M-E. (2003). *Developing spatial data infrastructures: from concept to reality*, CRC Press, Boca Raton.

Processus cognitifs experts vs étudiants lors de la réalisation de cartes de susceptibilité

Protocole expérimental pour une analyse comparée

Raffaella Balzarini, Paule-Annick Davoine, Muriel Ney

Laboratoire d'Informatique de Grenoble LIG,
681 rue de la Passerelle 38400 Saint Martin d'Hères, France
prénon.nom@imag.fr

RESUME. *Afin d'optimiser l'apprentissage des SIG dans l'enseignement supérieur des Sciences de la Terre et de l'Environnement, nous présentons une recherche en cours dont l'objectif est d'identifier les raisonnements et les concepts qui sous-tendent les processus de cartographie de susceptibilité avec les SIG. En s'appuyant sur des méthodologies d'analyse quantitative et qualitative des données verbales, notre recherche met en évidence les différences de stratégies entre les experts et les étudiants afin de proposer des instructions didactiques qui aident les étudiants à se rapprocher de démarches experts pour la résolution des problèmes spatiaux.*

ABSTRACT. *In order to enhance GIS learning and teaching in Earth Sciences higher education, we present an ongoing research that aims to identify the strategies and the concepts that underlie a suitability mapping process. Linking quantitative and qualitative verbal data analysis, our research shed light on the differences between experts and students and it suggests instructions that help students to move closer to expertise in problem-solving ability.*

MOTS-CLES : *processus cognitifs, analyses de données verbales, cartographie de susceptibilité, modèles de potentialité, analyse spatiale multicritères.*

KEYWORDS: *mental process, verbal data analysis, GIS-based land-use suitability mapping, multicriteria spatial analysis.*

1. Introduction

En géosciences deux champs d'application bénéficient particulièrement des apports des SIG : la numérisation cartographique et la cartographie de susceptibilité /vulnérabilité via les méthodes d'analyse spatiale multicritère (Balzarini *et al*., 2011). La maitrise de ces méthodes est donc un atout indispensable au géologue car les problématiques de terrain, demandant le croisement de données multi sources, et la cartographie de modèle d'aptitude (susceptibilité/vulnérabilité) sont de plus en plus fréquentes en Géosciences (Malczewski, 2004). D'un point de vue cognitif, la réalisation de cartes de susceptibilité comporte des processus relatifs à la représentation des données géographiques, leur interprétation, et leur traitement (mathématisation) nécessitant des connaissances complètes et complexes des SIG. La conceptualisation et l'analyse visuelle de ces cartes réalisées avec des méthodes multicritères, impliquent une interaction complexe de compétences et de schémas mentaux. Dans une perspective didactique, et afin de proposer des situations d'apprentissage qui poussent les étudiants vers une démarche d'expert (Petcovic et Libarkin, 2007), nous nous sommes intéressés : d'une part à analyser les processus cognitifs qui sous-tendent le raisonnement visuel (MacEachren, 1995), les stratégies et les compétences cartographiques mobilisés lors de la résolution de problèmes d'aptitude avec des méthodes d'analyse spatiale multicritère ; d'autre part, à identifier les différences entre experts et novices. Cette recherche s'appuie sur des méthodes quantitative et qualitatives d'analyse de données verbales, qui permettent d' identifier et formaliser la connaissance mobilisée dans une tâche cognitive (Chi, 1997).

Cet article présente les méthodologies de recherche et d'analyses adoptées, les résultats préliminaires issus du protocole expérimental et quelques pistes d'étayages didactiques qui aident les étudiants à avancer vers des compétences experts.

2. Cadres théoriques et Problématique

Les Sciences Cognitives ont une riche tradition de recherche sur l'expertise dans différents domaines. La capacité d'abstraction, de stockage et de rappel d'un large ensemble d'information, les stratégies de résolution de problèmes et la souplesse conceptuelle à l'intérieur d'un domaine de connaissance, illustrent ce que signifie d'être un expert (Petcovic et Libarkin, 2007).

Nous nous appuyons sur le modèle de structure mentale mobilisée lors de l'interprétation cartographique proposée par MacEachren's (1995) pour représenter et organiser les concepts ; ce modèle permet de relier les processus cognitifs issus de l'information cartographique, les rôles de la connaissance, de l'expérience, de la pratique et de l'apprentissage des lecteurs de la carte. Les experts et les novices activent des structures mentales très différentes : l'utilisation de la structure appropriée à une situation donnée, demande de l'apprentissage et de la pratique. Dans une tâche de lecture d'une carte qui représente un paysage connu, un expert activera une structure mentale qui permet non seulement de « coder » tous les objets visuels dans différents patterns, mais aussi d'intégrer toutes les relations qui existent

entre les objets. (Kent et Chang, 2008). Le lecteur de carte expérimenté exprimera les entités complexes et leurs relations avec un vocabulaire étendu et des labels (Edwards, in MacEachren, 1995). L'action de labéliser implique la capacité de décrire une catégorie et de la caractériser. Plus précisément, un expert est en condition de : 1. repérer les anomalies ; 2.identifier rapidement les procédés fondamentaux ; 3. vérifier la correspondance entre l'information géométrique ou symbolique fournie par la carte et les caractéristiques géographiques du terrain représenté (Chang et al, 1985). Une autre stratégie-expert, essentielle au processus de résolution de problème a été mise en évidence par Crampton (in MacEachren, 1995): l'expert mobilise une auto-analyse de prévention d'erreurs, active et itérative, pour cibler la solution.

Le processus de réalisation d'une carte de susceptibilité comporte des « obstacles » conceptuels que nous prenons en compte pour notre analyse et dont nous donnons une synthèse ci-de suite.

D'un point de vue de la procédure mathématique, la combinaison linéaire pondérée (*weighted linear combination WLC)* et les opérateurs Booléennes, AND d'intersection et OR d'union, sont les opérations d'analyse spatiale multicritères intégrées dans les SIG, les plus populaires. Contrairement aux opérations Booléennes, la WLC est une méthode compensatoire : elle permet de compenser, sur une même cellule, une valeur faible, induite par un critère, par une valeur plus forte, induite par un autre critère. Cette méthode semble relativement intuitive et simple à mettre en œuvre, notamment pour des décideurs, pas nécessairement experts en SIG. Néanmoins, elle est souvent appliquée sans une compréhension précise de deux aspects clés de cette approche : l'attribution des poids aux critères et les procédures de standardisation des données de différente nature, pour qu'elles soient commensurables (Malzcewski, 06).

D'un point de vue de la représentation, la carte résultant de la MCSA est une carte Choroplèthe: il s'agit d'une représentation de quantités (plethos) relatives à des espaces, ou aires géographiques (khorê), par le moyen d'une échelle de tons gradués. Sa réalisation repose d'abord sur le choix d'une méthode de discrétisation, c'est-à-dire de division de la série statistique que l'on veut cartographier en classes, ou intervalles (Béguin et Pumain, 2007). Dans cette carte, tous les éléments appartenant à la même catégorie sont représentés avec la même symbologie; cette perspective se fonde sur le consensus que dans les catégories, au sens générique du terme, chaque élément qui la constitue doit être représentatif de la catégorie. La compréhension d'une carte choroplète dépend de la capacité de l'utilisateur d'interpréter les catégories (MacEachren 1995).

D'un point de vue de l'interprétation et de l'usage de la carte de susceptibilité, nous nous appuyons sur les processus visuels de lecture de la carte établis par Marr (1985) ainsi que sur les concepts d'organisation perceptive, de catégorisation et jugement de MacEachren (1995). La vision humaine est très habile à extraire des formes d'une scène visuelle, à évaluer la profondeur et les dimensions relatives ainsi qu'à remarquer les mouvements. Une caractéristique déterminante du système visuel est sa capacité à emphatiser le contraste par rapport à l'illumination absolue et

l'acuité pour le ton d'un couleur (des centaines de différences sont discriminables) par rapport à la valeur d'une couleur (par ex. dans un camaïeu, quelques dizaines de différences sont discriminables). Pour cette raison la valeur et la saturation d'une couleur peuvent être ordonnées, tandis que le ton ne peut pas. Un autre aspect qui joue dans l'interprétation d'une carte est la capacité du système visuel-neurologique à grouper des éléments extraits de la scène visuelle en « objets » (MacEachren 1995).

D'un point de vue de la résolution d'un problème, la cartographie de susceptibilité présente une spécificité : elle est décisionnelle. Jankowski et Nyerges (2001) ont montré qu'une réduction de la charge cognitive est nécessaire pour faciliter le raisonnement visuel apporté par la carte. Pour que la décision soit prise, la cartographie d'aptitude demande une dernière attention : les cartes résultant des analyses doivent être « contrôlées » itérativement tout au long de la procédure ; leur validité est subordonnée à la corrélation avec la représentation de la réalité du terrain. Grâce aux technologies SIG, le raisonnement visuel sur les cartes de synthèse est facilité par leur superposition (drapées) sur des fonds de carte, que ce soit des cartes d'infrastructures, des modèles numériques de terrain représentant le relief, des globes virtuels, ou tout simplement des cartes comportant des limites administratives. Notre cas d'étude concerne plus particulièrement les superpositions avec le relief, créé par l'ombrage sur des MNT. Leur interprétation et leur lecture est alors liée à la perception de la profondeur sur une scène 2D ou 3D. Nous nous appuyions sur la taxonomie des indices de profondeur établie par Kraak (1988) et plus précisément sur les indices « pictural » qui nous renseignent sur la structure de l'objet et sur l'effet visuel produit.

Ces aspects théoriques nous ont permis de cadrer notre problématique sur des questions de perception et interprétation de la carte et, par conséquent, d'affiner nos analyses des structures et des objets cognitifs. Notre objectif est d'identifier les concepts, les actions et les stratégies qui supportent la tâche de cartographie de susceptibilité. Particulièrement, nous abordons une question fondamentale : quelles opérations et quels concepts sont utilisés par les experts et les novices lors de la résolution d'un problème de localisation de site avec des méthodes d'analyse spatiale multicritères ?

3. Le Protocole expérimental

Pour répondre à cette question, nous avons mis en place un protocole expérimental selon les paradigmes de la Design-Based Research (Edelson 2002). Ce champ de recherche utilise des multiples méthodes telles que l'évaluation, l'analyse des productions, les observations de terrain (enregistrements), les interviews, etc. Notre protocole expérimental est basé sur la conception d'une situation-problème la plus proche possible de la réalité professionnelle et un recueil de données sur l'observation de l'activité des étudiants et experts résolvant ce problème.

3.1. Les Participants

Nous avons mené l'expérimentation auprès de deux populations d'étudiants bien différentiées mais toutes deux au niveau bac +3. D'une part 15 étudiants en licence professionnelle en Sciences de la Terre et, d'autre part, 50 étudiants en d'Ecole d'Ingénieurs en sciences de l'environnement. Ces étudiants sont novices en analyse spatiale et SIG. Ils travaillaient par groupes de 5 à 7 étudiants. 3 spécialistes, un géophysicien, un cartographe et un ingénieur SIG, ont constitué la partie Experts de notre expérimentation. Ils ont été interrogés sur leur lieu de travail, seuls avec un chercheur, selon la méthode de la « tâche artificielle » : ils réalisent l'activité préconçue et verbalisent à voix hautes (« thinking aloud ») leurs actions. Les chemins empruntés par l'expert lors de l'exécution de la tâche fournissent un aperçu incomparable sur le raisonnement.

3.2. La situation-problème : localiser un site

La situation-problème est la séquence pédagogique qui nous a permis d'observer les étudiants et les experts. Le problème posé appartient à la famille de problèmes identifiée comme récurrente chez les professionnels (Authentic Learning, Simon, 1962), à savoir la cartographie de susceptibilité pour la prise de décision. En l'occurrence, les sujets devaient *mener une étude cartographique de susceptibilité du sol pour l'implémentation d'une station de ski dans le sud Isère*. Pour cela, ils ont effectué une analyse spatiale multicritères sur des critères géomorphologiques, hydrogéologiques, climatiques et géographiques.

Pour les étudiants, nous avons disposé de 17 h tutorées, divisées en 5 séances de 3 heures, en 1 séance de 2, distribuées sur 2 semaines. Pour les experts nous avons disposé de 4 heures. Le SIG est ArcGIS 10.0.

3.3. Le recueil de données

Nous avons recueilli trois types de données :

- *les productions* : les cartes des étudiants et des experts
- *les données verbales* : les interactions verbales des étudiants pendant les étapes –clé de la procédure de résolution (enregistrements vidéo et audio) ; les protocoles « thinking-aloud » des experts (enregistrements vidéo)
- trois *questionnaires* à questions ouvertes : l'analyse des réponses permet de repérer les principales difficultés ressenties par les étudiants.

Nous présentons ci-après la méthode utilisée pour l'analyse des données verbales qui ont été retranscrites. Nous avons analysé 3 heures d'enregistrements pour chacun des 3 expert et 2,5 heures pour chacun des 6 groupes d'étudiants (3 groupes « géologues » et 3 groupes «ingénieurs »).

6

3.4. La Méthode d'analyses des données verbales

Notre méthodologie d'analyse est conçue comme une approche intégrée (Chi, 1997) qui associe analyse quantitative et analyse qualitative du contenu des transcriptions. L'analyse verbale est une méthodologie pour quantifier le codage subjectif ou qualitatif des contenus des phrases. L'objectif de cette méthode est de mettre en évidence quelles connaissances le sujet utilise pour résoudre un problème, qu'elles soient correctes ou pas. Essentiellement, cette méthode permet, à travers le codage d'éléments verbaux (*verbatim*) propres à une idée, de concrétiser l'idée même, pour ensuite pouvoir en comptabiliser ou comparer les occurrences. Les instructions de Chi (1997) comportent 8 étapes :

1. Réduire le protocole: réduire les données en sélectionnant quelques activités.
2. Segmenter le protocole réduit.
3. Développer le schéma de codage (schéma des catégories taxonomiques).
4. Identifier les *verbatim* et les affecter aux catégories taxonomiques.
5. Représenter le formalisme du schéma.
6. Identifier le(s) pattern(s).
7. Interpréter le(s) pattern(s) et sa validité.
8. Effectuer un double codage.

De tout le processus de résolution problème de localisation de site, nous avons retenu 4 tâches fondamentales :

- T0 = la standardisation et le classement des données (discrétisation)
- T1= l'attribution des poids et l'agrégation des critères (sommes pondérées)
- T2= l'analyse visuelle des cartes résultant des sommes pondérées
- T3= le choix du site

Après avoir défini le corpus à analyser sur la base des 4 tâches retenues, il est nécessaire de segmenter les phrases pour dégager les unités de codage. La segmentation peut s'effectuer à plusieurs niveaux, révélant une granularité diversifiée telle que un propos, une idée, un échange, un épisode, un changement d'activité ou une impasse. Des caractéristiques sémantiques ou syntaxiques (mot-clé, connexions..) illustrent ces segments. Dans les segments nous avons identifié tous les verbatim ou les mots-clés, comme par ex. «montagne » ou « vallée », « station de ski » ou « nom de localité », « palette » « vert », « rouge », « seuils, %.. », pouvant constituer des groupes sémantiques, pour concevoir les catégories. Les catégories du schéma taxonomique sont donc définies selon un processus itératif dérivé des expressions et des explications des sujets (bottom-up process) et enrichi par des cadres théoriques (top-down process). Nous avons identifié 5 catégories d'objets conceptuels : algébrique, couleur, orographique, géographique et analytique (zones susceptibles). Une fois les catégories définies, les verbatim ont été isolés, codés et affectés à chaque catégorie pour être comptabilisés. L'opération de codage a été effectuée par deux codeurs différents. Tous les désaccords ont été ré-analysés

pour atteindre un taux d'accord final de 87 ,8% (moyenne).La Figure 1 formalise la démarche retenue pour l'analyse des données.

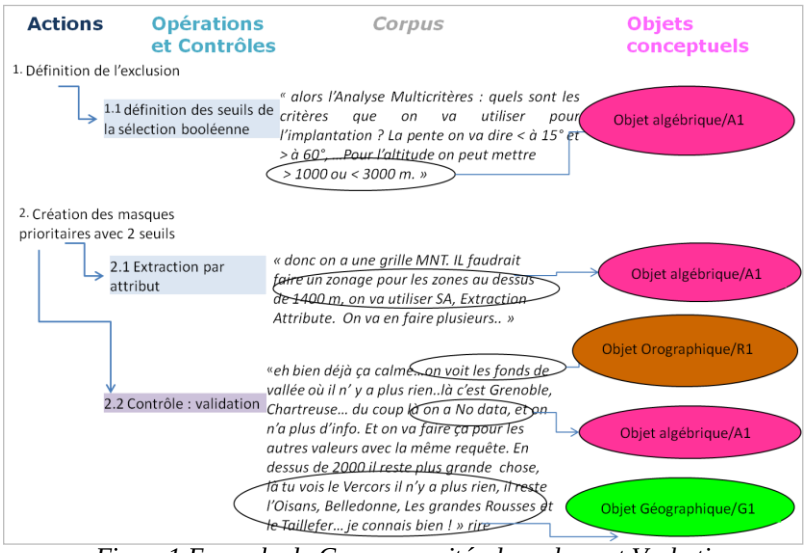


Figure1.Exemple de Corpus : unités de codage et Verbatim

4. Les Résultats de l'analyse qualitative

4.1. Les Objets Conceptuels : une proposition de taxonomie

Notre taxonomie est composée de 5 catégories, chacune représentative d'un objet conceptuel. Chaque objet est décliné par des descripteurs aux quels on a attribué un code. Il s'agit de :

- l'objet **Algébrique**, qui rassemble tous les éléments verbaux relatifs au domaine de la mathématisation de la donnée d'IG. Cet objet est caractérisé par 6 descripteurs ;

- l'objet **Couleur**, qui rassemble tous les éléments verbaux relatifs au domaine de la variable visuelle couleur. Cet objet comporte 5 descripteurs ;

-l'objet **Orographique**, qui rassemble tous les éléments verbaux relatifs à la description géomorphologique du terrain. Cet objet comporte 5 descripteurs ;

-l'objet **Géographique**, qui rassemble toutes les expressions relatives aux éléments géographiques d'une région analysée. Cet objet comporte 6 descripteurs ;

-l'objet **Zone Analysée** qui rassemble toutes les expressions qui montrent l'observation d'une ou plusieurs zones dans un but d'évaluation de l'aptitude. Cette évaluation se concrétise à travers 6 déclinaisons.
Le Tableau1 présente le schéma taxonomique.

1. Objet Algébrique	codes	A1	A2	A4	A5	A6	A7
	descripteurs	exclusion, sélection booléenne	seuils, nombre et ordre des classes, discrétisation, synthèse	hiérarchie des critères, représentation des poids	somme pondérée, combinaisons	mesures (areas, périmètres, pixel)	intersections vectorielles
	verbatim	*« la pente on va dire comprise entre 15° et 60° et l'alti entre 1000 et 3000 m »[e1]*	*« quelle échelle pour les classes, je ne me rappelle plus, 1,2,3? Je vais faire hiérarchique..»[e1*	*« Donc l'alti est important, 3, la Pente aussi, 3, la distance, 2, l'orientation, 1.»[e1]*	*« essaie une avec alti » « 40 (A) plus 30 (P) plus 20 (E) et plus 10 (CLC) »[g1]*	*« on pourrait définir un nombre minimal de pixel, vectoriser pour calculer l'area.. »[e1*	*« pour croiser les données on va utiliser des outils d'intersection.. »[e2]*
2. Objet Couleur		V1	V2	V3	V4	V5	
	descripteurs	le ton et la valeur	le contraste, la dissociation et le dégradé	l'harmonie des couleurs (la palette)	les limites, les bordures	l'affichage et la transparence	
	verbatim	*« les couleurs sont bien choisis, car le rouge indique « le haut, le fort », le vert le faible » [e2]*	*« Là c'est la plus nette, il y a beaucoup de bleu qui ressort, l'autre est plus flou mais il y a moins de bleu » [e3]*	*« pour le raster SP1 je choisi la palette Bleu-Jaune-Rouge » [e3]*	*« il faut bien démarquer entre le rouge clair et le rouge foncé » [g4]*	*« c'est au niveau des transparences que tu peux jouer, mais des transparences sur du noir et blanc »...[g3]*	
3. Objet Orographique	codes	R1	R2	R3	R4	R5	
	descripteurs	toutes les références aux connaissances du milieu (alpin)	la nomenclature courante : montagne, plaine, vallée, versants	la nomenclature spécialisée : crête, talweg, cône	les représentations du relief : isolignes ou MNT	la lithologie, la falaise	
	verbatim	*« l'Alpes du Grand Serre est 10 fois moins enneigé que le Col de Porte qui est encaissé » [g1]*	*« on n'a rien en montagne et tout en plaine » [g2]*	*« je ne crois pas qu'il soit bon de traverser une zones rouge, ça veut dire que tu traverses un talweg, une vallée » [g3]*	*« on n'a pas une carte des isolignes ? [g1] « je vais faire un ombrage pour voir le relief.. »[e1]*	*« Il (le calcul) enlève les falaises, mais regarde au Saint-Eynard il les laisse et c'est trop raide.. »[e1]*	
4. Objet Géographique	codes	G1	G2	G3	G4	G5	G6
	descripteurs	Les lieux connus : stations de ski, villes, localités,	le réseau routier	le réseau hydrographique	les représentations : carte Topo, carte de fond	les espaces verts : forêts, espaces protégés	les infrastructures : remontées mécaniques
	verbatim	*« là tu vois le Vercors il n'y a plus rien, il reste l'Oisans, Belledonne, Les Grandes Rousses et le Taillefer.... je connais bien ! » rire [e2]*	*«regarde juste les routes » [g2]*	*« parce que là tu es à côté de l'eau, tu peux mettre des pompes » [g5]*	*« il me faut un fond de carte, je mets le Scan 100 » [e2]*	*« il me faut les espaces protégés » [e2]*	*« tu ne peux pas mettre tes remontés comme ça et comment tu fais pour redescendre, il faut des trucs qui convergent vers les remontées » [g3]*
5. Objet Zone Analysée	codes	Z1	Z2	Z3	Z5	Z4	Z6
	descripteurs	zone observée et correspondance entre les critères, les poids	zone observée et correspondance avec la légende, les couleurs	zone observée et correspondance avec la nature du terrain	zones observées et sélectivité de la carte	zones observées comparées sur différentes cartes de sommmes	zone observée et besoin d'informations supplémentaires
	verbatim	*« quand les pentes sont très inclinées il faut que leur poids soit moins important » [e1]*	*« on a la valeur 8 (légende) en blanc et 1 en noir, low c'es pas bien et high c'est bien » [g3]*	*« là c'est Bourg d'Oisans, les 2 Alpes....d'ailleurs regarde, c'est là (Zoom) et c'est en vert! »[e1]*	*« cette carte a un peu moins de zones en rouge, plus précise, c'est un peu mieux,»[g5]*	*« l'autre est plus flou mais ça restreint plus les zones et fort heureusement les zones sont les mêmes » [e3]*	*« il faudrait faire encore un masque avec les zones à plus hauts valeurs et puis comparer avec la couche éboulis et espaces protégés » [e1]*

Tableau 1. Schéma Taxonomique des Objets Conceptuels

4.2. Les Actions et les Opérations de Contrôle : des catégories

Nous avons formalisé les séquences des actions et opérations effectués par les experts et les étudiants tout au long du processus cartographique, afin d'identifier un cheminement de résolution. Toutes les Actions des processus de solution experts et étudiants ont été répertoriées. Au total, 30 Actions ont été identifiées chacune caractérisée par une ou plusieurs Opération(s).

Nous nous sommes intéressés particulièrement aux opérations de contrôle, que nous présentons ci-après. Nous avons catégorisés les contrôles-erreurs nous basant sur la Théorie d'Ohlsson « Learning from Error », (1997). La cartographie de susceptibilité appartient à la catégorie « des tâches de choix, caractérisée par la sequentialité, la multiplicité et le résultat. La sequentialité signifie que la tâche demande une séquence d'actions dénombrables ; la multiplicité signifie que chaque action a lieu dans un contexte où plusieurs actions peuvent être concurrentes et que seulement peu d'entre elles sont appropriées, utiles ou correctes ; les actions pertinentes sont celles guidées par le résultat. […] Commettre une erreur dans une tâche de choix séquentiel, signifie faire une action qui est inutile et inappropriée dans un contexte spécifique. Elle entraine des conséquences sur le milieu, appelées 'signaux d'erreur'. Puisque la connaissance guide l'action, corriger une erreur permet d'améliorer une future performance par l'application d'une structure cognitive appropriée » (Ohlsson 97, traduction libre). Le Tableau 2 présente une synthèse des 3 catégories de contrôles d'erreurs que nous avons définies.

Codes	Type de Contrôle	Description	Exemples de Verbatim
CV	**Vérification :**	les opérations permettant de valider ou anticiper un choix, par ex. en comparant la même zone sur plusieurs cartes ou à la nature du terrain ou en vérifiant les résultats algébriques en légende.	*« ça marque plus la transition des altitudes, ce qui est artificiel. L'alti ressort trop, on aurait dû mettre plus de classes en alti pour atténuer » [e1]*
CD	**Diagnostic :**	les opérations permettant de reconnaitre les sources d'erreur, par ex. l'usage non approprié d'une fonctionnalité, d'une requête, de seuils des classes ou de l'attribution des poids.	*« il y a un problème là, on a toujours la même chose ça ne veut rien dire! Stop ! on a tout faux ! c'est celui-ci qui est mal classé ! » [g2]*
CC	**Correction :**	les opérations permettant d'intervenir sur l'erreur par ex. en modifiant la fonctionnalité, l'expression, l'ordre des classes ou la hiérarchie des critères.	*« Je reste persuadée qu'il faut refaire le classement de 1 à 9 pour chacune, car on ne travaille pas sur les mêmes valeurs » [e3]*

Tableau 2. Catégories des opérations de contrôle

Une fois la formalisation définie, les analyses quantitatives peuvent être réalisées.

5. Les Résultats préliminaires de l'analyse quantitative

Pour mettre en évidence les patterns, l'analyse quantitative a comporté les calculs de Présence/Absence, des Occurrences et des comparaisons entre experts et étudiants aussi bien pour les Objets que pour les Opérations-Contrôles et les Actions. Les graphiques en Fig.2 présentent une synthèse des comptages.

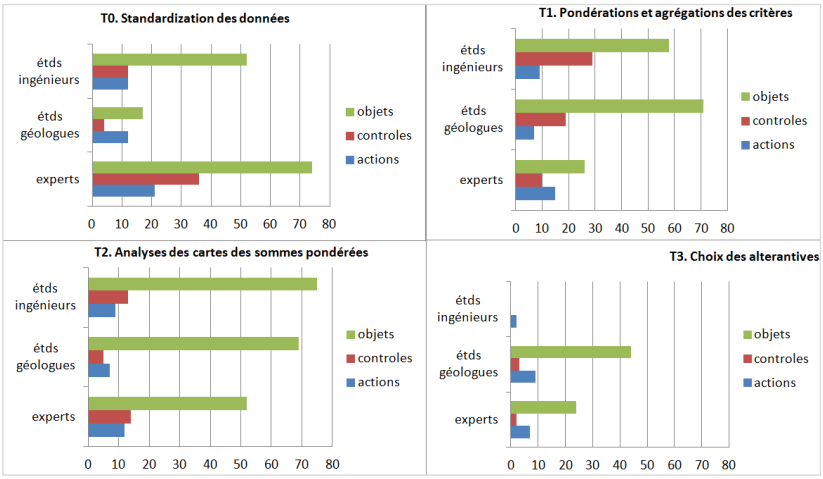


Figure 2. Synthèses des comptages des variables sur les 4 Tâches principales

Certes, les différences existantes entre Experts et Etudiants sont courantes. Toutefois, l'analyse quantitative précise les différences entre les types d'objets ou d'actions mobilisés ainsi que les écarts d'occurrences. On observe de manière générale que les experts mobilisent plus d'actions et d'objets que les étudiants dans la phase déterminante de préparation des données, la **T0,** standardisation et discrétisation des donnés. Afin de réduire les données à traiter les Experts agissent sur l'exclusion par rapport à des seuils, tandis que les étudiants ne montrent aucune démarche dans ce sens (T01Ex = 5 et T01étud= 0) (voir Fig.3*)*. Aussi, dès les premières étapes de la résolution du problème, les experts mettent en place beaucoup plus des contrôles que les étudiants, qu'ils soient de type vérification ou diagnostic (CEx= 37, CGéol= 4 et CIng= 12). L'activité- experts, plus intense dans cette phase que l'activité-étudiants, appelle un nombre plus importants et plus varié d'objets : principalement l'objet algébrique (AEx=45, AGéol=11, AIng=39) et géographique (GEx=9, GGéol= 0, GIng=3).

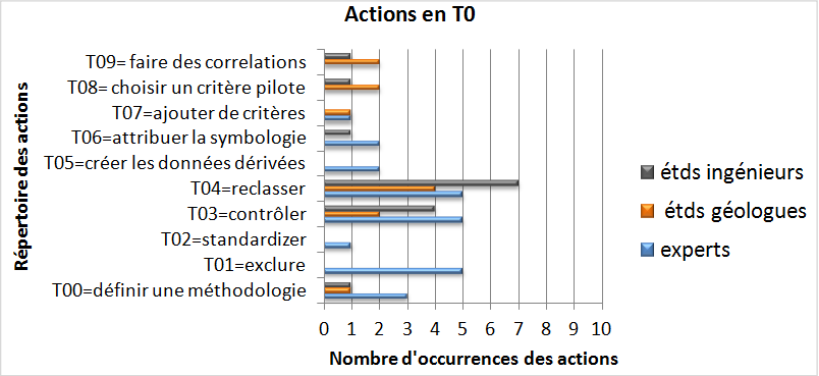


Figure3.Occurrences des 10 Actions dans la Tâche T0

Dans la phase d'allocation des poids et des sommes pondérées des critères, **(T1),** les étudiants effectuent plus d'opérations de contrôle que les experts car ils sont dans une démarche empirique d'essai-erreur. **En T2**, on remarque chez les experts l'usage d'outil d'effets visuels (zoom, transparence) qui améliorent l'analyse (T22 Ex=2 et T22étudts=0) et des actions de contrôle telles que la vérification avec le relief et le repérage d'effets seuils ; chez les étudiants les mêmes actions sont moins présentes. **En T3**, le choix d'un ou plusieurs site(s) se fait par rapport à des contrôles pour les experts surtout sur la base d'ajout d'informations et du repérage de lieux connus (CVEx=3 et CVétudts=1).

Une modélisation réalisée au moyen d'un logiciel de *mind mapping*, XMind, a été proposée. Elle représente l'enchainement d'actions et d'objets conceptuels sur les 4 principales Tâches de cartographie d'aptitude pour la résolution d'un problème de localisation. La Figure 4a et 4b montre le détail de la phase de standardisation et de classement des données (T0) pour un expert et pour un groupe d'étudiants.

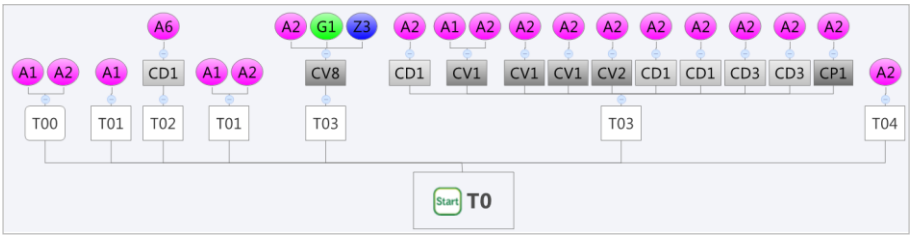


Figure 4a. Ligne Expert

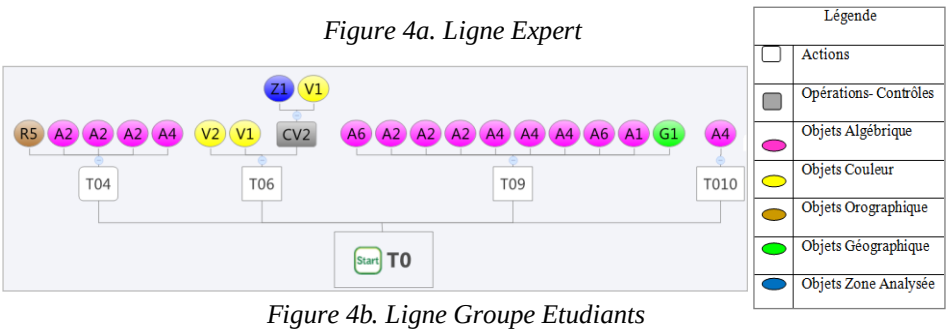


Figure 4b. Ligne Groupe Etudiants

6. Conclusions

Les résultats de notre étude, bien que préliminaires, nous ont permis de mettre en évidence les différences de stratégies (séries d'actions et d'opérations) entre les experts et les étudiants et de tracer de chemins de résolution de problème. Nous avons tiré de nos données une modélisation des objets conceptuels (en 5 catégories) et des opérations de contrôle (en 3 catégories) nécessaires à la résolution de ce problème. L'étude de l'activité de l'expert et de ses stratégies, qu'il s'agit d'une tâche de création ou d'interprétation de la carte, nous permet de finaliser les aides didactiques pour développer les compétences en analyses spatiale et visuelle

complexes. Les aides peuvent être de nature méthodologique (par ex. , réduire la masse de données et cibler les objectifs de résolution en créant des masques) thématique (par ex., fournir des règles et des notions de sémiotique (usage de la couleur), stratégique (par ex., stimuler le contrôle de conflits, erreurs ou anomalies ou technique (par ex., perfectionner la manipulation des outils d'effets visuels pour renforcer l'analyse visuelle-relief, transparence, globe-). La spécification de ces aides constituera la prochaine étape de notre travail. L'enseignement explicite de stratégies expertes, ainsi que l'usage de problématiques du monde professionnel et des approches collaboratifs pour développer la métacognition aident les étudiants à se rapprocher de l'expertise (Petcovic et Libarkin, 2007).

References

Balzarini R., Davoine P-A., Ney M. (2011). GIS in Geosciences Training: an experience in pedagogical-engineering. *In Proceedings of* International Cartographic Conference, Paris.

Béguin M., Pumain D. (2007). *La représentation des données géographiques*. Armand Colin, Paris.

Chang K.T., Lenzen T., Antes J. (1985). The effect of experience on reading topographic relief information: analyses of performance and eye movements. *The Cartographic Journal*, 22:88-94.

Chi T.H.M. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *The Journal of the Learning Sciences*, 6(3):271-315

Edelson D. (2002). Design research. *Journal of the learning Sciences*, 11(1):105-121.

Hmelo-Silver C.E. (2004.) Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16:235-266.

Jankowski P., Nyerges T. (2001). GIS-supported collaborative decision making: results of an experiment. *Annals of the Assosciation of American Geographer*, 91(1):48-70.

Kent R., Chang P. CH. (2008). Expertise in a Map Reading Task: The Role of Schemas in the Processing of Topographical Relief Information. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, Washington.

Kraak M.J. (1988.) *Computer-assisted cartographical Three-Dimensional Imaging Techniques*. Delft: Delft University Press.

MacEachren A.M. (1995). *How maps work: representation visualization and design*. New York: Guilford Press.

Malzcewski J. (2004.) GIS-based land-use suitability: a critical overview. *Progress in Planning*, 62:3-65.

Malzcewski J. (2006). GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographic Information Science*, 20(7): 703-726.

Marr D. (1985). *Vision: The philosophy and the approach. Issues in Cognitive Modeling*. London: Erlbaum.

Newell A., Simon H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall

Ohlsson S. (1996). Learning from Performance Errors. *Psychological Review*, 241–262.

Petcovic H.L., Libarkin J.C. (2007). Research in Science Education: The Expert-Novice Continuum. *Journal of Geoscience Education*, 55(4):333-339.

Usages des modèles spatiaux pour la prospective

Houet Thomas[1]

1. GEODE UMR 5602 CNRS, Université Toulouse le Mirail
5 allée Antonio Machado, 31058 Toulouse cedex
thomas.houet@univ-tlse2.fr

RÉSUMÉ.
Cet article théorique a pour objectif de faire un état de l'art sur l'usage des modèles spatiaux pour la prospective. Dans un premier temps, il présente un bref historique de la convergence implicite entre prospective et géographie. Dans un second temps, il aborde la question du choix du modèle en présentant les critères à prendre en compte. Dans un troisième temps, il présente la validation des modèles comme un moyen d'améliorer la plausibilité des scénarios à travers la combinaison de méthodes d'évaluation. Enfin, si on constate un usage de plus en plus important de modèles spatiaux en prospective, les méthodes évoquées sont loin d'être exhaustives et replacent la géoprospective comme une simple communauté de pratiques et de méthodes ayant un objectif commun : mieux explorer les futurs pour éclairer l'action présente. Au final, il apporte un éclairage sur l'apport des modèles aux démarches prospectives et vise à aider les géographes, les modélisateurs et/ou les prospectivistes à choisir un modèle approprié à leur problématique et à leurs objectifs afin de tirer partie de tous les avantages qu'ils offrent. Il tente également de clarifier certaines confusions sémantiques qui existent autour de l'usage des modèles couplés à des scénarios.

ABSTRACT.
This theoretical paper presents a state of the art on the use of spatial models dedicated to prospective studies. The first part summarizes the converging interests of future studies and geography. The second part gives helpful insights to answer the question of how choosing an appropriate model for exploring the future of territories/landscapes. The third part is dedicated to models validation tools and methods which combination may be useful to improve the plausibility of scenarios. These methods are not exhaustively described and illustrate that the geoprospective consists in a community of practices and tools with a common objective: better exploring the future to improve current decision making. This paper aims at highlighting the contribution of models for such purpose and helping geographer, modeler and/or prospectivists to choose a convenient model. It attempts to clarify some misunderstanding and confusions that already exists when combining scenarios and models.

MOTS-CLES : Typologie, Scénarios, Simulation dynamique et spatialement explicite, Validation, Plausibilité.

KEYWORDS: Typology, Scenarios, Spatially explicit and dynamic simulation, Validation, Plausibility

2

1. La place des modèles spatiaux dans une démarche prospective

1.1. Prospective et géographie, une convergence implicite

La prospective a connu un réel essor au début des années 1970 grâce au rapport Meadows (Meadows *et al.* 1972) qui a fait émerger le concept de développement durable. Déjà à l'époque, un modèle de simulation (World 3) était utilisé pour prendre en compte les interactions entre évolution de la démographie, des ressources et des déchets. La généralisation de ce concept aux projets d'aménagement du territoire a contribué, par incidence, à faire accroitre l'usage de la prospective dans ce domaine.

Une convergence implicite entre géographie et prospective s'est opérée avec la dimension spatiale comme dénominateur commun de représentation de la complexité. Dès 1971, la géographie s'est intéressée à l'exploration du futur à long terme (Phlipponneau, 1971). Dans le même temps, la prospective territoriale émerge dans le domaine de l'aménagement et de la planification du territoire, portée par la DATAR (Datar 1971), illustrant cette convergence. La prospective territoriale est définie comme ayant pour objet de définir des faits et des territoires, porteurs d'avenir, dans une perspective opérationnelle programmatique (Delamarre 2002). Par la suite, les développements de méthodes et d'outils suivant le paradigme de l'analyse spatiale a également contribué à renforcer cette convergence. L'analyse systémique, les méthodes de formalisation de la complexité et de simulation ont ainsi permis de renforcer la place de la dimension spatiale dans le processus prospectif (Voiron 2012 ; Emsellem et al 2012). La géoprospective constitue ainsi une communauté de pratiques visant à mettre la dimension spatiale, sous toutes ces acceptions, au cœur de la démarche prospective, depuis la caractérisation des dynamiques spatiales et temporelles du système, l'intégration des interactions multi-échelles jusqu'à la simulation spatialement explicite des changements futurs possibles pour l'évaluation qualitative et quantitative de leurs impacts. Les modèles géographiques et/ou spatialement explicites sont ainsi la pierre angulaire de la géoprospective, le lien entre « scénarios » et « participation » (Houet et Gourmelon 2013).

Parallèlement, le projet international LAND, ainsi que son prédécesseur LUCC (Lambin et Geist 2006), ont permis, riches des travaux de recherches et développements méthodologiques, de faire émerger la science du *Land change* (Turner et al 2007) ou encore la théorie du *Land System* (Kok et al 2004) où « modèle, scénario et acteur » doivent être appréhendés de concert pour comprendre les changements passés et anticiper ceux à venir. Leur finalité vise toujours l'évaluation quantifiée des changements et de leurs impacts possibles au travers le prisme de la dimension spatiale (surfaces, localisation, etc.) ainsi que l'aide à la décision pour une gestion à moyen ou long terme des ressources (Gourmelon et al 2008).

1.2. Scénarios prospectifs et modèles : usages, fonctions et validation

Du point de vue de la prospective, les modèles sont souvent utilisés de façon à traduire un scénario sous forme graphique. Plus généralement, la complexité des interactions à prendre en compte pour faire évoluer le système (territoire) étudié rend inopérant la plupart des modèles. En effet, seul l'exercice intellectuel permet de prendre en compte l'ensemble des processus et interactions entre des facteurs sociaux, démographiques, politiques, technologiques, économiques, culturels, historiques, géographiques ou encore environnementaux (Huss et Honton 1983, 1987). De plus, de par leur architecture, ils peuvent présenter certaines limites pour construire et explorer une plus grande diversité de futures contrairement aux méthodes de prospective. Il existe une multitude de méthodes, les scénarios étant une des plus répandues. Trois grandes 'écoles' de construction de scénarios, au niveau international, se distinguent (Amer *et al.* 2013 ; Bradfield *et al.* 2005) :

- l'école des *'intuitive logics'* (Wack 1985a, 1985b). Cette école est fondée sur les relations logiques de causalité entre les variables du système et n'utilise pas de modèles ou d'algorithmes mathématiques.
- l'école des *'probabilistic modified trends'*. Elle se distingue par les méthodes utilisées : les *trends impact analysis* et *le cross impact analysis* (Bishop *et al.* 2007), méthodes probabilistes servant à projeter des tendances à partir des tendances passées, déterminent les principales relations de causalités entre des variables, et permettent ainsi de modifier certaines projections.
- l'école française - *la prospective* (Godet 1986). Cette dernière, considérée comme une méta-méthode, englobe les deux écoles précédentes en s'appuyant sur un ensemble d'outils non délimité et une démarche méthodologique constituée d'étapes spécifiques dont le nombre varie selon les auteurs (Durand 1972 ; Godet, 1986 ; Mermet et al 2003).

La déclinaison de la prospective en géographie, visant à intégrer la dimension spatiale tout au long de la démarche prospective (Houet *et al.* 2008, 2010), confère aux modèles de simulation spatiale une place relativement limitée. Néanmoins, ils permettent de prendre en compte des interactions spatiales que la prospective territoriale n'intègre pas. Leur principal usage consiste ainsi à retranscrire spatialement les changements futurs d'occupation et d'usages des sols du territoire étudié tels que définis par des scénarios prospectifs à des fins d'évaluation. Plusieurs facteurs peuvent influer sur le choix du modèle à utiliser : le type de scénarios, la compréhension du modèle et de son fonctionnement, etc. Le modèle peut également avoir plusieurs fonctions (outil de diagnostic, de co-construction, etc.) et contribuer de diverses façons à l'exercice de prospective. Enfin, son choix va également dépendre des objectifs poursuivis par la démarche prospective. Piveteau (1995) distinguent trois types d'objectifs : l'apport de la connaissance sur le futur (simuler des évolutions tendancielles, anticiper des ruptures, explorer des conjectures), faciliter la médiation entre connaissance et action (favoriser le processus d'apprentissage de façon organisée, partagé par tous, de façon plus ou moins ludique et évolutif) et aider à l'action (fournir des données quantifiées menant à des

décisions transparentes, consensuelles ou négociées, innovantes et audacieuses). Ainsi, les modèles par leur usage, leur explicitation et les simulations vont pouvoir contribuer à répondre aux objectifs de la prospective.

La question de la validation des projections est aussi une question récurrente à laquelle il est difficile de répondre. En effet, comment peut-on valider le(s) futur(s) ? En prospective, cette question ne se pose pas, l'objectif étant d'explorer une diversité des futurs possibles et non de tenter de prédire quel sera le futur (Godet 1977). Néanmoins, la validation des scénarios produits reste indispensable et consiste à évaluer la démarche de construction et l'apport des scénarios *in fine*. La validation repose sur des critères dont le nombre peut varier selon les auteurs (Wilson 1998 ; Alcamo et Henrichs 2009 ; Durance et Godet 2010). En général, la validation vise ainsi à vérifier que les scénarios répondent aux critères suivants : vraisemblance, cohérence, pertinence, utilité, originalité. Au final, d'après les principes de la prospective, la validation du modèle utilisé dans une démarche prospective ne consiste pas à évaluer le pouvoir prédictif du modèle mais bien à démontrer que les critères d'évaluation cités ci-dessus sont respectés.

Cet article a pour objectif de faire un état de l'art des usages des modèles utilisés en géographie dans une démarche de prospective.

2. Le choix du modèle spatial

Le choix du modèle est déterminant et doit être réalisé tôt dans la démarche prospective. Il doit être fait selon les objectifs visés, c'est-à-dire l'apport que l'on attend du modèle et son aptitude à simuler dans le temps et dans l'espace les dynamiques considérées. Ce choix peut-être d'autant plus difficile si l'on attend du modèle de répondre à plusieurs de ces critères.

2.1. Les divers objectifs possibles liés à l'usage de modèles en prospective

D'après la littérature, si l'on considère les modèles spatiaux au sens large, plusieurs types d'apports peuvent être distingués :

- Traduire sous la forme de rendus spatiaux, les *images* d'un scénario à différentes dates. A travers la prospective territoriale, c'était l'objectif premier de l'usage de ces modèles.
- Favoriser la participation et l'apprentissage collectif lors de la construction de scénarios. L'usage de toute représentation simplifiée de la réalité pour identifier les dynamiques du système, les enjeux prospectifs, ou encore les jeux d'acteurs et les processus d'usages des sols lors d'une démarche prospective (construction de la base, définition de scénarios, etc.) peut ainsi constituer un objectif lié à l'usage du modèle. Toutes les méthodes de diagnostique territorial peuvent ainsi être considérées (Lardon et Roche 2008, Batton-Hubert et al 2008). La modélisation d'accompagnement peut également offrir des principes méthodologiques particulièrement bénéfiques

(Etienne 2010). L'explication du fonctionnement du modèle de simulation spatiale peut également favoriser la définition collective des processus et zones présentant des enjeux de préservation à l'avenir.

- Simuler de façon réaliste les processus spatio-temporels contribuant à l'évolution des territoires et leurs interactions. L'objectif des modèles concernés, principalement dynamiques et spatialement explicites, est de reproduire le plus fidèlement possible les processus d'occupation et d'usage des sols en tenant compte ou non des interactions entre l'agent et son environnement (le territoire), voire les interactions entre agents. Un fort réalisme renforcera la plausibilité des scénarios prospectifs.
- Contribuer à l'évaluation des scénarios. La simulation spatiale de changements futurs a pour principal intérêt d'identifier les espaces où ils pourraient se produire à l'avenir, et d'évaluer leurs impacts sur les ressources et l'environnement. Les simulations peuvent également permettre d'évaluer une incertitude quant à la probabilité que le changement intervienne à l'avenir. Dans le cas d'un scénario couplé à un modèle présentant des règles aléatoires, des simulations *monte carlo* rendent possible l'estimation de l'occurrence du changement. En couplant le résultat de plusieurs scénarios, on peut déterminer les zones présentant une forte probabilité qu'un changement se produise (ex : abandon de terres agricoles) dès qu'il est observé dans plusieurs scénarios (Verburg et al 2010).
- Contribuer à la gestion optimisée de ressources. Dans ce cadre, les modèles utilisés ont pour objectif de trouver la composition ou la configuration paysagère optimale sous contrainte d'une ressource existante (eau, biodiversité, etc.). Ainsi, ils permettent d'évaluer l'influence des mesures d'incitation ou de gestion sur les modes d'occupation et d'usages des sols, et par corollaire, leur faisabilité (Leenhardt *et al.* 2004).

Au final, le choix du modèle doit s'effectuer à la lumière des objectifs visés qui peuvent se distinguer à partir des questions suivantes. *Le modèle a-t-il pour objectif de représenter le territoire et son évolution de façon simplifiée ou réaliste ? A-t-il pour objectif de favoriser la participation ou d'affiner l'évaluation ? A-t-il pour objectif d'évaluer l'influence de l'interaction de multiples processus sur le devenir du territoire ou bien d'évaluer les processus à mettre en œuvre pour atteindre une situation donnée ?*

2.2. L'aptitude du modèle à simuler les dynamiques spatiales désirées / définies

Le choix du modèle doit également être effectué d'après son aptitude à simuler des dynamiques spatiales et temporelles spécifiques ou encore à intégrer des interactions entre ressources et acteurs.

Une approche comparative est dès lors indispensable pour évaluer l'aptitude d'un modèle à simuler les dynamiques spatiales et temporelles qui vont affecter l'évolution d'un territoire (Houet et Gaucherel 2007). Il faudra ainsi s'assurer que le

modèle est adapté à représenter la complexité du système étudié dans ces dimensions spatiale (unité spatiale élémentaire, étendue, interactions spatiales), temporelle (pas de temps, durée, boucles de rétroactions), et humaine (agent, groupes sociaux, relations hiérarchiques) (Agarwal *et al.* 2000). En effet, différentes dynamiques et interactions spatiales peuvent être observées, rendant inadaptés certains modèles suivant les thématiques / territoires étudiés (Houet et Hubert-Moy 2006). Certains peuvent présenter des limites à prendre en compte différents processus de diffusion, ou encore des relations multi-échelles, etc. De même, tous les modèles ne sont pas aptes à simuler et/ou combiner des processus présentant une dynamique temporelle continue, discrète ou cyclique.

L'évaluation de l'aptitude du modèle peut ainsi conduire à confirmer le choix d'un modèle ou, dans le cas contraire, à développer un modèle spécifique à l'aide de plateformes de modélisation par exemple. D'une façon générale, deux types de modèles de simulation se distinguent, et dont la dichotomie peut aider à l'analyse et au choix du modèle : les *process-based* versus *pattern-based* modèles (Fortin et al 2003). Les premiers ont pour objectif de simuler les processus influençant l'évolution d'un paysage (modèles de croissance, de diffusion d'espèces, règles d'usages du sol, etc.). Les seconds cherchent à simuler un patron paysager (habitat dispersé, contigu, etc.) qui imite les dynamiques spatiales observées.

3. La question de la validation du modèle

Le principe fondamental de la prospective consiste à explorer une multitude de futurs possibles pour éclairer l'action présente (Hatem 1993). En aucun cas, il s'agit de prédire le futur. Dès lors, la calibration du modèle à partir d'une période passée peut constituer un faux enjeu en modélisation prospective.

La validation du modèle a pour objectif de contribuer à rendre les scénarios plausibles, c'est-à-dire crédibles et vraisemblables. Cela consiste à s'assurer que la représentation spatiale et temporelle des processus simulés respecte les tendances, les processus de changements voire les patrons paysagers tels qu'ils sont préalablement définis dans les scénarios prospectifs.

3.1. La calibration, un faux enjeu ?

La plupart des modèles LUCC probabilistes existant et facilement utilisables s'appuient sur l'utilisation de données (cartes, règles expertes…) traduisant une évolution récente, observée. Leur fonctionnement « type » se résume en trois étapes : (1) à partir de ces données, ils estiment les tendances de changements futurs en terme de quantité de changements ainsi que les probabilités de transition d'une classe d'occupation du sol vers une autre ; (2) à l'aide de données géographiques traduisant les facteurs explicatifs de la localisation des changements, ils produisent, pour chaque transition et à l'aide de méthodes statistiques plus ou moins complexes (régression logistique, réseau neuronal, etc.), une carte des probabilités de localisation des changements ; (3) ils utilisent des méthodes dynamiques et spatialement explicites (automate cellulaire ou assimilé, etc.) pour simuler les

évolutions estimées et les allouer dans l'espace. La calibration d'un modèle à partir de données passées constitue un faux enjeu pour la modélisation prospective, sauf éventuellement dans le cas de scénarios tendanciels, et ce pour plusieurs raisons :

- les dates des données d'entrée ont une influence sur l'estimation des quantités de changements futurs (Houet et Hubert-Moy 2006). L'estimation des changements futurs (quantité) ne doit pas dépendre du modèle (et par corollaire des données d'entrée disponibles) mais être définie par le scénario prospectif;

- le mode probabiliste d'estimation des quantités de changements futurs implique une extrapolation des tendances de façon linéaire dans le temps. Ainsi, la simulation engendrera une évolution linéaire qui pourra ne pas être concordante avec le *cheminement* du scénario. Par exemple, si l'image finale peut atteindre la situation définie par le scénario, l'évolution linéaire pourra être en contradiction avec le récit.

- les probabilités de transition d'une classe d'occupation du sol à une autre sont généralement fixes dans le temps pour ce type de modèle. Dans le cas où un scénario prospectif prévoit une modification de ces probabilités (ex : l'urbanisation se fera plus dans des zones boisées que par le passé), ce type de modèle pourra être inadapté. De plus, l'usage de ce type de modèle prévoit que le poids des facteurs de changements est supposé constant sur la période de calibration et jusqu'à l'horizon temporel visé. Kolb *et al.* (2013) ont démontré que ceci n'était jamais démontré même dans le cas où aucun facteur de changement supplémentaire n'ait été constaté.

- à partir d'un même jeu de données et de règles, les modèles peuvent également produire des simulations spatio-temporelles différentes en raison des méthodes mobilisées (Mas *et al.* 2011) et donc engendrer une incertitude propre à chacun.

Au final, certains modèles de simulation spatiale, tels que les modèles probabilistes présentent des limites pour spatialiser tout type de scénarios prospectifs. Dès lors qu'il s'agira de simuler des scénarios tendanciels, leur utilisation devra être faite avec rigueur.

3.2. La validation pour accroître la plausibilité des scénarios prospectifs

Afin d'accroître la plausibilité des scénarios, la validation consiste à démontrer l'aptitude du modèle à simuler des changements, en termes de dynamiques et de processus, définis par des scénarios. L'approche proposée repose sur la combinaison de plusieurs démarches de validation :

- la réalisation de tests de sensibilité et d'intégration. Il s'agit de s'assurer que l'ensemble des processus sont convenablement simulés par le modèle et que leur(s) interaction(s) n'entrave pas son fonctionnement (Houet et Gaucherel 2007).

- la simulation sur une période passée. Si la calibration sur une période passée peut ne pas être obligatoire, la démarche de simuler des changements observés durant une période passée (entre t_{-2} et t_{-1}), et de comparer le(s) résultat(s) de la simulation avec une situation observée (à t), reste particulièrement intéressante (Houet et Gaucherel 2007). Dans le cas où le modèle est en mesure de reproduire, sur le plan de la composition et de la configuration du paysage, une évolution

semblable à celle qui s'est réellement produite, alors cela renforce la croyance que l'on peut avoir dans le modèle à produire des paysages vraisemblables.

- la simulation de dynamiques contrastées (en termes de localisation, de quantité et de processus). Si cette démarche est similaire à des tests d'intégration, elle s'en différencie quelque peu à travers la simulation de scénarios fictifs. Si le modèle est en mesure de gérer des dynamiques contrastées, avec une parfaite maîtrise de leur séquençage, tout en s'assurant que leurs interactions n'altèrent pas la production de paysages fidèles aux patrons paysagers et/ou aux processus d'usages des sols tels que définis par les scénarios, alors cela renforce la confiance que l'on peut avoir dans le modèle à produire des simulations plausibles.

Il n'existe pas d'indicateur synthétique sous forme statistique ou cartographique qui permette d'évaluer la plausibilité d'une simulation. De plus, les simulations produites reposent quasiment toujours sur des règles aléatoires (choix aléatoire de localisation par exemple), engendrant potentiellement diverses simulations possibles pour un même scénario. Les indicateurs reposant sur la comparaison d'une simulation avec une situation observée tels que le Kappa, la courbe ROC, etc. apparaissent inadaptés dès lors qu'ils sont pris individuellement. En effet, ils évaluent le pouvoir prédictif d'un modèle et non son aptitude à correctement simuler les processus d'usages des sols ou les patrons paysagers désirés (Brown *et al.* 2005). Pour chacune des démarches mentionnées, un ou plusieurs indicateurs peuvent être utilisés : indicateur statistique simple, ROC, Kappa et autres méthodes de validation par comparaison (Pontius 2002, Pontius et al 2008) ainsi que les métriques paysagères (Uuermaa *et al.* 2009), etc. C'est leur combinaison, à travers ces trois démarches de validation, qui importe. Néanmoins, rappelons que la plausibilité des simulations dépend essentiellement de la plausibilité des scénarios. Un modèle, une fois validé, peut en tout état de cause simuler de façon vraisemblable un scénario totalement incohérent.

4. Influence du type de scénario dans l'usage de modèle

De nombreuses typologies de scénarios existent dans la littérature (Greeuw et al 2000 ; Hatem 1993). La méthode de construction des scénarios, distinguant les scénarios exploratoires et normatifs, peut influer sur le type de modèle spatial utilisé (Houet et al 2008).

Les scénarios exploratoires parcourent le futur au fur et à mesure, sans idées préconçues ; la somme des relations de causalité produisant la situation future. Les *process-based* modèles sont particulièrement adaptés : la somme des règles et des processus, produit des *images* du futur à un pas de temps défini par le modélisateur, dont la juxtaposition dans le temps constitue l'évolution (*cheminement*).

Les scénarios normatifs partent d'une image du futur, désirée ou non, et le cheminement est construit de façon rétrospective, l'objectif étant d'identifier les leviers d'actions à mettre en œuvre (ex : règles d'usages des sols) pour arriver à cette situation. Ce type de scénario est particulièrement adapté dans le cas de la gestion spatiale d'une ressource, où l'organisation spatiale des types d'occupation des sols aura une forte influence sur la ressource. Les modèles spatiaux visant à

produire les *images* du futur sont nombreux (SIG, chorèmes, etc.). Les modèles de simulation spatiale permettant d'atteindre les objectifs futurs, tels que définis par des scénarios normatifs, peuvent être distingués en deux catégories : (1) lorsque l'objectif porte seulement sur la composition paysagère (quantité de changement à atteindre), les *process-based* et *pattern-based* modèles sont adaptés. Ils allouent spatialement les changements jusqu'à atteindre les proportions d'occupation et d'usages des sols attendus (Engelen *et al.* 2007) ; (2) lorsque l'objectif repose sur la configuration spatiale (une organisation spatiale à atteindre), les *pattern-based* modèles sont les plus adaptés dès lors qu'ils reposent sur des méthodes d'optimisation (Bamière *et al.* 2011 ; Gusdorf and Hallegatte, 2007). La difficulté est réelle dès lors que l'objectif consiste à reproduire une somme de processus cohérents – c'est-à-dire un cheminement – et une configuration spatiale – image du futur – désirée. Pour ce faire, il s'agit de coupler des méthodes d'optimisation avec un *process-based* modèle. Ce type de modèle est beaucoup plus rare (Benke *et al.* 2011).

Au final, on peut distinguer des démarches de prospective qui reposent sur des approches « modèle-dépendantes » ou « scénario-dépendantes ». Pour les premières, des hypothèses générales sont arrêtées, alimentant un modèle choisi de façon rigoureuse, et dont les simulations constituent le scénario prospectif. Les secondes prévoient la définition de scénarios desquels devront être déterminés les paramètres d'entrée du modèle.

5. La géoprospective : une communauté de pratiques et de méthodes

La convergence entre prospective et géographie est évidente et implicite depuis quelques décennies. La modélisation prospective a été identifiée comme un enjeu scientifique et méthodologique fort par le CNRS dans le projet de l'InSHS (2009, p.4) ainsi que lors de la « Prospective de l'Institut Ecologie et Environnement » de 2009 (InEE, 2009). La géoprospective constitue une communauté de pratiques et de méthodes visant à l'amélioration de l'exploration du futur (Houet et Gourmelon 2013). Les méthodes évoquées ne sont pas représentatives d'un recensement exhaustif des méthodes mobilisables.

De nombreuses confusions existent, tant les termes de scénarios et de modèles sont polysémiques. Les typologies de modèles, de scénarios, sont également nombreuses et ne permettent pas toujours d'évaluer l'aptitude d'un modèle à spatialiser tel ou tel autre scénario. Ils pâtissent également d'un effet de mode où l'usage de ces termes ne renvoie pas à un cadre méthodologique rigoureux. A travers cet article, l'objectif était de clarifier certains aspects sémantiques et méthodologiques visant à éclairer les futurs utilisateurs de modèles dans leur choix.

Remerciements

Cet article a bénéficié du soutien du projet ANR 2010 JCJC 1804 01 « MODE RESPYR » (http://w3.mode-respyr.univ-tlse2.fr/index.php).

Bibliographie

Agarwal C., Green G.L., Grove M., Evans T., Schweik C. (2000). *A review and assessment of land-use change models : Dynamics of space, time and human choice*, General Technical Report NE-297, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, 67 p.

Amer M., Daim T.U., Jetter A. (2013). A review of scenario planning. *Futures*, n° 46, p. 23–40.

Alcamo T., Henrichs T. (2009). *Towards guidelines for environmental scenario analysis, In: Environmental Futures: The Practice of Environmental Scenario Analysis*, first ed., Elsevier, Amsterdam, The Netherlands, pp. 13–35.

Bamière L., Havlík P., Jacquet F., Lherm M., Millet G., Bretagnolle V. (2011). Farming system modelling for agri-environmental policy design: the case of spatially non-aggregated allocation of conservation measures. *Ecological Economics*. vol. 70, n° 5, p. 891-899.

Batton-Hubert M., Joliveau T., Lardon S., (2008). Modélisation spatiale et décision territoriale participative. Conception et mise en œuvre dans des ateliers chercheurs-acteurs. *Revue Internationale de Géomatique*, vol. 18, n° 4, p. 549-569.

Benke K. K., Wyatt R.G., Sposito V.A. (2011). A discrete simulation approach to spatial allocation of commodity production for revenue optimisation over a local region. *Journal of Spatial Science*, vol. 56, n° 1, p. 89-101.

Bishop P., Hines A., Collins T. (2007). The current state of scenario development: an overview of techniques, *Foresight*, n° 9, p. 5–25

Bradfield R., Wright G., Burt G., Cairns G., Van Der Heijden K. (2005). The origins and evolution of scenario techniques in long range business planning, *Futures*, n° 37, p. 795–812

Brown D.G., Page S., Riolo R., Zellner M., Rand W. (2005). Path dependence and the validation of agent-based models of land use. *International journal of geographical information science*, vol. 19, n°2, p. 153-174.

Datar (Délégation à l'Aménagement du Territoire et à l'Action Régionale). (1971). *Scénarios d'aménagement du territoire. Essais méthodologiques*. Paris : La documentation française, collection Travaux et Recherches de Prospective.

Delamarre A. (2002), *La prospective territoriale*. Paris. La documentation française. Collection Territoire en mouvement, 102 p.

Durand J. (1972). A new method for constructing scenarios, *Futures*, n° 4, p. 325–330.

Durance P., Godet M. (2010). Scenario building: uses and abuses. *Technological forecasting and social change*, n° 77, p. 1488-1492.

Emsellem K., Liziard S., Scarella F. (2012). La géoprospective: l'émergence d'un nouveau champ de recherche? *L'espace Géographique*, n° 12, p. 154-168

Engelen G., Lavalle C., Barredo J.I., Meulen M., White R. (2007). The MOLAND modelling framework for urban and regional land-use dynamics. Chapter 17. *In*: Koomen, E., Stillwell, J., Scholten, H.J. and Bakema, A. (eds.), *Modelling land-use change; progress and applications*. Springer, Dordrecht, 297-319.

Etienne M., (2010) *La modélisation d'accompagnement : une démarche en appui au développement durable*. Quae editions, Paris, 367 p.

Fortin M.-J., Boots B., Csillag F., Remmel T.K. (2003). On the role of spatial stochastic models in understanding landscape indices in ecology, *Okois*, vol. 102, n°1, p. 203-212

Godet M., 1977, Crise de la prévision, essor de la prospective, Presses Universitaires de France, 188 p.

Godet M., (1986), Introduction to *la prospective*: seven key ideas and one scenario method. *Futures*, vol. 18, p. 134-157

Gourmelon F., Etienne M., Rouan M., Kerbiriou C., Charles M., Bioret F., Chlous-Ducharme F., Guermeur Y., Levrel H. (2008). *Éléments de prospective environnementale dans une réserve de biosphère*, *Cybergeo:* http://cybergeo.revues.org/20343

Greeuw S. C. H., van Asselt M. B. A., Grosskurth J., Storms C. A. M. H., Rijkens-Klomp N., Rothman D. S., Rotmans, J. (2000). *Cloudy crystal balls: an assessment of recent European and global scenario studies and models*, European Environment Agency, Copenhagen, Denmark, 96 p

Gusdorf F., Hallegatte S. (2007). Compact or spread-out cities: Urban planning, taxation, and the vulnerability to transportation shocks. *Energy Policy*, vol. 35, n°10, p. 4826–4838.

Hatem F., (1993). *La prospective : Pratiques et méthodes*, Economica, série "Gestion", Paris, 385 p.

Houet T., Hubert-Moy L., (2006). Modelling and projecting land-use and land-cover changes with a cellular automaton considering landscape trajectories : an improvement for simulation of plausible future states, *EARSeL eProceedings*, n°5, p. 63-76

Houet T., Gaucherel C., (2007). Simulation dynamique et spatialement explicite d'un paysage agricole bocager : Validation sur un petit bassin versant breton sur la période 1981-1998, *Revue Internationale de Géomatique*, vol. 17, n°3-4, p. 491-516

Houet T., Hubert-Moy L., Tyssot C., (2008). Modélisation prospective spatialisée à l'échelle locale : approche méthodologique, *Revue Internationale de Géomatique*, vol. 18, n°3, p. 345-373

Houet T., Loveland T. R., Hubert-Moy L., Napton D., Gaucherel C., Barnes C. (2010). Exploring subtle land use and land cover changes: a framework based on future landscape studies, *Landscape Ecology*, vol 25, n°2, p.249-266

Houet T.**,** Gourmelon F. (2013). La géoprospective – Apport de la dimension spatiale aux approches prospectives, *Cybergéo*, Sous presse

Huss W.R., Honton E.J., (1983) Scenario planning—what style should you use? *Long Range Planning*, vol. 20, p. 21–29

Huss W.R., Honton E.J., (1987) Alternative methods for developing business scenarios, *Technological Forecasting and Social Change*, vol. 31, p. 219–238.

InEE (2009). *Prospective de l'Institut Ecologie et Environnement du CNRS*, Compte rendu des journées des 12 et 13 mai, Rennes, InEE – CNRS, p. 66

InSHS (2009). *Projet scientifique et organisationnel, Institut Sciences Humaines et Sociales*, CNRS, 11p.

Kolb M., Mas J.F., Galicia L., (2013). Evaluating drivers and transition potential models in a complex landscape in southern Mexico. *International Journal of Geographical Information Science*, In press

Kok K, Verburg P., Veldkamp T. (2004). Integrated Assessment of the land system: The future of land use, *Land Use Policy*, vol. 24, p. 517-520

Lambin E.-F., Geist H.-J., (2006). *Land-Use and Land-Cover Change – Local processes and Global impacts*, Global Change – The IGBP Series, Ed. Springer, 222 p.

Lardon S., Roche S., (2008). Représentations spatiales dans les démarches participatives. Production et usages. *Revue Internationale de Géomatique*, vol. 18, n°4, p. 423-428.

Leenhardt D., Trouvat J-L., Gonzalès G., Pérarnaud V., Prats V., Bergez J-E., (2004). Estimating irrigation demand for water management on a regional scale. I. ADEAUMIS, a simulation platform based on bio-decisional modelling and spatial information. *Agricultural and Water Management*, 68(3) :207-232

Mas J.F., Kolb M., Houet T., Paegelow M., Camacho Olmedo M. (2011). Éclairer le choix des outils de simulation des changements des modes d'occupation et d'usages des sols. Une approche comparative. *Revue Internationale de Géomatique*, vol. 21, n°3, p. 405-430

Meadows D., Meadows D., Randers J., Behrens W., (1972). *The Limits to Growth*, Universe Books, New York, 205p.

Mermet L. (2003). *Prospectives pour l'environnement. Quelles recherches ? Quelles ressources ? Quelles méthodes ?* Collection Réponses environnement. La Documentation Française 107 p.

Phlipponneau M. (1971). *Géographie et perspectives à long terme*, Ouvrage de synthèse, IVth international Symposium of the International Geographical Union, Commission on Applied Geography, Editions Coconnier, Sablé, France, 467 p.

Piveteau V. (1995). *Prospective et territoire : apports d'une réflexion sur le jeu*. Cemagref Editions, collection Gestion des Territoires n°15

Pontius Jr. R.G. (2002). Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering & Remote Sensing*, vol. 68, p. 1041–1049.

Pontius Jr R.G., Boersma W., Castella J.C., Clarke K., de Nijs T., Dietzel C., Duan Z., Fotsing E., Goldstein N., Kok K., Koomen E., Lippitt C.D., McConnell W,, Sood A.M., Pijankowski B., Pidhadia S., Sweeney S., Trung T.N., Veldkamp A.T., Verburg P.H. (2008). Comparing the input, output, and validation maps for several models of land change. *Annals of Regional Science*, vol. 42, n°1, p. 11-27.

Turner B., Lambin E., Reenberg A. (2007). The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Science of the USA*, vol. 104, n°52, p. 20666-20671.

Uuemaa E., Antrop M., Roosaare J., Marja R., Mander U. (2009). Landscape Metrics and Indices: An Overview of Their Use in Landscape Research, *Living Reviews in Landscape Research*, 3, http://www.livingreviews.org/lrlr-2009-1

Verburg P., van Berkel D., van Doorn A., van Eupen M., van den Heiligenberg H. (2010). Trajectories of land use change in Europe: a model-based exploration of rural futures, *Landscape Ecology*, vol. 25, p. 217–232

Voirin C. (2012). L'anticipation du changement en prospective et des changements spatiaux en géoprospective, *L'espace Géographique*, n°12, p. 99-110

Wack P. (1985a). *Scenarios: uncharted waters ahead*, Harvard Business Review, n°63, p. 73–89.

Wack P. (1985b). Scenarios: shooting the rapids, Harvard Business Review, n°63, p. 139-150

Wilson I. (1998) Mental maps of the future: an intuitive logics approach to scenarios, *In*: L. Fahey, R.M. Randall (Eds.), *Learning from the Future: Competitive Foresight Scenarios*, first ed., John Wiley & Sons Inc., New York, pp. 81–108.

Itinéraire Méthodologique de Diagnostic Partagé des enjeux environnementaux.

Comment accompagner le changement dans l'ingénierie territoriale d'Etat ?

Sylvie Lardon[1], Aurélie Dumont[1], Cécile Lacroix[2], Marie-Neige Lebourg[2]

1. INRA & AgroParisTech, UMR Métafort, Clermont-Ferrand
24, Avenue des landais, CS 70054, 63178 Aubière

2. DREAL Poitou-Charentes, 15, Rue Arthur Ranc, 86020 Poitiers

RESUME. *Les acteurs des territoires sont souvent démunis pour concilier préservation de l'environnement et développement économique. La mission de la DREAL est de renforcer les approches transversales et l'intégration des enjeux environnementaux dans les projets de territoire. La méthode de diagnostic partagé (MDP) des enjeux environnementaux que nous avons expérimentée dans une collaboration entre des chercheurs et formateurs et des agents de l'Etat s'appuie sur l'usage de représentations spatiales. Elle permet de changer la façon de partager et de porter le Dire de l'Etat entre agents des services et auprès des acteurs des territoires. L'accompagnement par la recherche et par la formation constitue une plus-value appréciée aux différents niveaux concernés. L'appropriation des outils, l'auto-évaluation de leur usage et la formalisation d'itinéraire méthodologique renouvellent la question de l'accompagnement des acteurs dans la chaine d'ingénierie territoriale.*

ABSTRACT. *To conciliate environment preservation and economic development is a challenge for territory stake-holders. The mission of the DREAL (estate agency) is to give them knowledge to reinforce transversal approaches and integration of environmental stakes in territory projects. The shared diagnosis method (MDP) of environmental stakes, tested in partnership between researchers, trainers and state agents, uses spatial representations. For the state agents, it changes the way to share the state point of view and to give support to territory stakeholders. At different levels, the help of researchers and trainers is an added value. Tool appropriation, uses evaluation and methodological itinerary formalization renewed the stakeholders' accompaniment in the territory engineering chain.*

MOTS-CLES : *représentations spatiales, usage de l'information géographique, ingénierie territoriale, diagnostic, environnement.*

KEYWORDS: *spatial representations, use of geographical information, territorial engineering, diagnosis, environment*

1. Introduction

L'évaluation environnementale est dorénavant constitutive des projets de territoire et doit faire l'objet d'une attention particulière des acteurs des territoires, dans tous les plans, projets, programmes (MEDDE, 2011). Or si les acteurs des territoires sont souvent démunis pour concilier préservation de l'environnement et développement économique, les services de l'Etat, garants des enjeux environnementaux globaux, ont également du mal à les décliner à l'échelle des territoires de projet, et ce faisant, à se placer comme ambassadeurs *a priori* et non pas comme évaluateurs *a posteriori.* Pourtant, il est clair que la chaine d'ingénierie territoriale (Lardon, 2011) gagnerait à ce que l'analyse des enjeux environnementaux soit faite en amont des projets de territoire, pour pouvoir mieux la partager, entre services de l'Etat et la porter, auprès des acteurs des territoires, afin qu'ils s'en saisissent pour mieux prendre en compte l'environnement, non pas comme une contrainte, mais comme une ressource territoriale (Gumuchian et Pecqueur, 2007).

C'est l'objectif de la méthode de diagnostic partagé (MDP) des enjeux environnementaux que nous avons expérimentée dans une collaboration entre des chercheurs et formateurs d'AgroParisTech[1], UMR Métafort Clermont-Ferrand et des agents de l'Etat de la DREAL Poitou-Charentes[2]. La méthode est doublement originale. Elle s'appuie sur l'usage de représentations spatiales (Lardon et Roche, 2008) simplifiées pour analyser les enjeux environnementaux du territoire et les communiquer. Ces représentations spatiales deviennent des objets intermédiaires (Vinck, 2009) entre les acteurs parties-prenantes de la construction territoriale et contribuent aux changements de pratiques des agents dans leurs services. Co-construire une méthode de diagnostic partagé conduit alors à être soi-même acteur du changement. Ce double apprentissage est ci-dessous décrit et analysé, comme une expérience réussie d'accompagnement du changement au sein de l'ingénierie territoriale d'Etat (Piveteau, 2010).

Après avoir précisé le contexte institutionnel et le dispositif d'accompagnement réalisé (2), les différentes étapes de la méthode de diagnostic partagé sont présentées au long de l'itinéraire méthodologique et des changements de pratiques qu'il induit (3). La généricité de la méthode est évaluée non seulement au regard de la lisibilité des enjeux environnementaux exprimés pour les territoires, mais aussi en référence au déplacement qu'elle entraine dans les processus collaboratifs de production de représentations territoriales (4). Cette expérience met en évidence le double enjeu de l'usage des représentations géographiques (5).

---

1. AgroParisTech, Institut des sciences et industries du vivant et de l'environnement. Le Centre de Clermont-Ferrand est dédié à la formation et la recherche sur l'aménagement et le développement durables des territoires (http://www.agroparistech.fr/).
2. DREAL (Direction régionale de l'Environnement, de l'Aménagement et du Logement) Poitou-Charentes (http://www.poitou-charentes.developpement-durable.gouv.fr/).

2. Le dispositif d'accompagnement

2.1. L'évaluation environnementale par les services de l'Etat

Les enjeux environnementaux sont constitutifs des projets de territoire. Le développement durable amène à concilier les trois piliers de développement économique, social et environnemental, et ce, avec la participation des acteurs parties prenantes. Ainsi, dans les documents d'urbanisme tels que les SCoT[3] (Loudiyi, 2008) et les PLU[4], (Planchat-Hery, 2008), il importe de concilier les enjeux environnementaux avec les autres dimensions du développement.

C'est la mission des DREAL de renforcer les approches transversales et l'intégration des enjeux inhérents à l'environnement, à l'aménagement et au logement dans la gestion des territoires administratifs. Aussi, la DREAL Poitou-Charentes souhaitait-elle, dans le cadre de ses missions relatives à la connaissance, à l'analyse et l'évaluation des territoires, que ces préoccupations environnementales, soient mieux prises en compte dans les projets d'aménagements et de planification qu'elle suit sur des territoires communaux, intercommunaux et régionaux (encart 1).

Les services d'Etat font partie de la chaine d'ingénierie territoriale, même si la décentralisation et les recompositions territoriales en cours tendent à modifier leurs missions. S'ils n'interviennent plus directement dans la réalisation des diagnostics territoriaux, les services de l'Etat ont obligation de porter-à-connaissance les informations dont ils disposent pour les porteurs de projet. Pour les DREAL, cela demande d'exprimer clairement les enjeux environnementaux portés par l'Etat pour mieux les communiquer aux acteurs des collectivités territoriales et qu'ils s'en emparent pour élaborer leurs projets.

Encart 1. L'élaboration des au sein des DREAL

En Poitou-Charentes, les collectivités territoriales qui réalisent un document de planification urbaine (SCoT, PLU) sont étroitement accompagnées par les services déconcentrés de l'Etat. En effet, telle que la réglementation nationale le permet tout au long de la procédure d'élaboration de document de planification, dans cette région ces derniers ont pris l'habitude de fournir aux collectivités :

- un Porter à Connaissance (PAC) pour rappeler, dès le début de la procédure, les principales réglementations qui s'appliquent au territoire (loi SRU, loi littoral, etc) et lister l'ensemble des zonages et servitudes qui y sont présents (monuments historiques, sites Natura 2000, canalisations, etc) – Le PAC est produit par la Direction Départementale des Territoires (DDT), qui recueille pour ce faire les contributions de tous les autres services de l'Etat dans leurs domaines de compétence respectifs, dont celle de la Direction Régionale de l'Environnement, de l'Aménagement et du Logement (DREAL) ;

3. SCoT, Schéma de Cohérence Territoriale
4. PLU, Plan Local d'Urbanisme

- une Note d'enjeux (ou Document d'association) pour formaliser, à partir d'un diagnostic territorial multi-thématique, le « Dire de l'Etat » sur ce territoire et identifier les enjeux à prendre en compte pour y définir un projet durable d'aménagement urbain – La Note d'enjeux est produite par la DDT, avec la contribution éventuelle d'autres services de l'Etat, dont la DREAL ;
- et un Cadrage préalable d'Evaluation Environnementale (si la collectivité en fait la demande formelle), dans l'objectif de fournir, à partir d'un diagnostic territorial axé sur les thématiques environnementales (eau, air, paysage, nature, énergie et gaz à effets de serre, consommation d'espace, etc) et dès lors que son projet de territoire se précise, les enjeux pressentis en matière d'aménagement de l'espace, ainsi que les réflexions et pistes d'actions possibles sur son territoire pour que l'environnement soit pris en compte de façon pertinente – Le cadrage préalable est produit par la DREAL (pour le compte du Préfet de département, autorité environnementale) avec une forte contribution de la DDT pour assurer une totale cohérence avec la note d'enjeux préalablement réalisée par cette dernière.

2.2. La démarche de diagnostic prospectif participatif par la formation et la recherche

La démarche développée s'appuie sur une méthodologie désormais éprouvée d'usage des représentations spatiales dans les démarches participatives, le « jeu de territoire » (Lardon, 2013). Le jeu de territoire est un jeu d'expression pour construire une vision stratégique partagée du territoire (Angeon et Lardon, 2003). Il se joue en trois étapes (encart 2) en mobilisant les chorèmes pour extraire les principes organisateurs de l'espace et donner à voir les enjeux du territoire (Lardon et Piveteau, 2005).

Encart 2. Les trois étapes du jeu de territoire

Etape 1 : Construire la maquette des dynamiques du territoire
La première étape se réalise en ateliers parallèles de 6 à 7 joueurs. Elle consiste en un diagnostic partagé des dynamiques du territoire et une spécification des enjeux. Les joueurs construisent collectivement la maquette à partir des informations qu'ils détiennent dans les fiches à jouer, et rendent compte de leurs représentations du territoire. Les fiches de jeu sont réparties entre les joueurs et un petit temps est laissé pour la lecture des informations. A tour de rôle, les joueurs sélectionnent l'une des fiches qui leur ont été distribuées (4 à 5 cartes par joueur), énoncent les informations qu'elle contient, argumentent l'importance de ces informations pour le territoire, proposent une légende et dessinent sur la maquette commune les caractéristiques retenues du territoire. Ce choix peut être discuté par les autres, mais au final, seule compte la voix du joueur qui détient la carte. Cette première étape du jeu s'établit en deux tours. A l'issue de la première étape, les joueurs commentent la maquette produite et discutent des fiches non jouées. Ils récapitulent les principales structures et dynamiques observées et spécifient les enjeux identifiés.

Etape 2 : Imaginer des scénarios d'évolution pour le territoire (prospective à 15 ans)
La seconde étape du jeu est d'imaginer des scénarios sur la base de l'expression d'enjeux apparaissant déterminants aux acteurs. Ces scénarios d'évolution s'appuient sur une combinaison de dynamiques qui donnent à voir des horizons du futur. Concrètement, chaque joueur (ou petit groupe de joueurs) produit un scénario d'évolution du territoire (matérialisé par un titre, un dessin, une légende, une phrase d'explication du scénario). Les scénarios sont ensuite présentés à l'ensemble des participants de l'atelier. Les dessins sont affichés au mur en rapprochant ceux qui se ressemblent. Cette confrontation des différents scénarios est le support de débats sur ce que les acteurs veulent ou redoutent pour leur territoire.

Etape 3 : Enoncer les pistes d'actions à mener
La troisième étape du jeu se fait en session plénière, après regroupement des différents ateliers. Après une présentation rapide de la maquette et du scénario synthétisé par chacun des groupes, un débat est organisé, pour discuter des pistes d'actions à mener, pour assurer les dynamiques de développement territorial voulues par les acteurs. Ils spécifient les conditions entravant ou facilitant la mise en œuvre de ce scénario sur la base des observations et de l'analyse des caractéristiques territoriales mises en évidence auparavant. Les pistes d'actions sont reportées sur un tableau blanc et emportées par l'acteur partenaire pour s'en servir par la suite.

Le jeu de territoire repose sur quatre principes méthodologiques de l'approche du diagnostic prospectif (spatialiser les informations, articuler les échelles, assurer la transversalité, hybrider les connaissances) et trois principes de participation des acteurs (faciliter l'expression des acteurs, argumenter les choix, être créatif) (Lardon et LeBlanc, 2012). Nous avons expérimenté une version adaptée (sans prospective) dans le contexte spécifique de la constitution du Dire de l'Etat relatif aux enjeux environnementaux et produit des résultats au-delà des simples méthodes d'analyse spatiale, dans le changement de pratiques des agents des services.

2.3. Accompagnement méthodologique à l'expression spatialisée des enjeux environnementaux sur un territoire et à la transversalité des approches entre services

Ce fut donc l'objet d'une collaboration au long cours entre la DREAL PC et AgroParisTech, initialement pour construire ensemble une méthode de diagnostic pour spatialiser les enjeux environnementaux des territoires et les communiquer aux porteurs de projets dans les territoires et progressivement pour changer la façon de partager et de porter le Dire de l'Etat entre agents des services et auprès des acteurs des territoires. L'accompagnement par la recherche et par la formation a constitué une plus-value appréciée aux différents niveaux concernés.

La démarche a été élaborée pour anticiper la prise en compte des enjeux environnementaux dans les projets de territoire, en les formulant clairement en amont de la procédure d'évaluation environnementale, afin que les acteurs s'en emparent et les intègrent dans leur projet de territoire. Cela a impliqué deux services

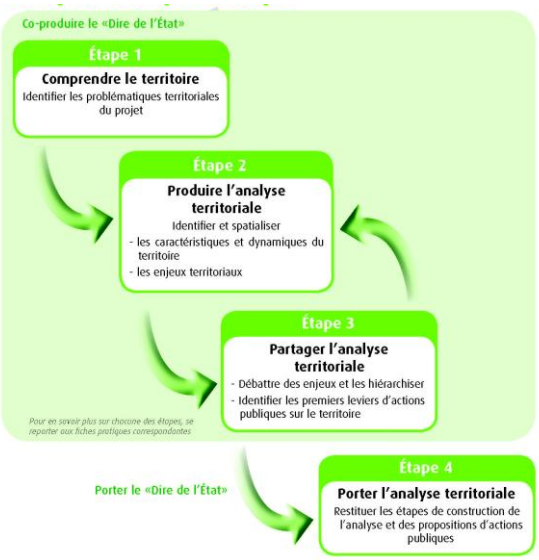
de la DREAL Poitou-Charentes, le service producteur des informations spatiales (DCAT) et le service instructeur des évaluations environnementales (DEE). Elle a pris appui sur les missions en cours des agents DREAL auprès de territoires de la région Poitou-Charentes, à différentes échelles (SCoT, PLU).

Cela s'est fait sur un temps relativement long (trois années) qui a permis l'expérimentation de la démarche, sa formalisation et son appropriation de façon intégrée dans les services. Ainsi, les différentes étapes sont construites et validées au fur et à mesure de leur mise à l'épreuve sur le terrain, faisant de l'intervention des chercheurs à la fois une recherche-action et une recherche participative et de celle des partenaires acteurs, ici les agents DREAL, des expérimentateurs et concepteurs de nouvelles pratiques au sein de leurs services. La formation, comme processus d'apprentissage collectif, a contribué à cette construction et a marqué les jalons de la progression.

3. La Méthode de Diagnostic Partagé

La méthode de diagnostic partagé mise au point vise à assurer la reproductibilité des productions de diagnostics territoriaux réalisés dans le cadre du portage du dire de l'Etat. Elle est conceptualisée en deux phases (co-produire le dire de l'Etat et porter le dire de l'Etat) et quatre étapes (comprendre, analyser, partager et porter) (Figure 1). L'analyse du territoire se fait à l'aide de représentations spatiales schématiques, permettant d'extraire les principes organisateurs de l'espace (structures et dynamiques) et de faire émerger les enjeux environnementaux du territoire les plus prégnants.

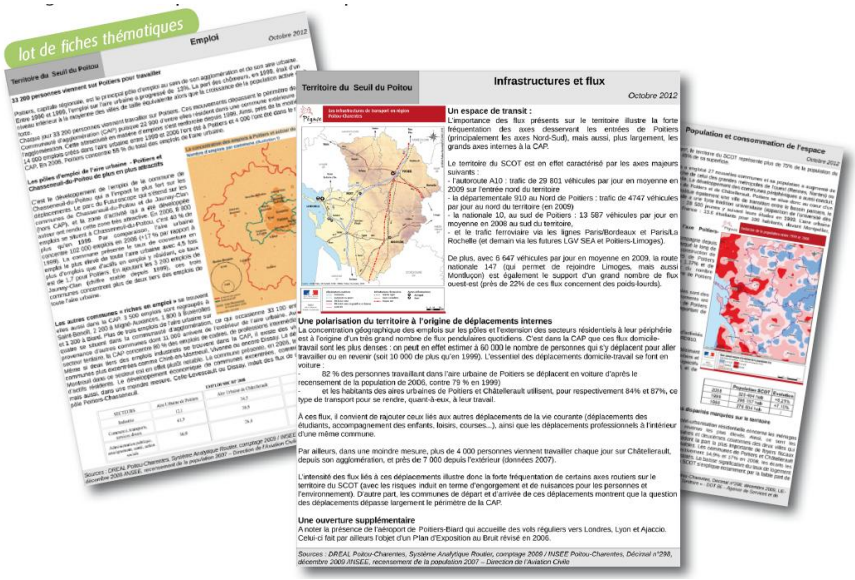
Figure 1. Les quatre étapes de la méthode de diagnostic partagé



3.1. Etape 1 - Comprendre le territoire

Il s'agit d'identifier les problématiques territoriales du projet. Des fiches thématiques sont élaborées par le service DCAT de la DREAL, à partir des données froides disponibles sous SIG : données géographiques (relief, occupation du sol, habitat …), données statistiques (population, emploi …), et des documents réglementaires et bibliographiques complétées par des données chaudes : recueil des connaissances et de la perception du territoire par les acteurs (vécu, usages, connaissance historique...). Ces fiches déclinent les principales caractéristiques du territoire étudié en fonction du projet concerné et des politiques publiques à l'œuvre, tels que définis par les agents du service DEE de la DREAL.

Figure 2. Exemple de fiches produites pour le SCoT « Seuil du Poitou en Vienne »



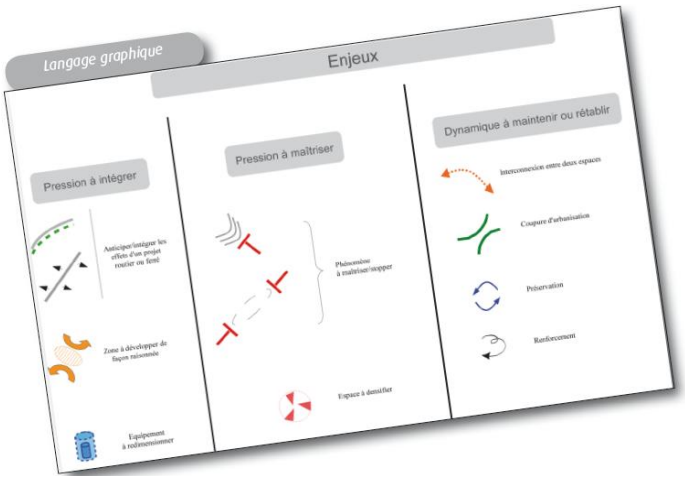
L'originalité de la méthode réside dans le fait que les fiches qui sont réalisées ne présentent plus uniquement une expertise de spécialiste sur les thématiques balayées, mais bien un savoir partagé entre les acteurs et ciblé sur le territoire étudié. En effet, d'une part les fiches combinent formalisation des connaissances scientifiques et façon dont les acteurs vivent et perçoivent leur territoire et, d'autre part, elles se concentrent directement sur les problématiques inhérentes au territoire et sur les marges d'action publiques existantes. De plus, ces fiches sont adaptables et peuvent être enrichies, réutilisées … en tant que de besoin. La méthode facilite donc également leur mise à jour et la construction progressive d'une culture commune des agents DREAL autour de l'information géographique et de son usage. La constitution des fonds de carte, la recherche des sources de données, la distribution des thématiques sont autant de moments d'échanges qui ancrent une appropriation commune des représentations spatiales et de leur interprétation.

3.2. Etape 2 – Co-construire l'analyse territoriale

Il s'agit, dans un premier temps, de spatialiser, de façon partagée, les caractéristiques et dynamiques du territoire afin de comprendre son organisation spatiale. Pour ce faire, les informations disponibles dans les fiches sont mutualisées, en explicitant tour à tour les informations jugées pertinentes pour la construction du diagnostic et en les dessinant sur un fond de carte commun, préalablement constitué. La représentation spatiale se fait en s'aidant d'un alphabet chorématique simplifié. Les structures mettent en évidence le découpage du territoire, les voies de communication et les réseaux qui le drainent et l'irriguent, la hiérarchie des différentes entités et leur rôle au sein du territoire, la spécialisation des différents lieux, les facteurs de ruptures et de discontinuités. Les dynamiques rendent compte des interactions entre les lieux (attractions, rayonnement...), des flux qui traversent le territoire et des circulations préférentielles (personnes, biens, informations...) et de la façon dont l'espace se transforme (processus de front, mitage, mosaïque...). On construit ainsi progressivement la maquette des principales structures et dynamiques, constituant une vision partagée du territoire.

Dans un second temps, la relecture de la maquette permet d'identifier, d'expliciter puis de schématiser, toujours de façon partagée, les enjeux du territoire afin d'affiner l'analyse et de cibler les orientations stratégiques. Un premier énoncé des enjeux relatifs aux dynamiques à l'œuvre est réalisé, puis les thématiques sont regroupées en quelques enjeux qui sont alors représentés spatialement. On s'attache à dessiner à la fois les objets spatiaux issus de la maquette des structures et dynamiques et leurs interactions. Ainsi, par exemple, sur l'enjeu de maitriser les risques d'inondation des zones urbanisées, seront soulignées les rivières à risque et les zones urbaines de proximité, et symbolisées la maîtrise des inondations ou de l'extension urbaine.

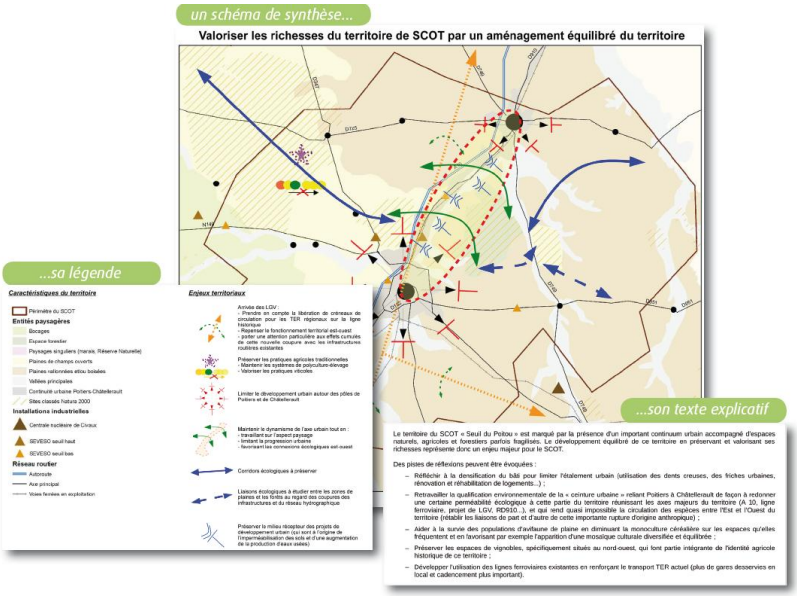
Figure 2. Alphabet chorématique d'enjeux environnementaux



Cette étape permet donc de mutualiser les informations qui ont été récoltées à l'étape précédente afin de produire des représentations schématiques simplifiées, d'identifier les caractéristiques et dynamiques territoriales, puis d'énoncer et de spatialiser les enjeux relatifs au territoire. Ces schémas sont produits sur un fond cartographique localisant les principaux éléments structurants du territoire (réseaux routiers, villes, relief…) à l'aide d'un langage graphique basé sur un alphabet et une grammaire (formes, tailles, couleurs) de chorèmes qui constitue une grille de lecture et d'écriture du territoire (Figure 2).

Ces schémas sont accompagnés de leur légende et du discours associé (Figure 3). Ils permettent ainsi de regrouper, combiner, synthétiser les questions qui se posent au regard de l'avenir du territoire (et cela en fonction des problématiques du territoire et des politiques publiques à l'œuvre) ; d'énoncer les enjeux territoriaux tels qu'ils ressortent de l'analyse des dynamiques ; et d'argumenter ces enjeux au regard des leviers d'action possibles sur le territoire.

Figure 3. Exemple des enjeux environnementaux identifiés pour le SCoT « Seuil du Poitou en Vienne »



L'originalité de la méthode réside donc dans le fait que cette étape permet une réelle intégration de l'environnement au territoire. En effet, ce dernier n'est plus considéré comme une contrainte à subir mais il est bien intégré dès le début du diagnostic en tant que véritable ressource territoriale. Par ailleurs, la méthode, parce qu'elle se base sur la co-construction de l'analyse, facilite une acculturation réciproque entre les agents des services de l'Etat chargés de l'évaluation environnementale et ceux travaillant à la valorisation de la connaissance.

3.3. Etape 3 – Partager l'analyse territoriale

Il s'agit de débattre des enjeux et de les hiérarchiser, pour partager la vision du territoire entre les services de l'État et d'identifier les premiers leviers d'action publique sur le territoire, mobilisables localement par les acteurs concernés. Le partage se fait avec les différents services de l'Etat concernés, sur la base des représentations spatiales produites à l'étape précédente.

Une présentation décompose, élément par élément, le diagnostic effectué, dans l'objectif :

- d'énoncer les critères de choix de hiérarchisation des enjeux au regard des politiques publiques en œuvre sur le territoire : critères réglementaires, acceptabilité du risque, valeur sociétale (principe de précaution, droit à la santé, réalité économique...), valeur patrimoniale ;
- de mettre en débat le diagnostic effectué ;
- de partager la vision du territoire (bonne compréhension, apport de nouvelles connaissances) ainsi que la priorisation des enjeux (s'accorder sur les éléments majeurs) en finalisant les représentations spatiales schématiques ;
- et de proposer les premiers leviers d'action publique qui répondent aux enjeux retenus (meilleure adéquation moyens/objectifs).

L'étape se concrétise par la rédaction d'un document comprenant les représentations schématiques partagées ainsi qu'un argumentaire mettant en avant les principes de pertinence (l'argument doit être légitime et facilement compréhensible), de transparence (la méthode doit être décrite, reproductible et les limites précisées) et de robustesse (les arguments doivent être objectivés et traduire notamment la recherche de précision des informations).

L'originalité de la méthode réside donc dans le fait de produire un Dire de l'Etat partagé et argumenté au regard des capacités d'actions existantes sur le territoire (leviers d'action publique). Le diagnostic ne constituera donc pas seulement un « porter à connaissance » pour les acteurs du territoire, mais plutôt un réel « porter à raisonnement ». Il est validé par le collectif.

3.4. Etape 4 – Porter l'analyse territoriale

Il s'agit de restituer les étapes de construction de l'analyse du territoire aux acteurs du territoire (élus, techniciens, bureaux d'études) et de leur proposer des leviers d'action publique mobilisables localement. Il s'agit bien d'un véritable portage auprès des acteurs territoriaux de l'analyse réalisée par les services de l'Etat. Le document rédigé à l'étape 3, et le diagnostic qu'il contient, décomposé élément par élément (afin de permettre un raisonnement progressif des acteurs et le ré-enrichir au besoin) est présenté dans l'objectif :

- d'inviter les participants à débattre à partir des représentations chorématiques afin de partager la vision du territoire et de ses enjeux ;
- de présenter les principaux leviers d'action publique proposés par l'État et d'inciter les acteurs à les mobiliser dans leurs actions.

L'originalité de la méthode réside donc dans le fait que cette étape permet d'assurer une pleine transparence du Dire de l'Etat sur les territoires. Les acteurs locaux sont intégrés dans la démarche de portage sur les territoires afin de leur garantir une visibilité maximale et une meilleure compréhension du point de vue des services de l'Etat. La concrétisation du projet de territoire concerné est au final facilitée par la capacité des acteurs locaux à réutiliser ce matériau et à s'impliquer en agissant ensuite selon les leviers d'action publiques ou collective disponibles.

4. Discussion

Ainsi, la démarche contribue au développement d'une « culture spatiale et territoriale » au sein des services de l'Etat et à son partage (vocabulaire, langage, construction...), autour des informations géographiques (Noucher, 2012). Les acteurs à associer dans la mise en œuvre de la démarche, sont les producteurs de données, les référents techniques (géomaticiens et cartographes, statisticiens, chargés d'études, experts) et les services de l'Etat accompagnateurs des porteurs de projet (collectivités, associations...). Cette démarche constitue un outil opérationnel de travail inter-services puisque des agents de différents services construisent ensemble le diagnostic et partagent les enjeux à porter sur le territoire. C'est aussi un outil de médiation avec les porteurs de projets de territoire, nécessitant un effort de communication et d'explicitation, voire l'acceptation de se confronter aux autres.

4.1. Les points clés du dispositif

L'appropriation de cette démarche comporte trois verrous à lever :

- Il n'est pas trivial de « jouer le jeu » des représentations spatiales, car tout le monde n'a pas une forme d'« intelligence de l'espace ». Cela demande un apprentissage.
- Il importe de bien comprendre les dynamiques territoriales dans lesquelles s'inscrivent les enjeux environnementaux. La phase amont est primordiale.
- Il est nécessaire d'accepter de se confronter aux acteurs et de réinventer les modalités d'intervention de l'Etat dans la chaine d'ingénierie territoriale.

Le temps long de la collaboration a permis de construire cet apprentissage de l'usage de l'information géographique et des représentations spatiales. Les pratiques de travail en commun entre les services ont évolué du fait de cet itinéraire méthodologique partagé. La place des agents DREAL dans la chaine d'ingénierie territoriale est en recomposition.

4.2. Les acquis pour les services de l'Etat

Les acquis de la démarche dépassent le simple intérêt d'avoir des productions de qualité, reproductibles et appropriables. Elle amène également un changement de pratiques en douceur, chemin faisant, et non pas en force, au sein des services de l'Etat, tant pour les agents, qu'elle accompagne dans leur professionnalisation, que pour les services, qui augmentent leurs compétences collectives en terme d'appui au

développement territorial (Lenormand, 2011). Ainsi, on constate non seulement la construction d'une culture spatiale commune, pour appréhender les enjeux environnementaux, quelles que soient les échelles, de la commune pour les documents d'urbanisme tels que les PLU au grand territoire pour les SCoT, jusqu'à la région pour le profil environnemental régional. Il y a évolution conjointe et amélioration de la chaine d'ingénierie territoriale, par l'apprentissage de relations de travail renouvelées entre les services de l'Etat et les collectivités territoriales, porteuses de projets de territoire.

4.3. Le rôle médiateur de la formation

La formation a eu un rôle crucial dans cette démarche même si ce n'est pas directement visible dans les productions. En effet, une formation continue menée à AgroParisTech sur l'usage des chorèmes et le jeu de territoire est à l'origine de la collaboration. Chemin-faisant les agents DREAL formés sont devenus formateurs dans les sessions annuelles suivantes. Au cœur du processus, c'est une ancienne étudiante de master en géographie qui a assuré l'accompagnement des agents DREAL afin qu'ils s'approprient la méthode, en l'intégrant dans leurs taches au quotidien. Nous avons jalonné l'appropriation institutionnelle par des sessions de formation-action sur site, pour apprendre aux agents de différents services de l'Etat à travailler ensemble en l'expérimentant en temps limité et avec un objectif clairement affiché. Nous comptons sur les retours d'expérience en formation pour approfondir la capacité collective et l'analyse réflexive sur l'action (Cayre, 2013). La façon dont les participants des formations précédentes mettent en œuvre la démarche dans leur situation spécifique, les difficultés rencontrées mais aussi les adaptations réalisées sont autant d'indices d'appropriation de la démarche.

5. Conclusion :

La méthode de diagnostic partagé présente plusieurs propriétés qui la rendent généralisable à d'autres situations pour territorialiser des enjeux environnementaux. Tout d'abord, elle combine des données chaudes avec des données froides, rendant ainsi possible de considérer l'environnement comme une ressource et pas seulement comme une contrainte. Ensuite, elle intègre les différentes dimensions du développement territorial dans l'analyse des enjeux environnementaux, garantissant ainsi la cohérence des actions à mener. Enfin, non seulement elle facilite l'appropriation en donnant aux acteurs des territoires des clés de lecture pour argumenter leurs propres choix, mais aussi elle garantit la légitimité des acteurs de l'Etat sur le terrain. L'accompagnement par la recherche assure la rigueur scientifique de la démarche.

Cette approche illustre les deux enjeux de l'usage des représentations géographiques. D'une part, la représentation de l'information prend différentes acceptions, de la figuration (fond, forme, support) à la représentation mentale de l'espace propre à chacun des destinataires. Ainsi, les enjeux liés au développement durable et à l'approche globale et articulée des problèmes sociaux, spatiaux et

environnementaux, posent plus que jamais la question de la connaissance territoriale et, partant, des informations permettant de la produire et de la partager (Feyt, 2011).

D'autre part, il n'y a pas seulement une multiplicité de producteurs et d'usagers des outils, mais aussi une diversité des traitements et usages de ces outils, selon les acteurs concernés. L'appropriation des outils, l'auto-évaluation de leur usage (avec les détournements ou dérives possibles), et la formalisation des méthodologies générées par les acteurs renouvellent la question de l'accompagnement des acteurs, dans des dispositifs d'ingénierie territoriale (Maurel, 2008).

Ce sont de nouvelles connaissances pour l'action qui sont à élaborer (Avenier et Schmitt, 2007). Elles renvoient au partage des savoirs, entre acteurs de mondes différents (Béguin et Cerf, 2009). Elles passent par la production d'objets intermédiaires (au sens de Vinck, 2009) qui scandent les étapes de la construction territoriale et fournissent des outils de gouvernance territoriale (Rey-Valette *et al.*, 2011). Gageons que cet itinéraire méthodologique produise bien des connaissances pour l'action, dans une perspective de développement durable.

Références bibliographiques

Angeon V., Lardon S., (2003). Dessiner et comprendre le territoire : quand le jeu devient un processus collectif d'apprentissage et de création. *Les figures du projet territorial*. Editions de l'Aube, Datar, Bibliothèque des territoires, p 245-257.

Avenier J-M, Schmitt C., (2007) La construction de savoirs pour l'action, L'Harmattan, 2007, Paris.

Béguin P., Cerf M., (2009), Dynamique des savoirs, dynamique des changements, Octarès, Toulouse.

Cayre, P., (2013). Former « au » métier, former « le » métier. La médiation pédagogique pour accompagner la recomposition du métier d'agriculteur dans l'enseignement agricole. Thèse de doctorat, AgroParisTech Paris.

Feyt, G., (2011) « Les visages et usages de l'information géographique dans le processus de décision territoriale », In Florent Joerin , Martin Schuler , Laurent Thévoz , Olivier Walser , Stéphane Joost , Bernard Debarbieux , Hy Dao, *Les SIG au service du développement territorial*, Presses Polytechniques et Universitaires Romandes.

Gumuchian H., Pecqueur B., 2007. La ressource territoriale, Paris, Ed. Economica. 252p.

Lardon S., Piveteau V., (2005). "Méthodologie de diagnostic pour le projet de territoire : une approche par les modèles spatiaux". *Géocarrefour*, Vol. 80 n° 2.

Lardon S., Roche S., (2008). Représentations spatiales dans les démarches participatives. Production et usages. *Revue Internationale de Géomatique*, n°18(4).

Lardon S, LeBlanc P., (2012). Le jeu de territoire : Bilan d'expériences françaises et québécoises. Colloque ASRDLF « Industries, villes et régions dans une économie mondialisée », Belfort, 9,10, 11 Juillet 2012.

Lardon S., (2011). Chaîne d'ingénierie territoriale : diversité des acteurs dans la conduite d'un projet de territoire. In DAYAN L., JOYAL., LARDON S., *L'ingénierie de territoire à l'épreuve du développement durable*. L'Harmattan, pp 145-161.

Lardon S., (2013). Le « jeu de territoire », un outil de coordination des acteurs locaux. *Revue FaçSADe*, Résultats de recherches du département Inra-Sad, 38, 4 p.

Lenormand P., (2011). *L'ingénierie territoriale à l'épreuve des observatoires territoriaux : analyse des compétences des professionnels du développement dans le massif pyrénéen*. Thèse de doctorat, Université de Toulouse Le Mirail.

Loudiyi S., (2008). Le SCoT, instrument de gouvernance territoriale ? la conduite locale de la concertation dans le pays du Grand Clermont. *Norois*, N°209, 4, pp. 37-56.

Maurel P., (2008). Apprentissage collectif pour décider de l'avenir du territoire de Thau : premiers retours d'une approche combinant politique publique et participation. Conférence OPDE (Les Outils pour Décider Ensemble) - 5-6 juin 2008 - Québec , 25p.

MEDDE, (2011). Les procédures d'évaluation environnementale. (http://www.developpement-durable.gouv.fr/Les-procedures-d-evaluation,12012.html)

Noucher M., (2012). La donnée géographique aux frontières des organisations : des réseaux géomatiques à l'intelligence territoriale, In *Revue Internationale de Géomatique*, 22/1, 2012. Planchat-Hery C., (2008). Les outils graphiques paysagers, révélateurs des enjeux agricoles dans un Plan Local d'urbanisme. *Norois*, N°209, 4, pp. 109-125.

Piveteau V., (2010), L'ingénierie territoriale : pour un parti pris géographique, Habilitation à diriger les recherches, Ecole Normale Supérieure Lettres et sciences humaines, Lyon, 114p.

Rey-Valette H. *et al*, (2011). Guide pour la mise en œuvre de la gouvernance en appui au développement durable des territoires © Cemagref, CNRS, Geyser, Inra, Supagro, Université Montpellier 1. Diffusion INRA-Montpellier, 155 p.

Vinck D., (2009). De l'objet intermédiaire à l'objet-frontière : Vers la prise en compte du travail d'équipement. *Revue d'anthropologie des connaissances*, vol. 3, n° 1, p. 51-72.

Analyse des patrimoines de données géographiques nationaux.

Comparaison de trois infrastructures nationales de données géographiques (France, Brésil, Bolivie)

Julie Pierson[1], Matthieu Noucher[1], Pierre Gautreau[2], Louca Lerch[3], Olivier Pissoat[1], Aurély Jautard[1], Sylvain Lesage[4]

1. UMR 5185 ADESS, CNRS et Université de Bordeaux
Maison des Suds, 12 Esplanade des Antilles, F-33607 Pessac Cedex, France
j.pierson@ades.cnrs.fr ; m.noucher@ades.cnrs.fr ; o.pissoat@ades.cnrs.fr ; a.jautard@gmail.com

2. UMR 8586 PRODIG, Université Paris 1 Panthéon Sorbonne
2 rue Valette, F-75005 Paris, France.
pierre.gautreau@univ-paris1.fr

3. Université de Genève, Département de Géographie et Environnement
Uni Mail, 40 Bd du Pont-d'Arve, CH-1211 Genève 4, Suisse
louca.lerch@unige.ch

4. GeoBolivia - Vicepresidencia del Estado Plurinacional de Bolivia
(jusqu'en janvier 2013)
severo@rednegra.net

RÉSUMÉ. Par le double effet de l'interopérabilité des systèmes et de l'évolution du cadre légal, les patrimoines de données géographiques institutionnelles tendent aujourd'hui à être de plus en plus facilement accessibles à travers la mise en place généralisée d'infrastructures de données géographiques (IDG). L'analyse comparée du contenu de trois IDG nationales (France, Bolivie, Brésil) permet alors de révéler quelques tendances. L'originalité de cette contribution est donc de faire des IDG non pas un support à la recherche d'information mais un véritable objet de recherche. L'analyse des métadonnées des trois catalogues nationaux nous permettra alors de réinterroger les logiques de formalisation et de diffusion des connaissances sur les territoires.

ABSTRACT. Due to the dual impact of systems interoperability and changes in the legal framework, institutional spatial data tends to become more and more accessible through widespread deployment of spatial data infrastructures (SDI). Comparative analysis of the content of three national SDI (France, Bolivia, Brazil) reveals a few tendancies. This contribution's originality lies in using SDI not as an aid to research but as a true object of study. Metadata analysis of three national catalogs offers a way to reconsider the logics of both formalisation and knowledge diffusion through territories.

MOTS-CLÉS : infrastructure de données géographiques ; métadonnée ; géocatalogue ; thésaurus ; emprise géographique ; ISO 19115.

KEYWORDS: spatial data infrastructure ; metadata ; geocatalog ; thesaurus ; bounding box ; ISO 19115.

1. Introduction

La diffusion croissante d'outils, de données, de méthodes et la mise en réseau des SIG rendent aujourd'hui leur analyse incontournable pour comprendre la constitution des politiques publiques en matière d'environnement, en particulier la genèse de certaines normes en matière de représentations et d'évaluation territoriale. Ainsi, le SIG devient un point de passage obligé (d'Alessandro-Scarpari et *al.*, 2008) (Amelot, 2012). Celui-ci tend aujourd'hui à se renforcer avec, sous le fait des évolutions législatives (illustrées par la mise en œuvre – en cours et jusqu'en 2020 – d'un « arsenal juridique » autour de la Directive INSPIRE) et de l'accroissement des capacités techniques (liées notamment aux efforts de normalisation), une diffusion exponentielle de flux de données, en particulier de données géographiques (Noucher et Gautreau, 2013).

Dans ce contexte, les infrastructures de données géographiques qui visent à favoriser la mise en réseau des SIG, en facilitant l'accessibilité aux données géographiques, semblent un objet de recherche intéressant pour comprendre, par l'analyse de leur contenu, les logiques de formalisation et de diffusion des connaissances sur les territoires. Nous proposons ici de revenir rapidement sur la notion d'IDG (chap. 2) puis de présenter les trois IDG nationales analysées (chap. 3) avant d'indiquer les éléments d'analyse des métadonnées qui alimenteront le poster présenté à l'occasion de la conférence SAGEO'13 (chap. 4).

2. Les infrastructures de données géographiques au centre des dynamiques de diffusion des connaissances territoriales institutionnelles.

Une infrastructure de données géographiques (IDG) a pour but de permettre aux utilisateurs un accès direct à des informations et services géographiques de différents fournisseurs. Une IDG se compose ainsi d'un ensemble de technologies, de politiques, de normes et de ressources humaines pour acquérir, traiter, stocker, distribuer et améliorer l'utilisation de l'information géographique (Rajabifard et *al.*, 2003) (Crompvoets et *al.*, 2004). Les IDG existent à différents niveaux (Masser, 2010) selon le principe de subsidiarité. Ainsi, en Europe l'infrastructure européenne de données géographiques repose sur des infrastructures nationales (en France, d'un point de vue technique, le niveau national est représenté par le couple 'géoportail IGN / géocatalogue BRGM') qui elles-mêmes reposent sur des IDG régionales (par exemple, PIGMA en Aquitaine), départementales (par exemple, IGECOM 40 dans les Landes) voire locales.

Ces IDG nous semblent aujourd'hui intéressantes à analyser pour, *a minima*, deux raisons. Premièrement, le mouvement de structuration et de diffusion de l'information géographique institutionnelle au sein des IDG a été initié depuis une dizaine d'années. Par conséquent, on dispose désormais d'un historique

potentiellement intéressant à considérer (analyse rétrospective de la constitution des bases de métadonnées, des emprises disponibles, des échelles de travail, des nomenclatures...). Deuxièmement, malgré l'apparition progressive des normes en vigueur (et des guides de bonnes pratiques qui y sont associées), les stades d'avancement sont encore très différents. Les IDG travaillent donc avec des approches et des priorités thématiques différentes selon les territoires et permettent de comparer des compréhensions / appropriations différentes des normes en cours de constitution.

3. La comparaison des INDG française, brésilienne et bolivienne pour réinterroger les données géographiques disponibles dans des contextes d'usage différents.

Selon les degrés de pénétration des techniques web et géomatiques, des niveaux de conflictualités et des pratiques de décision autour des données géographiques, pour une même thématique les productions de données peuvent radicalement différer. L'étude de contextes sociétaux différents permet donc d'ouvrir une perspective comparative apte à réinterroger les patrimoines de données géographiques disponibles dans les différents systèmes étudiés. Dans cette logique, nous avons étudié trois catalogues de données géographiques issus d'infrastructures nationales du Brésil, de Bolivie et de France. Le choix d'étudier simultanément et de façon comparée les usages sociaux des IDG dans un pays européen (la France) et des pays sud-américains (Bolivie, Brésil) doit permettre en particulier de repérer une grande diversité de bases, du fait d'histoires institutionnelles fortement contrastées, d'identifier des modes divers de circulation de l'information environnementale (des formes les plus institutionnalisées aux plus informelles selon un gradient France-Brésil-Bolivie). Ce choix est également lié à la disponibilité dans ces trois pays de catalogues de données géographiques nationaux dont la structuration des métadonnées repose, en partie, sur la même norme.

Pour chacun de ces catalogues, les fiches de métadonnées au format XML ont été extraites au moyen d'un script en Python créé pour l'occasion. Cette extraction est programmée pour se répéter automatiquement tous les trois mois, la première extraction ayant eu lieu au 1er avril 2013. Les fichiers XML extraits répondent aux normes internationales ISO 19115 (définition du contenu des métadonnées) et ISO 19139 (définition de la structure des métadonnées).

Tableau 1: Nombre de fichiers XML extraits par géocatalogue au 01/04/2013.

Pays	Géocatalogue	Nombre de fichiers XML
Bolivie	http://www.geo.gob.bo/catalogapp/	527
Brésil	http://www.inde.gov.br/	8466
France	http://www.geocatalogue.fr/	35303

On remarque d'emblée une très forte disparité dans le volume de ces catalogues (cf. Tableau 1) qui s'explique à la fois par des historiques différents (le géocatalogue français date de 2007 alors que le géocatalogue bolivien a été lancé en 2012) mais aussi par des modalités de saisie différentes et des organisations territoriales plus ou moins complexes. Par exemple, sur les 35 303 fiches du géocatalogue français, on compte plus de 23 000 fiches produite par la DGFIP[1] et qui concernent le cadastre à l'échelon communal, une couverture continue à très grande échelle sur l'ensemble du territoire. A l'autre extrême, le catalogue de l'IDG bolivienne révèle de grandes disparités en termes d'information disponible suivant que l'on se trouve sur des territoires privilégiés par les acteurs internationaux (coopération inter- et non-gouvernementale, multinationales) ou non. L'analyse des métadonnées ne peut donc se contenter d'un comptage des fiches pour essayer de comprendre des logiques informationnelles des pays étudiés.

4. L'analyse des métadonnées comme révélateur des logiques de formation et de diffusion des connaissances territoriales.

L'analyse du contenu des champs de métadonnées doit permettre de disposer d'une information plus fine susceptible d'alimenter l'approche comparative préconisée. Pour ce faire, des requêtes en langage XQuery permettent d'interroger les 90 balises XML identifiées comme similaires entre les trois catalogues. Trois axes d'analyse, dont les résultats seront présentés sous la forme de cartes et/ou de graphiques statistiques sur le poster, sont privilégiés :

Où : La couverture territoriale des données géographiques est-elle homogène ?

Pour répondre à cette question, nous analysons les emprises géographiques définies dans les fiches de métadonnées. L'extraction des 4 coordonnées des rectangles d'emprise (un même jeu de données peut posséder plusieurs emprises si la couverture est discontinue) est réalisée à partir d'un script en Python qui permet de récupérer des champs d'un catalogue CSW, d'en extraire les balises XML correspondantes aux emprises (<EX_GeographicBoundingBox>) en format CSV pour générer ensuite une couche SIG en format SHP. Un traitement dans un logiciel SIG, permet enfin, pour chaque pays, de produire une carte des densités de données géographiques. Cette *géographie de l'information géographique* révèle alors des répartitions hétérogènes de l'information et des discontinuités spatiales qui ne sont pas le seul fait de disparité démographique. Ainsi, en Bolivie par exemple, l'aire urbaine de Cochabamba apparaît comme beaucoup moins couverte que des secteurs quasi désertique comme l'Oruro ou le Potosi. L'analyse de la couverture géographique des données cartographiques institutionnelles permet donc de révéler des situations potentiellement intéressantes à approfondir pour comprendre les dynamiques territoriales et/ou politiques de constitution des patrimoines de données géographiques.

[1]Direction Générale des Finances Publiques

Quoi : Les thématiques prioritaires sont-elles les mêmes ?

L'analyse des thésaurus et surtout la comparaison des catégories affectées aux 40.000 jeux de données des trois catalogues nationaux étudiés permet d'identifier les thématiques prioritaires et de comparer leur répartition. Cette tâche s'est révélée complexe dans la mesure où les trois catalogues ne respectent que partiellement la norme ISO19915. Il a donc été nécessaire de concevoir une table d'appariement qui assure une jointure entre des méta-catégories et les catégories utilisées dans chaque catalogue national. Ainsi le hiatus entre les catégories INSPIRE et les catégories ISO a, par exemple, nécessité quelques regroupements. Une fois cet appariement réalisé la comparaison entre pays permet de *donner à voir* les préoccupations prioritaires. L'analyse dans le temps de ces comparaisons permettra, à terme, d'identifier l'évolution des thématiques et de révéler ainsi la transformation des questionnements.

Quand : La constitution des patrimoines s'opère-t-elle au même rythme ?

Enfin, l'analyse des informations temporelles permet d'identifier à la fois les grandes phases de constitution des patrimoines de données et de mises en partage. La datation de la constitution peut se faire par l'analyse de trois balises XML relatives à des champs de métadonnées qui concernent la date de création et/ou la date de révision et/ou la date de publication de la donnée. Cependant, comme l'illustre le tableau 2 la saisie de ces trois types de dates est très inégale d'une fiche de métadonnées à l'autre et d'un catalogue à l'autre. Dans ces conditions il devient complexe de ne sélectionner qu'un champ de métadonnées pour comparer les trois catalogues.

Tableau 2 : part des métadonnées renseignée par type de 'date'.

	France	**Bolivie**	**Brésil**
% de métadonnées avec **date de création** renseignée	90	4	54
% de métadonnées avec **date de publication** renseignée	26	71	33
% de métadonnées avec **date de révision** renseignée	20	8	1

Par ailleurs, l'analyse des grandes phases de mise en partage des données peut potentiellement s'opérer à partir des dates de création des métadonnées. Ces dates sont saisies automatiquement à chaque création de fiche. La comparaison sur des trois géocatalogues est donc possible. Elle révèle des rythmes de montées en charge des catalogues différents (progressif au Brésil, par palier en France et en Bolivie). Les ruptures observées peuvent alors s'expliquer par les évolution du cadre réglementaire (par exemple, la publication du règlement INSPIRE concernant les métadonnées le 3 décembre 2008 marque le début d'une montée en charge du

géocatalogue français) ou par des logiques organisationnelles (mise en place de d'ateliers de saisie de métadonnées en Bolivie à partir de mi-2012).

5. Conclusion

L'objectif des extractions et analyses en cours est de pouvoir caractériser l'information présente dans les catalogues de trois pays (Brésil, France, Bolivie), de manière reproductible dans le temps. Au-delà des schémas et cartes issus de ce travail qui seront présentés sur le poster l'intérêt de ce travail est aussi de montrer les difficultés méthodologiques associées à l'évaluation des infrastructures de données géographiques pour l'analyse des patrimoines de données géographiques institutionnelles. Les écarts aux normes internationales se révèlent alors être à la fois une difficulté technique pour engager la comparaison de catalogues différents et une opportunité pour analyser les interprétations divergentes et adaptations locales qui révèlent la variété des dynamiques organisationnelles et territoriales sous-jacentes à ces dispositifs socio-techniques.

Remerciements:

Cette contribution s'inscrit dans le cadre du projet de recherche BAGUALA dont l'objectif est d'analyser les impacts de la diffusion des données environnementales en libre accès sur les pratiques de gestion de l'environnement. Pour plus d'information : http://baguala.hypotheses.org

Bibliographie

Amelot X. (2012). Cartographie participative et web 2.0 pour le développement local et la gestion de l'environnement à Madagascar, empowerment ou impérialisme numérique ?, *Séminaire GRANIT « Cartographie et participation : quand la cartographie critique et la cartographie 2.0 se rencontrent. »,* UMR ADESS, Pessac.

Crompvoets J., Bregt A., Rajabifard A., Williamson I. (2004). Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science,* vol. 18, p. 665-689.

D'Alessandro-Scarpari C., Elmes G., Weiner D. (2008). L'impérialisme numérique. Une réflexion sur les Peace Parks en Afrique australe, *Géocarrefour*, vol. 83, n°1, p. 35-44.

Masser I. (2010). *Building European Spatial Data Infrastructures, Second Edition*, ESRI Press, Redlands, California.

Noucher M., Gautreau P. (2013). Le libre accès rebat les cartes. *Les Cahiers du Numérique,* vol.9, n°1, p. 57-83.

Rajabifard A., Feeney M., Williamson I.P., Masser I. (2003). Chapter 6, National SDI Initiatives. *Development of Spatial Data Infrastructures: from Concept to Reality.* London, Taylor & Francis, p. 95-109.

Capitalisation de l'information géo-historique dans le cadre de la Directive Inondation

Principes et développements de la Base de Données Historiques sur les Inondations (BDHI)

Marlène Villanova-Oliver, Anthony Hombiat, Cécile Saint-Marc, Paule-Annick Davoine, Jérôme Gensel

Laboratoire d'Informatique de Grenoble
681 rue de la Passerelle, Bat D Ensimag
38400 Saint-Martin-d'Hères, France
{marlene.villanova-oliver ; anthony.hombiat ; cecile.saint-marc ; paule-annick.davoine ; jerome.gensel}@imag.fr

RESUME. *La Base de Données Historiques sur les Inondations (BDHI), développée dans le cadre de la Directive Inondation, vise à centraliser les informations documentaires et phénoménologiques relatives aux événements passés. Nous présentons ici la démarche experte qui sous-tend la capitalisation de l'information géo-historique, et le modèle de données qui repose sur une démarche d'archivage historique tout en intégrant une dimension territoriale. Le modèle de données prend en compte la gestion des imprécisions tant sur les aspects spatiaux et temporels que phénoménologiques. La Base de Données est couplée à une interface web qui permet, au moyen de composants cartographiques et temporels, l'intégration et l'exploitation des données.*

ABSTRACT. *The Floods Historical Database (called 'BDHI') has been developed following the European Directive on the management of flood risks. In this poster, we present the expert approach adopted as far as capitalizing on geo-historical information is concerned, before describing the spatio-temporal data model, dedicated to the representation of the content of archives. This data model also handles some of the vagueness aspects unavoidably associated with such historical data. A web interface, built with cartographic and temporal graphical components, make it possible for flood experts to upload data and for a general audience to query the BDHI and visualize past flood events spatial footprints and damages.*

MOTS-CLES : *inondations, données géo-historiques, modèle de données, données imprécises, consultation spatio-temporelle multi-échelles*

KEYWORDS: *floods, geo-historical data, data model, uncertainty, imprecision. spatio-temporal and multi-level consultation*

1. Contexte, problématique et objectifs

La directive européenne 2007/60/CE du 23 octobre 2007 (JO UE, 2007) relative à l'évaluation et à la gestion des risques d'inondation[1] est une des conséquences d'une série d'inondations majeures survenues au début des années 2000 en Europe, et ayant occasionné des pertes humaines et économiques importantes (MEDDE, 2012). Elle établit un cadre commun au niveau européen pour l'évaluation et la gestion globale des risques d'inondations afin de réduire les conséquences négatives pour la santé humaine, l'environnement, le patrimoine culturel, l'activité économique, etc. Elle a donné lieu, au niveau français, au décret n° 2011-227 du 2 mars 2011 qui planifie les grandes actions à mener en termes d'évaluation et de gestion des risques d'inondation. La première étape, appelée Évaluation Préliminaire des Risques d'Inondation (EPRI), vise à assurer "*la description des inondations survenues dans le passé et ayant eu des impacts négatifs significatifs sur la santé humaine, l'environnement, les biens, dont le patrimoine culturel, ou l'activité économique (...)"* (MEDDE, 2012). Cette démarche de portée nationale s'accompagne de la constitution d'une Base de Données Historiques sur les Inondations (BDHI), dont l'objectif est de capitaliser la connaissance relative aux inondations du passé détenues dans ces documents d'archives. Cette capitalisation consiste dans un recensement systématique des documents d'archives et de leur contenu, afin de pouvoir les exploiter, en vue de contribuer à la définition de politiques de gestion du risque d'inondation.

Nous présentons, à travers ce poster, le travail réalisé autour de la BDHI : la démarche adoptée pour la capitalisation de l'information géo-historique liée aux inondations, l'approche suivie pour la conception de la BDHI qui vise à intégrer des données multi-sources (dispersées dans les divers organismes en charge de la gestion des territoires, des risques, du patrimoine, etc.), multiformes (ouvrages, gravures, témoignages oraux, coupures de presse, etc.), et imprécises (dates, localisations, emprises spatiales incertaines) et les développements réalisés.

2. Démarche de capitalisation et d'expertise adoptée dans la BDHI

La démarche de capitalisation des informations documentaires qui sous-tend la BDHI mobilise une certaine expertise aux niveaux historique et hydrologique :

– des techniciens des services publics nationaux (DREAL, DEAL, DDT, etc.) saisissent les données dans la BDHI à partir des documents archivés par leurs services, en renseignant de l'information brute sur chaque source documentaire et les éléments relatifs aux inondations ;

[1] Aussi appelée plus simplement « Directive Inondation »

– des experts analysent ensuite les informations renseignées dans la base et, grâce aux outils mis à disposition, produisent des synthèses sur les événements d'inondation ;

– les gestionnaires de territoires peuvent alors accéder à une information expertisée (recoupement, synthèse, etc.) et l'utiliser dans un processus d'aide à la décision.

3. Le modèle de données de la Base de Données Historiques sur les Inondations

Le modèle de données repose sur une démarche d'archivage historique, qui dissocie les données concernant le document source en lui-même (auteur, date de production, format, lieu de conservation) et les données relatives à son contenu, c'est-à-dire aux inondations décrites dans ce document (période de retour, zones concernées, durée, impacts humains, économiques et environnementaux, etc.). Ce choix est renforcé par le fait que plusieurs inondations concernant des zones ou des époques différentes peuvent être décrites dans un même document source. Le modèle de données de la BDHI s'articule donc autour de trois dimensions :

- les documents qui sont décrits sous la forme de *Fiches Document* ;

- un événement d'inondation décrit dans un document est archivé sous la forme d'une *Note Inondation*. Une *Note Inondation* n'est issue que d'une seule *Fiche Document,* mais une *Fiche Document* peut être rattachée à plusieurs *Notes Inondation* ;

- pour faire face à la multiplicité des sources documentaires décrivant une même inondation, il est nécessaire de pouvoir synthétiser les informations de plusieurs *Notes Inondation* se rapportant en fait à un seul et même événement. A cet effet, des *Fiches Synthèse* permettent d'associer plusieurs *Notes Inondation.*

Le modèle de données tient compte de la qualité des informations qui doivent être intégrées dans la BDHI, notamment relative à leur précision. A titre d'exemples, l'imprécision temporelle est gérée par la notion de *« Date approchée »* ; l'imprécision phénoménologique ou thématique repose sur l'utilisation d'*intervalle de valeurs* ; l'imprécision spatiale est gérée par la possibilité d'associer un événement à un ou plusieurs maillages géographiques de natures différentes : localisation du cours d'eau au bassin versant, à base de maillages administratifs ou encore de découpage du territoire selon la Directive Inondation.

4. L'interface web d'acquisition et de restitution des données de la BDHI

De façon à alimenter la BDHI et à en exploiter le contenu, une application web d'édition et de consultation des données, intégrant des composants cartographiques et temporels, a été développée. Elle repose sur une logique de localisation spatio-temporelle multicritères et multi-échelles de l'information.

L'interface en ligne de la BDHI s'appuie sur le cadriciel PHP Symfony2, qui structure l'application selon le schéma Modèle-Vue-Contrôleur (Reenskaug, 2007). Les données sont intégrées dans une base de données PostgreSQL avec son extension PostGIS. L'interface entre le logiciel et la base de données est assurée par des annotations via l'association objet-relationnel proposée par le système Doctrine2. Enfin, la partie client de l'application utilise le moteur de patrons Twig ainsi que des bibliothèques graphiques Javascript basées sur la sur-couche jQuery pour offrir plus de confort visuel à l'utilisateur.

Cette application ayant vocation à être mutualisée entre tous les acteurs nationaux de la gestion du risque inondation, l'interface web de la BDHI est bâtie sur une hiérarchie de rôles utilisateur qui attribue à chaque type de profil utilisateur un panel de fonctionnalités disponibles, allant de la simple consultation à l'intégration de données.

5. Conclusion

La BDHI est destinée à fournir aux spécialistes du risque inondation un outillage leur permettant de réaliser des expertises sur les événements, parfois très anciens, qui ont touché le territoire en confrontant diverses sources (i.e. différentes *Notes Inondations*). Ces expertises, appelées *Synthèses Inondations*, qui ont vocation à être stockées dans la BDHI, seront utilisées par les gestionnaires du risque dans le cadre de la Directive Inondation pour affiner et compléter leurs prises de décision, en se référant à la dimension historique, mais aussi géographique des événements. Enfin, les données de la BDHI viendront alimenter le portail Géorisques du MEDDE pour la partie inondation. Géorisques est une plate-forme à base de services web (WFS-WMS) consacrée à l'échange et au partage de données et d'informations cartographiques dans le domaine des risques naturels et technologiques.

6. Bibliographie

Ministère de l'Écologie, du Développement Durable et de l'Énergie (2012). M*ieux savoir pour mieux agir : Principaux enseignements de la première évaluation des risques d'inondation sur le territoire français,* http://catalogue.prim.net/190__evaluation-preliminaire-des-risques-d-inondation-nationale.pdf

Journal Officiel de l'Union Européenne, Directive 2007/60/CE du Parlement Européen et du Conseil du 23 octobre 2007 relative à l'évaluation et à la gestion des risques d'inondation, 6 novembre 2007.

Reenskaug T. (2007). *The original MVC reports,* https://www.duo.uio.no/bitstream/handle/10852/9621/Reenskaug-MVC.pdf

SESSION 3 :

Analyse spatiale

Cartographie historique : outil clé pour caractériser la continuité temporelle

Application à des travaux en écologie forestière

Ladet S.[1], Lopez J.[1], Grel A.[1], Valladares L.[2], Brin A.[2]

1. INRA, UMR 1201 DYNAFOR

24 Chemin de Borde Rouge- Auzeville, CS 52627 – 31326 Castanet Tolosan cedex – France

{sylvie.ladet, judicael.lopez, audrey.grel}@toulouse.inra.fr

2. Université de Toulouse, INP-EI Purpan, UMR 1201 DYNAFOR

75 voie du TOEC, BP57611 – 31076 Toulouse cedex 3 - France

{lionel.valladares, antoine.brin}@purpan.fr

RESUME. Les faibles capacités de dispersion de certaines espèces végétales ou animales conduisent à penser que l'ancienneté de l'état boisé est un facteur déterminant pour expliquer la structure actuelle des communautés d'espèces forestières. C'est ce que veut vérifier le projet de recherche Distrafor dans lequel les outils de la géomatique sont mis en œuvre sur des données cartographiques représentant la couverture forestière aux XIX et XXI siècles. Nous proposons d'identifier et cartographier les polygones de forêts à ces deux périodes. Puis par croisement spatial et application d'une analyse multicritère, nous soumettons aux écologues un plan d'échantillonnage stratifié sur les forêts tenant compte de la dépendance temporelle et spatiale et sur lequel des mesures de biodiversité seront effectuées.

ABSTRACT. As some plant or animal species are poor dispersers, the continuity of forest cover is expected to be a key driver of the structure of forest species communities. The research project Distrafor aims at investigating this hypothesis, using the geomatics tools on cartographic data representing the forest cover at the XIX and XXI centuries. We propose to identify and map forests polygons at these two periods. Then by spatial crossing function and implementation of a multicriteria analysis, we submit to the ecologists a stratified sampling design on the forests which takes account of the temporal and space dependence and on which mesures of biodiversity will be taken.

MOTS-CLES : cartographie historique ; continuité temporelle forestière ; carte d'Etat-Major ; forêt ancienne ou récente ; plan d'échantillonnage stratifié

KEYWORDS: landscape history ; wooded temporal continuity ; ordnance survey map ; ancient or recent forest ; stratified sampling design

1. Introduction

La mise en place de la trame verte et bleue (TVB) doit permettre aux espèces de pouvoir migrer au sein des paysages, en particulier sous la contrainte climatique (Bennett, 2003). Les éléments boisés en particulier par leur composition et leur configuration dans les paysages agricoles jouent un rôle essentiel pour la présence d'espèces d'habitat forestier (Bonthoux *et al*., 2012). Or, la distribution actuelle de la biodiversité serait plus le reflet du paysage passé et ce d'autant plus que les taxa sont peu dispersants, ce qui militerait pour la mise en place d'une sous-trame de forêts anciennes sous forme de puits de biodiversité (Dupouey *et al*., 2002 ; Hermy et Verheyen, 2007).

Le projet de recherche Distrafor[1] s'intéresse à la dispersion et la persistance de la biodiversité dans la trame forestière. Par exemple, il est émis l'hypothèse qu'une partie de l'entomofaune vivant dans des microhabitas tels que les cavités arboricoles et le bois mort soit dépendante des forêts anciennes (Emberger *et al.*, 2013). L'objectif général de ce projet est de mieux comprendre comment la trame forestière influence la dynamique spatiale de la biodiversité forestière à travers plusieurs volets et sur plusieurs sites en mobilisant notamment les outils de la cartographie historique. Deux grandes hypothèses sont explorées dans ce projet :

- dû à la dépendance spatiale (effet de la distance), la distribution actuelle de la biodiversité est expliquée par la surface et l'isolement de chaque tache forestière.

- dû à la dépendance temporelle (effet de la banque de graines dans le sol), la fréquence d'occurrence et l'abondance de nombreuses espèces dépendent de l'ancienneté du fragment forestier.

Cet article est un retour d'expériences en termes de géomatique sur la mise en place d'un plan d'échantillonnage cartographique sur l'un des sites d'étude du projet Distrafor. La prise en compte des aspects historiques a été réalisée par le recours à des cartes anciennes qui ont été croisées avec des cartes forestières récentes. Concrètement, il s'agissait de distinguer parmi les boisements actuels, ceux que l'on peut qualifier d'« anciens » car ils étaient déjà présents sur les cartes d'Etat-Major, de ceux que l'on peut qualifier de « récents » car absents de ces cartes anciennes. Le second élément de stratification du plan d'échantillonnage, la surface du boisement, a également pu été pris en compte dans notre démarche.

[1] Distrafor [2012-2014] : projet « Dispersion et persistance de la biodiversité dans la trame forestière » du programme de recherche MEDDE 2010 « Biodiversité, gestion forestière et politiques publiques », coordonné par l'Irstea de Nogent sur Vernisson.

2. Etat de l'art sur le thème du projet vue par une géomaticienne

Partant d'une bibliographie orientée vers l'écologie et la cartographie historique, où les caractéristiques spatio-temporelles sont prises en compte, nous avons retenu les points suivants :

- En géomatique, étudier le temps c'est définir les changements et les événements qui se produisent dans l'espace géographique étudié. Ici un changement peut alors être défini comme une différence significative d'état de notre objet d'étude à savoir la forêt (Plumjeaud, 2011).
- Utiliser des données géographiques (qu'elles sont récentes ou anciennes) au sein d'un projet SIG donne un système pilotable de la dimension spatiale à la dimension temporelle ou thématique, aidant les écologues à répondre aux questions classiques du triade de Peuquet (1994) :
    - Où se trouvait une forêt à un certain moment ?
    - Quand se trouvait la forêt à cet endroit ?
    - Quelle forêt se trouvait à cet endroit à ce moment-là ?
- Les forêts anciennes ont été définies comme ayant existées de façon continue depuis au moins une date spécifiée (date seuil), sélectionnées à partir de l'information sur l'occupation des sols historique disponible (Hermy et Verheyen, 2007) ;
- l'effet des forêts anciennes a souvent été étudiée pour un nombre limité de groupes taxonomiques le plus souvent floristique (Assmann 1999 ; Goßner *et al*., 2008 ; Cousins 2009 ; De Frenne *et al*., 2011). Ces études ont été notamment approfondies par Dupouey dans le Nord-Est de la France (Dupouey *et al*., 2002 ; Sciama *et al*., 2009).

3. Mise au point de la chaine de traitement

3.1 Méthode et site d'étude

La méthode proposée comporte deux grandes phases et démarre d'une base de données géographique constituée à partir de sources de données de l'Institut Géographique National (IGN) accessible gratuitement pour la Recherche Publique. Cette géodatabase exploitée dans le logiciel ArcGis for Desktop 10.1, est composée de deux données principales (i) les cartes anciennes qui sont les minutes d'Etat-Major et (ii) les cartes récentes qui sont la couche vecteur de végétation de la BD Topo. Dans la première phase, il s'agit des prétraitements spécifiques de chacune de ces données de départ; dans la seconde phase, un croisement géométrique suivi d'une analyse multicritère pour établir le plan d'échantillonnage final sont réalisés. En comparant les forêts de 1860 et celles de 2010, il est possible d'isoler les forêts anciennes de superficie plus ou moins grande, dont le couvert boisé est présent depuis plus de 150 ans. Le choix des paramètres imposés par la méthode se fait en fonction des connaissances écologiques.

4

Nous avons appliqué cette chaine de traitement sur la zone d'étude des Vallées et Coteaux de Gascogne du Bas-Comminges qui est un site labellisé LTER (Long Term Ecological Research). Dans cette zone, l'altitude varie peu (300-360 m) et la pente est faible (0-16°). Les forêts, d'une surface de 0,5 à 35 ha couvrant actuellement 18 % de la zone étudiée, appartiennent à un grand nombre de propriétaires privés locaux, souvent des exploitants agricoles, qui les gèrent eux-mêmes et qui sont transmises au cours des générations (Sourdril *et al*., 2011).

3.2 Prétraitements

3.2.1 La couche Végétation ancienne

Les cartes minutes d'Etat-Major (EM) représentent une couverture française du XIXème siècle (noté en abrégé 1800s dans la suite de l'article) de l'occupation des sols en couleurs qui sont numérisées et fournies gratuitement sur le Géoportail de l'IGN mais elles ne sont pas géoréférencées[2]. Les planches aquarellées scannées ont été dessinées par les cartographes officiers de l'époque selon une charte graphique très précise. Les forêts, les prairies, les vignes, … sont toujours figurés selon les mêmes conventions. Par exemple les forêts sont colorisées en vert à vert-jaune et sont souvent délimitées par un liseré noir sachant qu'elles étaient représentées avec soin à l'époque pour éviter les obstacles de progression des militaires. Ces "dessins-minute" fournissent ainsi un précieux témoignage sur l'occupation des sols ou la transformation du paysage. Pour couvrir notre zone d'étude des Coteaux, il nous faut télécharger 3 dalles de minutes en format jpeg2000, à la résolution de 600 points par pouce (ppp) et à l'échelle 1:40000ème (tableau 1). Les différentes étapes de prétraitement de cette donnée sont : (i) la vectorisation des contours d'usage des sols par digitalisation manuelle de type polygonale puis (ii) le géoréférencement de chaque carte et de chaque digitalisation réalisé en deux fois : d'abord par une méthode affine puis élastique (figure 1). Le géoréférencement est réalisé après la vectorisation car cela permet un confort de digitalisation sans qu'il y ait eu de déformation pixel due au géoréférencement du raster. Pour le détail de ces étapes, nous avons suivi scrupuleusement le manuel de digitalisation des cartes anciennes fournies par Favre *et al*. en 2012. Ce dernier est diffusé à la communauté scientifique pour aider ceux qui veulent réaliser, à partir de la carte d'Etat-Major, la carte des usages du sol dans une région particulière et harmoniser les pratiques de digitalisation de cette carte ancienne. De cette façon, les réalisations en cours dans d'autres organismes sur diverses régions de France peuvent concourir à la réalisation d'une carte nationale des usages anciens du sol. Par exemple Favre *et al.* (2012) préconise des bonnes pratiques formalisées dans le manuel, pour chaque étape :

---

[2] http://fr.wikipedia.org/wiki/Carte_d%27%C3%89tat-Major. [1848- 1872] Elle a été réalisée dans la projection de Bonne sur l'ellipsoïde de Plessis qui impose que les parallèles soient des cercles concentriques équidistants, et que l'échelle le long des parallèles soit constante, et égale à celle du méridien d'origine. Le parallèle origine était le 45e nord, et le méridien d'origine est celui de Paris.

- pour la vectorisation, des règles de digitalisation standardisée pour obtenir la couche vecteur polygone des occupations des sols ont été rédigées. En général chaque tâche forestière est délimitée par un trait noir, dans ce cas il faut suivre ce trait. Il est possible de rencontrer des cas particuliers pour lesquels des règles de digitalisation sont proposées (figure 2). Histoire de conserver cette information de cas particuliers, nous avons créé un champ booléen « qualité-digit » dans la table attributaire de la couche crée qui nous permettra de sélectionner ces polygones dans l'analyse multicritère (cf partie 3.3.1).

- pour le géoréférencement en deux temps, d'abord commencer avec une affine pour caler globalement chaque minute d'EM en utilisant les points géodésiques repris dans la légende de la carte puis avec un étirement élastique pour améliorer le calage de la digitalisation. La déformation élastique d'ajustement spatial sous ArcGis utilise deux réseaux de triangulation irréguliers (TIN) temporaires pour interpoler les modifications des coordonnées d'entités x (dX) et y (dY) le long des liens spécifiés par l'utilisateur (figure 3). Chaque TIN possède la même structure de triangulation. Le début des liens de déplacement et tous les liens d'identité représentent les angles des triangles TIN (nœuds). Pour pouvoir faire l'étirement élastique et visualiser la déformation correctement il a été plus pratique de digitaliser d'autres occupations des sols en plus des forêts à savoir des éléments du bâti ainsi que des routes et des rivières principales.

Tableau 1. Description des cartes minutes utilisées dans le projet DISTRAFOR.

Nom de la carte d'Etat-Major	Date des levés	Nom des cartographes ayant réalisés les levés
241 NE – St Gaudens	1849	Mr Péro et Mr Petit
230 SO - Toulouse	Inconnue (1848 selon l'IGN)	Mr de Boutaud et Mr Fourcade
229 SE - Auch	Inconnue (1849 selon l'IGN)	Mr Maurice et Mr Garédanne

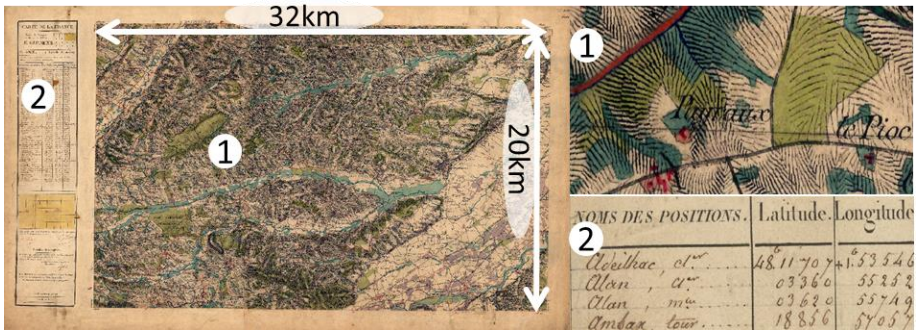


Figure 1. Illustration de la minute n°241NE de surface couverte 640km². L'encart n°1 est un extrait de la planche aquarellée où figurent ici en vert les forêts, en bleu les prairies et en rouge le bâti. Et l'encart n°2 est un extrait des points géodésiques de la carte utiles pour le géoréferencement affine.

6

<Sur les limites des polygones>
SI un linéaire délimite une forêt **ALORS** la limite sera placée au milieu du linéaire.
SI plusieurs linéaires délimitent une forêt **ALORS** la limite sera placée au milieu du linéaire adjacent à la forêt.
SI aucun linéaire ne délimite une forêt **ALORS** les sommets des polygones digitalisés seront placés aux limites des deux figurés surfaciques.
SI le coloriage est en désaccord avec un trait de délimitation **ALORS** la limite du polygone suivra le trait de délimitation.
SI la limite d'un polygone est sous un caractère de la toponymie **ALORS** ne pas mettre de sommet sur le caractère, les placer avant et après l'imprimé.
<Cas de juxtaposition avec une limite administrative (figuré linéaire épais rouge)>
S'il s'agit du seul figuré qui sépare 2 occupations du sol **ALORS** cette limite ne sera pas utilisée pour délimiter les polygones.
SI une tache forêt est traversée par ce figuré **ALORS** la forêt sera considérée comme une entité unique.
<Cas d'usages différents dans un polygone forêt>
SI un bâti ou autre usages est inclus **ALORS** on détoure du polygone « non forêt » etc.

Figure 2. Règles de digitalisation des minutes d'Etat-Major.

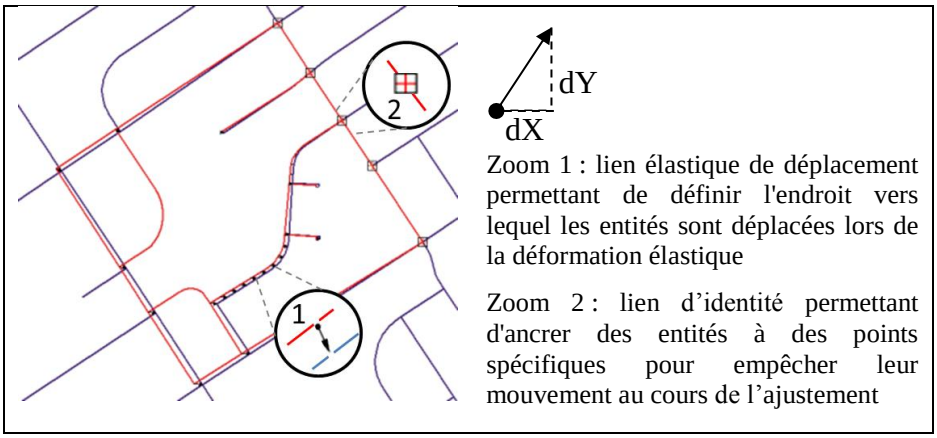


Figure 3. Principe du deuxième géoréférencement par déformation élastique d'ajustement spatial utilisant des liens de déplacement et des liens d'identité.

3.2.2 La couche Végétation actuelle

En parallèle, dans le but de comparer les forêts cartographiées en 1800s et les forêts actuelles, la vectorielle végétation de la BD TOPO version 1 (sans distinction de la nature des zones arborées) a été utilisée. Elle est obtenu par l'IGN par segmentation orientée objet automatique des photographies aériennes de la BD ORTHO (IGN) datant de 2010. Nous souhaitons exclure les éléments boisés tels que les haies, les arbres isolés pour lesquels il n'y pas d'équivalence cartographiée dans la carte de 1800s. Nous avons testé des outils fonctionnant en mode vectoriel (ex :

Polygon_Thickness de l'outil Easy Calculate 10), non concluant pour supprimer des haies longues et épaisses dont les caractéristiques de largeur et épaisseur de polygone se confondent avec des petites forêts. Pour localiser les polygones boisés de surface et de compacité suffisantes correspondant à des forêts, nous choisissons donc d'utiliser un outil en mode raster déjà disponible au laboratoire, GUIDOS[3] (Graphical User Interface for the Description of image Objects and their Shapes). GUIDOS distingue, à partir d'une carte forestière binaire, plusieurs types de couverts boisés selon leur configuration spatiale. Cette analyse est appelée MSPA (pour Morphological Spatial Pattern Analysis) et fait intervenir différents filtres morphologiques à paramétrer (traitement des formes) (Vogt *et al.*, 2007 ; Ostapowicz *et al.*, 2008 ; Soille et Vogt, 2009 ; Vogt, 2010). MSPA consiste en une séquence personnalisée d'opérateurs mathématiques morphologiques ciblés sur la description de la géométrie et la connectivité des éléments de l'image. Cette méthode identifie, en fonction des paramètres introduits et en adéquation avec la définition que nous avons adoptée, des bois de taille et de compacité suffisantes pour contenir une lisière (appelée edge) et un cœur de bois (appelé core). (Alignier *et al.*, 2011), ces 2 objets sont alors combinés en une même classe en sortie. Par cette méthode, sur la couche boisée actuelle, nous pouvons sélectionner les polygones de forêts actuelles (figure 4).

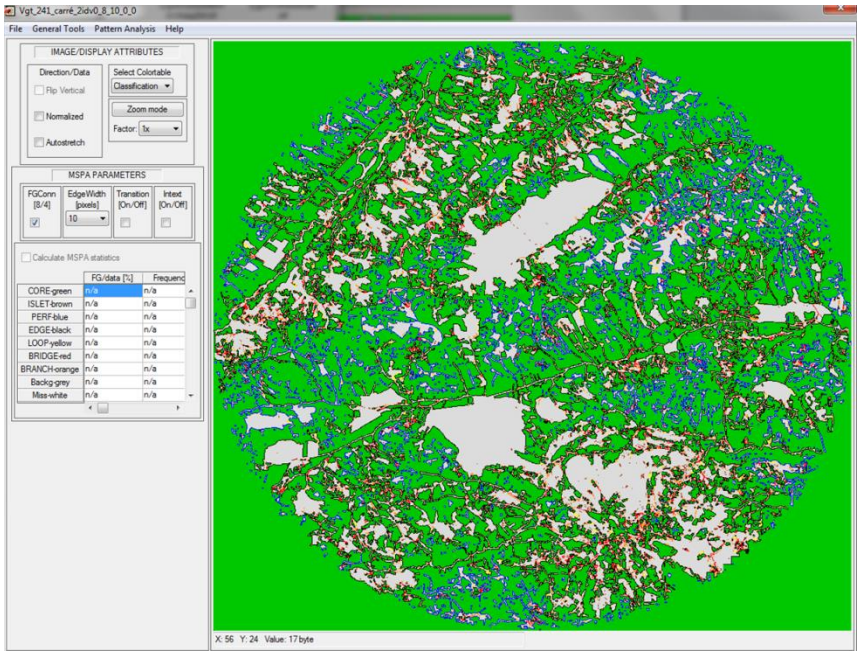


Figure 4. Interface de GUIDOS avec le résultat de l'analyse MSPA sur la couche végétation de la BD TOPO. En gris figurent les forêts. Les autres couleurs correspondent aux autres types d'éléments boisés.

[3] http://forest.jrc.ec.europa.eu/download/software/guidos/

3.3 Datation et typologie des forêts

3.3.1 Définition des critères de croisement et d'analyse multicritère

Le croisement des deux couches vecteur prétraitées est une union géométrique. Cette fonction nous permet d'aboutir à une carte de continuité historique à 3 classes : « continu » ou « noyaux forestiers anciens » pour les zones non défrichées depuis 1800s au moins et pouvant être considérées comme anciennes ; « récent » ou « autres boisements actuels » pour les zones présentes seulement après 1800s ; « disparu » ou « terrains déboisés à ce jour » pour les zones défrichées après 1800s (Vallauri *et al*. 2012).

Une analyse multicritère est enchainée à ce croisement et permet de sélectionner les forêts aptes à êtres échantillonnées selon les critères (surface, isolement, pourcentage d'ancienneté,…) définis dans le projet Distrafor (tableau 2).

Tableau 2. Description des paramètres utilisés dans les étapes du processus appliqués aux données des sites d'étude, détaillées ultérieurement. A=pourcentage d'ancienneté, S= surface des polygones (ha) et D= distance entre forêts (km).

Phases	Processus	Paramètres	Valeurs
PRETRAITEMENT	Digitalisation des polygones de forêts anciennes	Echelle de digitalisation	1:1500ème
	Géoréferencement	Erreur maximum en géoréférenecment affine	RMS inférieur à 50 mètre avec aucun RMS de point supérieur à 100 mètres, par minute de carte
		Nombre de points d'amers en géoréférencement élastique	Au moins 500 points par minute de carte au prorata de la surface géoréférencée
	Traitement sous GUIDOS : analyse morphologique MSPA sur une image raster de résolution 2 m obtenue à partir de la vectorielle végétation de la BD Topo	Pas de rastérisation de la couche vecteur	2 m
		Méthode de rastérisation	Maximum combined area
		Foreground Connectivity, ensemble de 3 x 3 pixels. Le pixel central est connecté aux pixels adjacents.	8 pixels
		Edge Width définit la largeur ou l'épaisseur en pixel des classes *non-cœur*. La distance réelle en mètres correspond au nombre de pixels du bord multiplié par la résolution en pixel de l'image.	10 pixels = 20 m

		Transition, pixels situés sur une lisière ou une perforation où le cœur est intersecté avec une boucle ou un pont. Si la transition est mise à 0 alors les bords des cœurs seront fermés.	0
		Intext permet de distinguer les caractéristiques internes de l'extérieur, où les éléments internes sont définis comme étant délimités par une perforation.	0
TRAITEMENT	Analyse multicritère	GA =grande forêt ancienne	9.5ha≤S≥40ha et A ≥ 60
		PA= petite forêt ancienne	0.8ha≤S≥2.5ha et A ≥ 60
		PR= petite forêt récente	0.8ha≤S≥2.5ha et A ≥ 5
		Isolement triplets GA-PA-PR	D<5km

4 Résultats

4.1. Du point de vue géomatique

Avec les outils de la géomatique nous avons pu obtenir 19 triplets de forêts, localisés sur les 3 cercles de prospection qui ont ensuite été validés sur le terrain et pour lesquels on a demandé l'autorisation aux propriétaires de faire nos mesures (figures 5 et 6 et tableau 3).

Tableau 3. Description des paramètres utilisés dans les différentes étapes du processus appliqués aux données des sites d'étude, détaillées ultérieurement.

Type de forêts	Surface moyenne	% d'ancienneté moyen	Distance moyenne 2 à 2 /triplet
GA	28 ha (± 18.7)	81.4 % (± 7.6)	1.3 km (± 0.5)
PA	1.8 ha (± 0.9)	70.7% (± 25.9)	
PR	1.6 ha (± 0.6)	0.9 % (± 3.4)	

10

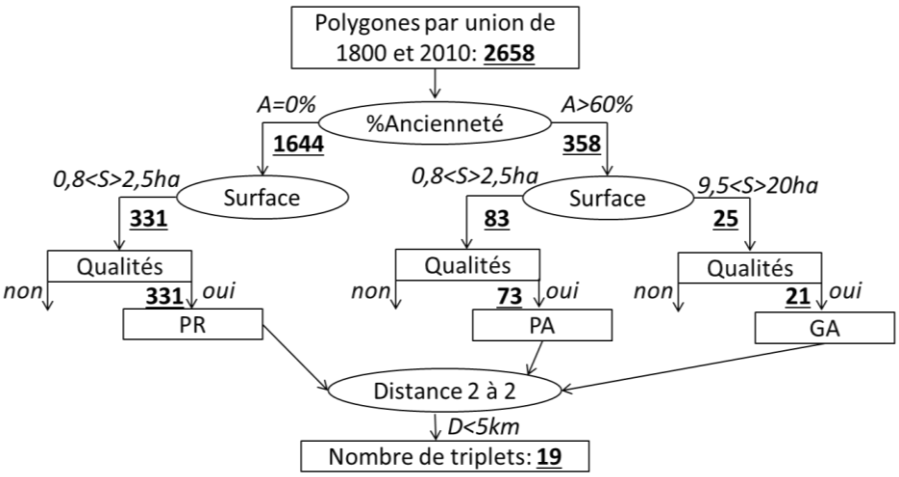


Figure 5. Diagramme d'effectifs de forêts le long de l'analyse multicritère avec A=pourcentage d'ancienneté, S= surface des polygones et D= distance entre forêts.

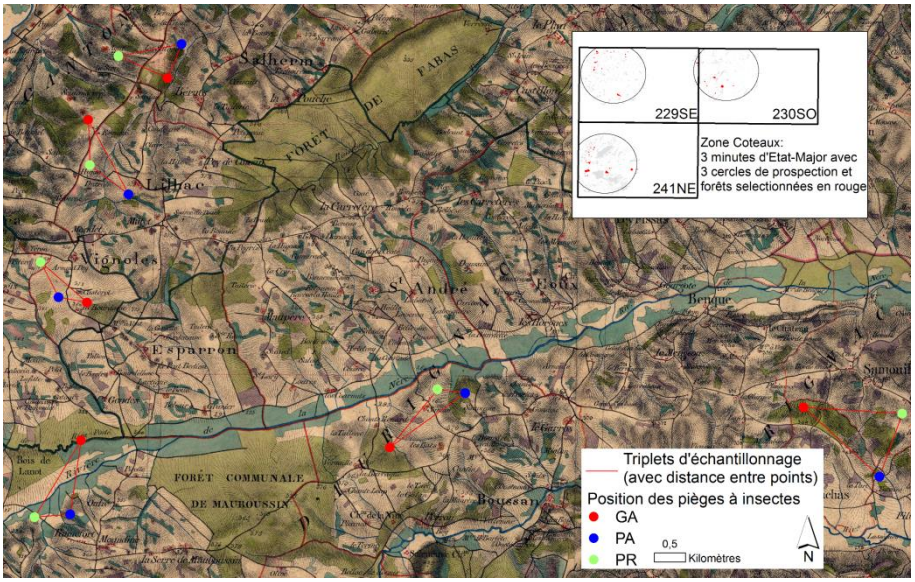


Figure 6. Illustration sur la minute n°241NE des 6 triplets sélectionnés parmi les 3 cercles cartographiés dans l'encart.

Le choix des minutes des cartes d'EM comme source historique est adaptée car (i) le type de projection choisi à l'époque, la projection de Bonne, est basé sur la conservation des surfaces d'où une prise d'assurance sur la comparaison possible de cette couche avec celle de maintenant (ii) les entités forêts de cette carte sont des objets facilement repérables et digitalisables manuellement et (iii) la période 1800s

correspond au minimum forestier connu (Dupouey *et al.*, 2007). Pour la couche forestière récente, la vectorielle de la BD Topo au moment de l'étude s'est avérée être la plus pertinente et la plus actualisée. Depuis peu, la fusion de l'IGN avec l'Inventaire Forestier National permet de bénéficier de la couche BD Forêt, plus précise sur les types de peuplements forestiers.

Des difficultés subsistent quant à la fourniture à cette méthode de données géographiques prétraitées fiables et au calage des paramètres de réglage (seuil de surface, paramètres MSPA,…). Il a été nécessaire de tester différents paramètres (i) pas et méthode de rastérisation de la couche vecteur récente pour l'analyse MSPA, par exemple pour conserver ceux qui altèrent le moins l'information sur l'objet « forêt » en termes de qualité et quantité ou (ii) seuils de classe pour l'analyse multicritère pour tenir compte du contexte forestier de notre zone d'étude (taille moyenne des forêts). Cette analyse de sensibilité n'est pas reportée dans ce document. La déformation élastique par étirement, classée dans les méthodes locales de transformation utilisée après une transformation globale (affine) permet d'augmenter la précision des polygones de forêts en 1800s (tableau 4).

Tableau 4. Paramètres des deux géoréferencements successifs avec RMS= erreur quadratique moyenne de chaque géoréferencement sur la référence SCAN 25 (IGN). Pour chacun est rapporté le nombre de points d'amers dont les points géodésiques présents dans le cartouche de la carte et le nombre de points de liens élastique.

	1er géoréferencement		2ième géoréferencement	
Carte	Nb amers total (dont g)	RMS_1 (m)	Nb liens élastique	RMS_2 (m)
229	51 (23)	44.6	755	6.8
230	36 (29)	28.7	525	2.2
241	55 (36)	37.3	1099	1.4

4.2. Du point de vue écologique

Après la validation terrain réalisée pour vérifier la nature et l'homogénéité du peuplement forestier et l'obtention de l'accord des propriétaires, nous avons retenu 15 triplets de forêts GA-PA-PR. L'objectif des analyses écologiques est d'évaluer, a posteriori, le plan d'échantillonnage au regard de la ressource locale en bois mort car il s'agit d'un élément déterminant pour expliquer la structure des assemblages de Coléoptères saproxyliques. Une première exploration graphique des données à l'aide de boîtes à moustache permet de constater qu'il y a de légères différences en termes de volume comme de diversité en bois mort entre les 3 types de forêts identifiés par analyses géomatiques (figure 7). Des ANOVA nous permettent de considérer cependant qu'il n'y a pas de différence significative (au risque alpha=0.05) entre les 3 types, que ce soit pour le volume (F=2.02, P=0.145) ou la diversité (F=0.925, P=0.404).

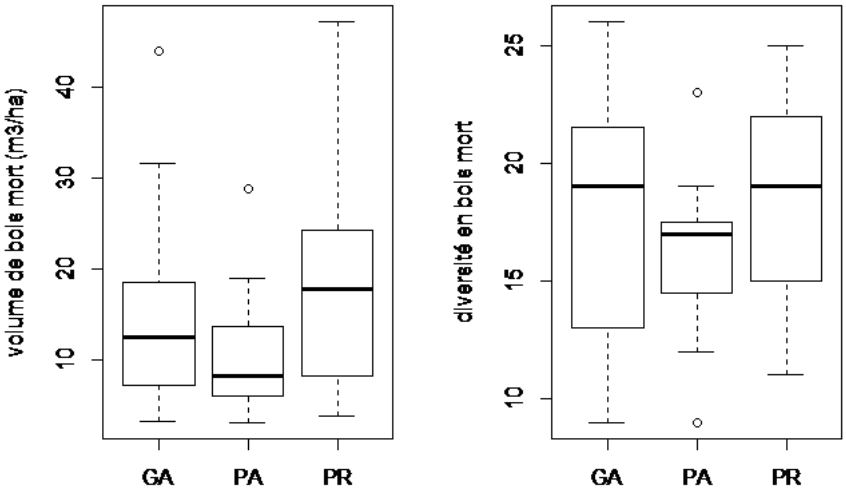


Figure 7. Volume de bois mort total (à gauche) ou diversité en bois mort (à droite) en fonction du type de forêt.

Une analyse multivariée permet de compléter l'exploration de ces données. On considère cette fois-ci le volume de chaque combinaison issue du croisement des facteurs « diamètre » et « stade de décomposition ». Lorsque l'on projette les sites dans le premier plan factoriel d'une ACP basée sur ces descripteurs de la ressource en bois mort, on constate que globalement, les 3 types de forêts ne s'individualisent pas sous la forme de 3 nuages de points distincts (figure 8).

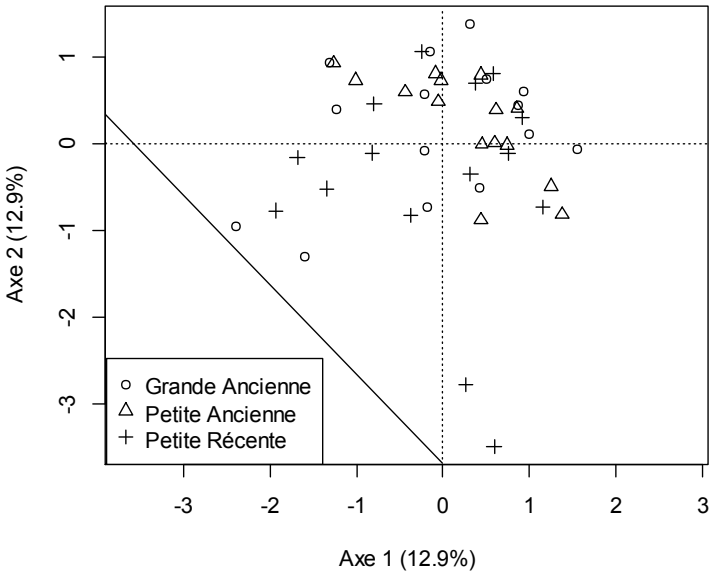


Figure 8. Projection des 45 sites dans le premier plan factoriel d'une ACP basées sur 36 descripteurs volumiques du bois mort au sol.

A ce jour, à l'issu d'une première phase d'identification des échantillons du premier relevé, 130 espèces saproxyliques ont pu être identifiées. Près de la moitié (60) de ces espèces n'ont été observées que dans 1 ou 2 sites. Si on ne tient pas compte de ces espèces « rares », ce sont près des 2/3 du pool régional qui est commun aux trois types de forêts (figure 9). On notera toutefois que seule la modalité « petite récente » n'abrite aucune espèce propre. Les deux autres modalités ont des contributions propres qui s'élève à 2 espèces chacune. Il s'agit de *Dissoleucas niveirsotris* et *Soronia grisea* pour les grandes anciennes et de *Bitoma crenata* et *Oedemera podagrariae* pour les petites anciennes. Huit espèces sont communes aux forêts anciennes tout en étant absentes des petites récentes. Cette première approche qualitative des résultats nécessite d'être complétée par des analyses quantitatives plus détaillées. Celles-ci seront conduites lorsque l'ensemble des relevés auront été traités.

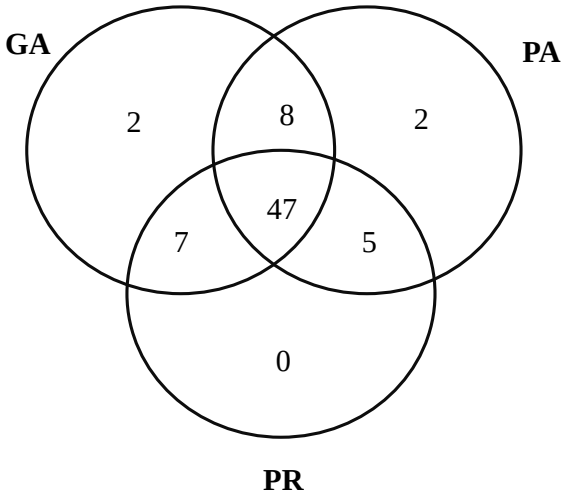


Figure 9. Diagramme de Venn illustrant la répartition des espèces entre les 3 types de forêts étudiées (GA : grande ancienne, PA : petite ancienne, PR : petite récente). Les espèces singletons et doubletons ont été exclues.

5 Discussion

Nous avons proposé une stratégie d'échantillonnage tenant compte des dépendances temporelles et spatiales, susceptible de guider la campagne de terrain mise en œuvre par la suite. Cela permet aussi d'accroitre l'efficience de cette dernière car ce travail cartographique préalable permet (i) de croiser dans le SIG ces informations géolocalisées avec la couche parcellaire (BD Parcellaire de l'IGN) et ainsi faciliter la recherche des propriétaires forestiers à contacter pour avoir les autorisations de piégeages et mesures in situ (ii) de construire un atlas

cartographique précis permettant un gain de temps dans la localisation pour les équipes techniques.

Cependant, la manipulation et le traitement concret de ces données à des fins de croisement posent des problèmes méthodologiques complexes, liés en particulier à la qualité et à leur précision géométrique et qui font référence à des difficultés liées à la dépendance spatio-temporelle. Les contours des bois ont été digitalisés en tentant de respecter des règles sur l'ensemble des zones étudiées, tout en tenant compte des différents cas de figures. Cette difficulté à établir parfois des limites des forêts digitalisées introduit une incertitude sur leur contour et donc sur leur surface calculée, difficile à estimer mais très localisée. Dans le cas présent, où la surface du boisement est un facteur étudié, quelles possibilités sont finalement offertes aux écologues par l'approche géomatique ? Autrement dit, quelle précision aurait-on pu envisager dans le gradient de surface ? Dans Distrafor dans un premier temps pour réduire les effets de dépendance spatio-temporelle, nous ne considérons seulement que deux catégories de surfaces bien distinctes pour définir des « petites « forêts (0.8ha<Surf≥2.5ha) et des « grandes » forêts (9.5ha<Surf≥40ha) et nous ne travaillons qu'à deux dates seuil extrêmes (1800s et 2010). La dépendance sémantique a peu été prise en compte dans cette étude. Comment s'assurer que chaque typon saisi pour la carte de 1800s est de même nature en 2010? Pour un sous-échantillon de forêts actuelles (anciennes ou récentes), une validation sur le terrain à dire d'expert permettrait de vérifier si ce sont bien des forêts et d'évaluer leur degré d'ancienneté (Emberger *et al.*, 2013).

6 Conclusion- Perspectives

Dans le domaine de l'écologie forestière, l'accès à ces données géographiques anciennes via l'utilisation de systèmes d'information géographique (SIG) a permis une meilleure appréhension de la dimension historique en offrant des nouvelles fonctionnalités de traitement et d'exploitation des cartes anciennes.

Plusieurs perspectives sont envisagées dans ce travail. Citons par exemple l'automatisation des méthodes de prétraitements (vectorisation par segmentation objet et géoréférencement automatique) des cartes anciennes (Herrault *et al*, 2013a ; 2013b) ce qui permettra de développer des études similaires sur de grandes étendues. D'autre part, la prise en compte plus systématique et quantitative de l'incertitude tout au long du processus cartographique mis en œuvre permettrait aux écologues forestiers d'intégrer ces données dans leurs analyses statistiques (Lechner *et al*., 2012). Etant donné que la période d'acquisition des minutes d'Etat-Major couvre plusieurs décennies et que les auteurs de ces cartes sont nombreux, il aurait été pertinent de vérifier l'état boisé de 1800s via l'usage du cadastre napoléonien qui date globalement de la même période et dans lequel la nature des parcelles était renseignée en étiquettes sur la planche cartographiée. Pour étudier l'évolution des forêts entre 1800s et 2010 nous pourrions utiliser plusieurs cartes intermédiaires (photo aérienne, cadastre rénové). Ceci permettrait de localiser de manière plus sûre les forêts anciennes n'ayant subi aucun déboisement et de confirmer leur contour. Dès lors, l'écologie historique sortira de l'étude de cas particulier qui, par leur

histoire et par les documents disponibles sur celle-ci, sont considérées comme exemplaires.

Remerciements
Ce travail s'insère dans le cadre du projet DISTRAFOR du programme de recherche MEDDE 2010 « Biodiversité, gestion forestière et politiques publiques » n° de convention 10-MBGD-BGF-2-CVS-103.

7 Bibliographie

Alignier A., Espy P., Deconchat M., Ladet S. (2011). CARTOLIS. Vers un outil géomatique pour identifier et caractériser les segments de lisières forestières. *Revue Internationale de Géomatique*, Numéro spécial « Analyse spatiale 2D et 3D : de l'écologie des paysages à la modélisation de processus urbains », vol. 21, n°4, p. 443- 467.

Assmann T. (1999). The ground beetle fauna of ancient and recent woodlands in the lowlands of north-west Germany (Coleoptera, Carabidae). *Biodiversity and Conservation*, vol. 8, p. 1499-1517.

Bennett A.F. (2003). *Linkages in the landscape. The role of corridors and connectivity in wildlife conservation*. IUCN, Gland Switzerland and Cambridge, UK. xiv + 254 p.

Bonthoux S., Barnagaud J.Y., Goulard M., Balent G. (2012). Contrasting spatial and temporal responses of bird communities to landscape changes. *Oecologia*, DOI 10.1007/s00442-012-2498-2.

Brin A., Brustel H., Jactel H. (2009). Species variables or environmental variables as indicators of forest biodiversity: a case study using saproxylic beetles in Maritime pine plantations. *Annals of Forest Science*, vol. 66, n°3, DOI 10.1051/forest/2009009

Cousins S.A.O. (2009). Landscape history and soil properties affect grassland decline and plant species richness in rural landscapes. *Biological Conservation*, vol. 142, n°11, p. 2752-2758.

De Frenne P., Baeten L., Graae B.J., Brunet J., Wulf M., Orczewska A., Kolb A., Jansen I., Jamoneau A., Jacquemyn H., Hermy M., Diekmann M., De Schrijver A., De Sanctis M., Decocq G., Cousins S.A.O., et Verheyen K. (2011). Interregional variation in the floristic recovery of post-agricultural forests. *Journal of Ecology*, 99: 600-609.

Dupouey J.L., Dambrine E., Laffite J.D., Moares C. (2002). Irreversible impact of past land use on forest soils and biodiversity. *Ecology*, vol. 83, n°11, p. 2978-2984.

Dupouey J.L., Bachacou J .Cosserat R., Aberdam S., Vallauri D., Chappart G.,Corvisier-de Villèle M.A. (2007). Vers la réalisation d'une carte géoréférencée des forêts anciennes de france. *Le Monde des Cartes*, vol. 191, p. 85-98.

Emberger C., Larrieu L., Gonin P. (2013) Dix facteurs clés pour la diversité des espèces en forêt. Comprendre l'Indice de Biodiversité Potentielle (IBP). Document technique. Paris : Institut pour le développement forestier, 56 p.

Favre C., Granier E., Cosserat-Mangeot R., Bachacou J., Dupouey J.L. (2012). Digitalisation des cartes anciennes. Manuel pour la vectorisation de l'usage des sols et le

géoréférencement des minutes 1:40 000 de la carte d'Etat-Major. Rapport INRA, Version 11.2, 41 p.

Goßner M., Engel K., Jessel B. (2008). Plant and arthropod communities in young oak stands: Are they determined by site history ? *Biodiversity and Conservation*, vol. 17, n°13, p. 3165-3180.

Hermy M., Verheyen K. (2007). Legacies of the past in the present-day forest biodiversity: a review of past land-use effects on forest plant species composition and diversity. *Ecological Research*, vol. 22, n°3, p. 361-371.

Herrault P.A., Sheeren D., Fauvel M., Monteil C.,Paegelow M. (2013a). A comparative study of geometric transformation models for the historical "map of France" registration. *Geographia Technica*, n°1, p. 34- 46.

Herrault P.A., Sheeren D., Fauvel M., Paegelow M. (2013b). Vectorisation automatique des forêts dans les minutes de la carte d'Etat Major du 19ème siècle, *Actes du colloque SAGEO 2013*, Brest, 14 p.

Lechner A.M., Langford W.T., Bekessy S.A., Jones S.D. (2012). Are landscape ecologists addressing uncertainty in their remote sensing data? *Landscape Ecology*, vol. 27 n° 9, p. 1249-1261.

Peuquet D. (1994). It's About Time : A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Annals of the Association of American Geographers*, vol. 83, n° 3, p. 441–461.

Plumjeaud C. (2011). Modèles et méthodes pour l'information spatio-temporelle évolutive Thèse en informatique soutenue en 2011 (Université de Grenoble), dirigée par Jérôme Gensel et Claude Grasland, 330 p.

Ostapowicz K., Vogt P., Riitters K.H., Kozak J., Estreguil C. (2008). Impact of scale on morphological spatial pattern of forest. *Landscape Ecology*, vol. 23, n° 9, p. 1107-1117.

Sciama D., Augusto L., Dupouey J.L., Gonzalez M., Moares Dominguez C. (2009). Floristic and ecological differences between recent and ancient forests growing on non-acidic soils. *Forest ecology and management*, vol. 258, n°5, p. 600-608.

Soille P., Vogt P., (2009). Morphological segmentation of binary patterns. *Pattern Recognition Letters*, vol. 30, p. 456-459.

Sourdril A., Andrieu E., Cabanettes A., Elyakime B., Ladet S. (2012). How to maintain domesticity of usages in small rural forests? Lessons from Forest Management Continuity through a French Case Study, *Ecology and Society*, 17, 2, 6. http://dx.doi.org/10.5751/ES-04746-170206

Vallauri D., Grel A., Granier E., Dupouey J.-L. (2012). Les forêts de Cassini. Analyse quantitative et comparaison avec les forêts actuelles. Rapport WWF/INRA. Marseille, 64 p + CD.

Vogt P. (2010). MSPA GUIDE Release: Version 1.3, European Commission, Joint Research Centre (JRC), February 2010, 7 p.

Vogt P., Riitters K.H., Estreguil C., Kozak J.,Wade T.G. (2007). Mapping spatial patterns with morphological image processing. *Landscape Ecology*, vol. 22, n° 2, p. 171-177.

Caractérisation des trajectoires d'évolution de l'occupation et de l'usage des sols

Approche multi-scalaire appliquée aux Pyrénées

Laure Vacquié, Thomas Houet et Christelle Vigneau[1]

1. Laboratoire GEODE, Université Toulouse 2 Le Mirail
5 allée Antonio Machado, 31000 Toulouse, France
laure.vacquie@univ-tlse2.fr, thomas.houet@univ-tlse2.fr, c.vigneau6@gmail.com

RESUME. *La reforestation des espaces montagnards est un phénomène global dont l'ampleur et la vitesse s'amplifient depuis plus d'un demi-siècle, affectant l'organisation paysagère des vallées Pyrénéennes. Cet article vise à caractériser les changements d'occupation et d'usage des sols à différentes échelles spatiales et temporelles. Une approche de classification semi-automatique (GEOBIA) est proposée pour la cartographie de l'occupation des sols depuis les années 1940 sur trois sites d'étude locaux. Les résultats témoignent d'une fermeture des espaces agro-pastoraux par des processus d'enfrichement et de reforestation spontanée. La distribution spatiale de ces changements présente toutefois des dynamiques différentes selon les sites. Même si les facteurs explicatifs de ces transformations sont les mêmes, des disparités locales apparaissent du fait de facteurs naturels et anthropiques spécifiques. Ces résultats sont ensuite confrontés aux tendances régionales afin d'analyser la représentativité des vallées par rapports aux dynamiques observées à plus large échelle.*

ABSTRACT. *Natural reforestation in mountain regions is a well-known phenomenon that has been developing for the last fifty years, affecting landscape dynamics of Pyrenean valley. This paper aims at assessing land-cover and land-use changes in the French Pyrenees at various spatial and temporal scales. A semi-automatic GEOBIA approach is applied to produce land cover maps since the 1940s in three local study sites. The results indicate a landscape enclosure due to forest encroachment and spontaneous reforestation. However, the spatial distribution of changes differs from a site to another. Although main drivers of encroachment are the same, local disparities appear because of specific natural and anthropogenic factors. Results are then compared to regional LUCC trajectories to assess valley's representativeness of regional trends.*

MOTS-CLES : *Photographies aériennes, dynamiques d'occupation et d'usages des sols, reforestation, géomatique, Pyrénées, données multi-scalaires.*

KEYWORDS: *Aerial photographs, land use and land cover dynamics, reforestation, GIS, Pyrenees, multi-scaled data.*

1. Introduction

A partir des années 70, il a été admis que l'occupation et l'utilisation des sols a une influence sur les échanges énergétiques entre terre et atmosphère, impactant à la fois sur le climat (Pielke, 2002), la qualité de l'eau ou des sols, les cycles biogéochimiques, la biodiversité, etc. (Steffen et *al.*, 2004, Gillson, 2009) à des échelles globales, régionales et locales (Lambin et *al.*, 2006). Les milieux montagnards sont représentatifs de ces enjeux en ce sens qu'ils ont connu au cours des derniers siècles une évolution marquée (Métailié, 2006), cet héritage ainsi que la diversité des activités agro-pastorales actuelles influant encore aujourd'hui sur l'organisation des paysages de vallée. Mais, même si ces activités ont de tout temps modelé et entretenu ces espaces, on observe depuis les cinquante dernières années des dynamiques d'afforestation et de fermeture des milieux liées à un déclin des activités de production en zone de montagne et un exode rural massif (Galop et *al.*, 2011).

L'objectif de cet article est de caractériser les trajectoires d'évolution de l'occupation et de l'usage des sols à une échelle locale, en comparant trois sites, et de les comparer aux dynamiques régionales afin (1) de montrer l'intérêt d'une approche multi-scalaire dans la caractérisation des processus de fermeture de paysage et pour l'identification des facteurs explicatifs de ces changements. Ce travail se base sur des cartographies issues de photographies aériennes des soixante dernières années permettant, à un pas de temps décennal, d'analyser les dynamiques d'occupation des sols. L'utilisation d'indicateurs cartographiques et statistiques permet de quantifier et de spatialiser les changements afin de caractériser l'évolution des vallées. On part de l'hypothèse que des disparités locales existent à l'échelle du massif, les changements d'occupation et d'usage des sols étant influencés à la fois par des facteurs naturels (climat, géomorphologie, pédologie…) et anthropiques (pression urbaine, usages…).

2. Sites d'étude

Cette recherche est réalisée à l'échelle de trois sites d'étude (figure 3) suivant un gradient d'ouest en est le long de la chaîne Pyrénéenne : le Davantaygue (76m²), le Haut-Vicdessos (244m²) et les Garrotxes (85m²). Le premier site est situé à la périphérie du Parc National des Pyrénées. Le second site comprend plusieurs communes du Parc Naturel Régional d'Ariège, et le troisième site se situe dans le Parc Naturel Régional des Pyrénées Catalanes. Le choix de ces trois sites résulte de la convergence d'intérêts scientifiques (projets de recherches) et techniques (disponibilité de données satellitaires, de relevés terrains et de cartographies d'occupation des sols).

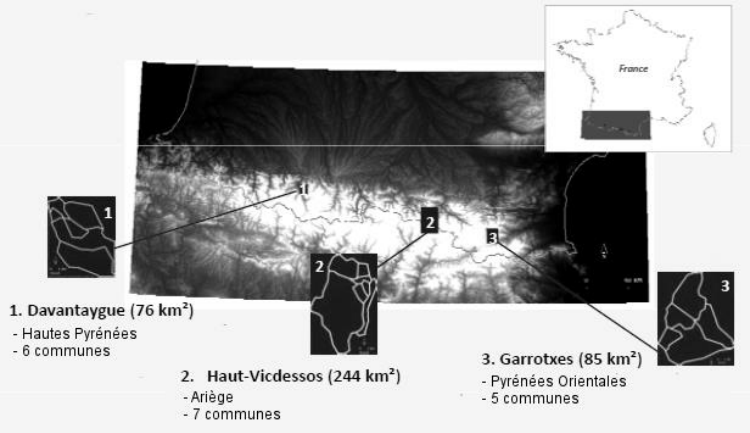


Figure 1. Localisation des sites d'étude locaux

3. Matériels et méthodes

3.1. Données disponibles

Les changements d'occupation du sol ont été identifiés à partir de photographies aériennes. Les photographies aériennes panchromatiques couvrent la période des années 1940 aux années 1990, et sont complétées pour la période récente (1990-2010) de photographies couleurs. Chaque zone d'étude compte entre six et huit orthophotoplans historiques, présentant un pas de temps quasi décennal (tableau 1).

Tableau 1. Caractéristiques des photographies aériennes

	Années	Nb de clichés	Emulsion	Echelle	Résolution
Davantaygue	1948/1959/1971/ 1978/1989/2001	67	Panchro / Couleur	1 :15000 à 1 :30000	0.5 m
Vicdessos	1942/1953/1962/1976/ 1983/1993/2003/2008	286		1 :25000 à 1 :30000	
Garrotxes	19421962/1980/ 1989/2000/2009	104			

3.2. Cartographie des modes d'occupation et d'usages des sols

La méthode développée pour produire chaque cartographie d'occupation des sols se base sur les principes du GEOBIA (GEOgraphic Object-Based Image Analysis), une méthode de classification qui considère les entités géographiques comme des

objets contextuels homogènes (Blaschke, 2010). Les traitements appliqués consistent à (1) segmenter chaque orthophotoplan en objets homogènes, ou polygones, selon des critères de teinte, de texture, de taille et de forme puis à (2) classifier l'image en affectant manuellement chaque objet à la classe d'occupation du sol correspondante. Les polygones contigus d'un même type d'occupation du sol sont ensuite fusionnés afin de permettre une meilleure représentativité des entités de même classe. Afin de limiter les éventuelles erreurs de sur-détection liées aux petits décalages persistants entre les orthophotoplans, la segmentation pour une date *t* tient compte des limites vectorielles de la carte classifiée à la date *t-1* (figure 2). La méthodologie est présentée en détail dans Houet *et al.* (2012). Les classifications sont ensuite validées en utilisant des données de terrain (pour les dates récentes) ou des vues historiques obliques (pour les périodes plus anciennes) lorsqu'elles sont disponibles.

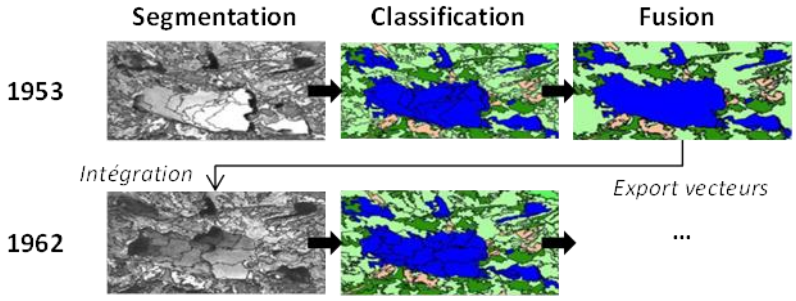


Figure 2. Méthodologie GEOBIA de classification à partir d'orthophotoplans historiques destinée à éviter les erreurs de sur/sous détection liées aux décalages spatiaux inhérents au contexte topographique montagnard

3.3. Analyse des dynamiques d'évolution

3.3.1. Une approche comparative par zonage altitudinal

Les espaces agro-pastoraux se structurent autour de zones d'usages spécifiques : les zones de fonds de vallées, les zones intermédiaires et les estives. Leurs limites, principalement altitudinales, sont spécifiques à chaque vallée, et varient selon la période considérée (Houet et *al.*, 2012). Afin d'appréhender plus finement leur évolution, il est nécessaire de distinguer ces secteurs géographiques. On choisit donc de subdiviser chaque vallée en trois zones (altitude, versant, fond de vallée) les plus représentatives des espaces fonctionnels, à la date la plus ancienne. Les zones d'altitude sont assimilées aux zones de pâturages estivaux (présence de cabanes pastorales et d'orris). Les zones de versant regroupent les secteurs de granges, utilisés à l'intersaison par les éleveurs tandis que les zones de fond de vallées correspondent aux bâtiments d'exploitation. La définition des zonages se base ainsi sur des critères à la fois géographiques (altitude, pente, exposition) et agronomiques (type d'utilisation des espaces) déterminés par les cartes d'occupation des sols les

plus anciennes et par des recherches bibliographiques. Ces critères ont été établis de manière générique pour chaque vallée afin d'assurer au maximum la comparabilité des espaces agro-pastoraux entre les sites et leurs dynamiques d'évolution.

3.3.2. Analyse des changements d'occupation des sols

L'analyse des changements d'occupation et d'usages des sols s'effectue en deux temps : (1) le calcul d'indicateurs retranscrivant l'ampleur et la vitesse de fermeture des milieux et (2) la caractérisation des changements et de leur répartition spatiale.

L'ampleur et la vitesse de fermeture des paysages sont évaluées grâce à deux indicateurs simples calculés à partir des statistiques d'occupation des sols. L'ampleur est mesurée par l'évolution de la part de surface boisée par rapport à la zone considérée (zone entière, d'altitude, de versant et de fond de vallée). La vitesse est mesurée à travers deux indicateurs, calculés pour chaque site et pour chaque zone :

– le taux annuel de changements forestiers (R) mesure un taux moyen annuel de gains ou de pertes nettes des surfaces forestières sur une période donnée, une valeur positive traduisant un phénomène d'afforestation, et inversement (FAO, 1995).

$$R = \left[\left(\left(\frac{A_{n+1}}{A_n}\right)^{1/t}\right) - 1\right] * 100 \qquad (1)$$

Où A_n= surfaces en forêt à la date n
A_{n+1} = surfaces en forêt à la date $n+1$
T = durée de la période $(n - n + 1)$

– le taux d'accroissement moyen annuel (A2, en %) des surfaces boisées permet de témoigner d'une accélération ou non de la fermeture des paysages. Ce taux est normalisé par la taille de la zone concernée ce qui permet de comparer les dynamiques de fermeture entre les différentes vallées.

4. Résultats

A l'échelle des trois vallées, on observe une modification des paysages, passant de milieux agricoles et pâturés à des étendues recolonisées par la forêt. Toutefois, l'ampleur et le rythme de ces changements diffèrent selon les sites et les zones d'usages.

La vallée des Garrotxes est l'exemple classique de l'évolution des vallées montagnardes subissant les effets de l'exode rural et de la faible exploitation de ses ressources depuis les années 50 : les terres labourées sont d'abord converties en près de fauche puis progressivement laissées au pâturage extensif avant d'être abandonnées et aujourd'hui colonisées par la forêt à tous les étages. L'indicateur R (figure 3) rend compte d'une modification dans la vitesse de fermeture. A l'échelle du site, on observe des dynamiques similaires à tous les étages jusqu'au début des années 1990 avec un taux annuel de changements forestiers en baisse constante (+0.70% entre 1942/1962, +0.55% entre 1962/1980 et +0.19% entre 1980/1989). Si

l'on considère les zones de versant, on observe une forte progression de ce taux durant la période 1989-2000 (+1.03%), suivie d'un très net ralentissement pour la période 2000-2009 (+0.54%). A l'inverse, les zones d'altitude connaissent une faible évolution jusqu'en 2000 (R inférieur à +0.05%/an soit moins de 1ha de forêts en plus par an) pour atteindre +0.28%/an après 2000 soit un gain net de 5.8ha/an de surfaces boisées. Cette relative stabilité s'explique par le fait que plus de deux tiers des surfaces en 1942 sont déjà occupées par des forêts, laissant peu d'espace à une extension supplémentaire. Ces dynamiques d'abandon s'expliquent principalement par l'extrême isolement de cette vallée où l'accessibilité constitue un réel frein à l'exploitation des ressources, qu'elles soient sylvicoles ou pastorales (Paegelow et *al.*, 2004).

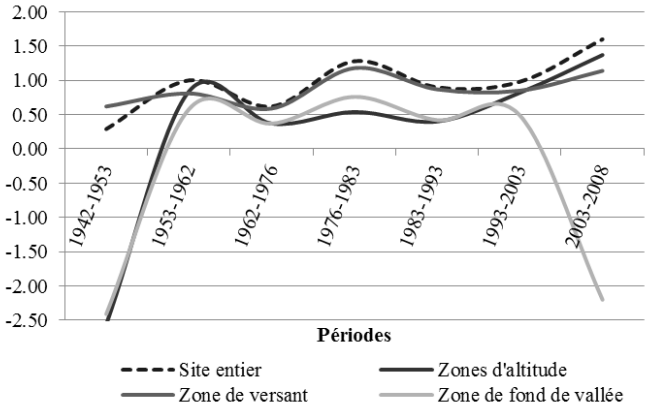


Figure 3. Taux annuel de changements forestiers (R en %) dans les Garrotxes

Dans le Vicdessos (figure 4), la fermeture des zones de versant est progressive et suit les tendances observées à l'échelle de la vallée. Des coupes forestières effectuées entre 1942 et 1953 expliquent le faible taux de R à cette période dans les zones d'altitudes et de fond de vallée. Au niveau de ces dernières, la fermeture est plus importante à partir des années 60 (R entre +0.37% et +0.58%). Cet enfrichement marque le début de l'abandon des petites parcelles agricoles, s'exprimant par un phagocytage progressif par les haies bocagères les délimitant. La diminution de R entre 2003 et 2008 s'explique par le court intervalle de temps entre les deux cartographies et par un taux de surfaces boisées déjà élevé dès les années 2000 (66%). Les zones d'altitude présentent une faible évolution avant 1962. On note cependant une accélération à partir des années 90 (R entre +0.82% et +1.37%/an). La fermeture des estives est, de façon relative aux espaces considérés, plus forte que l'évolution observée à l'échelle du site entier sur certaines périodes, avec un décalage d'environ 30 ans. Globalement, on observe à l'échelle de la vallée une homogénéisation des paysages avec des versants forestiers qui empiètent à la fois sur les estives et les fonds de vallées dû à un ralentissement des activités agro-pastorales et l'extension des arbres issus des travaux de reboisement (Carré, 2010).

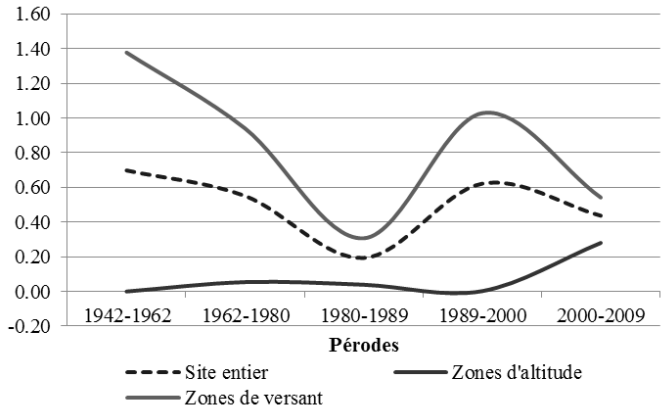


Figure 4. Taux annuel de changements forestiers (R en %) dans le Vicdessos

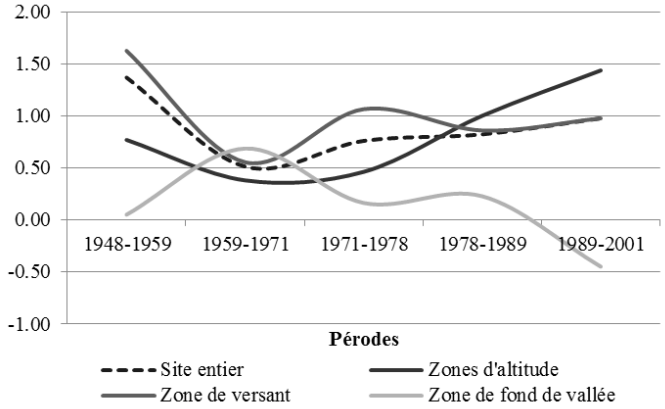


Figure 5. Taux annuel de changements forestiers (R en %) dans le Davantaygue

Même si le paysage du Davantaygue reste relativement stable depuis les années 50 (figure 5) les dynamiques d'enfrichement tendent à s'amplifier depuis les années 70 principalement en zones de versants. Ces dernières affichent des taux de changements importants sur la période 1971-1978 (+1.07%) qui resteront relativement stables jusqu'en 2001 (+0.86% et +0.98%). Ces dynamiques vont de pair avec un abandon progressif de ces espaces pentus, difficilement exploitables et mécanisables, et parfois éloignés des sièges d'exploitation (Gibon et *al.*, 2010). Tandis qu'en zone de fond de vallée le taux de changement forestier moyen diminue

(+0.17% entre 1971/1978, +0.23 entre 1978/1989 et -0.45% entre 1989/2001), il tend à augmenter de manière constante en zone d'altitude pour atteindre des tendances similaires à celles observées à l'échelle de la vallée (+1.44% entre 1989/2001). Les effets de l'extensification de l'utilisation des sols entrainent ainsi des dynamiques de fermeture des paysages irréversibles (Couvreur et *al.*, 1999) en zone de versant et d'altitude. Dans le même temps, le maintien des activités agro-pastorales et la ré-intensification des pratiques dans les années 80 suite notamment à la mise en place de mesures agro-environnementales (réformes de la PAC) encourageant l'exploitation des prairies de montagne (Mottet, 2005) permettent d'entretenir et de conserver l'organisation traditionnelle du paysage montagnard dans les fonds de vallées et les bas de versant.

5. Discussion

5.1. Sur la méthode

La méthodologie suivie pour la réalisation des cartes d'occupation des sols, en traitant de manière semi-automatique des zones homogènes et contextuelles, offre des résultats satisfaisants ; les méthodes de classifications automatiques étant encore trop peu adaptées au traitement d'anciennes photographes aériennes panchromatiques. Cette approche a déjà montré son efficacité pour la cartographie des dynamiques d'enfrichement à partir d'images aériennes panchromatiques et d'images satellites à haute résolution spatiale (Laliberte et *al.*, 2004). L'utilisation des délimitations de chaque entité géographique à la date antérieure rend cette méthode d'autant plus robuste qu'elle permet également de s'affranchir d'erreur de classification due à des décalages entre unités géographiques identiques lors des segmentations du fait de différences de teinte ou de texture entre photographies multi-dates. En reposant sur une nomenclature développée de manière concertée avec les différents chercheurs spécialistes des Pyrénées, prenant en compte les différents modes d'occupation des sols de chaque espace agro-pastoral, cette méthode est générique et donc transposable à n'importe quelle vallée Pyrénéenne.

En outre, le choix d'utiliser des indicateurs statistiques simples relève ici d'un besoin de quantifier l'ampleur et la vitesse des changements. En effet, l'objectif n'était pas ici de mesurer l'évolution de l'organisation spatiale de la fermeture paysagère, qui ne témoigne pas d'un *pattern* spécifique du fait de l'influence de la topographie, mais bien de caractériser les trajectoires d'évolution des types d'occupation et d'usages des sols. Ainsi, contrairement aux métriques paysagères qui cherchent à quantifier l'organisation de la mosaïque paysagère et son hétérogénéité, les indicateurs utilisés permettent d'évaluer et de comparer les dynamiques de la fermeture des vallées Pyrénéennes, rendant possible la détection des décalages temporels entre chaque site.

5.2. Des tendances similaires à celles observées à l'échelle du massif Pyrénéen

Les dynamiques de déprise et d'afforestation observées à l'échelle des vallées, relevant de facteurs naturels et anthropiques, reflètent les tendances globales liées à l'évolution démographique et économique du massif. Ces tendances, combinées aux difficultés que connaissent les activités agro-pastorales depuis les dernières années, ont un impact très marqué sur l'espace Pyrénéen. Les activités traditionnelles sont en concurrence foncière pour l'usage de l'espace. Le fonctionnement des exploitations est, lui, perturbé par le tourisme qui renchérit les coûts. Ces difficultés sont aggravées par la nécessité pour les éleveurs de montagne d'avoir à leur disposition des surfaces exploitables à plusieurs niveaux d'altitude, la confiscation foncière de l'une de celle-ci entraînant la disparition d'exploitation, laissant en déshérence les terres agricoles aux autres altitudes (Comité de massif des Pyrénées, 2006). Avec la diminution du nombre d'exploitants, le vieillissement de la population, et la mise en place d'activités concurrençant l'agro-pastoralisme, la société se réorganise et les dynamiques paysagères se modifient.

Toutefois, le système Pyrénéen se présente comme un système territorial hétérogène où la diversité des modes d'occupation et d'usage des sols est à l'image de la diversité des orientations agro-pastorales. La productivité des surfaces agricoles, dépendante des conditions pédoclimatiques d'ouest en est de la chaîne, détermine les surfaces moyennes des exploitations agricoles et *a fortiori* la distribution spatiale des cheptels (figure 6) et la place des activités de production. En parallèle, les dynamiques de fermeture des vallées vont être directement corrélées avec l'intensité des activités agro-pastorales.

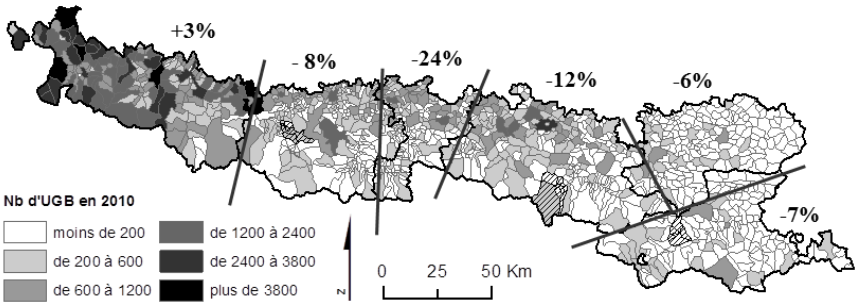


Figure 6. Nombre d'UGB (Unités Gros Bétail) en 2010 et évolution par département (1988-2010) (Données : RGA 2010)

10

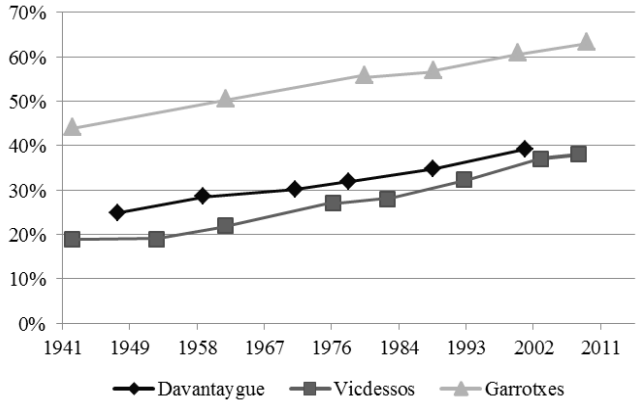


Figure 7. Evolution du taux de surfaces boisées à l'échelle des trois sites (en %)

Force est de constater que les périodes de rupture des systèmes agro-pastoraux de chacune des vallées reflètent les dynamiques du massif Pyrénéen suivant un gradient d'est en ouest de la chaîne. Un potentiel agronomique traditionnellement faible et un nombre d'UGB en baisse (-7%) va de pair avec les dynamiques de déprise les plus anciennes (dès les années 40) dans les Garrotxes (figure 7). La vallée du Davantaygue dont la période de rupture dans les années 70 est la plus tardive - et malgré une baisse du nombre d'UGB (-8%) - bénéficie de l'influence et du dynamisme des Pyrénées Atlantiques. La vallée du Vicdessos située entre ces deux pôles tant d'un point de vue géographique que temporel (rupture dans les années 60), présente des dynamiques de déprise qui pourraient donc potentiellement entrainer une fermeture complète du paysage dans les années à venir, à l'image des dynamiques récentes observées dans les Garrotxes.

5.3. De l'intérêt d'une approche multi-scalaire

Il est admis que les dynamiques d'occupation et d'utilisation des sols sont le résultat de processus multiples qui s'opèrent à différentes échelles (Turner II et *al.*, 1995). Même si globalement des dynamiques similaires s'observent à l'échelle du massif (enfrichement, afforestation), l'évolution récente des modes d'occupation et d'usage des sols affecte de manières différenciées l'organisation paysagère des vallées Pyrénées. L'analyse de ces dynamiques à des échelles fines nécessite l'utilisation de données cartographiques à des résolutions spatiales adaptées. Les cartographies régionales de l'occupation des sols actuellement disponibles, telles que Corine Land Cover, bien qu'offrant une base de connaissance non négligeable sur les tendances du massif, ne permettent pas de retranscrire finement l'évolution des paysages observée à l'échelle des vallées (figure 8). Entre 1990 et 2006, le taux de surface boisé y demeure stable tandis qu'on observe une légère diminution des surfaces agricoles de versant (-5%), contrebalançant une augmentation (+6%) des

pelouses. Cette stabilité montre que, de, manière générale, l'utilisation de données aériennes et/ou satellitaires présentant une trop faible résolution spatiale ne permet de détecter que partiellement les espèces végétales dominantes et de caractériser de manière succincte les types d'occupation du sol. Détectés comme relativement homogènes à une résolution spatiale large, ils peuvent s'avérer hétérogènes à une échelle spatiale plus fine (Gibson et *al.*, 2000).

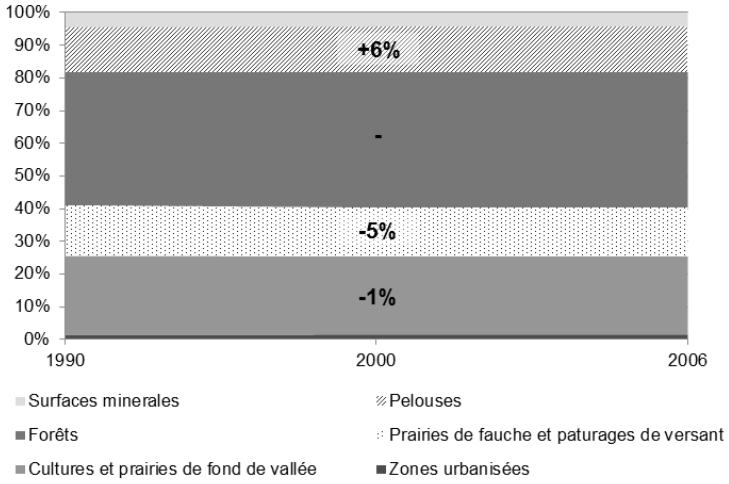


Figure 8. Evolution de l'occupation du sol à l'échelle du massif Pyrénéen (en %) (Source : Corine Land Cover)

Ces observations montrent d'une part la nécessité de cartographies fines à l'échelle du massif afin de réduire au maximum l'hétérogénéité au sein de chaque entité de végétation et d'autre part l'importance de la prise en compte des échelles pour la compréhension des phénomènes d'enfrichement. Dans un milieu complexe où les facteurs d'évolution sont dépendants de l'échelle d'étude considérée, l'échelle spatiale choisie pour l'observation des changements d'occupation et d'usage des sols a une influence significative sur leur organisation spatiale. A ce titre, les démarches de modélisation à des échelles spatiales indépendantes posent des problèmes d'interprétation et de compréhension des dynamiques d'évolution des milieux.

L'utilisation conjointe de données régionales et locales permet alors d'appréhender les facteurs d'évolution d'un système dans son intégralité en prenant en compte l'interaction de processus à des échelles spatiales différenciées (Bürgi et *al.*, 2004). Cette approche permet d'analyser la diversité interne d'un paysage à l'échelle de la vallée, mais aussi de mesurer sa singularité par rapport au milieu environnant à une échelle plus petite (Germaine et Puissant, 2007) afin d'analyser la représentativité d'une vallée par rapport aux dynamiques observées à l'échelle du massif Pyrénéen.

6. Conclusion et perspectives

L'analyse des changements d'occupation des sols observés sur ces trois vallées montrent des dynamiques globalement similaires : les espaces fermés (zones de recolonisation, forêts de conifères et de feuillus) succèdent aux espaces ouverts (cultures, prairies de fauche) à des rythmes cependant différenciées selon les sites. La distinction entre les secteurs agro-pastoraux des zones de fonds de vallées, de versant et d'altitude nous a permis d'identifier plus finement la complexité des changements d'occupation et d'usages des sols. Même si globalement les dynamiques de fermeture des vallées sont dues à une diminution de l'intensité des pratiques agro-pastorales, les processus d'enfrichement et de reforestation varient selon les sites. Ces décalages temporels proviennent principalement de facteurs environnementaux, climatiques et socio-économiques locaux, reflétant les dynamiques observées à l'échelle du massif Pyrénéen. Afin de retranscrire plus précisément les changements fins de l'occupation des sols à l'échelle régionale, des cartographies précises sont cependant nécessaires. Les résultats obtenus renforcent l'intérêt d'une approche multi-scalaire pour appréhender les interactions entre processus locaux et régionaux et *a fortiori* l'étude des dynamiques paysagères à l'échelle de vallées.

Remerciements

Ce travail s'inscrit dans le projet MODE-RESPYR (Modélisation rétrospective et prospective des changements d'occupation des sols dans les Pyrénées) qui s'attache à l'étude des dynamiques d'occupation et d'usages des sols dans les Pyrénées à différentes échelles spatiales et temporelles. Cet article a bénéficié du soutien du projet ANR 2010 JCJC 1804 01 « MODE RESPYR » (http://w3.mode-respyr.univ-tlse2.fr/index.php) et de l'Observatoire Hommes-Milieux (OHM) du Haut-Vicdessos (http://w3.ohmpyr.univ-tlse2.fr/), intégré au Labex DRIIHM, soutenu par l'Institut Ecologie et Environnement (InEE) du CNRS.

Bibliographie

Blaschke T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol 65, n°1, p. 2-16.

Bürgi M., Hersperger A.-M., Schneeberger N. (2004). Driving forces of landscape change – current and new directions. *Landscape Ecology*, vol. 19, p. 857-868.

Carré J. (2010). *Le temps des paysages. Evolutions paysagères et gestion durable des territoires en montagne Pyrénéenne (hautes vallées du gave de Pau et du Vicdessos*. Thèse de Géographie, Université Toulouse II Le Mirail.

Comité de massif des Pyrénées (2006). *Schéma interrégional d'aménagement et de développement des Pyrénées*. Rapport interne. décembre 2006.

Couvreur M., Mitteault F. et Pech M. (1999). Les mesures agri-environnementales mises en œuvre en France. *Economie rurale*, vol. 249, p. 6-9.

FAO. (1995) *Forest resources assessment*, Global Synthesis. Food and Agricultural Organization, Rome.

Galop D., Houet T., Mazier F., Leroux G. et Ruis D. (2011). Grazing activities and biodiversity in the Pyrenees : new insight on high altitude ecosystems in the framework of a Human-Environment Observatory. *PAGESNews*, vol. 19, n°2, p. 53-55.

Germaine M.-A., Puissant A. (2007). Cartographie et caractérisation quantitative des paysages de « vallées ordinaires » : l'exemple de la Seulles, *Acte du colloque international de Géomatique et d'Analyse Spatiale - SAGEO*, 2007, Clermont-Ferrand.

Gibon A., Sheeren D., Monteil C., Ladet S. et Balent G. (2010). Modelling and Simulating change in reforesting mountain landscapes using a social-ecological framework, *Landscape Ecology*, vol. 25, p. 267-285.

Gibson C.-C., Ostrom E. et Anh T.-K. (2000). The concept of scale and the human dimensions of global change : A survey, *Ecological Economics*, vol. 32, p. 217-239.

Gillson L. (2009). Landscapes en Time and Space, *Landscape Ecology*, vol. 24, p. 149-155.

Houet T., Ribière O., Vacquié L., Vidal F. et Galop D. (2012). Analyse spatiale de l'évolution des modes d'occupation et d'usages des sols sur le Vicdessos de 1942 à nos jours, *Revue du Sud-Ouest Européen*, Presses Universitaires du Mirail, n°33, p. 41-57.

Houet T., Loveland T., Hubert-Moy L., Gaucherel C., Napton D., Barnes C. et Sayler K. (2010). Exploring subtle land use and land cover changes: a framework for future landscape studies, *Landscape Ecology*, vol. 25, p. 249–266.

Laliberte A.S., Rango A., Havstad K.M., Paris J.F., Beck R.F., McNeely R. et Gonzalez A.L. (2004). Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico, *Remote Sensing of Environment*, vol. 93, p. 198-210.

Lambin E.-F. et Geist H.-J. (2006). Land-Use and Land-Cover Change - Local processes and Global impacts, *Global Change – The IGBP Series*, Ed. Springer.

Métailié J.-P. (2006). La dégradation des montagnes au XIXe siècle dans les Pyrénées. *Temps et espaces des crises de l'environnement. Sociétés et ressources renouvelables*, Paris, Quæ, p.191-210.

Mottet A. (2005). *Transformations des systèmes d'élevage depuis 1950 et conséquences pour la dynamique des paysages dans les Pyrénées. Contribution à l'étude du phénomène d'abandon de terres agricoles en montagne à partir de l'exemple de quatre communes des Hautes-Pyrénées*. Thèse de Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries, INP Toulouse.

Paegelow M., Villa N., Cornez L., Ferraty F., Ferré L. et Sarda P. (2004). *Modélisation prospective de l'occupation du sol. Le cas d'une montagne méditerranéenne*, http://cybergeo.revue.org/2811

Pielke R.A. (2002). *Mesoscale meteorological modeling*. Academic Press, San Diego.

Steffen W., Sanderson A., Tyson P.D., Jäge J., Matson P.A., Moore B., Oldfield F., Richarson K., Schellnhuber H.J., Turner B.L. et Wasson R.J. (2004). Global Change and the Earth System: A Planet under Pressure, *Global Change – The IGBP Series*, Ed. Springer.

Turner II B.L., Lambin E.F. et Reenberg A. (1995). The emergence of land change science for global environmental change and sustainability, *PNAS*, vol. 104, n° 25.

Appariement hiérarchique de réseaux hydrographiques imparfaits

Benoît Costes[1]

1. Université Paris-Est, IGN/SR, COGIT
73 avenue de Paris, 94160 Saint-Mandé, France
benoit.costes@ign.fr

RESUME. *Cet article se focalise sur la problématique de l'appariement de réseaux hydrographiques imparfaits. Les spécificités inhérentes à ces données rendent peu efficaces les méthodes classiques d'appariement de réseaux. Nous proposons une approche originale d'appariement hiérarchique, en changeant le mode de représentation des réseaux par la reconstitution de strokes, des structures naturelles de bonne continuité. Chaque stroke est classé en fonction de son importance dans le réseau en termes de structure hydrographique. La sélection hiérarchique des candidats à l'appariement, ainsi que l'utilisation d'une méthode multi-critères de prise de décision (AHP) constituent une stratégie d'appariement adaptée aux réseaux hydrographiques imparfaits. Notre méthodologie est testée sur des données anciennes datées du XVIIIème siècle extraites des cartes de Cassini,*

ABSTRACT. *This article focuses on the process of matching imperfect hydrographic networks. The inherent specificities of such data make conventional methods for matching networks inefficient. We propose a novel hierarchical approach to match hydrographic networks, by extracting natural structures called strokes. Each stroke is then classified according to its weight in terms of hydrographic structure within the graph. The hierarchical selection of candidates and the use of a multi-criteria decision making process (AHP) constitute a good strategy to match imperfect hydrographic networks. We assess our approach on vector data extracted from old Cassini maps (XVIIIth century).*

MOTS-CLES : *appariement de réseaux hydrographiques, imprécision, incomplétude, strokes, multi-critères, appariement hiérarchique, intégration*

KEYWORDS: *hydrographic network matching, imprecision, incompleteness, strokes, multi-criteria, hierarchical matching, integration*

Introduction

Aujourd'hui, un SIG est non seulement un outil d'analyse spatial sophistiqué, mais permet aussi de répondre aux questions de disciplines connexes aux sciences de l'information géographique de plus en plus variées, telles que l'informatique, les mathématiques, mais aussi l'histoire, la géographie et l'archéologie. Pour l'historien par exemple, un SIG est un outil de compréhension des phénomènes sociaux. Le développement croissant de nouvelles méthodes d'acquisition de données, telles que les SIG collaboratifs, combiné à la diversification des communautés d'utilisateurs rendent possible l'acquisition et la mise en base de données disparates décrivant la même réalité. Cette multiplicité de données spatiales homologues et hétérogènes pose le problème de leur qualification et de leur intégration. En effet, les données spatiales sont par nature imparfaites (Goodchild, 1995 ; Hunter, 1998). Des imperfections courantes peuvent par exemple être caractérisées par d'importants décalages géométriques entre deux bases de données différentes décrivant la même réalité. Les plans anciens en sont une illustration pertinente. En effet, le dessin des cours d'eau pouvait être modifié volontairement afin de ne pas gêner la cartographie d'autres entités, comme par exemple celles liées à l'information religieuse, jugée de plus haute importance. Il est également fréquent de constater des problèmes d'imprécision et d'incomplétude attributaire, comme des toponymes non renseignés ou ambigües (par exemple deux lieux-dits proches ayant le même nom). Beaucoup d'erreurs inhérentes aux données géographiques proviennent de la saisie des données et des processus cartographiques utilisés (vectorisation, généralisation, etc.), mais également des sources elles mêmes (triangulation et symbologie peu précises, erreurs de relevé, etc.). De plus, les données saisies ne sont généralement pas directement intégrables dans un modèle commun, pour des raisons de différence de point de vue et d'objectif de représentation, de spécification de saisie, etc. (Abadie, 2012). Une étape indispensable pour l'intégration de ces données est donc leur mise en correspondance, c'est à dire leur appariement.

Cet article se focalise sur l'appariement de réseaux hydrographiques imparfaits. Il existe de nombreuses taxonomies des imperfections, et différentes terminologies sont ainsi couramment utilisées pour qualifier les données géographiques, telles que l'imprécision (Worboys, 1998), l'incertitude (Fisher, 2003), l'exactitude (Devogele et al., 2002), etc. Si de nombreuses théories mathématiques permettent de formaliser les imperfections des données, comme la théorie des possibilités (Zadeh, 1965), la théorie de l'évidence (Dempster, 1967 ; Shafer, 1976) ou encore la théorie des sous-ensembles flous (Zadeh, 1965 ; Bouchon-Meunier, 1995), nous choisissons d'adopter plutôt une approche géométrique proche de l'analyse spatiale.

Nous organisons l'article comme suit : la première section fait un état de l'art sur les techniques d'appariement de réseaux, la deuxième présente l'approche hiérarchique adoptée, la troisième section est consacrée à l'évaluation de notre approche sur des données présentant plus de 250 ans d'écart.

1. L'appariement de réseaux dans la littérature

Devogele (1997) propose dans sa thèse une approche d'appariement adaptée aux réseaux présentant des niveaux de détail différents. L'appariement des arcs des réseaux repose sur un pré-appariement des nœuds, sur des critères de distance géométrique, de similarité topologique et de plus courts chemins. Cette approche est testée et améliorée par Mustière (2006) et Mustière et Devogele (2008) sur des bases de données à différentes résolutions (1:25000 et 1:50000). Lüscher et al. (2007) adoptent une méthodologie similaire pour apparier des données à des échelles sensiblement différentes (1:25000 et 1:200000), en combinant des critères topologiques et angulaires pour l'appariement des nœuds. Walter et Fristch (1999) proposent une méthode statistique d'appariement de bases de données de même résolution. Les candidats à l'appariement sont sélectionnés par une technique de buffer grandissant et une batterie d'indicateurs est calculée pour chacun : longueur, orientation de la ligne de base, etc. La décision est prise par un algorithme d'optimisation local. Zhang et al (2005) utilisent également une approche par buffer grandissant, dont le rayon est auto-adaptatif. Toutes ces méthodes sont essentiellement basées sur des critères géométriques et topologiques. Elles ne sont donc pas adaptées dans le cadre de l'appariement de réseaux présentant par exemple d'importants décalages géométriques, comme dans l'exemple de la figure 1. A notre connaissance, il existe peu d'approches d'appariement de réseaux adaptées aux données imparfaites. Les méthodes multi-critères basées sur des fonctions de coûts (Costes et al., 2012) ou sur la théorie de l'évidence (Olteanu, 2008) permettent de gérer l'imprécision des données, mais trouvent leur limite lorsque l'imperfection concerne l'ensemble des critères. De plus, la complexité algorithmique de l'approche d'Olteanu (2008) constitue un problème majeur. Enfin, ces méthodes nécessitent généralement des connaissances supplémentaires pour définir les paramètres utiles à l'appariement comme dans Costes et al. (2012) et Walter et Fritsch (1999). Nous donnerons une évaluation comparative de ces méthodes dans la section 3.

Figure 1 : Situation posant problème à une approche d'appariement classique

2. Proposition d'appariement hiérarchique multi-critères

Nous proposons une nouvelle approche multi-critères basée sur un appariement hiérarchique des tronçons, en identifiant les cours d'eau principaux appelés strokes par la suite, classés selon une relation d'importance prenant en compte les ramifications du réseau. Cette approche comprend 3 principales étapes (cf. fig. 2):

- l'enrichissement change le mode de représentation des réseaux pour les rendre plus comparables ;
- la sélection hiérarchique des candidats respecte l'organisation des réseaux;
- la prise de décision permet de choisir le meilleur candidat à l'appariement.

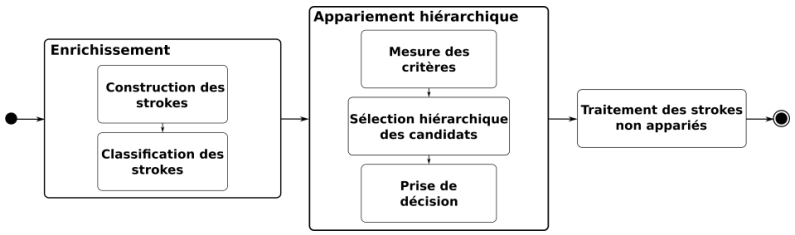


Figure 2 : Schéma global de l'approche d'appariement hiérarchique multi-critères

2.1 Enrichissement des réseaux

Dans une première étape, nous rendons les réseaux plus comparables en changeant leur mode de représentation. Pour cela, nous dégageons les structures naturelles continues, appelées « strokes » (Thomson, 1999) ou encore « natural roads » (Jiang, 2008) reprenant le principe de continuité de la Gestalt, constituées d'une succession de tronçons et représentant au mieux les objets hydrographiques : fleuves, rivières, etc. Si à l'échelle du tronçon, les niveaux de détail et la géométrie des réseaux peuvent différer, ils sont relativement proches quand on raisonne à l'échelle des strokes.

2.1.1 Sélection des strokes

Nous construisons les strokes des réseaux hydrographiques en adoptant une approche qui étend l'algorithme "every-best-fit" de Jiang et al. (2008). Si leur approche, basée sur des critères angulaires, est adaptée à la reconstruction de structures ayant une bonne continuité visuelle, elle ne permet pas de prendre en compte les imperfections topologiques des réseaux, inhérentes aux problèmes de saisie, de représentation et d'imprécision des sources (c.f figure 3). L'utilisation de l'information sémantique et toponymique permet la construction des strokes représentant au mieux les fleuves, rivières, ruisseaux, etc. (c.f annexe 1 pour l'algorithme détaillé).

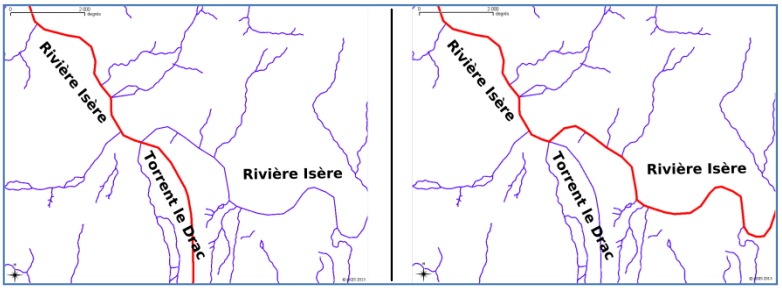


Figure 3 : Un stroke d'après (Jiang et al, 2008) à gauche et notre méthode à droite

2.1.2 Relation d'ordre sur les strokes

Thomson et Brooks (2000) utilisent pour généraliser des réseaux hydrographiques une classification de Horton (Horton, 1945), qui attribue à chaque stroke un degré d'importance (son ordre) en fonction de son nombre d'affluents : « tout cours d'eau sans affluent est d'ordre 1, tout cours d'eau ayant un affluent d'ordre *n* est d'ordre *n + 1*». Cette classification est fortement dépendante du niveau de détail des réseaux. En effet, la classification démarre des feuilles. Le stroke principal (sans affluent) d'un réseau détaillé aura donc un ordre potentiellement différent du stroke homologue dans une base à granularité plus faible. Nous reprenons ce type de classification en initialisant l'algorithme sur les racines du réseau et non sur les feuilles. Nous attribuons l'ordre 1 aux strokes les plus importants, qui ne sont affluents d'aucun autre stroke. L'ordre du stroke reflète son importance en termes de structure hydrographique. Cette classification est cependant sensible à l'emprise des données. Nous devons nous assurer que le découpage du réseau n'induit pas d'artefacts dans la classification effectuée. Pour la suite de cet article, nous appelons stroke parent d'un stroke d'ordre *n* le stroke d'ordre *n-1* dont il est affluent (cf. fig. 4).

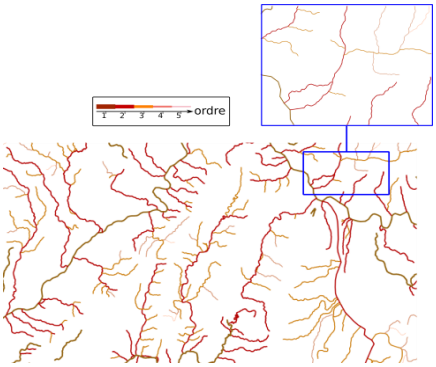


Figure 4 : Relation d'importance des strokes basée sur la ramification du réseau

2.2 Sélection hiérarchique des candidats

Le processus d'appariement travaille sur les données enrichies. Pour une passe donnée du processus, nous sélectionnons les n candidats à l'appariement du second réseau $E = \{c_1, \ldots, c_n\}$ avec le stroke de référence s_{ref} du premier réseau de la manière suivante : *pour* $i \in \{1, .., n\}$, soit s_{ref} est apparié avec le parent de c_i, soit le parent de s_{ref} est apparié avec c_i, soit le parent de s_{ref} est apparié avec le parent de c_i (voir fig. 5). L'algorithme de sélection des candidats à l'appariement est ainsi dépendant de la structure hiérarchique du réseau ainsi que des appariements précédents.

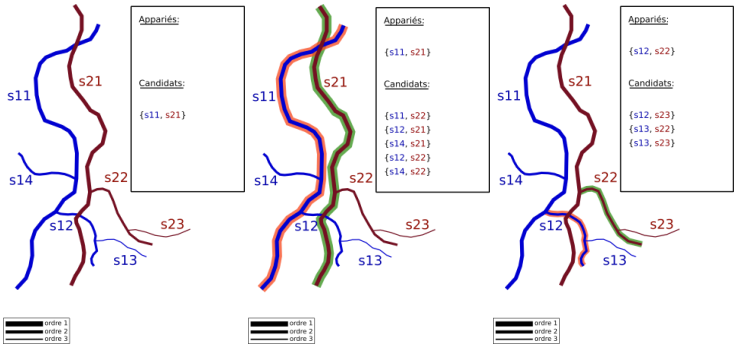


Figure 5 : Sélection des candidats et appariement hiérarchique

2.3 Prise de décision

Chaque stroke possède potentiellement plusieurs candidats à l'appariement. Nous proposons pour choisir le meilleur candidat une technique d'analyse multi-critères, l'Analytic Hierarchy Process (AHP), mise au point par Thomas L. Saaty (Saaty, 2008). L'AHP est un outil performant dans le cadre de la prise de décisions complexe, basé sur une analyse hiérarchique de différents critères et candidats permettant d'atteindre un objectif (voir fig. 6).

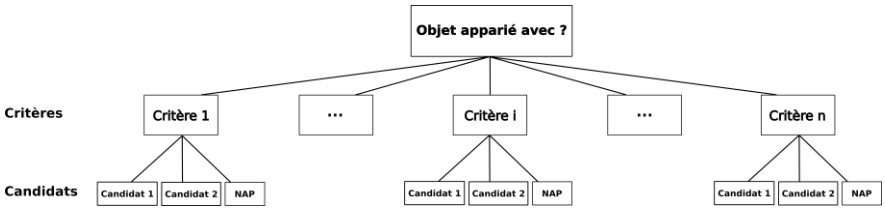


Figure 6 : Exemple de hiérarchie d'une AHP dans le cadre de l'appariement

2.3.1 Critères utilisés

Nous proposons d'utiliser quatre critères pouvant être utilisés quelque soit la source des réseaux hydrographiques à apparier :

- un critère géométrique mesurant l'éloignement entre deux strokes candidats par le calcul de leur distance de Fréchet discrète, une approximation de la distance de Fréchet pour les polylignes calculée en temps polynomial (Eiter et Mannila, 1994) ;

- un critère d'orientation qui mesure l'angle entre les directions générales des deux strokes (Hangouët, 1998) ;

- un critère de proximité globale qui permet de s'assurer que deux géométries ne s'éloignent pas trop l'une de l'autre globalement. En effet, l'utilisation seule d'un critère de distance ne suffit pas, car deux géométries globalement proches mais s'éloignant en une extrémité peuvent avoir une distance de Fréchet importante alors que les strokes se suivent bien dans l'ensemble. Ce critère se base sur l'indicateur développé par Goodchild (1997) pour mesurer les incertitudes des écarts géométriques des polylignes en mesurant le taux d'inclusion d'une polyligne à contrôler dans la bande epsilon de la polyligne de référence (Bel Hadj Ali, 2001) ;

- un critère de distance toponymique, reposant sur une amélioration de la distance toponymique définie dans (Samal et al., 2004) et utilisant la distance de Damerau-Levenshtein (Damerau, 1964), une extension de la distance de Levenshtein (Levenshtein, 1965) intéressante, car elle prend en compte plus de 80 % des fautes d'orthographes humaines (Damerau, 1964).

2.3.2 Principe de l'AHP

Nous illustrons le principe théorique de fonctionnement de l'AHP dans notre cadre d'appariement. L'objectif est de choisir le meilleur candidat à l'appariement s'il existe. L'approche s'effectue en quatre temps.

2.3.2.1 Regroupement des candidats

Tous les candidats à l'appariement d'un objet étudié sont sélectionnés afin d'être comparés mutuellement. Nous appelons pour la suite *alternative* le choix d'un candidat particulier. Nous construisons également une alternative fictive prise en compte dans l'analyse: « l'objet n'est pas apparié », notée NAP (comme « Non APparié), qui devra être choisie si aucun candidat ne se démarque particulièrement.

2.3.2.2 Confrontation des candidats

Pour chaque critère, les scores des candidats sont confrontés deux à deux. Lors de la confrontation d'un candidat ayant comme score m_i à l'hypothèse NAP, nous associons à celle-ci le score $1\text{-}m_i$, rendant ainsi compte dans quelle mesure nous ne croyons pas que ce candidat soit apparié pour sa valeur de score mesurée. Les préférences des alternatives sont établies par un décideur (*the Fundamental Scale for Pairwise Comparisons*, (Saaty, 2008)) qui spécifie le degré d'importance d'une alternative par rapport à une autre en fonction du score de chacune pour le critère

étudié (c.f tableau 1). Notons que si a_{ij} est l'importance du candidat *i* par rapport au candidat *j*, alors $a_{ji} = \frac{1}{a_{ij}}$.

Tableau 1: exemple de décideur simple d'après le principe de The Fundamental Scale for Pairwise Comparisons (Saaty, 2008).

The Fundamental Scale for Pairwise Comparisons		
Importance	**Définition**	**Explication**
1	Importance égale	*Deux éléments contribuent également à l'objectif*
3	Importance modérée	*La comparaison favorise modérément un candidat par rapport à l'autre*
5	Importance forte	*La comparaison favorise fortement un candidat par rapport à l'autre*
7	Importance très forte	*La comparaison favorise très fortement un candidat par rapport à l'autre*
Les importances 2, 4 et 6 sont utilisées pour exprimer des valeurs intermédiaires. Les importances 1.1, 1.2, etc.. sont utilisées pour des alternatives d'importance relatives très proches.		

Le résultat de cette comparaison est stocké dans une matrice de taille *(n+1)*(n+1)* où *n* est le nombre de candidats confrontés. Un vecteur de priorités est ensuite calculé rendant compte de l'importance relative des candidats pour ce critère. Mathématiquement, ce vecteur correspond au vecteur propre principal de la matrice. Un des avantages de la méthode AHP est qu'elle calcule un indicateur de cohérence (appelé ratio de consistance et noté CR), permettant d'évaluer les calculs effectués. En d'autres termes, le CR donne une indication sur l'inconsistance faite dans le jugement des confrontations, en vérifiant si les valeurs des importances attribuées par le décideur sont cohérentes entre elles. En pratique, nous cherchons à nous assurer que $CR < 0.10$ (Saaty, 2008). Dans l'exemple illustré dans le tableau 2, c'est le candidat 1 qui l'emporte logiquement pour le critère étudié.

Tableau 2 : Exemple de matrice de comparaison pour un critère arbitraire

Scores (entre 0 et 1)	Matrice de comparaison				Priorités
		Candidat 1	Candidat 2	NAP	
Candidat 1 : 0.7	Candidat 1	1	3	4	0.63
Candidat 2 : 0.4	Candidat 2	1/3	1	1/2	0.15
	NAP	1/4	2	1	0.22
					CR=0.09

2.3.2.3 Confrontation mutuelle des critères

L'AHP permet de pondérer chaque critère en fonction de l'importance que nous lui accordons par rapport à chaque autre critère. Autrement dit, l'AHP nous permet

de définir une matrice de poids de taille *m*m*, avec m le nombre de critères (ici m=5), ce qui ajoute beaucoup de finesse dans la définition des importances relatives des critères, généralement faite par des experts. À partir de la matrice de confrontation des critères, un vecteur de priorités et un ratio d'inconsistance sont calculés comme précédemment.

2.3.2.4 Prise de décision

Nous calculons pour chaque candidat une priorité globale, combinant les priorités du candidat pour chaque critère et tenant compte des priorités relatives des critères les uns par rapport aux autres. Soient $(C_1, ..., C_N)$ N candidats, $(\alpha_{i,1}, \dots, \alpha_{i,N})$ le vecteur de priorités des candidats pour le i^{ème} critère, et $(p_1, ..., p_m)$ le vecteur de priorités issus de la confrontation des critères. La priorité globale $P(C_j)$ du candidat C_j est donnée par: $P(C_j) = \sum_{i=1}^{m} p_i * \alpha_{i,j}$. Nous choisissons comme solution le candidat ayant la plus grande priorité globale. Si c'est l'alternative NAP qui est ici choisie, alors l'objet est non apparié.

2.4 Traitement des strokes non appariés

Les strokes non appariés par le processus hiérarchique sont traités séparément, et pour chacun d'entre eux nous appliquons l'algorithme d'appariement de Mustière et Devogele (2008). Nous nous restreignons au sous-réseau constitué du stroke étudié, et des strokes qui lui sont connectés afin d'éviter d'introduire des effets de bord vu la simplification topologique qui est effectuée ici. Les liens issus de cette étape sont qualifiés d'incertains et nécessitent une vérification manuelle en post-traitement.

3. Évaluation des résultats

3.1 Mesures d'évaluation de l'appariement

Il est courant d'évaluer un processus d'appariement en termes de précision P_{app} et de rappel R_{app} des liens d'appariement (resp. P_{nap} et R_{nap} pour les objets non appariés). Nous appelons classiquement: vrais positifs les liens d'appariements correctement établis par le processus automatique (*vp*), faux positifs les liens établis par erreur (*fp*), vrais négatifs les objets correctement non appariés (*vn*) et faux négatifs les objets non appariés par erreur (*fn*). Notons N_{app} le nombre de liens d'appariement attendus, et N_{nap} le nombre d'objets non appariés attendu. Nous utilisons une mesure qui combine équitablement la précision et le rappel, appelée *Fscore*.

$$\left\{P_{app} = \frac{vp}{vp + fp}, R_{app} = \frac{vp}{N_{app}}\ P_{nap} = \frac{vn}{vn + fn}, R_{nap} = \frac{vn}{N_{nap}}\ , Fscore = 2 * \frac{precision * rappel}{precision + rappel}\right.$$

3.2 Tests et évaluation

Nous testons notre approche d'appariement sur les données issues des cartes de Cassini et vectorisées dans le cadre du projet GeoPeuple[1]. La spécificité de l'approche proposée permet de prendre en compte les difficultés liées à l'appariement des données anciennes: une différence de temporalité majeure entre les données (250 ans) impliquant des transformations des entités cartographiées, une différence de niveau de détail et de représentation, et des imperfections dues à l'imprécision des fonds de cartes. L'appariement est effectué avec la BDCarto de l'IGN. Nous donnons à titre indicatif le temps de calcul nécessaire à l'appariement, les strokes ayant été constitués au préalable, effectué sur une machine 4 cœurs (4*2.83GHz, 4 Go RAM).

Tableau 3 : Résultats quantitatifs de l'évaluation de trois approches d'appariement

	Liens d'appariement			Objets non appariés			Temps de calcul (en secondes)
Zones	**Précision**	**Rappel**	**Fscore**	**Précision**	**Rappel**	**Fscore**	
	Approche de Mustière et Devogele (2008)						
Reims	73%	85%	78%	44%	80%	57%	13.5
Grenoble	58%	76%	66%	89%	85%	87%	12.5
St Malo	63%	91%	74%	86%	71%	77%	6.1
	Approche de Costes et al. (2012)						
Reims	90%	87%	89%	57%	80%	67%	13.6
Grenoble	67%	82%	74%	97%	85%	90%	15.1
St Malo	48%	68%	56%	86%	71%	77%	11
	Approche hiérarchique multi-critères						
Reims	89%	93%	91%	50%	80%	62%	16.6
Grenoble	89%	71%	80%	90%	97%	93%	8.1
St Malo	96%	73%	83%	92%	100%	96%	4.2

Les trois méthodes étudiées comportent de nombreux paramètres, et nous comparons les meilleurs résultats obtenus par chacune des approches en les faisant varier. Comme l'illustre la figure 7, les grandes structures hydrographiques sont correctement appariées. Les résultats du tableau 3 soulignent que notre approche permet d'améliorer sensiblement les résultats par rapport aux approches classiques: +17% (resp. +14%) en moyenne pour le *FScore* des liens d'appariement (resp. des objets non appariés) par rapport à l'approche de Mustière et Devogele (2008) ; +19% (resp. +7%) en moyenne pour le *FScore* des liens d'appariement (resp. des objets non appariés) par rapport à l'approche de Costes et al. (2012). Le temps de calcul est amélioré en moyenne.

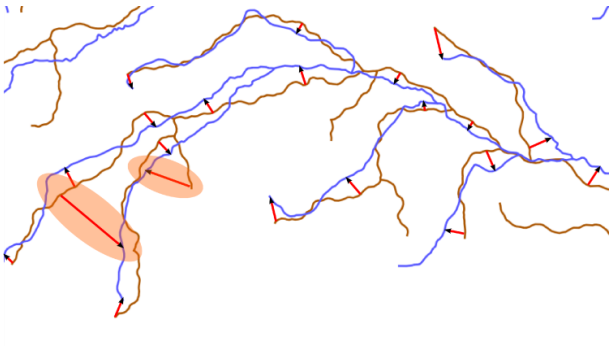


Figure 7 : Exemple d'appariement sur la zone de Grenoble (2 erreurs en orange)

3.3 Étude préliminaire de sensibilité aux paramètres

Le paramétrage de la matrice de confrontation des critères peut s'avérer délicat, et nous nous posons la question de la sensibilité de l'algorithme à ces paramètres. Pour l'étude, nous considérons que chaque critère a la même importance (i.e. les poids de la matrice valent tous 1). Le tableau 4 montre, à notre grande surprise, que les résultats sont significativement améliorés si le critère de distance géométrique n'est pas utilisé. Nous interprétons ce résultat comme la conséquence des décalages géométriques entre les réseaux étudiés : deux strokes proches géométriquement ne sont pas nécessairement homologues.. Le critère toponymique apparait essentiel pour l'appariement de nos données d'évaluation. Les graphiques de la figure 8 sont obtenus en faisant varier le poids d'un critère relativement à un autre, entre *1/5* et *5*. La comparaison des critères d'orientation et de proximité globale montre qu'ils sont d'importance égale. Globalement, l'algorithme semble être peu sensible à la variation des poids des critères, tant que le critère géométrique n'est soit pas utilisé, soit d'importance faible : le F-score des liens d'appariement reste supérieur à 80%.

Tableau 4: scores obtenus en supprimant certains critères.

	F-Score liens d'appariement	**F-Score objets non appariés**
Tous critères	83%	40%
Sans critère géométrique	86%	53%
Sans critère toponymique	42%	17%
Sans critère d'orientation	71%	27%
Sans critère de proximité	73%	28%

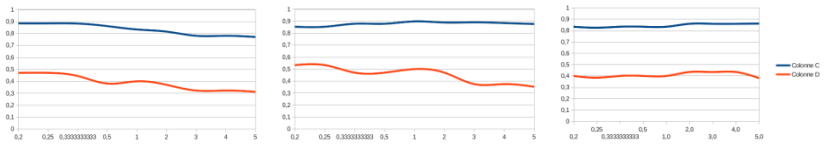


Figure 8 : évolution du F-Score des liens d'appariement (bleu) et des objets non appariés (rouge) en fonction du poids, de gauche à droite, des critères : orientation / distance, proximité globale / orientation, proximité globale / toponymie.

Conclusion

Les imperfections des données géographiques rendent difficile leur appariement. Des décalages géométriques majeurs, des topologies peu similaires, des imprécisions géométriques importantes, des échelles et niveaux de détails différents, des incomplétudes sémantiques significatives sont autant d'erreurs qui rendent inefficaces les méthodes classiques d'appariement de réseaux. Nous proposons une approche originale d'appariement de réseaux hydrographiques basée sur le respect de la structure hiérarchique des réseaux. Par une étape d'enrichissement, nous changeons leur mode de représentation en détectant les strokes, des structures naturelles de bonne continuité représentant réellement les entités hydrographiques telles que les rivières et les fleuves. L'agencement hiérarchique ainsi défini sert de pivot à l'appariement, en ne retenant parmi les candidats potentiels que ceux dont les parents, ou eux-mêmes, ont été préalablement appariés La sélection hiérarchique des strokes candidats à l'appariement combinée à l'utilisation d'une technique efficace de décision multi-critères (AHP) permettent d'améliorer sensiblement les résultats obtenus par les méthodes basées uniquement sur ces critères géométriques et topologiques et les approches par analyse multi-critères non hiérarchiques. L'évaluation est faite sur des données anciennes, présentant près de 250 ans d'écarts, issues des fonds de cartes de Cassini (18ème siècle). Si notre approche est moins efficace sur les feuilles des réseaux très ramifiés, les structures hydrographiques principales sont correctement appariées. La méthode de prise de décision par AHP pose le problème de son paramétrage. Une étude de sensibilité préliminaire de l'algorithme aux poids de la matrice de prise de décision semble montrer que, d'une part, les résultats s'améliorent lorsque le critère de distance géométrique n'est pas utilisé, et d'autre part, que notre approche est peu sensible à la variation des paramètres. Le critère toponymique reste cependant le plus important sur nos données d'évaluation. Nous pensons qu'une paramétrisation empirique faite par des experts est suffisante pour obtenir un appariement satisfaisant. Il serait intéressant à court terme, de continuer l'étude de sensibilité en faisant varier l'ensemble des paramètres de la matrice de confrontation simultanément. A plus long terme, une adaptation de notre approche à d'autres types de réseaux est envisagée.

Remerciements. Nous remercions vivement Julien Perret et Eric Grosso pour leurs relectures et précieux conseils.

Abadie N. (2012). Intégration des bases de données à partir de la formalisation de leurs spécifications. Thèse en Sciences de l'Information Géographique, Université de Marne-La-Vallée

Bel Hadj Ali A. (2001). Qualité géométrique des entités géographiques surfaciques. Application à l'appariement et définition d'une typologie des écarts géométriques. Thèse en Sciences de l'Information Géographique, Université de Marne-La-Vallée

Bouchon-Meunier B. (1995). La logique floue et ses applications. Paris : Addison-Wembley France, p. 257

Costes B., Grosso E., Plumejeaud C. (2012). Géoréférencement et appariement des données issues des cartes de Cassini. In: Actes du Colloque International de Géomatique et d'Analyse Spatiale (SAGEO 2012)

Damerau F.J.A (1964). Technique for computer detection and correction of spelling errors Commun. ACM, vol. 7, p. 171-176

Dempster A. (1967). Upper and lower probabilities induced by multivalued mapping. Annals of Mathematical Statistics, vol. AMS-38, p. 325-339

Devogele T. (1997). Processus d'intégration et d'appariement de bases de données Géographiques, Applications á une base de données routières multi-échelles. Thèse en Sciences de l'Information Géographiques, Université de Versailles

Devogele T., Badard T., Libourel t. (2002). La problématique de la représentation multiple. In: Ruas, A. (éd), Généralisation et représentation multiple. Paris, Lavoisier, 390 p.

Eiter T., Mannila H. (1994). Computing discrete Fréchet distance. Christian Doppler Laboratory for Expert Systems , rapport CD-TR 9464

Fisher P.F. (2003). Models of uncertainty in spatial data. Geographical Information System, vol. 1, p. 191-203

Goodchild M.F. (1995). Sharing imperfect data. In H.J. Onsrud and G. Rushton, editors, Sharing Geographic Information, New Brunswick, NJ: Rutgers University Prress, 1995, p. 413-425

Goodchild M.F., Hunter G.A, (1997). A simple positional accuracy measure for linear features. International Journal of Geographical Information Systems, vol. 11, p. 299-306

Hangouët J.F (1998). Approche et méthodes pour l'automatisation de la généralisation cartographique;application en bord de ville. Université de Marne-La-Vallée

Horton, R.A. (1945). Erosional development of Streams and their drainage basins: hydrophisical approach to quantitative morphology. Geological Society of America Bulletin, vol. 56, p. 275-370

Hunter G.J. (1998). Managing uncertainty in GIS. NCGIA Core Curriculum in GISScience, fév. 1998

Jiang B., Zhao S., Yin J. (2008). Self-organized Natural Roads for Predicting Traffic Flow: A Sensitivity Study. Journal of Statistical Mechanics: Theory and Experiment

Levenshtein V. (1965). Binary codes capable of correcting deletions, insertions and reversals. Doklady Akademii Nauk SSSR, vol. 4, p. 845-848

Lüscher P., Burghardt D. Weibel R. (2007). Matching road data of scales with an order of magnitude difference. In : 23th International Cartographic Conference

Mustière S. (2006). Results of experiments on automated matching of network at different scales. ISPRS, vol. 26.

Mustière S., Devogele T. (2008). Matching networks with different levels of detail. GeoInformatica, p. 435-453

Olteanu A.M. (2008). Fusion de connaissances imparfaites pour l'appariement de données géographiques. Proposition d'une approche s'appuyant sur la théorie des fonctions de croyance.Thèse en Sciences de l'Information Géographique, Université Paris-Est

Saaty T. (2008). Decision making with the analytic hierarchy process. International Journal of Services Sciences

Samal A., Seth S., Cueto K.A. (2004). A Feature-based approach to conflation of geospatial sources. International Journal of Geographical Information Sciences, vol. 18, p. 459-489

Shafer G. (1976). A mathematical Theory of Evidence. Princeton: Princeton University Press, 1976

Thomson R., Richardson D. (1999). The 'Good Continuation' principle of perceptual organization applied to the generalization of road networks. In : 19th International Cartographic Conference

Thomson R., Brooks R. (2000). Efficient generalization and abstraction of network data using perceptual grouping. In : 5th GeoComputation

Walter F., Fritsch D. (1999). Matching spatial data sets : a statistical approach. International Journal of Geographical Information Science, vol. 13:5, p. 445:473

Worboys M., Duckham M. (2004). Computation with Imprecise Geospatial Data. Computers, Environment and Urban systems, vol. 22, p. 85-106

Zadeh L.A. (1965). Fuzzy sets. Information and Control, vol. 8, p. 338-353

Zhang M., Shi W., Meng, L.A (2005). A generic matching algorithm for line networks of different resolutions. In: ICA Workshop on Generalisation and Multiple Representations

Annex: algorithme de construction des strokes

Cet algorithme étend l'algorithme « every-best-fit » de Jiang et al. (2008) en prenant en compte la sémantique et la toponymie de cours d'eau lorsqu'elles sont disponibles, en plus des critères angulaires. Les modifications apportées sont surlignées.

Input: the current segment, the search direction

Output: the next segment

```
buildStroke(currentSegment, direction){
    search the segments intersected with currentSegment in the good direction;
    Remove already processed segments;
    If(there are no intersected segments)
        Return currentSegment;
    End if
    If(currentSegment has toponym)
        Return checkAttributeContinuity(currentSegment, "toponym");
    End if
    Exclude the segments which have the same toponym
    If(there are no remaining segments)
        Return currentSegment;
    Else if(there is only one remaining segment)
                Return that segment;
    End if
    If(currentSegment has nature attribute)
        Return checkAttributeContinuity(currentSegment,"nature");
    End if
    Exclude the segments which have the same nature attribute
    If(there are no remaining segments)
        Return currentSegment;
    Else if(there is only one remaining segment)
        Return that segment;
    End if
    Calculate the delfection angles (a1) bewteen currentSegment and every segment   in the remained set;
    Calculate the deflection angle (a2) of every pair in the remained set;
    Select the segments wich meet with the condition a1 < a2;
    If(there are no selected segments)
        Return currentSegment;
    End if
    Get the minimum deflection mina1 angle from a1 and its corresponding segment;
    If(mina1 < thresold)
        Change the status of that segment to be processed;
        Return that segment;
    Else
        Return currentSegment;
    End if
}
```

Input: the current segment, the attribute to check
Output: the next segment with the best attribute continuity

```
checkAttributeContinuity(currentSegment, attribute){
    Select the segments with the same attribute to get a remained set;
    If(there are no remaining segments)
        Return currentSegment;
    Else if(there is only one remaining segment)
        Return that segment;
    Else
    Calculate the delfection angles (a1) bewteen currentSegment and every segments in the remained set;
    Calculate the deflection angle (a2) of every pair in the remained set;
    Select the segments wich meet with hte condition a1 < a2;
    If(there are no selected segments)
        Return currentSegment;
    End if
    Get the minimum deflection mina1 angle from a1 and its corresponding segment;
    Return that segment
}
```

Visibilité et prévisibilité : cartographie de la part cachée d'un paysage urbain

Etude comparée de quatre indicateurs qualifiant le champ visuel en milieu urbain

Thomas Leduc[1], Guillaume Chauvat[2]

1. Laboratoire CERMA, UMR CNRS 1563, Ecole Nationale Supérieure d'Architecture de Nantes
6 quai François Mitterrand, BP 16202, F-44262 Nantes cedex 2, France
thomas.leduc@cerma.archi.fr

2. Ecole Nationale Supérieure d'Architecture de Nantes
6 quai François Mitterrand, BP 16202, F-44262 Nantes cedex 2, France
chauvat.guillaume@gmail.com

RESUME. Pour la théorie de la perméabilité, la configuration des lieux et plus précisément la présence d'occlusions dans notre environnement visuel immédiat, peuvent nous inciter à son exploration. L'étude présentée dans cet article a pour but d'analyser de manière systématique et reproductible la part d'inconnu, d'incertitude de l'environnement urbain immédiat d'un piéton. Pour évaluer cette propriété du visualscape en tout point, nous rappelons trois indicateurs permettant de quantifier la prévisibilité de la forme ambiante (entropie), le poids des masses visuelles (drift) et la part d'occlusions (occlusivité) avant d'en proposer un quatrième relativisant le poids de chaque occlusion (l'anticipation). Nous concluons cette analyse par une référence aux techniques de partitionnement convexe de l'espace ouvert.

ABSTRACT. According to permeability theory, surroundings' configuration, and more precisely, the existence of occlusions in our enclosure may invite us to explore our environment. This article aims to analyze, in a systematic and reproducible way, the ratio of mystery in one's surrounding visualscape. To assess this property in each place, we recall three indicators relative to the boundary previsibility (entropy), the weight of visual heap (drift), and the occlusions ratio (occlusivity). We also introduce a fourth indicator which weights each occlusion (anticipation). At last, we refer to some convex partition technics of urban open spaces.

MOTS-CLES : occlusivité, drift, entropie, exploration environnementale, mystère, anticipation, morphologie, isovist.

KEYWORDS: occlusivity, drift, entropy, environmental exploration, mystery, morphology, isovist.

1. Introduction

Comme le rappelle (Brayer, 2013), la façon dont nous construisons les villes peut avoir un effet profond sur la manière dont nous les vivons. De ce constat a émergé un ensemble de préoccupations contemporaines portées aujourd'hui par la notion d'ambiance. Cette notion prend en compte les dimensions culturelles et conjoncturelles d'un lieu pour saisir son identité, parce que les pratiques d'un lieu ne sont pas réductibles à ses seules caractéristiques construites. Dans le même registre, pour (Péneau, 2000, p. 379), la seule approche des dimensions physiques ne suffit pas à rendre compte de la consistance phénoménologique des ambiances. Le sujet-citadin est ainsi marqué par toute une gamme de différenciateurs culturels, psychologiques et sociaux, représentant autant de filtres à la perception des phénomènes objectifs tels que la température ambiante, les rayonnements, les vitesses d'air, etc. Pour le concepteur d'environnements urbains, ce constat est lourd d'implications. Pour le bureau d'étude spécialisé dans certaines des dimensions phénoménales impliquées dans le processus de projet ou le modélisateur, il est implacable. La modélisation et la simulation d'une ou plusieurs dimensions physiques d'un environnement urbain ne suffisent pas à expliquer, seules, la complexité intrinsèque du *genius loci*.

Cette précaution ayant été prise, nous pouvons néanmoins reconnaître avec (Stamps III, 2010) que la configuration géométrique et topologique des lieux peut encourager ou inciter le piéton à l'exploration de son environnement immédiat. Pour (Kaplan, 1973), évaluer la part de mystère d'une scène signifie être en mesure de répondre à la question suivante : *to what degree do you think you would learn more if you could walk deeper into the scene* ? Plus précisément, pour (Cullen, 1961), il y a une promesse de gain d'information si le sujet a la possibilité de rejoindre un meilleur point de vue, c'est-à-dire encore si l'objet occlusif qui s'intercale entre le point de vue et l'arrière plan caché peut-être rejoint et offrir ainsi une nouvelle perspective sur la part d'ombre du paysage (ce qu'il appelle the *sense of mystery*). En fait, plus même que de mystère, la bonne terminologie, pour (Cullen, 1961), semble être l'anticipation. Il met ainsi en perspective le proche (l'« ici ») et le lointain (le « là-bas ») où l'« ici » est connu et le « là-bas » ne l'est pas et énonce que s'engager sur un nouveau tronçon est plus motivant si nous ne savons pas à l'avance ce qui nous attend (*The next stretch is more inviting if we don't know completely what awaits us*).

L'occlusion, la part d'inconnu d'un paysage, n'est cependant pas que stimulante ou attirante. Elle peut aussi être génératrice d'anxiété. (Stamps III, 2005) développe ainsi une théorie de la perméabilité selon laquelle le contour de l'espace ouvert immédiatement environnant est étroitement corrélé à l'impression de sécurité du sujet-piéton. Il distingue alors les frontières matérielles ou solides (en l'occurrence les façades opaques qui arrêtent le regard et préviennent toute pénétration dans l'environnement visuel immédiat), des frontières occlusives (immatérielles) qui résultent d'un effet de masque et cachent une part d'espace imprévisible.

L'étude que nous restituons ici a pour but d'analyser de manière systématique, objective et reproductible la part d'imprévu, d'incertitude d'un paysage urbain. Les

enjeux de cette recherche sont triples : le diagnostic d'abord, par l'identification des potentialités associées aux spécificités locales de forme, la compréhension des corrélations et des phénomènes ensuite, la conception cohérente d'environnements urbains exploitant cette dimension enfin. Cette analyse, pour pouvoir être transposée à d'autres terrains, met en œuvre un modèle qui exploite des jeux de donnée vectorielle standards, dans le contexte d'un Système d'Information Géographique (SIG). L'outil méthodologique que nous avons retenu est le *visualscape* défini dans (Llobera, 2003, p. 30) comme la représentation spatiale de toute propriété visuelle générée par, ou associée à, une configuration spatiale des lieux. On attribue traditionnellement les premières formalisations et représentations de *visualscape* à (Benedikt, 1979) qui a introduit la notion d'isovist (polyèdre - ou polygone si l'on travaille dans un plan 2D - formé de l'ensemble des portions de l'espace visibles depuis un lieu d'observation donné) et de champ d'isovists. Un champ d'isovists est, pour Benedikt, un champ scalaire, qui associe à tout point de l'espace (discrétisé) une grandeur qui représente l'environnement visible immédiat. Comme le remarque (Morello & Ratti, 2009), le champ d'isovists offre une perspective synthétique, sous forme cartographique, de l'ensemble des valeurs d'une même propriété visuelle en tout point du terrain d'étude

Pour analyser la part d'inconnu du *visualscape* en tout point de l'espace d'étude, nous utilisons quatre indicateurs. Trois d'entre eux sont issus de la bibliographie du domaine. Ils permettent de quantifier respectivement la prévisibilité de la forme (notion empruntée à la théorie de l'information), le « poids des masses visuelles » ou la part d'occlusions dans le bassin de visibilité. Le quatrième, développé au cours d'un stage réalisé au sein du laboratoire CERMA, s'attache à mesurer l'importance visuelle de ce qui échappe à notre champ de vision, par un contraste de plans. Ce dernier a plus particulièrement été formulé en vue d'une analyse des motivations intrinsèques au déplacement du piéton en situation de « mouvement naturel » (Hillier *et al.*, 1993). Dans cet article, après une présentation détaillée et motivée de la notion de champ d'isovists et du mode de calcul retenu pour l'évaluer, nous introduirons les quatre indicateurs que nous avons choisis pour analyser la part d'inconnu d'un environnement urbain. La méthodologie ayant été exposée, nous l'appliquerons à un terrain d'étude dans le centre de Nantes (France), puis discuterons leurs apports respectifs. Nous conclurons cette analyse par une référence aux techniques de partitionnement de l'espace de (Peponis *et al.*, 1997).

2. Méthode

L'analyse systématique de *visualscape* urbain que nous voulons conduire requiert un outil logiciel robuste et performant. Nous utilisons, pour le traitement spatial, l'instance de SIG Gearscape et, pour la visualisation, Quantum GIS (QGIS). Concernant l'environnement urbain ou plus exactement sa maquette, nous exploitons la couche bâtiments de la BD TOPO®. Nous faisons l'hypothèse que toutes les façades sont opaques et n'intégrons, dans notre modèle, ni les transparences, ni les divers obstacles au champ de vision comme la végétation, le mobilier urbain, etc. Le champ de vision que nous modélisons est périphérique 2D, il n'est en aucun cas

directionnel dans la mesure où nous cherchons à qualifier un ressenti d'espace ouvert immédiatement environnant plus qu'un champ de vision réaliste. Il est néanmoins envisageable de le formuler en 3D sous réserve de disponibilité de données parfois imprécises (les hauteurs à l'épannelage) ou difficiles à obtenir et à mettre à jour (les types de toits, pentes des toitures, etc.).

Pour (Joliveau, 2004, p. 121), des fonctions d'analyse de visibilité ont été implémentées dans les SIG sous le nom de *viewshed analysis*. Elles utilisent couramment des MNT (Modèles Numériques de Terrain) ou des TIN (réseau de triangles irréguliers) et évaluent l'espace visible depuis des cibles matérialisées sous forme de points vecteurs ou de cellules de grille raster. En environnement urbain, plusieurs approches pour délimiter automatiquement l'espace visible d'un piéton reposent plutôt sur des données vectorielles (Suleiman *et al.*, 2012). La plus connue, le *ray tracing*, consiste à évaluer l'intersection entre un ensemble de rayons équi-répartis autour du point de vue (3600 pour l'implémentation de (Rana, 2006) dans ArcView) et les premiers obstacles de l'environnement proche. Une variante de cette méthode, dite de l'agent en marche (Batty & Jiang, 1999), consiste non seulement à discrétiser l'azimut comme dans la méthode du *ray tracing*, mais aussi à discrétiser la distance au point de vue. (Turner *et al.*, 2001) ont proposé une troisième méthode consistant à discrétiser préalablement l'espace d'étude et à procéder en tout point à des évaluations d'intervisibilités avec l'ensemble des autres points du maillage.

2.1. Calcul « exact » de l'isovist en 2D

Comme le constate (Suleiman *et al.*, 2012), l'évaluation du bassin de visibilité en milieu urbain par l'une des approches mentionnées ci-avant dépend du nombre de rayons lancés, de la précision du « pas de l'agent », du nombre de points de maillage, ou de la résolution du pixel. L'environnement urbain étant susceptible (dans les tissus organiques du moins) de présenter une grande complexité du déroulé de façades, nous avons opté pour une approche « exacte ». Par « exacte », nous entendons ici que la méthode retenue ne procède pas par discrétisation de l'espace, par discrétisation par abscisse angulaire, ou par discrétisation des arêtes opaques. Nous travaillons dans un espace vectoriel continu (aux limitations de l'arithmétique flottante IEEE 754 de la machine virtuelle Java près). Calqué sur le constat de (Suleiman *et al.*, 2012), notre implémentation fait l'hypothèse que tout fragment urbain peut être assimilé à un ensemble de segments sans intersection sinon éventuellement aux extrémités. Elle reprend et adapte ensuite l'algorithme de détermination des arêtes cachées d'une chaîne graphique classique :

- élimination des faces ou arêtes cachées à partir de l'orientation des normales aux faces (*backface culling*),

- élimination des faces ou arêtes hors du champ ou cône de vision (*viewing frustum culling*). Comme la vision est périphérique sur 360°, nous ne pouvons en l'occurrence pas parler de cône de vision. Sont donc éliminés lors de cette phase tous les objets trop distants (pour des raisons d'efficacité, nous exploitons ici les index spatiaux de JTS). Au cours de cette même phase, nous procédons à un *clipping*. Celui-ci consiste à découper les arêtes partiellement visibles ;

- élimination des faces ou arêtes qui se cachent mutuellement (*occlusion culling*). D'un point de vue pratique, nous recherchons les points minimum (minimum de potentiel) dans une représentation polaire (centrée sur le point de vue) des arêtes.

La quatrième et dernière phase d'élimination des objets minuscules (*contribution culling*) n'est volontairement pas implémentée dans notre processus.

2.2. L'entropie de Shannon de la fonction de distances radiales comme indicateur de prévisibilité du bassin de visibilité

L'isovist, associé à un point de vue, est un polygone qui présente la particularité de contenir tout segment reliant chacun de ses points à ce même point de vue (appelé aussi point de génération). Cette propriété topologique permet de simplifier l'étude de l'isovist, en la réduisant à l'analyse de son contour. Nous pouvons donc caractériser ce polygone par l'étude de la fonction d'une variable réelle 2π-périodique qui, à un azimut donné (l'abscisse angulaire), associe la longueur du rayon reliant le point de vue au point de contour correspondant (cf. figure 1).

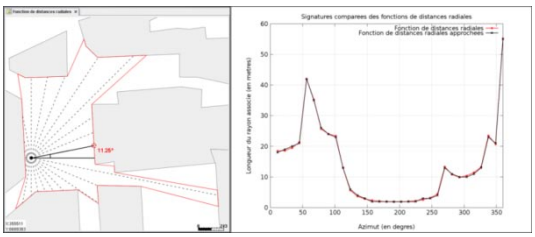


Figure 1. (gauche) Représentation en plan, pour un fragment urbain (les empreintes polygonales de bâtiments sont en gris clair), d'un point de vue et de son isovist. (droite) Développé en fonction de distances radiales correspondant.

Cette fonction mathématique d'une variable réelle peut alors être étudiée à l'aide d'un outil de la théorie de l'information. Comme rappelé par (Salat, 2011, p. 491), la formule de l'entropie de Shannon, en quantifiant la quantité d'information contenue dans une distribution, est une façon de quantifier la complexité de cette dernière.

L'entropie de la fonction de distances radiales de l'isovist d'un observateur placé au centre d'un espace ouvert parfaitement circulaire est nulle. En effet, dans cette configuration, tous les rayons visuels sont de même longueur, l'éloignement du point de vue au contour est parfaitement prévisible (car il est constant). En pratique, pour pouvoir calculer cette entropie, nous avons procédé à une discrétisation préalable des longueurs de rayons. Comme le montrent les signatures comparées de la figure 1, un échantillonnage avec une précision métrique ne modifie pas significativement la fonction des distances radiales.

6

2.3. Le drift comme révélateur du « poids des masses visuelles »

Le *drift* introduit par (Conroy Dalton & Dalton, 2001), correspond à la distance euclidienne séparant le point de vue du centre de masse de l'isovist correspondant (figure 2). Cette grandeur permet de mesurer l'éloignement du centre visuel au centroïde de l'espace ouvert correspondant, c'est-à-dire de quantifier l'impression éventuelle de l'observateur d'être ou non situé au centre de son bassin de visibilité.

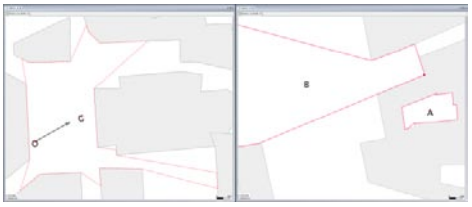


Figure 2. (gauche) Le drift de O est la distance qui le sépare de G (centre de gravité de l'isovist correspondant). (droite) Deux situations extrêmes pour l'indicateur drift.

Pour (Leduc & Kontovourkis, 2012, p. 70), le *drift* est non seulement une mesure de l'excentricité du point de vue dans son bassin visuel, mais aussi une mesure de l'effort requis pour superposer ce point au centre de gravité de l'espace ouvert immédiatement associé.

2.4. L'occlusivité comme ratio révélateur de la part d'inconnu du paysage

L'occlusivité, (Benedikt, 1979), correspond au cumul des longueurs d'arêtes adjacentes à des portions d'espace ouvert invisibles depuis le point de vue (figure 3). C'est un cumul de longueurs d'arêtes qui ne sont pas matérielles (et donc pas opaques). En pratique, la présence d'une occlusion se traduit par une concavité dans le bassin de visibilité correspondant. Cette notion est aussi régulièrement utilisée en aéraulique urbaine par des auteurs tels que (Adolphe, 2001; Edussuriya *et al.*, 2011). Dans le registre de la visibilité, un tel indicateur permet de mesurer l'importance de ce qui échappe à notre champ de vision (donc de quantifier le poids des parties cachées d'un paysage).

Figure 3. (gauche) L'occlusivité de O est la somme des longueurs des arêtes AiBi qui sont occlusives car elles séparent le bassin de visibilité de zones d'espace ouvert invisibles depuis O. (droite) Deux situations extrêmes pour l'indicateur d'occlusivité.

2.5. L'anticipation ou attraction des masses cachées du paysage

Ce quatrième indicateur, appelé « anticipation » et développé au cours d'un stage réalisé au sein du laboratoire CERMA (Chauvat, 2012), s'attache, comme l'occlusivité, à mesurer l'importance visuelle de ce qui échappe à notre champ de vision. Néanmoins, contrairement à l'occlusivité, il relativise l'importance de chaque occlusion en fonction de sa distance au point de vue d'une part, de sa longueur d'autre part et de la distance du point de vue à l'arrière plan enfin. La formulation mathématique de cet indicateur est donnée en équation 1 (la norme utilisée est la norme euclidienne), comme on peut le constater, la distance du point de vue à l'occlusion est au dénominateur car l'impact visuel de l'occlusion décroit avec la distance. Une illustration de cet indicateur dans le cas d'une configuration urbaine nantaise réelle est donnée en figure 4.

$$A = \sum_{i=0}^{n-1} \frac{\|\overrightarrow{OB_i}\| \times (\|\overrightarrow{OB_i}\| - \|\overrightarrow{OA_i}\|)}{\|\overrightarrow{OA_i}\|} = \sum_{i=0}^{n-1} \left(\frac{\|\overrightarrow{OB_i}\|^2}{\|\overrightarrow{OA_i}\|} - \|\overrightarrow{OB_i}\| \right) \quad (1)$$

Cet indicateur est dimensionné (il s'exprime en mètre). Il n'est pas normalisé.

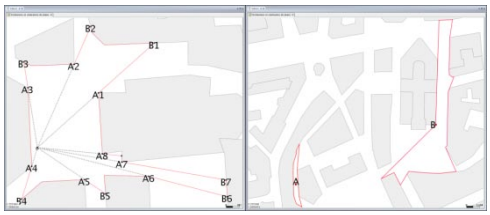


Figure 4. (gauche) Dans l'isovist associé au point de vue O, les points Ai matérialisent les huit occlusions associées à O et les points Bi les arrières plans correspondants. (droite) Deux situations extrêmes pour l'indicateur d'anticipation.

3. Cas d'application

Dans le cadre du stage (Chauvat, 2012), déjà mentionné, a été identifié un ensemble de terrains adaptés au « mouvement naturel » du piéton (dans le sens où les obstacles à cette mobilité sont limités), d'altimétrie et d'élévation à l'épannelage relativement constantes, ainsi que de composition architectonique homogène. Ces quatre terrains, tous situés dans le centre de Nantes, correspondent aux secteurs de la place Royale, des places du Change et de Sainte Croix, de la place du Bouffay et enfin de la place du Commerce. Ces places sont de configurations diverses (Morin, 1995), ce qui permet d'éprouver les indicateurs avec différentes morphologies.

Pour le calcul du bassin de visibilité, étant donné le contexte urbain de cette étude, nous avons arrêté la portée du rayon de visibilité à 100 m. Nous estimons en effet empiriquement que l'évaluation du poids des masses cachées par l'indicateur anticipation n'est pas significativement modifiée par la prise en compte d'une plus grande portée du champ de vision, au regard des rapports de distances observables dans cet environnement.

8

Pour analyser les valeurs des quatre indicateurs nous avons procédé à un échantillonnage orthogonal inframétrique (la résolution planimétrique est de 50 cm par 50 cm) des quatre secteurs. Les valeurs des points (de 135000 pour le secteur du Change et de Sainte Croix à 340000 pour celui de la place Royale) ont été « rasterisées » avant d'être visualisées à l'aide de QGIS. La planche de la figure 5 (seul le secteur de la place Royale est reproduit ici en raison de contraintes éditoriales) constitue une synthèse sous forme de cartes thématiques choroplèthes, par secteur, de ces traitements. Pour améliorer les contrastes, nous avons opté pour la méthode d'étirement jusqu'au MinMax (cas de l'entropie, du *drift* et de l'occlusivité) et pour la méthode d'étirement par écart-types, à la valeur 2, calculés sur la distribution des valeurs (cas de l'anticipation). Enfin, nous avons inversé les palettes de couleurs des cartes de *drift* et d'occlusivité.

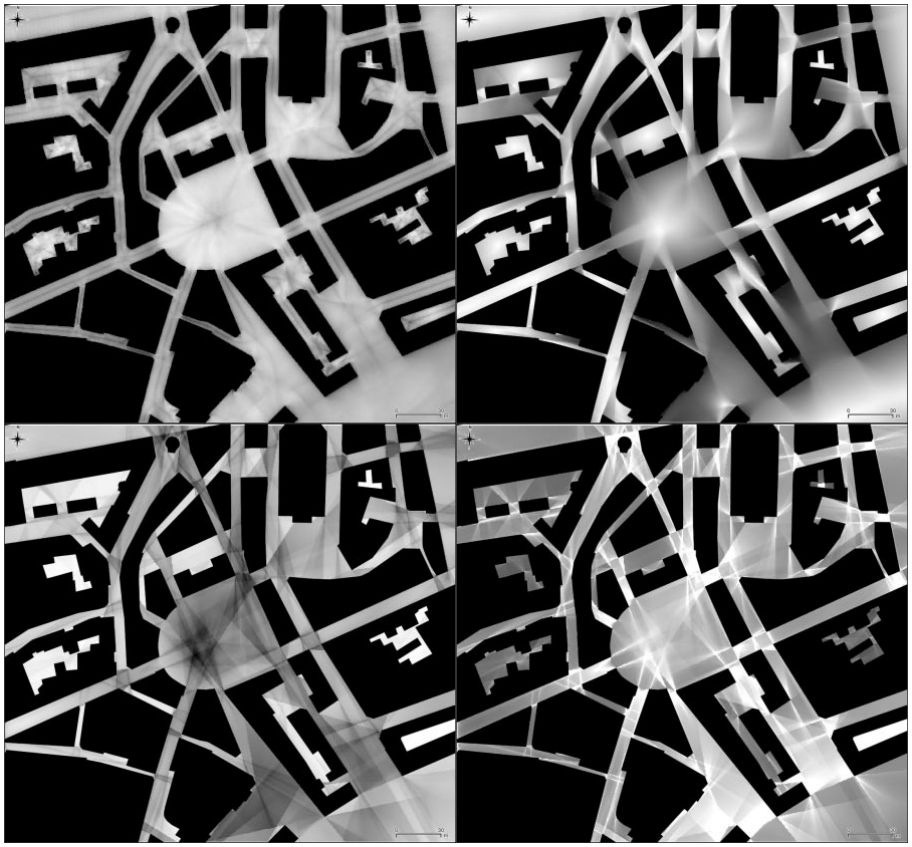


Figure 5. De haut en bas et de gauche à droite, pour le secteur de la place Royale, représentation de l'entropie, du drift, de l'occlusivité et de l'anticipation.

4. Discussion

La qualification du bassin visuel observable depuis un point est susceptible d'influer sur le processus de conception d'un projet urbain ou architectural, de par l'évaluation de l'impact visuel d'un aménagement, la diversité des vues offertes, etc.

Les indicateurs présentés ici ont été choisis en raison d'une intuition initiale de cohérence entre la prévisibilité, le poids des masses visuelles et la part d'incertitude inhérente au contour immédiatement environnant d'une part, et le déplacement d'un piéton en situation de dérive (au sens de la théorie éponyme du situationniste (Debord, 1956)) d'autre part. Il s'agit d'une hypothèse forte mais également réductrice, un modèle de déplacement piéton en site urbain ne pouvant se suffire de la seule forme de l'environnement immédiat. Une première observation des résultats fait ressortir quelques saillances. L'entropie, dessine une forme de squelette topologique de l'espace ouvert. Les points du squelette, plus foncés que ceux qui les environnent, sont ceux pour lesquels l'entropie est minimale (la prévisibilité du contour y est alors maximale). Ces points, sortes de promontoires au sens de la visibilité, sont, localement, ceux qui exposent le plus l'observateur qui s'y trouve mais ce sont aussi ceux qui lui permettent de mieux « anticiper » les éventuelles intrusions.

Les zones matérialisées par le *drift* sont moins précises. Elles révèlent des centralités ou, par le biais de lignes de fuite, des corridors de visibilité. Elles semblent aussi marquer des zones de stabilité de l'environnement visuel immédiat, zones qui seront affinées par les indicateurs d'occlusion (occlusivité et anticipation).

L'occlusivité, mais plus encore l'anticipation, dessinent des lieux de fortes transitions. C'est-à-dire des lieux aux approches desquels la promesse d'un « bouleversement » de l'environnement visuel immédiat croît substantiellement. Ainsi, toute rue reliant la place Royale comprend des points chauds à son extrémité (l'anticipation y est maximale). Le sujet-piéton est alors « poussé par la promesse » d'une découverte d'informations visuelles qui lui sont cachées. *A contrario*, les zones de minimum de potentiel d'anticipation, plus sombres, correspondent aux endroits où l'environnement immédiat est plus lisible et mieux maîtrisé en termes de perception des formes (la part de « surprise » possible y est moindre). Les tendances de l'indicateur anticipation révèlent une certaine pertinence quant à l'identification à la propension au mouvement. Une analyse croisée avec les cartes de *drift* pourrait préciser l'intentionnalité en fixant une direction.

Les cartes tracées par l'occlusivité et l'anticipation mettent en évidence une structuration très géométrique de l'espace. Elles renvoient aux travaux liés au partitionnement convexe de l'espace ouvert (Peponis *et al.*, 1997) dans lesquels sont définies les notions de *s-line* (extension des angles réflexes) et *e-line* (extension des diagonales). Le e-partitionnement permet de diviser l'espace en zones homogènes, zones pour lesquelles le champ visuel d'un piéton reste inchangé (au sens où aucune nouvelle opacité n'apparaît ou ne disparaît de son champ de vision). Le tracé de l'anticipation ne fait cependant pas apparaître toutes les *e-lines* et *s-lines* et il souligne davantage certaines d'entre elles. On pose ici l'hypothèse d'une hiérarchie

entre ces différentes lignes. Certaines transitions seraient plus signifiantes dans l'analyse d'un espace et de sa structuration. Il existerait alors une corrélation entre l'anticipation et la formation d'une s-partition.

5. Conclusion

L'étude présentée dans cet article s'est fixé pour but d'analyser de manière systématique et reproductible la part d'inconnu, d'incertitude, de l'environnement urbain immédiat d'un piéton. Elle a consisté en la mise en œuvre, dans le contexte d'un SIG, d'un modèle d'analyse (en tout point de l'espace ouvert) du *visualscape*, analyse utilisant des jeux de données vectorielles standards. Ce modèle repose plus précisément sur l'évaluation d'un champ d'isovists partiels (la portée du champ de vision étant limitée) à l'aide de quatre indicateurs liés à la théorie de l'information (entropie de Shannon), au poids des masses visuelles (*drift*) ou à la part d'occlusions dans le bassin de visibilité (occlusivité et anticipation).

Les résultats obtenus présentent diverses structurations de l'espace. La squelettisation, révélée par l'entropie, met en évidence des « promontoires » visuels. Le *drift* dessine des corridors visuels et des zones de stabilité au sens de la visibilité. L'anticipation enfin trace les contours de zones où la promesse d'un réel changement de l'environnement visuel peut orienter la mobilité d'un piéton en situation de dérive. Ce dernier découpage de l'espace, associé à l'analyse des occlusions et de leurs dispositions, révèle par ailleurs des similitudes aux techniques de partitionnement convexe de l'espace ouvert de (Peponis *et al.*, 1997).

Les combinaisons de ces indicateurs diffèrent selon les enjeux (sécurité visuelle, déplacement, lisibilité d'une séquence urbaine, surprise, etc.) et permettent une analyse plus approfondie de l'ambiance d'un espace urbain. La formulation d'un indicateur global de « qualité visuelle » reste cependant difficilement envisageable en pratique, du fait de ces enjeux contradictoires et du caractère subjectif de certains paramètres.

Cette analyse comportementale contextualisée d'un fragment urbain pourrait utilement être validée par un dispositif expérimental *in situ* (processus d'enquête et préférences déclarées ou préférences révélées par les trajectoires (Foltête & Piombini, 2010)) en prenant garde à la complexité et à la multiplicité des motivations intrinsèques au déplacement. Cette validation nuancerait, à titre annexe, les expérimentations décrites dans (Stamps III, 2010) établissant une corrélation entre le nombre de *s-lines* et le temps de parcours d'un espace. Nous formulons effectivement l'hypothèse qu'au-delà de l'importance du nombre de *s-lines*, une forte valeur d'anticipation au point d'observation de la scène figurerait une réduction de ce temps de parcours.

Bibliographie

Adolphe, L. (2001). A simplified model of urban morphology: application to an analysis of the environmental performance of cities. *Environment and Planning B: Planning and Design*, *28*(2), 183–200. doi:10.1068/b2631

Batty, M., & Jiang, B. (1999). *Multi-agent simulation: new approaches to exploring space-time dynamics in GIS*. London (UK).

Benedikt, M. L. (1979). To take hold of space: isovists and isovist fields. *Environment and Planning B: Planning and Design*, *6*(1), 47–65.

Brayer, L. (2013). Filmer l'ambiance urbaine: Les dispositifs vidéographiques à l'œuvre chez William H. Whyte dans La vie sociale des petits espaces urbains. *Ambiances [En ligne], Représentation - Traduction - Ecriture*. http://ambiances.revues.org/335

Chauvat, G. (2012). *Cartographier le mystère: caractérisation d'un indicateur figurant la propension au mouvement d'un individu en situation de dérive* (p. 72). Nantes, France.

Conroy Dalton, R., & Dalton, N. (2001). OmniVista: an application for Isovist field and path analysis. *3rd International Space Syntax Symposium*. Atlanta, Georgia, USA.

Cullen, G. (1961). *Townscape* (p. 315). New York: Reinhold Pub. Corp.

Debord, G.-E. (1956). Théorie de la dérive. *Internationale Situationniste*, *2*, 19–23.

Edussuriya, P., Chan, A., & Ye, A. (2011). Urban morphology and air quality in dense residential environments in Hong Kong. Part I: District-level analysis. *Atmospheric Environment*, *45*(27), 4789–4803. doi:10.1016/j.atmosenv.2009.07.061

Foltête, J.-C., & Piombini, A. (2010). Deviations in pedestrian itineraries in urban areas: a method to assess the role of environmental factors. *Environment and Planning B: Planning and Design*, *37*(4), 723–739. doi:10.1068/b35015

Hillier, B., Penn, A., Hanson, J., Grajewski, T., & Xu, J. (1993). Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B: Planning and Design*, 20(1), 29–66. doi:10.1068/b200029

Joliveau, T. (2004). *Géomatique et gestion environnementale du territoire. Recherches sur un usage géographique des SIG*. Université de Rouen.

Kaplan, R. (1973). Predictors of environmental preference: designers and clients. In W. F. E. Preiser (Ed.), *EDRA 4: Fourth International EDRA Conference - The Environmental Design Research Association* (Vol. 1, pp. 265–274).

Leduc, T., & Kontovourkis, O. (2012). Towards a mixed approach combining visibility and mobility studies to analyze the eleftheria square, Nicosia (CY). In R. Billen, M. Caglioni, O. Marina, G. Rabino, & R. S. José (Eds.), *3D Issues in Urban and Environmental Systems* (pp. 67–77). Bologna, Italy: Societa' Editrice Esculapio.

Llobera, M. (2003). Extending GIS-based visual analysis: the concept of visualscapes. *International Journal of Geographical Information Science*, *17*(1), 25–48.

Morin, M. (1995). *Lecture de la tour Bretagne: développement d'un outil de lecture de la ville* (p. 69). Nantes, France.

Morello, E., & Ratti, C. (2009). A digital image of the city: 3D isovists in Lynch's urban analysis. *Environment and Planning B: Planning and Design*, *36*(5), 837–853.

Péneau, J.-P. (2000). Les ambiances urbaines. In M.-F. Mattei & D. Pumain (Eds.), *Données Urbaines* (pp. 375–386). Paris, France: Anthropos.

Peponis, J., Wineman, J., Rashid, M., Hong Kim, S., & Bafna, S. (1997). On the description of shape and spatial configuration inside buildings: convex partitions and their local properties. *Environment and Planning B: Planning and Design*, *24*(5), 761–781.

Rana, S. (2006). Isovist Analyst - An Arcview extension for planning visual surveillance. *ESRI International User Conference. (on CD-ROM)* (Vol. 1, p. 9). Redlands, USA.

Salat, S. (2011). *Les villes et les formes: Sur l'urbanisme durable* (p. 544). Hermann.

Stamps III, A. E. (2005). Isovists, enclosure, and permeability theory. *Environment and Planning B: Planning and Design*, *32*(5), 735–762.

Stamps III, A. E. (2010). Slines, entropy, and environmental exploration. *Environment and Planning B: Planning and Design*, *37*(4), 704–722. doi:10.1068/b35117

Suleiman, W., Joliveau, T., & Favier, E. (2012). Une nouvelle méthode de calcul d'isovist en 2 et 3 dimensions. In R. Billen, M. Binard, P. Hallot, & J.-P. Donnay (Eds.), *Actes de la Conférence internationale de Géomatique et Analyse Spatiale - SAGEO 2012* (pp. 366–386). doi:2268/133126

Turner, A., Doxa, M., O'Sullivan, D., & Penn, A. (2001). From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design*, *28*(1), 103–121.

Reconstruction 3D de bâtiments à partir de cartes postales anciennes

Vers un SIG 4D collaboratif

YOUNES Lara, ROMANIUK Barbara, DESJARDIN Eric, BITTAR Eric

CReSTIC-SIC, Université Champagne-Ardenne
IUT de Reims Châlons Charleville
Rue des Crayères, BP 1035, 51687 Reims Cedex 2, France
{Lara.Younes, Barbara.Romaniuk, Eric.Desjardin, Eric.Bittar}@univ-reims.fr

RESUME. De nombreuses approches de reconstruction 3D des milieux urbains exploitent aujourd'hui la richesse qu'offrent les Systèmes d'Information Géographique (SIG). Ces systèmes permettent de rassembler des informations de natures variées et présentent une structure dynamique pouvant être facilement enrichie. C'est dans ce contexte que nous abordons la 4D : l'intégration du temps dans le problème de reconnaissance, d'extraction et de reconstruction de bâtiments géolocalisés dans la ville de Reims à partir de données anciennes constituées de cartes postales et de plans cadastraux. Nous souhaitons proposer un outil permettant aux citoyens de s'approprier l'histoire de leur ville à travers une navigation spatio-temporelle dans la ville de Reims nous orientant ainsi vers un SIG 4D collaboratif.

ABSTRACT. Nowadays many approaches of 3D urban reconstruction exploit the possibilities offered by the Geographic Information Systems (GIS). These systems collect various types of information and have a dynamic structure that can be easily enhanced. We address in this paper the problem of recognition, extraction and 3D reconstruction of geolocated buildings in Reims city (France) from old data composed of postcards and cadastral plans. We aim to offer a tool allowing citizens to appropriate the history of their city by providing a spatiotemporal navigation in the city of Reims and steering us to a 4D collaborative GIS.

MOTS-CLES : Reconstruction 3D, Reconstruction spatio-temporelle, Evolution, SIG collaboratif, Reims, Cartes postales anciennes, Cadastres, Géolocalisation, Descripteur, SIFT, Extraction de Caractéristiques, Structure from Motion.

KEYWORDS: 3D Reconstruction, Spatio-temporal Reconstruction, Evolution, Collaborative GIS, Reims (France), Old postcards, Cadastral, Geolocation, Descriptor, Feature Extraction, Structure from Motion.

1. Introduction

Les Systèmes d'Information Géographique (SIG) sont au cœur de nombreux projets permettant de rassembler des informations sur l'espace urbain de natures variées. De nos jours, ils permettent d'enrichir et faire évoluer de manière efficace les modèles qui mènent à une reconstruction 3D géolocalisée de monuments, bâtiments ou encore de villes entières. De nombreuses applications ont recours à ces reconstructions. Nous pouvons citer pour exemple la planification urbaine, l'archéologie, le tourisme virtuel ou encore la restauration du patrimoine.

Dans notre projet nous abordons le problème 4D de reconnaissance, d'extraction et de reconstruction 3D de l'évolution au cours du temps de bâtiments géolocalisés dans la ville de Reims à partir de données anciennes. Nous souhaitons développer un outil permettant aux citoyens de s'approprier l'histoire de leur ville en leur proposant une navigation spatio-temporelle dans la ville de Reims. Il s'agit d'intégrer dans un navigateur web un survol interactif de la ville virtuelle, accompagné d'un curseur temporel et de modes de visualisation adaptés. Reims ayant tenu un rôle central dans l'histoire de la France, de nombreux documents de natures différentes attestent de l'évolution de la ville. Cette évolution a été fortement impactée par les deux derniers conflits mondiaux, à l'origine de nombreuses destructions urbaines dans la ville de Reims. Détruite à plus de 60% lors de la première guerre, elle fut même appelée la "Ville martyre" et devint alors un symbole pour la France entière. Dans le contexte historique du centenaire de première guerre mondiale les données autour desquelles s'articule notre projet de reconstruction sont principalement constituées de cartes postales anciennes couvrant la période de la fin du dix-neuvième au milieu du vingtième siècle ainsi que des plans cadastraux couvrant cette même période. Les cartes postales témoignent de l'évolution de l'espace urbain en termes de destruction, restauration et reconstruction. Ces images non géo-localisées de faible résolution contiennent par ailleurs du texte, des timbres et des tampons. Les plans cadastraux napoléoniens et actuels reflètent la morphologie de l'organisation urbaine d'avant la première guerre mondiale et d'après la période de reconstruction de la ville dans les années 50. Reconstruire Reims dans le contexte d'une évolution temporelle à partir de ces données s'apparente donc à un vrai défi, et l'originalité de notre approche réside dans l'utilisation de ces données éparses et imparfaites, que nous souhaitons étoffer par le biais d'un SIG 4D collaboratif où chacun pourrait apporter ses propres sources iconographiques et les intégrer aisément dans le système d'information historique à dimensions spatiales et temporelles.

2. Approches existantes

La reconstruction 3D à partir d'importantes bases de données d'images actuelles prises au sol a été abordée dans de nombreux travaux récents. La reconstruction 3D de Rome (Italie) à partir d'une base d'images redondante et dense constituée de photographies issues du site de partage *Flickr* a été proposée par Agarwal *et al.*, (2010). Pour appréhender le problème de reconstruction à partir d'images non

structurées, non calibrées prises à différentes échelles et sous différents angles, et dont les coordonnées 3D ne sont pas connues, les auteurs ont utilisé une approche automatique classique dans le domaine de la vision par ordinateur appelée *Structure from Motion* (Robertson and Cipolla, 2010), qui permet une reconstruction tridimensionnelle dans le cas d'images très similaires.

(Bioret et al., 2009) mettent en correspondance des données vectorielles d'un SIG 2D avec des photographies pour la géolocalisation des bâtiments dans un système urbain. Les contours des façades sont extraits dans les images et des hypothèses permettent l'estimation d'une géométrie partielle du bâtiment (angles entre les façades et rapport de longueurs). La géométrie estimée sert à interroger le SIG à la recherche des ensembles de façades qui correspondent le plus à la situation ainsi définie et par la suite à la définition de la pose de la prise de vue. Cette approche nécessite des connaissances à priori sur la géométrie du bâtiment. Des limitations de la méthode peuvent être l'ambiguïté pour le choix de la meilleure pose parmi les poses estimées, ainsi que la difficulté de l'extraction des contours des façades. La présence de deux ou plusieurs façades au complet dans l'image pour l'estimation de leur corrélation en termes d'angle et de rapport de longueur étant nécessaire, cela implique une limitation de cette méthode dans le cas complexe où une seule façade est difficilement identifiable dans l'image comme le cas de la figure 1.

Figure 1. Dans cette vue, les limites des façades sont difficilement identifiables.

Le calcul de la pose de la prise de vue est effectué dans (Zhu et al., 2012) par une mise en correspondance de courbes. Pour une image, une courbe qui sépare le ciel des bâtiments est identifiée et est comparée à une courbe d'une vue panoramique d'un extrait 3D de bâtiments dans un SIG. Cette approche nécessite un SIG complet de tous les modèles 3D des bâtiments, dont nous ne disposons pas dans notre cas d'étude.

A partir d'un SIG 3D et d'une nouvelle photographie, Suleiman *et al.*, (2012) ont proposé une méthodologie pour reconnaître les façades des bâtiments photographiés et localiser la prise de vue. Ils utilisent des images satellitaires provenant de *Google Maps* et la sélection de points de repère sur le terrain, conjointement au SIG pour évaluer la qualité de la géolocalisation qu'ils calculent. Ils ont ainsi mesuré que la façade doit être photographiée quasiment de face, à une distance de moins de 100 mètres, pour obtenir une erreur de positionnement de moins de 3 mètres et une erreur d'orientations de moins de 15°.

Ne pouvant pas imposer de telles contraintes aux prises de vue de cartes postales existantes, nous avons choisi d'intégrer à notre approche la sélection interactive de points de repère sur les images et dans le SIG. Nous présentons notre méthodologie dans la section suivante.

3. Méthodologie

Nous décrivons comment, pour exploiter un ensemble de cartes postales anciennes, nous avons conçu un processus de création d'une collection de bâtiments 4D qui a une dimension collaborative et une visée universelle. Notre méthode concerne la création et l'enrichissement de cette collection.

Chaque image est représentée par une signature identifiant son contenu, établie à partir d'un ensemble de points caractéristiques et de descripteurs locaux associés. Suleiman *et al.*, (2012) indiquent utiliser une imagette composée d'une partie (typiquement 1%) de l'image comme signature. Leur approche a été de restreindre l'identification aux textures des façades, alors que nous calculons les points caractéristiques sur l'ensemble de l'image. Cette signature sert à identifier le bâtiment apparaissant dans une nouvelle carte postale en mettant en correspondance le contenu de la carte postale avec les images géoréférencées du même bâtiment dans la base de données. Ceci permet de calculer la position de la nouvelle prise de vue et par la suite de géoréférencer les points caractéristiques de la carte. Le traitement suivant consiste à exploiter les informations ainsi obtenues en complémentarité avec de données multimodales pour enrichir la modélisation spatio-temporelle de la ville.

Nous identifions deux types de données dans notre système d'information : des données de référence et des données qui servent à l'enrichissement du système. Nous effectuons un traitement au moment de la création du système pour géolocaliser des cartes postales de référence. Cet ensemble est utilisé par la suite pour l'enrichissement du système lors de l'introduction d'une nouvelle carte postale. Nous détaillons les étapes de notre approche qui sont illustrées dans la Figure 2.

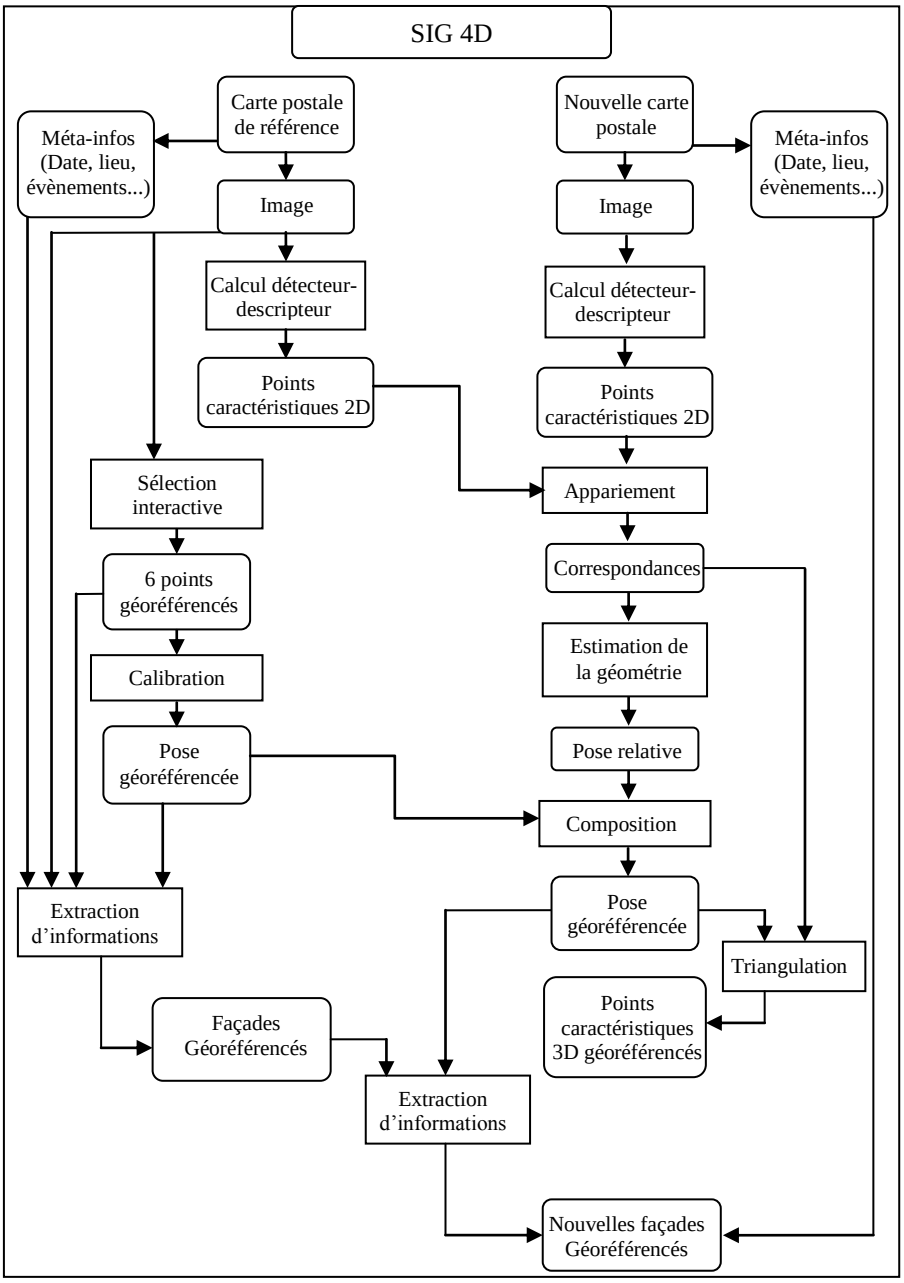


Figure 2. Le schéma de la méthodologie suivie

6

3.1. Géolocalisation des données

Nous établissons d'abord un lien entre des vues de référence et le SIG.

3.1.1.Géoréférencement des images

Cette étape consiste à géolocaliser les données contenues dans les images. Nous disposons pour ce faire d'une base de données urbaine géoréférencée qui indique la dimension et la hauteur des bâtiments, et de plans cadastraux de différentes époques.

Nous nous appuyons ainsi sur la connaissance des dimensions réelles de certains bâtiments de l'image afin de calculer pour chaque image le placement de l'appareil d'acquisition par rapport à l'objet. Il s'agit de retrouver la matrice de projection utilisée pour la projection d'un point 3D de l'espace sur le plan de l'image obtenue lors de l'acquisition. En pratique, nous sélectionnons dans chaque image 6 points qui représentent la projection de points 3D identifiés dans la base de données urbaine (Figure 3).

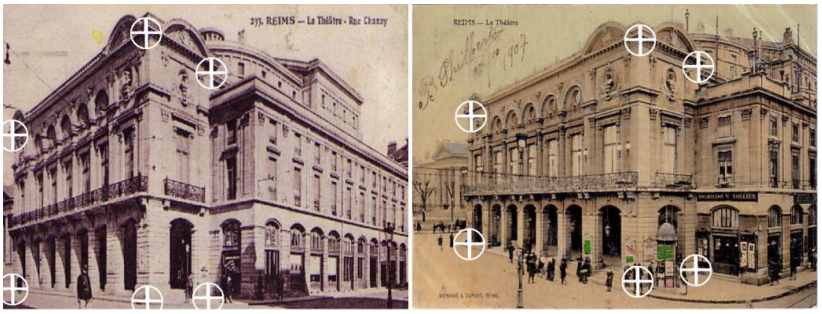


Figure 3. Sélection de 6 points géoréférencés pour la calibration de chaque vue du théâtre

Nous estimons les paramètres intrinsèques de l'appareil photo qui est à l'origine de la prise de vue. Dans le modèle du sténopé appelé aussi *pin-hole*, il s'agit de la distance focale f, des facteurs d'agrandissement de l'image k_u et k_v, des coordonnées de la projection du centre optique de la caméra sur le plan image c_u et c_v, et de la prise compte éventuelle la non-orthogonalité des lignes et des colonnes.

Il reste à calculer les paramètres extrinsèques, qui représentent le placement (translation et rotation) de l'appareil de prise de vue par rapport à la scène. Pour le faire, nous utilisons la méthode de Zhang (2000). Une transformation linéaire directe permet de calculer la matrice de projection de la caméra à partir du couplage 2D/3D des points par la technique des moindres carrés. Les paramètres extrinsèques ainsi obtenus sont raffinés par la méthode itérative de Levenberg-Marquardt d'optimisation non-linéaire.

Chaque image peut alors être mise en correspondance avec le modèle existant dans la base de données, comme illustré dans la Figure 4 où l'on retrouve l'identification des façades du théâtre dans les deux cartes postales.

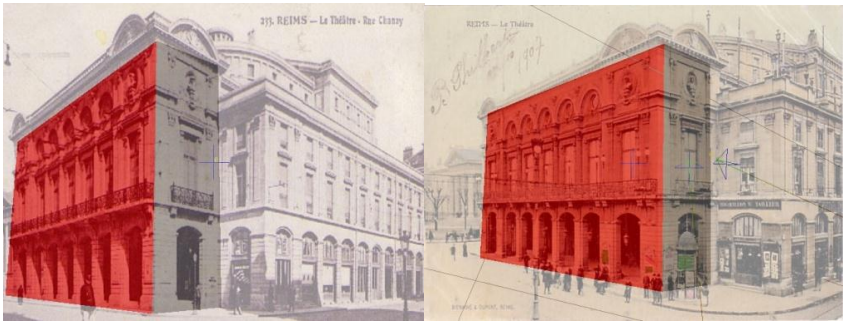


Figure 4. Correspondance entre les images et le modèle 3D

3.1.2. Détection des points caractéristiques et calcul de leurs descripteurs

Les points caractéristiques sont des points d'intérêts pouvant être identifiés dans différentes vues d'un même bâtiment, comme les coins des fenêtres, par exemple. La problématique est de caractériser ces points d'intérêt d'une manière suffisamment robuste pour être identifiés à différentes échelles d'images (avec différents zooms et résolutions du capteur), ainsi qu'avec des cadrages, des angles d'observation et des paramètres d'exposition (variations de luminosité) différents. Pour cela, une signature du point caractéristique appelée descripteur est calculée à partir d'un environnement local du point dans l'image, appelé fenêtre du descripteur.

En nous inspirant des travaux de Mikolajczyk et Schmid, (2005), nous avons réalisé une étude d'évaluation (Younes *et al.*, 2013) de couples de détecteurs-descripteurs qui assurent une invariance à la rotation, l'échelle, les transformations affines ainsi que les changements d'illumination. Nous avons imposé des transformations affines de rotation, changement d'échelle ainsi que d'occlusion à un ensemble d'images dans le cadre de notre base de données qui nous a permis de calculer la précision en terme de pourcentage de bonnes mise en correspondances par rapport aux transformations imposées. Dans notre base de données nous disposons des cartes postales qui illustrent généralement les sites marquants de la ville. Il peut s'agir de bâtiments historiques qui occupent la globalité du plan de la photographie, comme il peut s'agir de façades qui apparaissent des deux côtés d'une rue. Pour les deux catégories de contenu d'image, la combinaison détecteur-descripteur SIFT de Lowe (2004) a donné les meilleurs résultats.

Pour évaluer la qualité de la mise en correspondance, nous avons estimé manuellement la transformation épipolaire de référence pour un ensemble de couples d'images. Pour un point dans une image nous avons calculé la distance de son point correspondant dans l'autre image à la ligne épipolaire qui lui est associée

par la transformation épipolaire estimée. Le seuillage de ces distances nous a permis de définir le taux de bonnes correspondances parmi celles détectées automatiquement. La combinaison détecteur-descripteur SIFT de Lowe (2004) a également donné les meilleurs résultats.

Chaque vue de référence géolocalisée possédant ses points caractéristiques, le calcul des points et des descripteurs est également appliqué à chaque nouvelle image, pour établir des correspondances entre paires d'images.

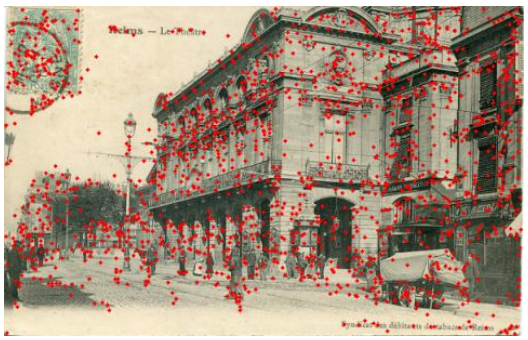


Figure 5. *Les points extraits par les détecteurs SIFT (gauche) sur une carte postale du Grand Théâtre de Reims.*

3.2. Reconstruction 3D par triangulation

3.2.1. Sélection de deux images

La base du processus est l'emploi de deux photographies contenant un même bâtiment, prises selon des points de vue différents et des époques différentes. Picard *et al.*, (2009) ont étudié le coût important que représente la recherche des images correspondant à une image donnée. Le Bris et Paparoditis, (2010) ont choisi l'ortho-rectification des images dans un plan parallèle aux façades pour gagner en efficacité. Nous proposons d'accélérer la recherche par l'exploitation d'éléments sémantiques d'indication du quartier ou du bâtiment photographié, lorsqu'ils sont disponibles. En effet ces éléments sont généralement présents lorsqu'il s'agit de cartes postales.

3.2.2. Appariement des points

Une mesure de similarité entre les vecteurs descripteurs des points caractéristiques permet de chercher pour chaque point caractéristique d'une image I1 s'il existe un correspondant dans une image I2. La méthode utilisée (et illustrée dans Figure 4) est celle du rapport des deux plus proches voisins (Lowe, 2004).

La détection automatique des correspondances produit des données aberrantes (*outliers*) qui doivent être éliminées par filtrage. L'algorithme ORSA (Moisan and

Stival, 2004), variant de l'algorithme RANSAC (Fischler and Bolles, 1981) permet d'identifier les bonnes correspondances (*inliers*) (Figure 2). Le principe consiste à échantillonner aléatoirement des sous-ensembles de correspondances pour étudier leur cohérence, en modélisant une transformation épipolaire entre les deux images. Cette méthode utilise le critère de discrimination a contrario qui permet l'adaptation du seuil de décision le long du processus itératif de filtrage. Ceci assure qu'aucun sous ensemble d*'inliers* n'est retenu si son degré de confiance n'est pas suffisamment significatif. L'algorithme ORSA permet d'apporter un bon filtrage des correspondances même quand le taux des *inliers* n'est que de 10%.

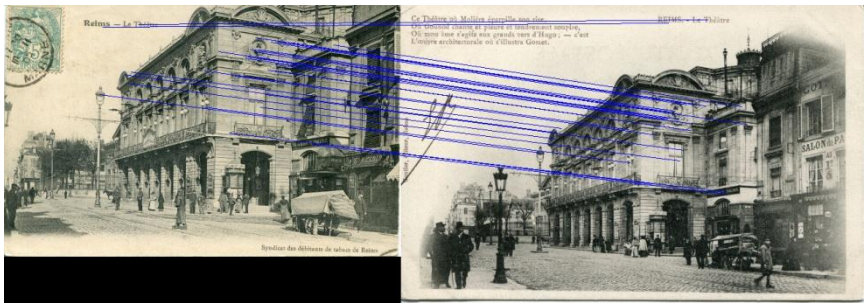


Figure 6. Exemple d'appariements obtenus par la méthode SIFT
Le texte descriptif des cartes postales est source de correspondances aberrantes.

3.2.3. Calcul de la position 3D relative des points caractéristiques

Comme le rappellent (Robertson and Cipolla, 2010), lorsque la correspondance est connue entre au moins sept points distincts dans les deux vues, la géométrie épipolaire permet de calculer la position et l'orientation relative des prises de vue, ainsi que les positions 3D de ces points, à un facteur d'échelle près. Cette méthode permet ainsi de calculer la position 3D relative de tous les points caractéristiques dont l'appariement est validé.

A la fin de cette étape, les éléments reconstruits en 3D ne peuvent pas être intégrés dans un SIG, car leur position n'est pas connue dans un référentiel absolu.

3.2.4. Géolocalisation des points caractéristiques

Reprenons le processus global de notre approche. En ayant géo-référencé une image I1, nous connaissons la transformation M1 qui définit le positionnement et l'orientation de la prise de vue (cf.Figure 7). Lors de l'étape 3.2.3 de reconstruction 3D par triangulation, nous avons calculé la matrice fondamentale M12 entre deux images I1 et I2, et les coordonnées 3D des points caractéristiques communs et appariés dans un repère relatif 3D.

Nous pouvons maintenant, en combinant ces deux sources d'information, établir le géoréférencement des points caractéristiques.

10

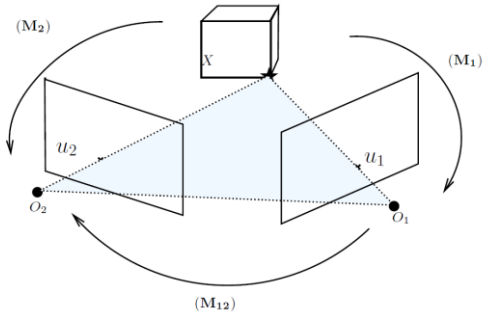


Figure 7. Modélisation géométrique des prises de vue.

3.3. Enrichissement du SIG

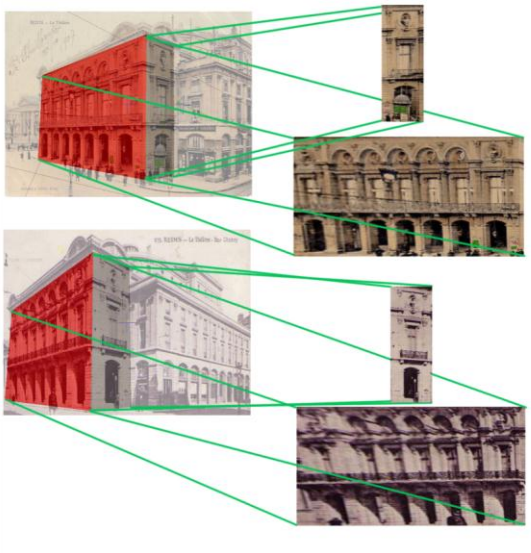


Figure 8. Obtention de nouvelles informations

De nouvelles informations contenues dans les images peuvent alors être extraites pour venir compléter le SIG. Il peut s'agir de données correspondant aux vues des bâtiments, comme illustré dans la Figure 8.

Etant donné que les images contiennent des informations spatiales et temporelles qui leur sont propres, l'enrichissement du SIG produit des éléments qui permettent

une prise en compte de l'évolution de la ville au cours du temps. Cela permet la comparaison des bâtiments à différentes époques, à l'instar du travail du collectif Reims Avant[1], qui fusionne des photographies actuelles avec des vues anciennes. Il peut s'agir également de fusionner des données multi-modales, comme la représentation de la modélisation de la ville sur un plan cadastrale ancien (Figure 9).

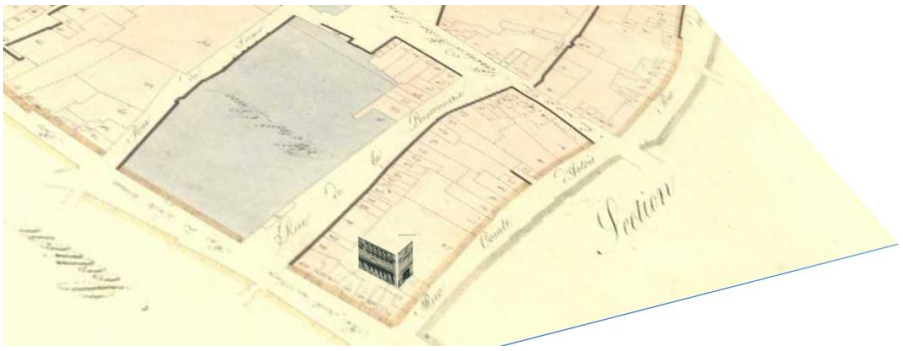


Figure 9. Le bâtiment 3D positionné sur un cadastre ancien.

4. Conclusion

Nous avons proposé une méthodologie pour reconstruire en 3D des bâtiments photographiés dans des cartes postales anciennes, matériau qui n'est pas disponible avec la même profusion que les photographies touristiques numériques actuelles. Notre approche exploite les informations disponibles dans les bases de données urbaines pour enrichir ces bases, et préparer la création d'un SIG 4D, qui intègre des informations spatiales datées.

La visualisation des éléments de ce SIG 4D ouvre des perspectives nouvelles de combinaison des éléments d'époques différentes pour une meilleure compréhension de l'évolution de la ville. L'évolution temporelle sera accessible à travers un curseur positionné sur une ligne de temps et les bâtiments situés à des époques proches du temps sélectionné seront visibles selon des modes de représentation adaptés (ex : variation de la transparence en fonction de l'éloignement temporel).

Notre démarche a pour objectif d'être ouverte et collaborative, afin de proposer à travers un site internet à toute personne intéressée l'intégration de nouvelles photographies anciennes.

[1] http://www.reimsavant.com

5. Bibliographie

Agarwal, S., Furukawa, Y., Snavely, N., Curless, B., Seitz, S.M., and Szeliski, R. (2010). Reconstructing Rome. Computer *43*, 40–47.

Bioret, N., Moreau, G., and Servières, M. (2009). Géolocalisation en milieu urbain par appariement entre une collection d'images et un SIG 2D. Ingénierie Systèmes Inf. *14*, 107–131.

Le Bris, A., and Paparoditis, N. (2010). Matching terrestrial images captured by a nomad system to images of a reference database for pose estimation purpose. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. Iaprs *vol. 38*, 133–138.

Fischler, M.A., and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun Acm *24*, 381–395.

Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vis. *60*, 91–110.

Mikolajczyk, K., and Schmid, C. (2005). A Performance Evaluation of Local Descriptors. Ieee Trans Pattern Anal Mach Intell *27*, 1615–1630.

Moisan, L., and Stival, B. (2004). A Probabilistic Criterion to Detect Rigid Point Matches Between Two Images and Estimate the Fundamental Matrix. Int. J. Comput. Vis. *57*, 201–218.

Picard, D., Cord, M., and Valle, E. (2009). Study of Sift Descriptors for Image Matching Based Localization in Urban Street View Context. In ISPRS Workshop CMRT 2009. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, (Paris, France), pp. 193–198.

Robertson, D.P., and Cipolla, R. (2010). Structure from Motion. In Practical Image Processing And Computer Vision, (Wiley-Blackwell (an imprint of John Wiley & Sons Ltd)),.

Suleiman, W., Favier, E., and Joliveau, T. (2012). Reconnaissance de bâtiments et localisation de photographies au moyen d'un descripteur de texture. Rev. Int. Géomatique *22*, 439–459.

Younes, L., Romaniuk, B., and Bittar, E. (2013). Can feature points be used with low resolution disparate images? (Application to postcard data set for 4D city modeling). In Eighth International Conference on Computer Vision Theory and Applications (VISAPP), pp. 481–486.

Zhang, Z. (2000). A flexible new technique for camera calibration. Ieee Trans. Pattern Anal. Mach. Intell. *22*, 1330–1334.

Zhu, S., Pressigout, M., Servières, M., Morin, L., and Moreau, G. (2012). Skyline matching: A robust registration method between Video and GIS. T. Leduc, G. Moreau, and R. Billen, eds. (EDP Sciences), p. 03007.

SESSION 4 :

Mobilités

CTRD : une base de données spatio-temporelle pour l'analyse de la mobilité des chercheurs dans les firmes multinationales

Mahfoud Boudis[1], Virginie Jacquier-Roux[2], Claude Paraponaris[3]

1. Laboratoire d'Informatique de Grenoble, Université de Grenoble
681 *rue de la Passerelle, F*-38402 *Saint Martin d'Hères cedex*
Mahfoud.Boudis@imag.fr

2. Centre de Recherche en Economie de Grenoble, Université de Grenoble
1241 *rue des Résidence*, 38400 *Saint-Martin-d'Hères*
Virginie.Jacquier-Roux@upmf-grenoble.fr

3. Institut de Recherche en Gestion, Université Paris Est Marne-la-Valée
5 *Boulevard Descartes*, 77420 *Champs-sur-Marne*
Claude.Paraponaris@univ-mlv.fr

RESUME. Cet article présente une application de la géomatique dans le domaine de l'économie de la connaissance. Il aborde la thématique de l'analyse des dispositifs de partage des savoirs dans les entreprises et leur dimension Ressources Humaines. Nous en présentons tout d'abord le contexte et les enjeux avant de décrire la base de données CTRD, un outil spatio-temporel spécifique que nous avons développé pour l'analyse de la mobilité des chercheurs dans les entreprises en France. L'objectif est de montrer l'apport d'un tel outil pour comprendre les spécificités des stratégies des firmes multinationales en matière de mobilité du personnel de la R&D à des échelles territoriales nationale et régionale.

ABSTRACT. This paper presents an application of geomatics in the field of the knowledge economy. It addresses the subject of the analysis of mechanisms for sharing knowledge in firms and human resources dimension. We first present the context and issues before describing the data base CTRD, a specific spatio-temporal tool we have developed for the analysis of the mobility of researchers in companies in France. The aim is to show the contribution of such a tool for understanding the specific strategies of multinational firms in staff mobility of R & D at national and regional spatial scales.

MOTS-CLES : Base de données spatio-temporelle, firmes multinationales, partage des connaissances, réseaux d'innovation

KEYWORDS: Spatio-temporal Data Base, Multinational Corporations, Knowledge Sharing, Innovation Networks

1. Introduction

Les stratégies des firmes multinationales (FMN) mobilisent l'attention des sciences économiques et de gestion depuis longtemps. Une place croissante est notamment faite aux analyses de l'internationalisation de la Recherche & Développement (R&D) par les FMN. C'est en effet l'aspect le plus récent des stratégies de globalisation des FMN. Il a donné lieu à une abondante littérature, fondée sur des exigences élevées d'observation empirique et de mesure du phénomène.

Pour constater l'internationalisation de la R&D des firmes, des indicateurs simples suffisent : nombres de laboratoires de FMN implantés hors du pays d'origine, nombres de brevets inventés par des chercheurs dans les filiales à l'étranger, part de l'effort national de R&D ou des brevets inventés dans un territoire dus aux FMN étrangères implantées sur ce territoire, etc. (Sachwald, 2008).

Pour décrire les modalités de l'internationalisation de la R&D des firmes, il a fallu en revanche s'appuyer sur des enquêtes comparatives menées au sein des FMN, sur des échantillons plus ou moins fournis (Gassmann et Von Zedwitz, 1999 ; Zander, 2002). Cette méthode est rendue difficile du fait de l'accès restreint qu'offrent les entreprises aux investigateurs dès lors qu'il est question de R&D.

Lorsqu'il s'agit enfin de comprendre les motivations de fond de cette stratégie, un détour par les concepts et les analyses de l'économie de la connaissance est nécessaire. Notre recherche se situe alors en amont de la problématique de l'innovation dans les firmes. Nous voulons identifier les ressorts stratégiques et cognitifs de l'internationalisation de la R&D des entreprises par l'analyse des dispositifs Ressources Humaines (RH) de partage des savoirs dans les firmes multinationales. Pour en donner une illustration sur des bases empiriques, on doit alors disposer de faits observables, qui montrent les comportements spécifiques des FMN en matière de production, de circulation, d'accumulation de savoirs via la gestion de leurs RH dédiées à l'innovation. C'est l'objet de notre démarche de constitution d'une base de données utilisable dans ce but.

Cet article expose dans sa première partie les dispositifs de partage des savoirs dans les FMN via la dimension RH. La seconde partie est consacrée à la description de la base de données spatio-temporelle CTRD, les informations qu'elle fournit et les usages que l'on peut en faire pour l'analyse des connaissances tacites en R&D des firmes. La troisième partie illustre, par des exemples, l'utilisation de cette base de données pour l'analyse de la mobilité des chercheurs dans les firmes.

2. Les dispositifs de partage des savoirs dans les FMN et leur dimension RH

Depuis plusieurs décennies maintenant, les grandes firmes multinationales déploient des laboratoires de R&D hors de leur pays d'origine, dans une démarche de course à l'innovation au niveau mondial. Cette stratégie d'internationalisation de la R&D des firmes vise notamment la réduction des coûts et la conquête de nouveaux marchés mais elle contribue surtout au processus d'apprentissage global des grandes firmes, que ce soit pour exploiter des connaissances sur des nouveaux marchés, ou pour explorer des champs cognitifs mal représentés dans leur pays d'origine et au contraire maîtrisés à un niveau d'excellence dans d'autres pays (Kuemmerle, 1997 ; Patel et Vega, 1999). Ce faisant, ces firmes assument le risque de transfert de leurs savoirs (Criscuolo, 2002) ou de dispersion de leurs ressources (Lahiri, 2010) sur les territoires où elles se déploient, en échange de l'objectif poursuivi d'augmentation de leurs bases de connaissances. Le repérage de pôles d'excellence dans l'espace cognitif mondial suscite un objectif d'accès à des connaissances stratégiques qui sont aussi souvent des connaissances tacites co-produites au sein de réseaux sociaux implantés dans des territoires.

2.1. Du transfert de connaissances au partage des savoirs

C'est en partant d'une analyse épistémique des connaissances que l'on peut éclairer d'une lumière nouvelle la stratégie d'internationalisation de la R&D des firmes. La distinction entre connaissances tacites et connaissances codifiées (Polanyi, 1967) est au cœur de cette analyse. Les connaissances tacites sont considérées aujourd'hui comme une ressource centrale des processus d'innovation (Nonaka et Takeuchi, 1995). Elles se distinguent des connaissances codifiées notamment par une mobilité réduite, un processus de création souvent collectif, contraint par des relations fondées sur la proximité et le contact direct répété. Selon Tsoukas (2003) toutefois, les connaissances tacites ne s'approprient pas, ne se transfèrent pas et ne se convertissent pas. Elles se manifestent dans des contextes, sont coproduites par des acteurs qui s'y rencontrent et travaillent à améliorer les dispositifs susceptibles de les produire par de meilleures interactions sociales.

Au-delà des connaissances tacites elles-mêmes, les FMN visent dans leur démarche d'internationalisation de R&D les réseaux mis en place afin de les créer. Ces réseaux interpersonnels, inter-organisationnels, qui sont des réseaux sociaux, ne se déplacent pas, et c'est en leur sein que se co-produisent des connaissances tacites elles aussi « visqueuses » (Von Hippel, 1994).

Ainsi les FMN restent dépendantes de cette intransférabilité conjointe des processus de co-production de connaissances tacites et des connaissances tacites elles-mêmes. Leur localisation sur place leur permettra de participer à ces réseaux sociaux, mais à la condition d'adopter une attitude contributive, relevant d'une démarche de territorialisation (Allison et Scott, 1987). Elles doivent aussi renoncer à l'objectif de s'approprier les connaissances tacites qu'elles co-produiront. Par contre, cette expérience leur permettra d'accumuler des savoirs et des informations qui eux sont appropriables.

Toutefois, avant que cette appropriation soit menée à son terme, les FMN passent d'abord par une étape de partage de savoirs au sein de ces réseaux sociaux, qui s'appuie sur l'existence de dispositifs spécifiques.

2.2. Dispositifs de partage des savoirs et réseaux de production de connaissances

Le concept de dispositif de partage des savoirs est proposé ici comme représentation centrale des moyens de produire des connaissances tacites au sein d'un réseau. La notion de dispositif est développée à partir de celle de dispositifs cognitifs fournie par Poitou : *« ensembles organisés et finalisés d'objets intellectuels, articulés entre eux et distribués dans l'espace à des fins de production de biens ou de connaissances »* (Poitou, 1997). Ces dispositifs sont avant tout des dispositifs organisationnels qui rassemblent des outils de gestion articulés entre eux et distribués au sein de l'organisation dans la perspective de faciliter la création et la capitalisation des connaissances, et de partager les savoirs.

La littérature en économie et sciences de gestion a recensé ces dispositifs en mettant l'accent sur les moyens de mutualiser, classer, mettre à jour et interconnecter les savoirs (Jacquier-Roux *et al*., 2012). On y trouve : l'organisation des structures de R&D via des systèmes d'information, la codification des expériences dans des bases de données techniques, les communautés techniques pour la gestion du patrimoine technologique, l'évaluation et le management des compétences.

Toutefois, ces dispositifs ne sont valides pour produire de l'information et partager des savoirs qu'au sein de chaque réseau de co-production de connaissances tacites. Il reste à la FMN la tâche de connecter ces réseaux pour intégrer leur diversité, produisant ainsi un méta-dispositif de partage des savoirs. La question de l'articulation des dispositifs se pose alors.

2.3. La FMN comme articulation de réseaux et de dispositifs de partage des savoirs

Une FMN qui déploie sa R&D à l'étranger s'inscrit en fait dans deux contextes cognitifs : celui du réseau global « interne » (maison-mère et filiales) et celui du réseau social localisé des pays d'accueil. Dans ces contextes, deux dimensions majeures s'opposent : la diversité et l'intégration. Une grande diversité peut faire qu'une firme dispose d'unités R&D beaucoup plus proches d'unités concurrentes que d'unités internes (Colovic et Mayrhofer, 2008). La qualité des liens cognitifs entre unités de la FMN est alors déterminante pour contenir ces effets et assurer l'intégration. En pratique, les deux contextes cognitifs mentionnés ont été décomposés en quatre types de réseau de R&D (tableau 1). Chaque réseau comporte des dispositifs de partage des connaissances adaptés au type de connaissances qui y sont produits : droits de propriété intellectuelle, contrats, supports informationnels pour les connaissances codifiées ; proximité située, culture et pratiques collectives pour les connaissances tacites. La FMN organise sur cette base l'architecture de son

système de production de connaissances, attribuant au laboratoire central et aux laboratoires à l'étranger des rôles précis et potentiellement évolutifs (Zander, 2002).

Tableau 1. Diversité des réseaux de la R&D globalisée des firmes

Indicateur de distance / Indicateur d'internalité	Relations du laboratoire avec son territoire local d'implantation à l'étranger	Relations du laboratoire avec des acteurs à grande distance géographique
Relations du laboratoire avec les autres unités de la firme	**Réseau intra-firme local**	**Réseau intra-firme global**
Relations du laboratoire avec des acteurs extérieurs à la firme	**Réseau externe local**	**Réseau externe global**

(Source : Jacquier-Roux et Le Bas, 2008)

On observera que seuls trois de ces quatre réseaux se structurent à partir d'une proximité, ou géographique, ou organisationnelle. En effet, le réseau externe global s'abstrait de toute proximité de ce type. En réalité nous le mentionnons pour mémoire, mais il concerne principalement la circulation de connaissances codifiées (via les marchés de connaissances, de brevets, les co-dépôts de brevets, les alliances technologiques formalisées, etc.). Il ne saurait par conséquent entrer dans le cadre d'une analyse de la démarche d'intégration et de partage de savoirs propre à la co-production de connaissances tacites, qui demande justement une proximité.

Ainsi l'internationalisation de la R&D des FMN comporte des opportunités d'enrichissement technologique via la participation aux réseaux de co-production de connaissances tacites. L'existence de dispositifs de partage des savoirs constitue donc une nécessité non seulement au niveau de chacun de ces réseaux, mais aussi à l'intersection de ces réseaux, afin de tirer parti dans l'ensemble de la tension entre diversité et intégration.

2.4. La dimension RH des dispositifs de partage des savoirs dans les FMN : capital humain ou mobilité des chercheurs ?

Les recensements effectués des différents dispositifs de partage des savoirs et de leur contenu sont surtout axés sur des outils de gestion des connaissances. Notre analyse dans cet article propose d'élargir ces recensements en intégrant de manière plus centrale la ressource humaine sur laquelle ces outils s'appuient.

En effet les dispositifs de partage des savoirs ne peuvent ni se structurer ni se connecter par eux-mêmes. Le facteur humain est le pivot essentiel de ces tâches. Au sein de chaque réseau, le dispositif de partage des savoirs repose sur des comportements et routines humaines à établir implicitement ou explicitement. De même, au niveau global, des individus dans la FMN doivent évoluer dans plusieurs dispositifs parallèlement pour réaliser l'intégration. L'analyse des dispositifs comme

de leur agencement se déplace donc vers les Ressources Humaines, au cœur du partage des savoirs.

L'ensemble de l'organisation d'une FMN représente potentiellement l'espace général d'articulation des dispositifs de partage des savoirs. La Gestion des Ressources Humaines (GRH) de la FMN est alors associée de façon cruciale à cet enjeu. Elle s'opère au sein d'une tension stratégique entre, d'une part, la nécessité de stabilité des RH dédiées aux savoirs, afin de « former communauté » dans l'organisation, et, d'autre part, la nécessité pour ces RH de participer aux différents réseaux de connaissances dans une évolution temporelle et spatiale, afin de multiplier rencontres et expériences.

En d'autres termes, une question essentielle se pose à la FMN : la GRH concernant les chercheurs doit-elle privilégier la constitution d'un « capital humain », et donc la stabilité des équipes, ou la mobilité géographique et organisationnelle, et donc un renouvellement soutenu de ces équipes ?

La formulation de cette alternative nous amène à proposer de poursuivre l'analyse en observant les comportements des entreprises. Y a-t-il chez les FMN un choix clair en faveur de la constitution d'un capital humain via la stabilité, ou en faveur de la mobilité des chercheurs ? Peut-on constater des spécificités propres aux FMN dans ce domaine, lorsqu'on les compare aux autres entreprises ?

Une telle analyse doit de plus se conduire à un niveau différencié sur le plan spatial. En effet, étant donnée l'importance du caractère territorialisé de plusieurs des réseaux de connaissances des FMN, le choix entre stabilité et mobilité des RH dédiées à l'innovation est susceptible de dépendre fortement de la performance technologique et scientifique de la région d'accueil de la FMN. Notamment, l'abondance ou la rareté des chercheurs et ingénieurs dans cette région peut jouer un rôle déterminant.

Une telle démarche empirique ne peut être effectuée que si l'on dispose de données fiables sur la GRH des entreprises concernant leur personnel de R&D, dans une optique à la fois spatiale et temporelle. Ces données faisant défaut de manière générale, restait à les rassembler et à les organiser. C'est l'objectif que nous avons poursuivi en nous dotant de la base de données CTRD.

3. La base de données CTRD

Pour répondre aux questionnements formulés à propos des dispositifs de partage des savoirs dans les FMN via leur dimension RH nous avons constitué une base de données spatio-temporelle dénommée CTRD pour l'analyse des Connaissances Tacites en Recherche et Développement des entreprises implantées en France.

3.1. Architecture de la base de données

La structure de la base de données CTRD est conçue selon une architecture à trois dimensions : thématique, spatiale et temporelle.

La dimension thématique, présentée en détail dans la section 2, concerne les connaissances tacites de l'entreprise appréciées en termes de ressources humaines consacrées à la R&D (Recherche fondamentale ou appliquée, et Développement expérimental) dans le but d'accroître son capital de connaissances et son utilisation pour de nouvelles applications. Les ressources humaines prises en compte ici se limitent aux chercheurs dans le sens du personnel de la R&D (hors ouvriers, techniciens et administratifs). Pour chaque entreprise identifiée, on s'intéresse tout particulièrement aux données relatives aux effectifs en termes de stock et des flux d'entrée et de sortie.

- pour *le stock*, on retient, les effectifs des chercheurs par genre qui sont en plus détaillés : par leur nationalité, par classe d'âge, par qualification au sens du diplôme le plus élevé, et par discipline d'exercice dans l'entreprise ;
- le *flux d'entrée* comptabilise le nombre d'arrivées de chercheurs dans l'activité R&D de l'entreprise en provenance d'une autre activité que la R&D de la même entreprise ou d'une activité (dont la R&D) d'une entreprise (du même groupe ou hors groupe) située au niveau national (territoire local) ou international (territoire global) ;
- le *flux de sortie* compte les effectifs de chercheurs partants de l'activité de R&D de l'entreprise à destination d'une autre activité dans la même entreprise ou d'une activité (dont la R&D) d'une entreprise (du même groupe ou hors groupe) située à l'échelle locale ou globale ;

La dimension spatiale retenue dans la base CTRD est dictée à la fois par les contraintes liées aux données collectées et par les objectifs de leur utilisation en termes d'analyses. Les données collectées concernent la population des entreprises implantées en France métropolitaine qui peuvent être localisées à l'échelle communale sur ce territoire. Le maillage élémentaire retenu est la commune tout en gardant dans la base de données trois niveaux hiérarchiques du territoire français (communes, départements, régions) avec leur code officiel au sens de l'INSEE[1] et les données géographiques associées de l'IGN[2]. Cela permettra d'effectuer des analyses territoriales agrégées et multi-échelles. Par ailleurs, pour permettre l'étude de la mobilité des chercheurs au sein du réseau local ou global d'une firme, nous avons délimité les frontières de ces réseaux par la nationalité de la firme. Celle-ci est déterminée par le pays d'origine de l'actionnaire qui contrôle directement ou indirectement (via ses filiales) la majorité des actions de la firme.

La dimension temporelle des données exploitées varie selon leur source : les données sur les stocks et les flux de chercheurs dans les entreprises sont disponibles par période de deux ans depuis 1992 ; celles sur la localisation et la nationalité de ces entreprises peuvent être obtenues pour les dix dernières années. La dimension temporelle de la base de données CTRD résulte de l'intersection, au sens de la logique temporelle [Allen et Hayes, 1989], des intervalles temporels de ces deux sources. Elle couvre une période de dix ans à partir de 2002 avec une échelle de

1. http://insee.fr
2. http://data.ign.fr

temps annuelle à pas variable en fonction des données disponibles. Ainsi, le choix de cette dimension permettra (entre autre) d'effectuer des analyses sur l'évolution temporelle de la mobilité des chercheurs au sein des entreprises en France pendant une décennie.

De par sa structure et son contenu, la base de données spatio-temporelle CTRD constitue un outil de recherche qui doit permettre en effet la mise en évidence des spécificités de la GRH des FMN étrangères implantées en France concernant leurs chercheurs, en comparaison avec les entreprises sous contrôle français sur le même territoire.

3.2. Cas d'utilisation de la base de données

En s'intéressant à l'étude de la mobilité des chercheurs dans les entreprises en France, des analyses peuvent être effectuées en distinguant deux catégories de firmes selon leur nationalité : LECE (Laboratoire ou Entreprise sous Contrôle Etranger) et LECF[3] (Laboratoire ou Entreprise sous Contrôle Français). L'échantillon de notre étude est constitué de l'ensemble des entreprises employant du personnel de R&D et qui sont localisées en France métropolitaine qu'elles soient sous contrôle français ou étranger. Ainsi, dans cet échantillon, des entreprises qui n'ont aucune entrée et/ou sortie de chercheurs sont également prises en compte pour permettre d'analyser à la fois la mobilité et la stabilité du personnel de R&D. La base de données CTRD, instanciée avec un tel échantillon d'entreprises, permettra d'étudier finement le partage des connaissances dans ces entreprises à partir de leur personnel de R&D et de son évolution à travers l'espace et/ou temps. Elle peut être utilisée pour diverses analyses permettant d'étudier des hypothèses concernant les effectifs, les entrées et les sorties de chercheurs dans les entreprises.

Analyses des effectifs chercheurs : l'hypothèse générale à analyser est que les LECE constituent en France une équipe alimentée par le système éducatif local, et par le tissu de la recherche locale, privée et publique. Même si cette recherche locale s'enrichit de plus en plus de chercheurs étrangers, elle reste dominée par des chercheurs français. Les LECF ont plus tendance à chercher à panacher leur équipe pour l'ouvrir à des enrichissements étrangers. Ces hypothèses sont plus « probables » dans le cas de laboratoires installés de longue date, dans lesquels les effectifs retracent effectivement ce type de stratégie, que dans des laboratoires nouvellement installés, où les influences de la nationalité d'origine de la firme sont fortes (et donc les LECE seront plus marqués par des chercheurs de cette nationalité d'origine, et les LECF seront plus marqués par des chercheurs français). De même, si on suppose que les firmes multinationales étrangères viennent faire de la R&D en France pour profiter de l'excellence nationale ou régionale dans certains domaines spécifiques dominants, alors les LECE doivent avoir une plus forte proportion de chercheurs dans ces disciplines dominantes que les LECF (qui elles ont un spectre plus large). L'analyse de ces hypothèses concernant l'effectif de chercheurs passe

3. Une LECF peut être une firme multinationale ou pas

par la comparaison des LECE aux LECF concernant l'origine des chercheurs (français/étrangers), leur qualification et leur discipline d'exercice dans l'entreprise.

Analyse des entrées de chercheurs : cela peut se manifester par le fait que les LECE recrutent plus de jeunes diplômés, plus de chercheurs hors de la firme, plus de chercheurs locaux, et s'ils recrutent en interne ils privilégient aussi les circuits locaux. Dans l'ensemble il s'agit d'accélérer le maillage dans les réseaux locaux via les recrutements. Pour ce type d'analyse, on peut comparer les LECE aux LECF selon plusieurs critères : le taux de recrutement de chercheurs, la part des jeunes diplômés dans ce recrutement, la qualification des jeunes diplômés recrutés, recrutement intra ou inter-organisationnel, recrutement externe local ou global, recrutement intra-firme local ou global.

Analyse des sorties de chercheurs : on peut s'attendre à constater que les LECE cherchent à irriguer le réseau intra-firme global (selon la logique du « transfert inverse » (Criscuolo, 2002)) par le biais des sorties de chercheurs, et préfèrent donc, quand les sorties sont en intra-firme, les sorties globales. Mais ils cherchent aussi à alimenter la dynamique du réseau externe local, et favorisent donc les sorties locales dans le cas des sorties en externe.

En utilisant la base de données spatio-temporelle CTRD, toutes ces analyses thématiques concernant la GRH des chercheurs et ingénieurs dans les firmes multinationales peuvent être spatiales et/ou temporelles : réalisées sur une ou plusieurs années et/ou à différentes échelles territoriales (nationale, régionale, départementale ou communale).

Les analyses qui seront présentées par la suite (section 4) sont fondées principalement sur l'enquête spéciale sur les chercheurs dans les entreprises, réalisée en 2007 par le Ministère de l'Enseignement Supérieur et de la Recherche[4] concernant les données sur les effectifs chercheurs et les flux d'entrées/sortie. Ces données sont croisées avec des informations relatives à la localisation géographique (commune d'implantation) et à la nationalité des entreprises (celle de leur actionnaire majoritaire) qui sont collectées, auprès de Diane[5], en 2012 et supposées inchangées depuis 2007.

Ces analyses sont effectuées à l'échelle régionale du territoire français et concernent un échantillon de plus de 1600 entreprises (dont 42% LECE) qui cumulent en 2007 : un effectif de 60000 chercheurs (dont 33% dans des LECE), 7000 entrées (dont 35% vers des LECE) et 5000 sorties (dont 31% depuis des LECE). On s'intéresse particulièrement à l'analyse de la mobilité des chercheurs qui sera appréciée notamment en termes des recrutements (taux d'entrée).

4. http://www.enseignementsup-recherche.gouv.fr/reperes/public/formul/default.htm
5. http://www.bvdinfo.com/Products/Company-Information/National/Diane.aspx

4. Analyse de la mobilité des chercheurs dans les entreprises

L'intuition que nous avons développée est que les Laboratoires d'Entreprises sous Contrôle Etranger (LECE), représentant dans notre échantillon les acteurs qui portent la stratégie d'internationalisation de la R&D des FMN étrangères, cherchent à la fois à s'insérer dans le réseau externe local du territoire d'accueil, et à articuler cette insertion avec la dynamique des autres réseaux, notamment intra-firme local et global. Nous faisons donc l'hypothèse que ces LECE utilisent la mobilité des chercheurs et ingénieurs au sein des différents réseaux comme un dispositif indispensable et organisé de partage des savoirs.

Ces hypothèses donnent lieu à divers traitements grâce à la base CTRD. Avant d'aborder (4.2) l'analyse de la mobilité des chercheurs au sein des réseaux des firmes, comparons d'abord cette mobilité dans les entreprises LECE et LECF.

4.1. Analyse comparative de la mobilité des chercheurs des LECE et des LECF

Une première étape de vérification doit donc porter sur la comparaison à effectuer entre les LECE et les LECF quant à la mobilité (turnover) de leurs RH dédiées à l'innovation. La logique sous-jacente à la stratégie des firmes doit donc se manifester par une tendance nette à une plus forte mobilité de ces RH dans les LECE que dans les LECF. Nous avons conduit cette vérification en l'enrichissant par une analyse régionale différenciée, dont les enseignements sont essentiels.

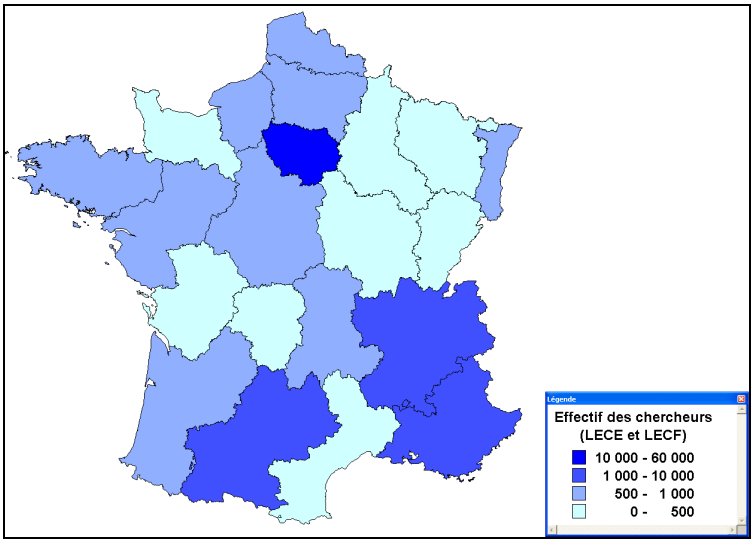
La démarche que nous présentons repose sur le principe d'une comparaison, pour toutes les données observées, entre les résultats concernant les LECE d'une part et les LECF d'autre part. Cette analyse comparative peut s'appuyer sur quatre indicateurs : l'effectif des chercheurs, le taux d'entrée/sortie des chercheurs par rapport à leur effectif et le taux de turnover = (taux d'entrée + taux de sortie) / 2.

En observant le taux moyen de turnover à l'échelle de la France en 2007 qui est de 11,9% dans les LECE et de 13,3% dans les LECF, on constate que dans l'ensemble, les LECE ne manifestent pas une tendance spécifique à la mobilité de leurs chercheurs et ingénieurs, et qu'ils sont même légèrement en retrait sur ce point par rapport aux LECF. Il est donc impossible à ce stade de conclure que les entreprises préfèrent la mobilité des chercheurs à la stabilité des équipes comme dispositif de partage des savoirs.

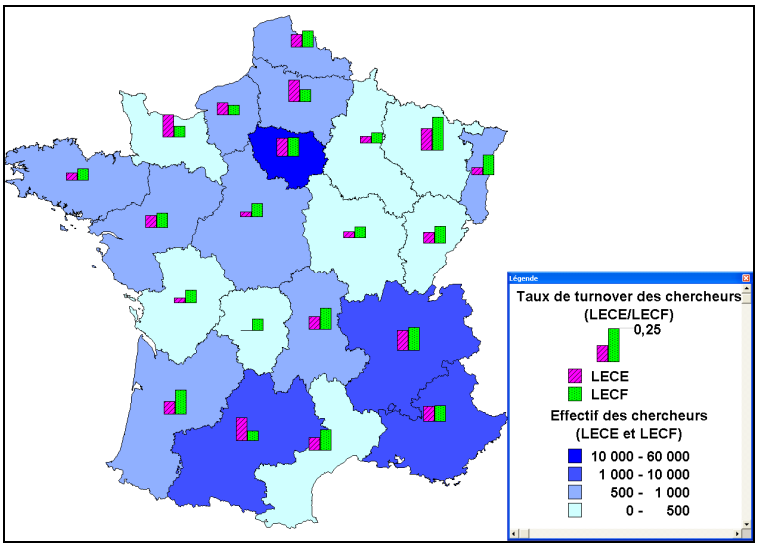
Ce résultat non conforme à nos attentes nous a incités à affiner l'analyse pour la prise en compte des dynamiques territoriales et spatiales au niveau infra-national, au moins régional car l'hétérogénéité entre les régions françaises, sur le plan technologique et scientifique, est considérable. On peut s'en rendre compte en distinguant les régions selon l'effectif de chercheurs et ingénieurs présent dans les entreprises (carte 1).

Il est donc important de mener l'analyse comparative LECE/LECF sur la mobilité des RH dédiées à l'innovation au moins à cette échelle régionale, car on peut faire l'hypothèse que les entreprises ne se comportent pas de la même manière

dans une région à fort potentiel technologique et scientifique et dans une région très mal pourvue dans ce domaine. En d'autres termes, nous proposons l'idée que c'est dans les régions à fort potentiel que la logique de partage des savoirs via la mobilité des chercheurs et ingénieurs prend tout son sens. On peut donc s'attendre à des résultats de plus forte mobilité pour les LECE dans ces régions-là.



Carte 1. Effectif total des chercheurs des LECE et des LECF en 2007



Carte 2. Taux de turnover des chercheurs dans les LECE et les LECF en 2007

La carte 2 montre de ce point de vue des tendances assez nettes et convergentes. En effet, pour les taux d'entrée des chercheurs et ingénieurs, on constate que la plus forte mobilité des LECF observée au niveau moyen national s'explique essentiellement par des résultats beaucoup plus élevés dans les régions à faible effectif de chercheurs et ingénieurs dans les entreprises. Par contraste, on constate au contraire que dans les quatre régions les plus dotées en chercheurs et ingénieurs (Ile de France, Rhône-Alpes, PACA et Midi-Pyrénées), les LECE ont dans l'ensemble ou bien des taux comparables à ceux des LECF, ou bien des taux nettement plus élevés que ceux des LECF (Midi-Pyrénées).

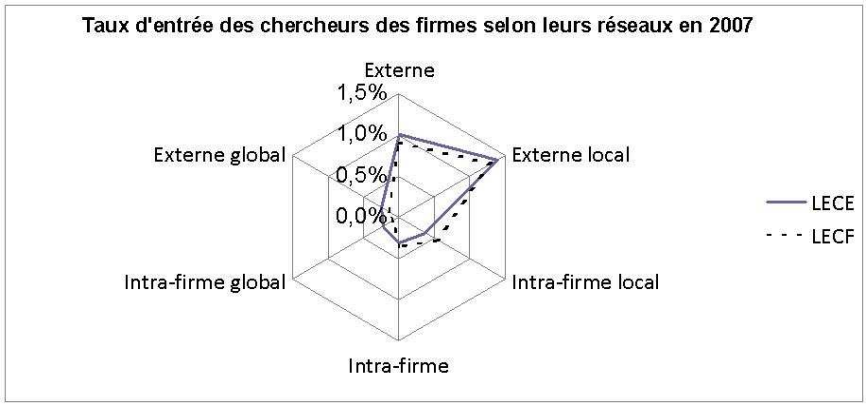
Ces résultats sont une étape importante et permettent de donner un fondement empirique à l'idée que nous proposons, selon laquelle les FMN utilisent la GRH de leurs chercheurs et ingénieurs dans les laboratoires qu'elles implantent à l'étranger, comme moyen de partager les savoirs au sein de leurs réseaux de connaissances, via la mobilité de ces chercheurs et ingénieurs.

L'analyse de la mobilité des chercheurs doit bien sûr être affinée, afin d'intégrer la diversité des réseaux de connaissances des entreprises (tableau 1).

4.2. Analyse de la mobilité des chercheurs au sein des réseaux des firmes

Nous nous limiterons dans cette section aux entrées des chercheurs en présentant, à titre d'exemple, une analyse comparative des recrutements des firmes en 2007 dans leurs réseaux externe (global/local)[6] et intra-firme (global/local). Pour apprécier ces recrutements, nous considérons comme indicateur le taux moyen d'entrée des chercheurs.

A l'échelle nationale, le graphique 1 montre les taux moyens d'entrée des chercheurs des LECE et des LECF selon leurs réseaux de connaissances.

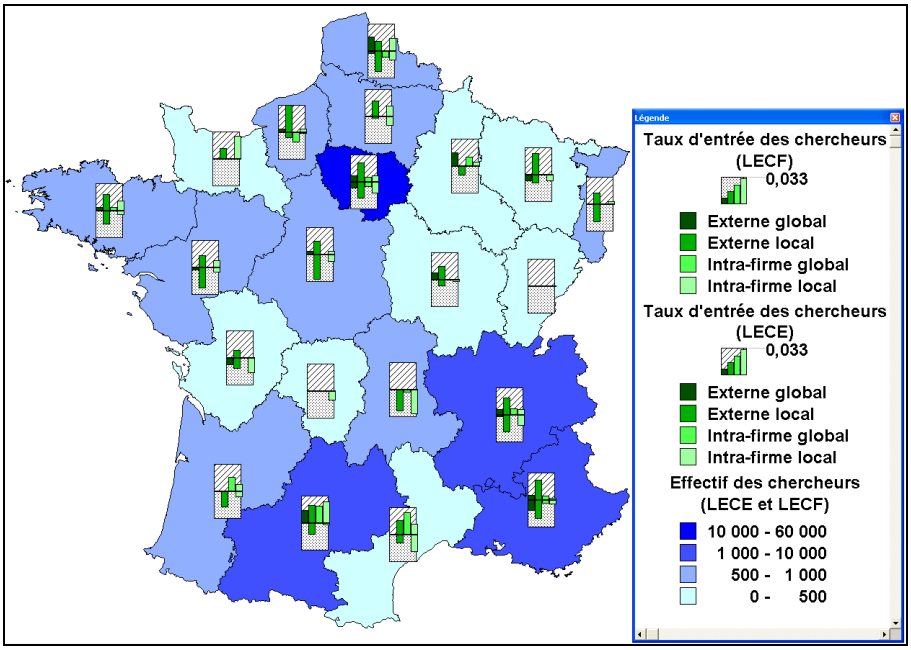


Graphique 1. Taux d'entrée des chercheurs des firmes selon leurs réseaux en 2007

6. Les LECF qui ne sont pas des FMN ne possèdent pas de réseau global.

Avant de procéder à l'analyse *comparative* des stratégies des LECE et des LECF dans le recrutement des chercheurs au sein de leurs réseaux, on analyse d'abord cette stratégie au sein de chaque type de firme. Globalement la stratégie de recrutement des chercheurs est la même pour les LECE que pour les LECF. Qu'elles soient sous contrôle français ou étranger, les firmes recrutent plus de chercheurs en externe qu'en intra-firme ; en plus, leur réseau local est privilégié à la fois dans leur recrutement en externe que dans leur recrutement en intra-firme (graphique 1).

Cette stratégie de préférer le recrutement en externe se confirme également à l'échelle régionale (carte 3) dans 74% des régions tant pour les LECE que pour les LECF. En plus. Cette préférence est d'autant plus nette dans les régions à fort potentiel de R&D. Concernant le recrutement en intra-firme au niveau des régions, les LECE et les LECF ont également la même préférence pour leur réseau local. Leur taux de recrutement intra-firme local est supérieur à leur taux de recrutement intra-firme global dans plus de 73% des régions dont on recrute en intra-firme.



Carte 3. Taux d'entrée des chercheurs LECE et LECF par provenance en 2007

Si toutes les firmes adoptent la même stratégie de recrutement au sein de leurs réseaux de connaissances, les LECE recrutent davantage au sein de leurs réseaux externes comparativement aux LECF. Cela est aussi valable au niveau global (graphique 1) qu'à l'échelle des régions (carte 3).

Recrutement intra ou inter-organisationnel : Les LECE recrutent plus en externe à la firme que les LECF. Au niveau national (graphique 1), le taux moyen de

recrutement externe représente 1,01% dans les LECE contre 0,92% dans les LECF. Au niveau régional (carte 3), cette stratégie est également vraie dans 63% des régions où les firmes recrutent des chercheurs en externe. Cette tendance est d'autant plus visible dans les régions à fort effectif de R&D ;

Recrutement externe local ou global : Dans leur recrutement en externe, les LECE privilégient plus les recrutés locaux, en comparaison avec les LECF. Cette stratégie est globalement vraie au niveau national (graphique 1) où le taux de recrutement externe local est de 1,40 % pour les LECE contre 1,30 % pour les LECF. Elle est également vérifiée au niveau régional, comme le montre la carte 3, dans 53% des régions où les firmes recrutent en externe ;

Recrutement intra-firme local ou global : On peut penser que dans leur recrutement en intra-firme, les LECE privilégient les recrutés locaux, en comparaison avec les LECF. Ce n'est pas le cas ni au niveau national (graphique 1) où le taux de recrutement intra-firme local représente 0,36% dans les LECE alors qu'il est de 0,54% dans les LECF, ni au niveau régional (carte 3) où cette hypothèse n'est vérifiée que dans 40% des régions qui recrutent en intra-firme.

Ainsi, les analyses effectuées à partir de la base de données CTRD ont permis d'esquisser les spécificités régionales des stratégies des firmes multinationales concernant la mobilité des chercheurs au sein de leurs différents réseaux de connaissances. Cette base de données permettra, moyennant l'utilisation d'outils géomatiques adaptés d'effectuer des analyses plus élaborées concernant la R&D des FMN: spatiales (Pumain et Saint-Julien, 2010), spatio-temporelles (Plumejeaud et al., 2011) et/ou multi-échelles (Le Rebrus, 2011).

5. Conclusion et perspectives

Nous avons présenté dans cet article la base de données spatio-temporelle CTRD conçue en général pour l'analyse des dispositifs RH de partage des savoirs dans les firmes multinationales. Nous avons montré en particulier, à titre d'illustration, son utilisation pour l'analyse de la mobilité des chercheurs et ingénieurs au sein des réseaux de connaissances de ces firmes. Cette analyse a révélé notamment des spécificités régionales concernant la stratégie des FMN en matière de mobilité de son personnel dédié à l'innovation.

Ce travail a démontré l'intérêt de la géomatique pour cette thématique et ouvre plusieurs perspectives de prolongement.

D'abord, enrichir le contenu de la base de données par l'intégration d'autres aspects de l'économie de la connaissance comme les brevets, les projets de recherche, les investissements R&D... Aussi, une meilleure prise en compte de la dimension temporelle dans la base de données en considérant l'évolution dans le temps de la population des FMN étudiées (création, changement de nom et/ou de nationalité par fusion et/ou acquisition, changement de localisation géographique par délocalisation/relocalisation, liquidation…).

Ensuite élargir le spectre des analyses potentielles à propos des stratégies des firmes multinationales : étude de similarité et/ou d'hétérogénéité des territoires, corrélation spatiale, analyse multi-échelle (département, commune) et multi-zonages (bassins d'emploi, zones de formation…).

Enfin, l'utilisation de nouveaux outils et de nouvelles techniques issus de la géomatique, comme la cartographie interactive ou l'analyse spatio-temporelle, permettra d'exploiter pleinement la base de données CTRD pour mieux comprendre et analyser l'évolution des stratégies des firmes multinationales relatives à la R&D.

Bibliographie

Allen J., Hayes P., (1989). Moments and points in an interval-based temporal logic. *Computationnal Intelligence*, 5(4), p. 225-238.

Allison P. Scott L., (1987). « Interuniversity mobility of academic scientists ». *American Sociological Review*, Vol. 52, p. 643-652.

Colovic A., Mayrhofer U., (2008). « Les stratégies de localisation des firmes multinationales. Une analyse du secteur automobile ». *Revue Française de Gestion*, n°184, p. 151-165.

Criscuolo P. (2002). « Reverse Technology Transfer: A Patent Citation Analysis of the European Chemicals and Pharmaceutical Sectors ». *MERIT-Infonomics, Research Memorandum series*, 2002-027.

Gassman, O., Von Zedtwitz, M., (1999). « New concepts and trends in international R&D organization », *Research Policy*, Vol. 28, n° 2-3, p. 231-250.

Jacquier-Roux V., Le Bas C., (2008). « Localisation des activités de RD des firmes multinationales, modes d'organisation en réseau et transfert transnational des connaissances : un cadre d'analyse ». *Région et Développement*, n° 28, p.11-38.

Jacquier-Roux V., Montana N., Paraponaris C., (2012). Diversité et partage des connaissances dans la R&D des firmes multinationales, des logiques de processus aux logiques de situation. Revue française de gestion, vol. 38, n° 221, pp. 129-148.

Kuemmerle, W. (1997). « Building effective R&D capabilities abroad ». *Harvard Business Review*, March-April, p. 61-70.

Lahiri, N. (2010). « Geographic distribution of R&D activity: how does it affect innovation quality? ». *Academy of Management Journal*, Vol. 53, n° 5, p. 1194-1209.

Le Rubrus B. (2011). *Cartographie et analyse territoriale multiscalaire Réingénierie des logiciels HyperAtlas et HyperAdmin*. Mémoire d'Ingénieur CNAM en Informatique, Conservatoire National Des Arts Et Métiers, Grenoble.

Nonaka I., Takeuchi H., (1995). *The Knowledge Creation Compagny*, New York, Oxford University Press, 284 p.

Patel P., Vega M., (1999). «Patterns of Internationalization of Corporate Technology: Location versus Home Country Advantage», *Research Policy*, Vol. 28, n° 2-3, p. 145-155.

Plumejeaud C., Hélène Mathian H., Gensel J., Grasland C., (2011). Spatio-temporal analysis of territorial changes from a multi-scale perspective. *International Journal of Geographical Information Science*, 25(10), p. 1597-1612.

Poitou J-P. (1997). « La gestion collective des connaissances et la mémoire individuelle », in J.M. Fouet (Coord.) *Connaissances et savoir-faire en entreprise. Intégration et capitalisation*. Paris, Hermès, p. 157-178.

Polanyi M. (1967). *The Tacit Dimension*, Doubleday, New York, 108 p.

Pumain D., Saint-Julien T., (2010). *Analyse spatiale : les localisations*, Armand Colin, Paris, 192 p.

Sachwald F., (2008). *Réseaux mondiaux d'innovation ouverte, systèmes nationaux et politiques publiques*. Ministère de l'enseignement supérieur et de la recherche, Paris 2008.

Tsoukas H. (2003). « Do we really Understand Tacit Knowledge ? », in Easterby-Smith Mark et Lyles Marjorie A. (sous la direction de), *The Blackwell Handbook of Organizational Learning and Knowledge Management*, Blackwell, p. 410-427.

Von Hippel E. (1994). « Sticky Information and the Locus of Problem Solving : Implications for Innovation ». *Management Science*, Vol. 40, n° 4, p. 429-439.

Zander, I. (2002). « The formation of International Innovation Networks in the Multinational Corporation: an Evolutionary Perspective », Industrial and Corporate Change, Vol. 11, n° 2, p. 327-353.

Dispositif mobile et géoweb pour la collecte de données d'inventaire sismique du bâti en milieu urbain

L'application Isibat

Paule-Annick Davoine[1], Laurent Poulenard[1], Philippe Gueguen[2], Jérôme Gensel[1], Jose Bringel-Filho[1]

1 Laboratoire d'Informatique de Grenoble – équipe STeamer
681 Rue de la Passerelle, BP 72
38 400 Saint-Martin d'Hères
Nom.Prenom@imag.fr

2 Institut des Sciences de la Terre, équipe Risques, & IFSTAR
Rue de la Chimie
38 400 Saint-Martin d'Hères
Philippe.Gueguen@ujf-grenoble.fr

RÉSUMÉ. *Réaliser un inventaire sismique du bâti constitue un enjeu majeur pour la connaissance et la gestion du risque sismique. Compte tenu de la difficulté à collecter en quantité suffisante, et de façon détaillée, les paramètres structuraux à l'échelle du bâtiment, nous proposons une approche basée sur les technologies offertes par le Géoweb dans le but de faciliter la co-production de données liées à l'évaluation de la vulnérabilité et des dommages sismique d'une ville. Notre approche repose sur une application associant un dispositif mobile pour la collecte in situ de données (partie client), et une application web pour le stockage, la visualisation et la diffusion des données collectées (partie serveur).*

ABSTRACT. *In the field of seismic risk, taking stock of the urban built is one of the main challenging tasks, considering the large number and the heterogeneity of the objects to be studied in every city. In this paper, we present Isibat, a client-serveur application which makes use of both wireless networks and technologies and geoweb software, dedicated to the collect of seismic data. On the client side, through an Iphone application, mobile users on the field are invited through screens, menus and maps, to collect data that are used to evaluate either the vulneraibility of surrouding buildings (in a pre-seismic phase) or the damage caused (in a post-seismic phase).On the server side, data are stored but also can be queried and visualized through adapted vulnerability or damage maps.*

MOTS-CLÉS : *Application mobile, Collecte In-situ, GéoWeb, Inventaire Sismique, Sismologie Citoyenne, co-production de données, Volunteered Geographical Information.*

KEYWORDS:,*Volunteered Geographical Information, Geoweb, Mobile software, Seismic riks, citizen sismology*

1. Introduction

La vulnérabilité sismique d'un milieu urbain est caractérisée par la capacité des bâtiments et des structures à supporter les secousses sismiques. De nombreux travaux de recherche ont montré l'existence d'une forte corrélation entre les caractéristiques de la construction (nature des matériaux, forme de toit, nombre d'étages, irrégularité du bâtiment …) et le niveau d'endommagement pour un niveau de sollicitation sismique (Spence and Lebrun 2006 ; Calvi et al. 2006 ; Gueguen et al., 2007). La connaissance du milieu urbain et, plus encore, du bâti qui le compose, constitue un élément important à maîtriser pour gérer, prédire et évaluer sa vulnérabilité et son intégrité post-sismique. A l'origine, différentes méthodes d'estimation de la vulnérabilité sismique du bâti ont été proposées (HAZUS 1997[1] ; GNDT 1993, RiskUE 2003), ayant conduit à la constitution d'un programme global de réflexion sur la vulnérabilité à l'échelle mondiale (Global Earthquake Model, GEM). Toutefois, un espace urbain est vaste et hétérogène, et l'une des difficultés revient à collecter en quantité suffisante, et de façon détaillée, les données nécessaires.

Définir des méthodes et des outils permettant d'effectuer un inventaire sismique du bâti d'une ville pour obtenir des données détaillées sur la structure et les caractéristiques des bâtiments devient un enjeu pour la connaissance et la gestion du risque sismique. Aujourd'hui, les technologies des téléphones mobiles intelligents de type *Smartphone*, associées aux technologies du Web 2.0, et plus particulièrement du Geoweb, constituent une approche intéressante pour l'acquisition d'informations in-situ. Les *smartphones* sont devenus omniprésents et familiers. Ils sont dotés de technologies de localisation de type GPS permettant de relever les positions des utilisateurs ou d'objets situés à la surface de la terre. Par ailleurs, le développement des applications mobiles offre la possibilité d'annoter sémantiquement ces objets localisés, renseignant ainsi leurs caractéristiques (Viana, 2010).

Dans le cadre du projet Urbasis[2], une application, basée sur l'utilisation de *Smartphones* de type Iphone, associée à une approche basée sur les concepts de la coproduction de données et de la *Volunteered Geographical Information* (VGI) (Goodchild, 2007), permet à une communauté d'utilisateurs de collecter, in situ, un maximum de données intervenant dans l'évaluation du risque sismique. Les données d'inventaire collectées sont mises à disposition des utilisateurs pour compléter une banque d'information du bâti et contribuer ainsi à l'estimation de la vulnérabilité en période pré-sismique, ou de l'endommagement, en période post-sismique.

Dans cet article, après avoir rappelé les méthodes d'évaluation de la vulnérabilité sismique et les enjeux d'un inventaire sismique du bâti urbain (section 2), nous traitons de l'intérêt du Géoweb pour sa réalisation (section 3), avant de présenter les

---

[1] http://www.fema.gou/hazus

[2] ANR-Urbasis, Sismologie Urbaine : Evaluation de la vulnérabilité et des dommages sismiques par des méthodes innovantes. ANR -09-Risk-009, http://users.isterre.fr/pgueg/URBASIS/Accueil.html

principales caractéristiques de l'application Isibat (Inventaire SIsmique du BAti) dans la section 4.

2. De la nécessité d'un Inventaire sismique du bâti

L'estimation de la vulnérabilité sismique du bâti est basée sur l'échelle macrosismique européenne EMS-98[3] (2001), qui propose une classification en fonction des caractéristiques de la construction (Figure 1). L'enjeu est de construire une cartographie fine de la vulnérabilité de la ville, ou de l'endommagement causé lors d'un séisme, afin d'améliorer la connaissance et la gestion du risque sismique.

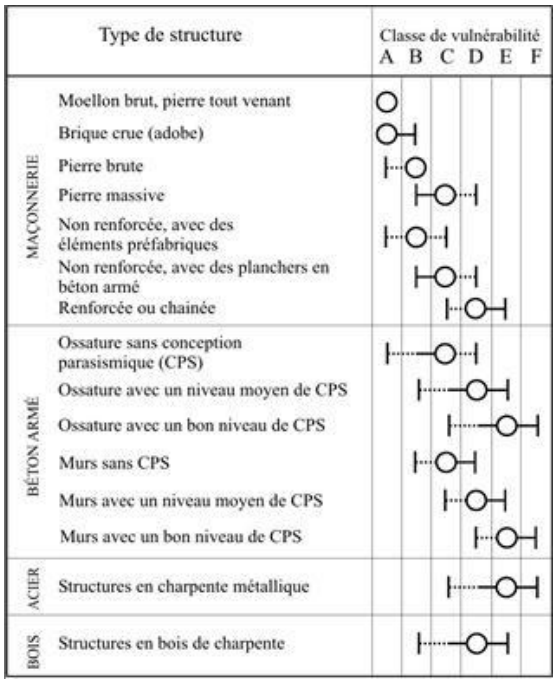


Figure 1 : Grille d'évaluation de la vulnérabilité selon l'échelle européenne EMS 98
Les bâtiments sont groupés par types de construction sur la base de leur conception. Le cercle indique la valeur la plus probable de vulnérabilité (A très vulnérable, F pas vulnérable), les traits continus l'intervalle le plus probable et les traits pointillés l'intervalle le moins probable.

Actuellement, les principales approches sont basées sur l'analyse d'orthophotos et de MNT couplés à des données géographiques cadastrales ou sur l'utilisation de données urbaines définies à l'échelle des quartiers. (Gueguen, 2007). Dans le

[3] Echelle d'intensité macrosismique : http://www.franceseisme.fr/ems98.html
.

premier cas, on obtient une image de la diversité de la morphologie urbaine qui, *in fine*, est caractéristique de sa composition. Un des critères retenus ici est la "rugosité urbaine", soit le rapport entre la hauteur moyenne des constructions et la densité urbaine (surface occupée par le bâti sur la surface totale de la zone). Or d'autres critères, tels que le type de matériaux de la construction, la forme du toit, l'irrégularité des bâtiments en élévation ou en plan, la nature et la pente du sol sur lequel sont construit les bâtiments, etc... sont tout aussi importants à prendre en compte. Dans le second cas, les études de vulnérabilité urbaine s'appuient sur des données spatiales agrégées ou des paramètres urbains déduits (Gueguen & al., 2007). Toutefois, ces agrégations correspondent à des résumés statistiques plus ou moins pertinents, simplifiant et généralisant les caractéristiques spatiales, sans prendre en compte les variations locales.

Lors de l'estimation des dommages, qui s'effectue en période post-sismique, le recueil de données in-situ est privilégié. Il s'agit d'effectuer par auscultation visuelle, un relevé de l'état des constructions en se basant sur la grille d'évaluation standard EMS 98 (figure 2). Ces relevés sont effectués par plusieurs enquêteurs répartis sur le site sinistré, au moyen de fiches papiers consignés dans des carnets de terrain. Une fois les relevés effectués, les données sont saisies sur ordinateur, puis traitées. Cette procédure longue et coûteuse, ne permet pas une cartographie immédiate des dommages, ni d'identifier in-situ, les enjeux. Or, l'objectif ici est à la fois de construire une cartographie des dommages causés, dans une problématique de gestion efficace de la crise, mais aussi de conduire à l'estimation d'un niveau d'intensité d'événement sismique selon l'échelle des dommages et des intensités EMS98. Par ailleurs, l'endommagement sur un bâtiment donné influençant sa vulnérabilité, il est important de pouvoir procéder à sa ré-évaluation entre deux répliques sismiques

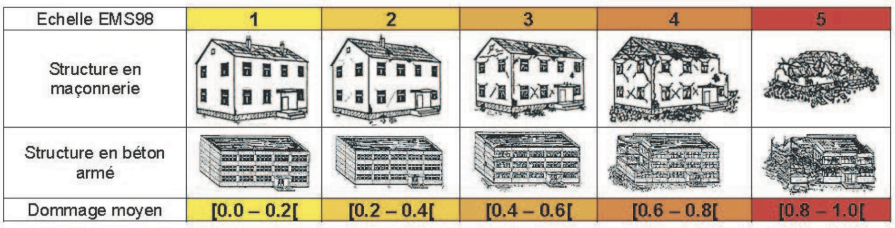


Figure 2 : Grille d'évaluation des dommage selon l'échelle EMS98

Quel que soit l'objectif, les différentes méthodes et données disponibles à ce jour possèdent plusieurs limites : elles ne permettent pas une connaissance fine du bâti, ni d'établir de façon pertinente des corrélations Construction/Vulnérabilité/ Endommagement pour un niveau de séisme donné ; Elles rendent aussi difficile l'identification de bâtiments spécifiques à forts enjeux structuraux, à l'intérieur d'un zonage. Pour pallier ces limites, il est nécessaire d'obtenir une grande quantité

d'informations géo-référencées au niveau du bâtiment, et de faciliter leur collecte et leur réutilisation.

3. Le Géoweb au service de la collecte de données de sismologie urbaine

3.1. Intérêt

Le Géoweb ou web géographique qui désigne la convergence des Technologies de l'Information Géographique, des Technologies de l'Information et de la Communication, notamment avec le Web, Internet, la téléphonie mobile et les systèmes de géo localisation (Merisckay & al, 2011) offre de nouvelles opportunités dans le domaine de la collecte de données in-situ.

D'un coté, les dispositifs mobiles tels que les *smarphones* ou les tablettes disposant de fonctionnalités de géolocalisation, peuvent être utilisés comme des carnets électroniques de terrain. Les informations sont alors directement saisies sous forme numérique et automatiquement géoréférencées. L'étape de retranscription, qui génère souvent de nombreuses erreurs est alors supprimée, améliorant ainsi la qualité des données. La saisie in-situ des données géoréférencées, nécessite de mettre en place des protocoles et des formulaires améliorant l'homogénéisation des données produites. Ces dispositifs mobiles permettent aussi de contextualiser l'information en associant aux données géoréférencées des métadonnées de divers formats multimédias : des photos, des vidéos, des enregistrements audio, des commentaires textuels, etc. (Viania, 2010).

De l'autre, le *crowdsourcing*, la *Volunteered Geographical Information* qui désignent l'exploitation des outils pour créer, assembler, et diffuser des données géographiques fournies volontairement par des individus ou citoyens (Goodchild, 2007) sont devenus un moyen puissant de co-production d'information géographique (Noucher 2009). Une communauté d'acteurs, guidée par un objectif commun, peut alors produire et partager des données géographiques sur une thématique donnée (Noucher 2011). Via une interface Web et des fonctionnalités spécifiques, les utilisateurs deviennent des contributeurs actifs dans le processus de production d'information. OpenStreetMap [4] est l'une des applications issues de cette approche les plus connues. Au cours de ces dernières années, de nombreuses initiatives se sont développées, tant dans le domaine des risques naturels (Ushahidi[5] ...) – notamment pour collecter rapidement des données afin d'évaluation des dégâts en période de crise –, que dans celui de l'observation de l'environnement (Ruitton-Allinieu, 2011 ; Mericskay &al 2001)

Procéder à un inventaire le plus exhaustif possible des bâtiments d'une ville et de ses caractéristiques structurales est une tâche longue, et coûteuse aussi bien en temps, qu'en ressources humaines. Les enquêteurs parcourent un espace urbain

[4] http://openstreetmap.org
[5] www.ushahidi.com

vaste, observent de façon visuelle le bâtiment, renseignent sa position géographique et la liste des paramètres ... Recourir à une approche basée sur la co-production de données, en mettant à disposition des contributeurs, un environnement associant un dispositif mobile pour l'acquisition des données et une application web pour le stockage, la visualisation et l'utilisation nous semble être une solution adaptée à la réalisation d'inventaires sismiques des bâtiments et au recueil de renseignements sur les paramètres structuraux, nécessaires à l'évaluation du risque sismique. Au delà d'une simple contribution d'experts-enquêteurs, nous émettons l'hypothèse que toute personne peut, à sa manière, contribuer à l'inventaire sismique d'un bâti urbain. Cette hypothèse nécessite de s'interroger sur les modalités de contribution.

3.2. Quelles contributions dans un contexte d'inventaire sismique et comment ?

Notre objectif est de permettre à toute personne, disposant sur son Smartphone ou sa tablette d'une application dédiée, de collecter des données sur les paramètres structuraux des bâtiments situés dans son environnement urbain. Pour cela, le niveau de compréhension des paramètres à renseigner doit être accessible à tous. Dans notre contexte, la liste de ces paramètres, sont ceux définis par l'EMS 84 et sont présentés dans le tableau 1..

Vulnérabilité	Dommage
Type de strucutre	Niveau de dommage
Epoque de construction	Fissures fines
Nature du sol	Fissures larges
Toiture	Chute petits morceaux
Nombre d'étages	Ecroulement de morceaux
Régularité en élévation	Fissures aux joints
Défaut de rigidité	Chutes aux joints
Hauteur des étages	Effrondrement partiel de plancher
Position de la structure	Effondrement d'un étage
Régularité en plan	Chute de partie de cheminée
Renforcement-Qualification	Chute de tuiles
Etat d'entretien	Effondrement partiel du toit
Diaphragmes horizontaux	Effondrement total du toir

Tableau 1 : Paramètres structuraux pour l'évaluation de la vulnérabilité et des dommages sismiques (niveau général) (Gueguen 2007)

A priori, ces paramètres ne nécessitent pas un haut niveau d'expertise. Toutefois, dans la nomenclature EMS 84, ces paramètres sont organisés de façon hiérarchique allant du niveau le plus général (niveau 1) au niveau le plus précis (niveau n). Par exemple, pour le paramètre *Toiture*, le niveau 1 consistera à renseigner uniquement si *la toiture est en terrasse ou non. Les autres niveaux consisteront à préciser la forme et pente de la toiture, et ses caractéristiques de construction.* Selon les

spécifications techniques, ces paramètres permettent d'affiner les méthodes d'évaluation. Si la compréhension de ces paramètres détaillés demande, elle, un certain niveau d'expertise, pour autant, l'évaluation de la vulnérabilité pourra être réalisée, même si ces derniers ne sont pas renseignés.

Dans le contexte de la sismologie urbaine, , selon les paramètres, un certain niveau d'expertise peut être requis, ce qui rend plus délicat la collecte de données dans une démarche citoyenne. En conséquence, nous proposons de mettre en place un processus de co-production de données, à la fois experte et citoyenne, en lien avec la complexité et l'organisation des données proposées par EMS 84. On distingue alors :

- La collecte citoyenne qui concerne essentiellement les données d'ordre général, de niveau 1, et destinée à des contributeurs non experts sensibilisés à la problématique du risque sismique.

- La collecte experte, permettant de renseigner les paramètres plus complexes et qui s'inscrit plutôt dans une démarche de co-production de données. Cette collecte peut se faire pendant des campagnes de collecte ou de façon volontaire.

Une autre contrainte à prendre en compte, est celle liée à la qualité des données, qui peut impacter le potentiel de réutilisabilité des données. En effet, différents travaux de recherche (Gires 2010 ; Ruitton-Allinieu, 2011) ont montré que les données issues de la VGI pouvaient facilement être sujettes à de nombreuses erreurs, tant géométriques que sémantiques. Dans notre cas, la dimension géographique porte sur la localisation du bâtiment à inventorier, (coordonnées géographiques), et non pas sur sa forme géométrique, ni sur sa précision planimétrique. Outre certains bâtiments caractéristiques et à forts enjeux, les autres peuvent tolérer une certaine imprécision. En revanche, les données sémantiques doivent faire l'objet de spécifications particulières afin de faciliter leur saisie et de limiter les erreurs. Dans cet objectif, nous proposons un dispositif de saisie qui repose sur l'usage d'items de type booléens (oui/non), mais également sur l'usage de liste d'items correspondant aux différentes valeurs que peuvent prendre les paramètres structuraux. Ces valeurs peuvent être qualitatives ou quantitatives mais aussi correspondre à des pictogrammes. Pour faciliter la compréhension et l'identification des paramètres à renseigner, nous proposons d'associer aux paramètres des éléments explicatifs, soit sous forme de texte, soit sous formes de schéma.

La motivation des contributeurs est un aspect important dans le processus de la VGI ou géographie participative (Noucher 2009). Si la motivation des contributeurs experts ne fait aucun doute, compte tenu des enjeux, en revanche on peut s'interroger sur ce qui peut pousser des citoyens à procéder volontairement et de façon spontanée à une collecte de données de ce type. Compte tenu du succès que rencontre, ce que Bossu (Bossu 2011 a, b)) appelle la *sismologie citoyenne* on peut penser qu'en période post-sismique, la contribution citoyenne sera importante ; en revanche, en période pré-sismique, il est nécessaire de trouver des éléments de motivation, d'autant plus que renseigner sur les caractéristiques structurales et sismologiques des bâtiments peut aussi avoir un impact négatif sur son appréciation, voir sur sa valeur vénale. L'idée est donc de donner au citoyen le sentiment de

contribuer à un projet. Nous proposons, qu'après traitement, et généralisation cartographique, les données produites serviront à publier des cartes thématiques d'information sur le risque sismique. D'une façon plus générale, nous proposons de définir des modalités différentes d'accès aux données selon que le contributeur s'inscrit dans une collecte citoyenne ou experte.

4. L'application Isibat

Isibat est une application qui permet à des utilisateurs de co-produire des informations géo-référencées sur les bâtiments d'une ville afin d'estimer leur vulnérabilité ou leur endommagement. Isibat est à la fois un dispositif mobile (partie client) et une application Web (partie serveur). Côté client, sur le dispositif mobile en situation de collecte, l'application IsibatMobile, permet l'acquisition *in situ* des données relatives aux bâtiments recensés ainsi que l'évaluation en temps réel de leur niveaux de vulnérabilité ou d'endommagement. Côté serveur, sur une station de travail par exemple, l'application Web, IsibatOnline, assure le dépôt, le stockage, la visualisation et la diffusion des données acquises. L'application web doit aussi offrir des fonctionnalités de traitement et d'analyse relatives à l'évaluation du risque sismique à l'échelle de la ville.

4.1. Principes de l'application

L'application Isibat repose sur les principes suivants :

- Une acquisition interactive et multi niveaux des données sémantiques : l'utilisateur saisit, au moyen de son dispositif mobile, les paramètres urbains caractérisant les bâtiments, selon différents niveaux de précision en fonction de ses compétences et de ses connaissances.

- Une acquisition automatique de données contextuelles : les coordonnées géographiques associées aux bâtiments, ainsi que celles liées à la date et l'heure de collecte sont fournies automatiquement par le dispositif mobile.

- Une utilisation contextualisée : différents modes de collecte sont proposés selon que l'on se situe en période pré-sismique pour l'évaluation de la vulnérabilité ou post-simisque, pour l'évaluation de l'endommagement.

- Une visualisation cartographique : les applications IsibatMobile et IsibatOnLine adoptent toutes deux une approche cartographique aussi bien dans le processus de collecte, que dans la restitution et la diffusion des données.

- Une co-production et un partage de l'information : l'application est conçue pour permettre à plusieurs contributeurs ayant des niveaux de connaissance variés et des objectifs différents, de collecter des données soit de façon spontanée, soit dans le cadre de campagnes d'acquisition spécifiques. Une fois les données collectées, celles-ci sont déposées sur un serveur, puis mises à disposition, selon différentes modalités, via l'application Web.

4.2 IsibatMobile

Téléchargeable gratuitement depuis l'*Apple Store*, l'application fonctionne en mode *Vulnérabilité,* ou bien en mode *Dommage* : un même bâtiment peut être édité alternativement dans l'un des deux modes. Chacun de ces modes possède des fonctions de saisie, de suppression et de modification des paramètres nécessaires à l'évaluation de la vulnérabilité et de l'endommagement. Les données sémantiques sont saisies au moyen de listes déroulantes contenant les items à sélectionner *(mode vulnérabilité)* ou « de case à sélectionner » (*(mode dommage)*. L'enquêteur peut aussi, à l'aide de son dispositif mobile, prendre des photos ou effectuer des enregistrements sonores.

- En mode *Vulnérabilité*, les valeurs des paramètres structuraux sont organisées hiérarchiquement permettant une saisie progressive des données. Lorsque l'utilisateur sélectionne une valeur au niveau 1 (figure 3 (1)), les valeurs possibles pour le niveau 2 se déroulent, et ce, jusqu'au dernier niveau (figure 3 (2)). Cette approche permet à l'utilisateur d'affiner l'inventaire sismique en fonction de son niveau d'expertise. Compte tenu de la complexité des paramètres et afin d'éviter les erreurs de compréhension, un texte explicatif, accompagné parfois de figures, est associé à chaque paramètre sélectionné (figure 3 (3)). Dés qu'un bâtiment est inventorié, ses caractéristiques sismologiques sont calculées et affichées selon l'échelle EMS-98 (figure 3 (4)).

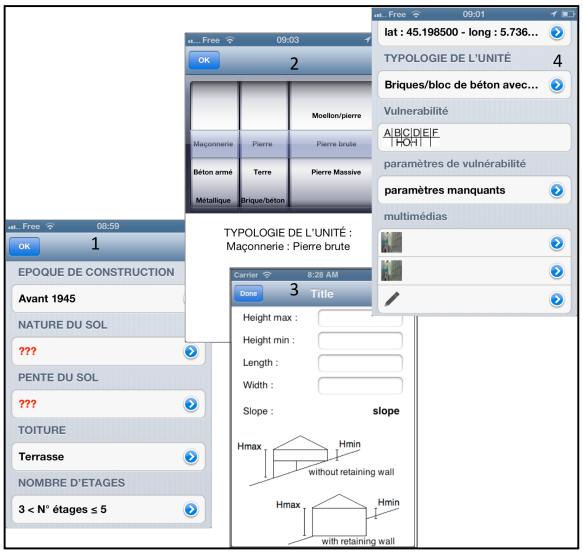


Figure 3 : IsibatMobile, principales fonctionnalités en mode Vulnérabilité :
Edition des paramètres niveau 1 (1) et de niveaux supérieurs (2) ; Exemple d'aide (3) ;
Calcul du niveau de vulnérabilité du bâtiment et enregistrements multimédia (4).

- En mode *Dommage,* les paramètres correspondent à une constatation visuelle que l'enquêteur réalise in-situ. Ces paramètres peuvent prendre trois valeurs : *oui, non ou non renseigné*. Le niveau de dommage est déterminé visuellement en sélectionnant une des images proposées par la grille EMS 84 (figure 4).

Figure 4 : IsibatMobile, principales fonctionnalités en mode Dommage : Edition des paramètres (figure gauche) ; Sélection du type de dommage (figure droite) .

L'application mobile est dotée d'un composant cartographique qui permet de localiser les bâtiments répertoriés et de cartographier leur niveau de vulnérabilité et /ou d'endommagement (figure 5).

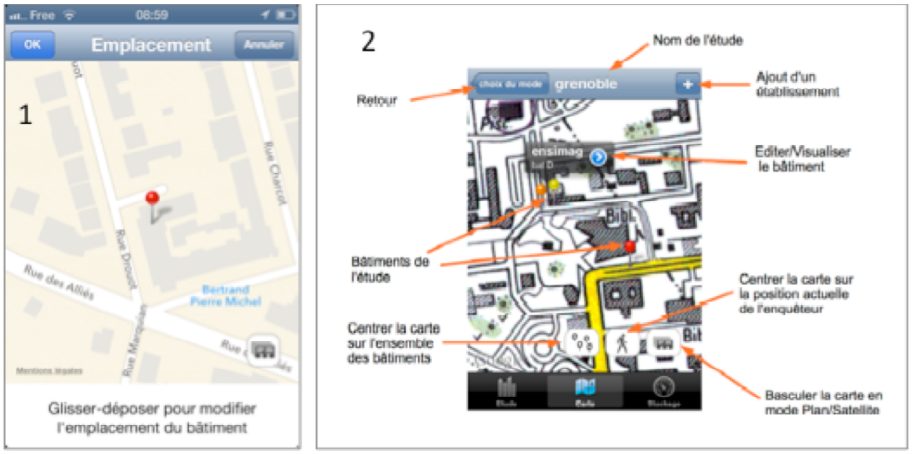


Figure 5 : IsibatMobile, composant cartographique

La localisation associée au bâtiment recensé peut être affinée de façon interactive.(Figure 5 (1)). Par défaut les données cartographiques utilisées sont celles proposées par Google, mais l'utilisateur à la possibilité d'intégrer ses propres fonds de carte, qu'ils soient en mode vecteur ou matriciel.

4.3 IsibatOnline

L'application Web, IsibatOnline a en charge la gestion, la visualisation et le traitement cartographique des données collectées. L'application est composée de plusieurs modules (figure 4):

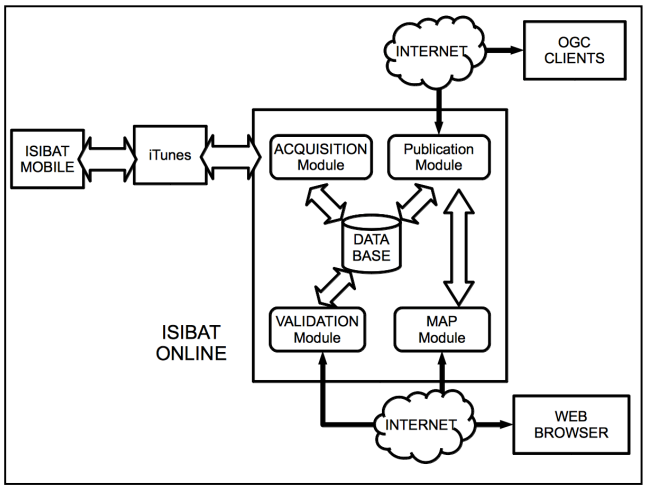


Figure 4 : Architecture de l'application Isibat

- Un module de gestion qui assure le dépôt ou le téléchargement des données collectées : un flux montant permet d'envoyer les données collectées vers le serveur de base de données ; un flux descendant permet à un contributeur de télécharger les données d'une collecte qu'il souhaite poursuivre.

- Un module de base de données de type PostgreSQL/PostGIS pour le stockage des données collectées.

- Un module de validation, des données déposées par un contributeur. Des modifications sur des données déjà existantes peuvent être réalisées selon les droits attribuer à l'utilisateur.

- Un module de publication des données sous forme de services Web OGC (WMS, WFS…). Ce module permet d'afficher les bâtiments recensés sur les cartes de l'application Web, ou de les exporter vers des applications clientes gérant les standards OGC.

- Un module de visualisation (MAP) pour la consultation des données une interface multidimensionnelle proposant différents modes d'affichage : cartographiques (carte en point, en aplat ou carte de densité), graphiques (diagrammes en bâtons, en secteurs ...) ou tableaux de données (figure 5). La consultation peut s'effectuer selon différents critères correspondant aux paramètres structuraux ou sismologiques (niveau de vulnérabilité et de dommage). On peut, par exemple, créer la carte des bâtiments construit avant 2000, ou celle de la vulnérabilité correspondant aux bâtiments construit avant une date sélectionnée. Le module de publication permet ensuite de construire dynamiquement la carte correspondant à ces filtres.,

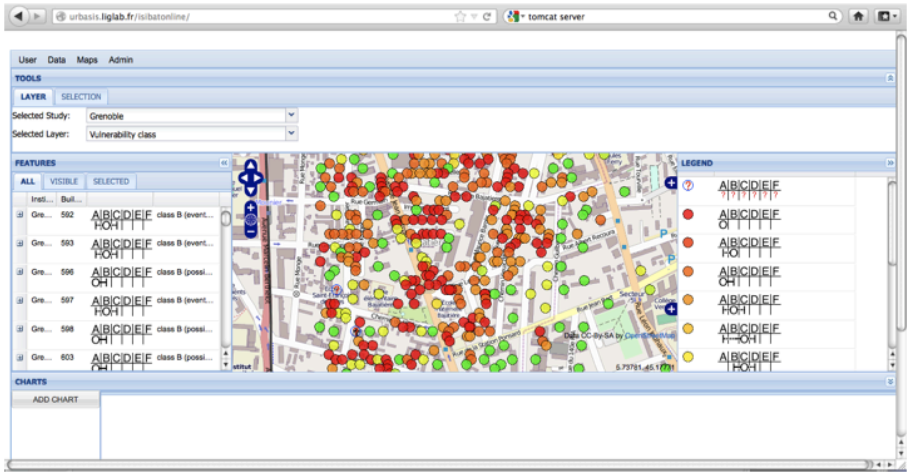


Figure 5 : IsibatOnLine ; Interface Web de visualisation

4.3 Aspects technologiques et fonctionnels

IsibatMobile repose essentiellement sur des technologies Apple ©. Elle a été développée avec le *framework* de développement XCode, et est basée sur le langage orienté objet Objective C. La partie cartographique de l'application utilise les bibliothèques Apple MapKit.framework (cartes interactives) et CoreLocation.framework (géolocalisation). Les données sont stockées à l'aide de SQLite qui est un moteur de base de données sans serveur, embarqué sur l'IPhone. SQLite fait partie du domaine public.

IsibatOnline, l'application Web, repose essentiellement sur le framework Google Web Toolkit (GWT). Cette solution permet d'écrire le code à l'aide de Java lors de la phase de développement, puis de traduire ensuite ce code en Javascript lors de la compilation. Ce choix a été effectué pour deux raisons : il facilite l'écriture de tests unitaires ou fonctionnels automatisés, et il permet de s'affranchir des problèmes liés à la compatibilité avec les différents navigateurs, la compilation générant plusieurs versions Javascript, adaptées aux différents navigateurs (Firefox, Chrome,

IExplorer…). Les cartes interactives de l'application utilisent un composant GWT basé sur ceux d'OpenLayers. Par défaut, le fond de carte utilise des couches OpenStreetMap ou Google Maps, mais d'autres données cartographiques ou géographiques peuvent être intégrées (au choix de l'utilisateur). Sur ce fond de carte, on ajoute des couches OGC (WMS essentiellement) affichant les données récoltées lors des collectes effectuées avec IsibatMobile. Les diagrammes statistiques et l'interface utilisateur ont été réalisés à l'aide de composants GXT, une bibliothèque OpenSource. Un serveur PostgreSQL/POSTGIS assure la persistance et l'accès aux données. L'application est installée sur un serveur Tomcat.

Les échanges entre les deux applications se font à l'aide de fichiers XML, basés sur un schéma propriétaire, ou tout simplement des fichiers images pour les photos. Chacune des deux applications traduit ce fichier et l'insère dans la base de données (SQLite ou PostgreSQL) qu'elle utilise. Pour récupérer ces fichiers ou les copier sur l'IPhone, il faut utiliser l'application iTunes. Les téléchargements vers ou depuis l'application Web se font ensuite de façon classique.

5. Conclusion et perspectives

Les progrès des technologies mobiles et réseaux sans fil d'un côté, des composants cartographiques de l'autre, permettent aujourd'hui d'envisager des plates-formes pour la collecte in-situ et l'exploitation de données géoréférencées diverses, sur la base d'architectures client-serveur performantes, et en tirant parti de la multiplicité des capteurs désormais embarqués sur la plupart des dispositifs mobiles (*Smartphones,* tablettes PC, …) du marché. Dans ce contexte, nous avons présenté l'application, Isibat, dédiée à la collecte in-situ de données d'inventaire sismique du bâti en milieu urbain. Celle-ci est composé de deux applications : La première, IsibatMobile, est une application mobile cliente, disponible sur Iphone ou Ipad, destinée à assister de manière intuitive et interactive les contributeurs lors de campagne de collecte, en période pré-sismique ou post-sismique. La seconde, IsibatOnLine, est une application Web, qui fait office de serveur dédié à la gestion des informations collectées via IsibatMobile. IsibatOnLine offre un accès aux données collectées et, à travers des composants cartographiques dynamiques et interactifs, permet une interrogation multi-critère des zones inventoriées.

A ce jour Isibat a été utilisée, par des contributeurs-enquêteurs, ayant une certaine expertise du domaine, sur deux études de cas : la ville de Grenoble (France) pour l'évaluation de la vulnérabilité urbaine, et la région de Ferrara au nord de l'Italie, pour l'évaluation de l'endommagement suite au séisme qui a frappé cette région en 2012. Toutefois, bien qu'étant tout à fait utilisable par des utilisateurs non experts, la mise à disposition de l'application (client *et* serveur) n'est pas encore effective, à l'heure actuelle. Elle pose des questions relatives à qualité et validation des données (sa représentation, son évaluation, sa prise en compte dans les analyses et traitements futurs, etc.) qui méritent d'être étudiées de façon approfondie et, qui, en fait, sont celles posées par toute application adoptant une approche de participation citoyenne, de type VGI.

Remerciements

Cette recherche s'inscrit dans le projet ANR-Urbasis, du programme RiskNat de l'Agence Nationale de la recherche : ANR -09-Risk-009.

Bibliographie

Bossu R, Gilles S., Mazt-Roux G., Roussel F. 2011, Citizen Seismology or How to Involve the Public in Earthquake Response, in Miller D.M., Rivera J., Comparative Emergency Managment : Examining Global and régional Responses to Disasters, Auerbach / Taylor and Francis Publishers, 237-259.

Bossu R, Communiquer sur l'Aléa Sismique dans les Régions à Sismicité Modérée, 5èmes Rencontres Géographes et Assureurs face aux Risques Naturels, 07/04/2011

Calvi G, Pinho R, Magenes G, Bommer J, Restrepo-Velez L, Crowley H. Development of seismic vulnerability assessment methodologies over the past 30 years. Indian Society Journal of Earthquake Technology 2006; 43(3):75–104.

EMS98, 2001, "L'Echelle Macrosismique Européenne 1998", Conseil de l'Europe, Cahiers du Centre Européen de Géodynamique et de Séismologie, Vol. 19, p124.

Guéguen P., Michel C., LeCorre L. 2007. A simplified approach for vulnerability assessment in moderate-to-low seismic hazard regions: application to Grenoble (France), Bull Earthq. Engng., 4(3), 467-490.http://www.springerlink.com/content/14hmjgn512805344/

Mericskay B., Roche S., 2012, « Cartographie 2.0 : le grand public, producteur de contenus et de savoirs géographiques avec le web 2.0 », *Cybergeo : European Journal of Geography* [En ligne], URL : http://cybergeo.revues.org/24710 ; DOI : 10.4000/cybergeo.24710

GNDT, 1986 Instruzioni per la Compilazione de lla Sceda di Relivamento Esposizione e Vulnerabilità Sismica Degli Edifici. Gruppo Nationale per la Difesa dai Terremoti, Regione Emilia Romagna y Regione Toscan, Italy, inItalian.

Gires, J-F, Touya G., Quality Assessment of the French OpenStreetMap Dataset, Transactions in GIS, 14: 435-459.

Goodchild, M.F ;, 2007, Citizens as voluntary sensors : spatial data infrastructure in the word of web 2.0, International Journal of Data Infrastructure Research, vol 2, p 24_52.

Noucher M., 2009, La donnée géographique aux frontières des organisations. Approches socio-cognitive et systémique de son organisation Thèse de doctorat de l'Ecole Polytechnique Fédérale de Lausanne, 225 p

Noucher M. Coproduction of spatial data : from compromise to argumentative consensus. Conditions and participatory processes for producing spatial data together, In International Journal Of geomatics and Spatial Analysis, Special issue, 2011.

Ruitton-Allinieu A.M, The Crowdsourcing of geoinformation : data quaity and possible applictions, Alto university, School of Engineering, department of surveying, 2011.

RiskUE, An Advanced approach to earthquake risk scenraios with applications to différents european towns », Projet Européen, EVK4-CT-2000-00014, 2003.

Spence R, Lebrun B (ed) (2006) Earthquake scenarios for European cities – the risk-UE project. Bull, Earthquake Eng 4(4): special issue.

Windson Viana de Carvalho, *Mobilité et sensibilité au contexte pour la gestion de documents multimédias personnels : CoMMediA*, thèse de l'Université Joseph Fourier, 2010.

Proposition d'une méthode de génération de matrices de flux temporelles issues de l'activité de teléphones mobiles

Ana-Maria Olteanu-Raimond[1], Françoise Bahoken[2], Thomas Couronné[3], Zbigniew Smoreda[3]

1. Laboratoire COGIT, IGN,
4 Avenue Pasteur, 94165 Saint Mandé, France,
ana-maria.raimond@ign.fr

2. Université Paris-Est / IFSTTAR et UMR 8504 Géographie-Cités
14-20 Boulevard Newton, Cité Descartes, Champs-sur-Marne F-77447 Marne-la-Vallée Cedex 2. françoise.bahoken@ifsttar.fr

3. Orange Labs, department Sociology and Economics of Networks and Services,
38 Avenue du Général Leclerc, Issy-les-Moulineaux, France
(thomas.couronne, zbigniew.smoreda)@orange.com

RESUME.
Nous présentons une méthode de génération de matrices de flux Origine-Destination réalisée à partir de l'exploitation de traces spatio-temporelles de téléphones portables. La méthode s'appuie sur une procédure de filtrage temporel fondée sur un algorithme de décomposition empirique du signal (Empirical Mode Decomposition) ; elle tient compte de l'hétérogénéité des données liée à une distribution intra urbaine des antennes. La typologie des matrices temporelles générées selon l'activité réelle des utilisateurs (appels et SMS) permet la mise en évidence de profils d'activités de communication propres à des mobilités domicile-travail et cela, en fonction de tranches horaires spécifiques aux jours de la semaine ou du week-end.

MOTS-CLES : Mobilité individuelle, matrice Origine-Destination, données de téléphone portable, filtrage temporel, Empirical Mode Decomposition (EMD)

ABSTRACT.

A method of designing Origin-Destination flow matrices from the exploitation of spatiotemporal mobile phones data is proposed. The method is based on the Empirical Mode Decomposition (EMD) algorithm that provided temporal filtering related to the heterogeneity of data. The typology of the matrices generated from real user activity (calls and SMS) allows the identification of specific profiles of activities during commuting and that, depending on the specific time intervals, weekdays or the weekend.

KEYWORDS: Human migration, Origin-Destination matrix, Mobile Phone Data, Temporal Filtering, Empirical Mode Decomposition (EMD)

1. Introduction

Les données des matrices de flux Origine-Destination (OD) sont traditionnellement utilisées dans les domaines du transport ou de la socio-économie, de l'aménagement du territoire, de la géographie ou encore de la démographie. Elles résultent d'agrégations spatiotemporelles réalisées sur des données collectées au niveau de déplacements individuels – réalisés dans un espace et sur une période donnée – par un dispositif spécifique.

La construction des flux OD est réalisée à partir de données issues d'enquêtes par comptages ou d'enquêtes rétrospectives. Ces flux estiment les changements de lieux des individus, généralement appréhendés à l'échelle quotidienne. Ils restituent des motifs de déplacement observés depuis le lieu de résidence : domicile-travail (ou études), de loisirs, d'achats ou domicile-accompagnement ; ou encore de mode de transport. La matrice O/D concerne le nombre de déplacements d'un individu (la personne de référence du ménage, par exemple) ou de plusieurs (l'ensemble des membres du ménage) réalisés pendant une période $[t_1 – t_2]$, sur un espace $(S_1…S_n)$. En démo-géographie, cette mesure est effectuée sur un (ou plusieurs) individu(s) ou encore sur des événements, en l'occurrence des migrations, principalement dans le cadre d'enquêtes ou de recensements de population. Elle est réalisée de manière transversale, c'est-à-dire à une date (t) précise ou longitudinale, sur une période $[t_1 – t_2]$ et porte sur le (ou les) déplacement(s) réalisés par l'individu – considéré comme mobile – entre le jour de l'enquête et un jour particulier.

Ces déplacements individuels collectés de manière rétrospective font ensuite l'objet d'une série d'opérations statistiques d'anonymisation, de filtrages thématiques, temporels et spatiaux, avant d'être transmis sous une forme agrégée aux utilisateurs. Ces opérations conduisent à un formatage des données qui peut être appréhendé comme une dégradation d'autant plus importante que la taille des mailles est lâche (un pas de temps important, par exemple), leur niveau de généralisation étant fonction du dispositif de mesure. La nécessaire application de filtres lors de la mesure des flux entraîne la collecte d'une sélection de déplacements puisqu'il n'est pas possible de les saisir de manière exhaustive (Bavoux *et al.*, 2005). Les données qui en résultent sont ainsi, par essence, lacunaires (ou généralisées) soit parce qu'elles ne portent que sur un seul motif de déplacement (domicile-travail, par exemple), soit qu'elles ne concernent qu'une période (l'heure de pointe définie sur un intervalle), soit, enfin, qu'elles ne sont lisibles qu'à une seule échelle (les communes, par exemple). Notons que la construction des matrices OD est une opération onéreuse : sur le plan financier mais aussi technique et pratique. La mise en place d'un dispositif de collecte de données individuelle est lourde et chronophage, en raison des nombreux tableaux à renseigner. Ces données sont dans bien des cas datées ; la périodicité de leur renouvellement n'est pas fine, ce qui interdit certaines analyses des flux contemporains à leur mesure.

Les caractéristiques propres aux dispositifs classiques de collecte ont notamment conduit à développer des méthodes alternatives de mesure des déplacements individuels. Celles-ci recourent à des marqueurs à partir desquels sont reconstitués des déplacements individuels. C'est le cas, par exemple, des fichiers d'abonnements

et de désabonnements à un fournisseur d'énergie (électricité, eau), de poste ou de télécommunication, pour reconstituer des mobilités résidentielles ou quotidiennes. Notre propos porte ici sur l'utilisation de données issues de la téléphonie mobile pour saisir des mobilités individuelles quotidiennes. Pour cela, il est postule qu'à un mobile correspond un et un seul individu (ce qui en pratique n'est pas toujours le cas). Il convient de noter que le recours au téléphone mobile comme marqueur de déplacements individuels ne permet pas de disposer du cheminement exact de l'individu dans l'espace (x,y) et dans le temps (t) mais plutôt de reconstituer celui-ci. La trajectoire individuelle ainsi reconstituée permet alors de déduire des positions dans l'espace et dans le temps du mobile de l'individu dont il est question.

Nous nous intéressons à la génération de matrices Origine-Destination temporelles, selon l'activité de téléphones mobiles. Elles représentent des flux réalisés à l'échelle de pôles urbains. Nous présentons dans un premier temps, un état de l'art des travaux liés à la génération de matrices à partir de traces issues de téléphones portables. Nous proposons ensuite une méthode de constitution de matrices de flux dont l'originalité réside d'une part, dans la prise en compte de l'hétérogénéité des données et d'autre part, dans l'utilisation d'une procédure de filtrage temporel fondée sur un algorithme de décomposition empirique (*EMD - Empirical Mode Decomposition*), les aspects liés au filtrage spatial n'étant pas développés ici, le lecteur intéressé peut consulter la proposition méthodologique de Bahoken et Olteanu-Raimond (2013). Nous générons ainsi plusieurs matrices OD temporelles, selon l'activité téléphonique estimée. Le cas d'application porte sur les traces de téléphones mobiles, collectées sur la région de Picardie en 2007.

2. Etat de l'art

Les principales lacunes des méthodes classiques d'estimation des matrices de flux OD résident dans leur incomplétude – nonobstant le fait que les flux sont toujours lisibles à des niveaux de résolution temporelle et spatiale déterminés – mais aussi dans la lourdeur administrative, technique et financière de leur mise en œuvre. Ces méthodes nécessitent, en outre, le déploiement de traitements statistiques complexes qui ne permettent pas de les répéter à des fréquences élevées. Ce contexte limitant la production de données sur des mobilités spatiales à des niveaux de résolution fins, a conduit au développement de méthodes basées sur l'utilisation de données issues de la téléphonie mobile.

L'un des avantages méthodologiques de l'utilisation de ces données réside, en effet, dans la possibilité de générer des matrices de flux en quasi temps réel, à des niveaux de résolutions spatiales et temporelles relativement[1] fins. Les traces de téléphones sont susceptibles d'être collectées automatiquement et sur un large échantillon d'utilisateurs. Elles présentent toutefois l'inconvénient non négligeable

[1] A la précision près de la position de l'antenne réceptrice, laquelle est très importante en milieu intra-urbain.

de ne pas permettre la caractérisation sociodémographique des individus, et se doivent d'être étudiées conjointement à d'autres sources, notamment contextuelles.

Au cours des dernières années, les téléphones mobiles sont devenus l'un des principaux capteurs de la mobilité humaine à grande échelle. De nombreux modèles de déplacements ont étés proposés à deux niveaux conceptuels : celui de l'agrégat (Gonzalez *et al*., 2008, Sevtsuk et Ratti, 2010) et celui du déplacement individuel (Reades *et al.*, 2007 ; Ahas *et al*., 2008 ; Gonzalez *et al.*, 2008 ; Olteanu-Raimond *et al.*, 2012). Ces approches répondent à différents objectifs tels que, notamment, la détection d'activités pendant les phases de mobilité (Phithakkitnukoon *et al.*, 2010 ; Olteanu-Raimond *et al.*, 2012) ; la prédiction de la mobilité générale (Calabrese *et al.,* 2010 ; Song *et al.*, 2010 ; Schneider *et al.*, 2013) ou touristique (Asakura et Takamasa, 2007 ; Ahas *et al.*, 2008 ; Olteanu-Raimond *et al.*, 2011) ; l'analyse de la dynamique urbaine, notamment au regard de la formation de communautés spatialisées (Blondel *et al.*, 2008 ; Ratti *et al.*, 2010). Certains modèles ont également permis des estimations de trafic (Caceres *et al.*, 2007; Smith *et al.*, 2001).

De nouvelles approches ont également été mises en œuvre pour inférer des matrices de flux OD à partir de données issues de capteurs mobiles tels que des GPS ou des téléphones (Zhang *et al.*, 2010 ; Frias-Martinez *et al.*, 2012). La proposition de Friaz-Martinez *et al.* (2012) permet de générer des flux en utilisant directement les traces des mobiles pour reconstituer des flux ; celle de Byeong-Seok *et al.* (2005) s'appuie sur des durées de présence en un lieu défini par l'aire de couverture de l'antenne réceptrice pour générer le flux. L'approche de Calabrese *et al.* (2011) consiste, quant à elle, à identifier les trajets et les stations[2] qui composent chacune des trajectoires, puis à générer un flux pour chaque trajet. Quelle que soit la méthode, ces trajets sont ensuite agrégés par l'opérateur somme, en flux OD. Les valeurs des flux sont, par conséquent, dans une fenêtre temporelle donnée, définies *a priori*, et sur un espace défini par les lieux d'origine et de destination.

Notre proposition méthodologique s'inscrit dans la lignée des travaux de Caceres *et al*. (2007) et de Giannotti *et al*. (2011). Elle porte sur la première étape de la procédure, en l'occurrence le filtrage temporel, et consiste à estimer des matrices de flux OD sur une période déterminée par l'activité téléphonique. De plus, notre approche permet de prendre en compte les hétérogénéités des traces de téléphones mobiles (Olteanu-Raimond *et al.*, 2012).

3. Estimation des matrices de flux OD

L'objectif de cette section est de décrire la méthode de construction de matrice de flux OD mais, avant cela, nous présentons brièvement la structure spécifique du réseau de télécommunications.

[2] La station ne concerne pas seulement la localisation de l'antenne mais aussi son voisinage.

3.1. Les données de télécommunication

Les traces de téléphones portables sont produites dans l'espace formé par la structure d'un réseau de radio-télécommunication. Ce réseau correspond à une partition du territoire en zones de couverture radio appelées « cellules ». Chaque cellule est rattachée à une station de base (BTS) sur laquelle se trouvent une ou plusieurs antennes, chacune d'elle desservant une cellule.

Nous entendons par données de téléphones mobiles, les données de localisation enregistrées de manière automatique par le réseau de téléphonie pour des raisons de facturation ou d'estimation de la qualité du réseau. Plusieurs types de données sont distingués : les comptes rendus d'appel (CRA), les données de sondes et les données Wi-Fi. Dans ce papier, nous ne nous intéressons qu'aux CRA. Une présentation récente des autres types de données est disponible dans Smoreda *et al.* (2013).

Les données de localisation destinées aux CRA sont enregistrées de manière automatique lorsqu'un événement se produit entre le réseau et le terminal mobile : la communication (appel entrant ou sortant) et l'envoi (ou la réception) d'un SMS. A chaque événement est associé un enregistrement contenant l'identifiant de l'antenne a laquelle le terminal est connecté, l'horodatage, l'identifiant de la carte SIM (anonymé) ; le type d'événement (appel ou SMS) et la durée de communication.

3.2. La procédure générale de construction de flux OD

Soit une trajectoire spatiotemporelle T_k composée d'un ensemble de points consécutifs, notée

$$T_k = \{p_1, p_2, \ldots, p_i, \ldots, p_{n-1}, p_n\} \quad (1)$$

où : $p_i = (x_i, y_i, t_i)$ est un point caractérisé par ses coordonnées (x, y) à un moment t_i, $t_1 < t_2 < \ldots t_i < \ldots t_{n-1} < t_n$, et i= 1..n représente le nombre de points qui compose la trajectoire T_k.

Pour générer un flux, nous considérons les deux premiers points et les deux derniers points composant la trajectoire. Le premier et le dernier point sont utilisés pour tester s'ils appartiennent, respectivement, à une zone d'origine (O) et de destination (D), d'un ensemble O/D donné. Le deuxième et l'avant dernier point de la trajectoire sont utilisés pour vérifier s'ils sont exclusivement inclus dans une zone Voronoï voisine, respectivement d'une zone d'Origine et de Destination. Ainsi, étant donné que chaque point p_i est associé à une BTS, il est construit pour chacun d'entre eux une liste de BTS, notée V*BTS*, qui contient les BTS voisines du BTS auquel p_i appartient. La notion de voisinage est, en effet, très utile dans le cas des O/D discontinues. Elle permet de prendre en compte l'hétérogénéité des données, et plus précisément l'imprécision de la localisation et la cadence aléatoire des enregistrements. Il est, par exemple, possible qu'un utilisateur soit situé dans une zone d'origine (O) ou de destination (D), mais que le premier point de sa trajectoire soit localisé dans une zone voisine d'ordre 1 (parce que l'utilisateur se situe à la limite de deux zones Voronoï), alors que le second point de sa trajectoire est bien

localisé dans une zone O ou D. Sans cette notion de voisinage appliquée au deuxième point pour l'origine et à l'avant dernier point d'une trajectoire quelconque pour la destination, la génération d'un flux O/D ne serait pas possible (voir sur la Figure 1, la trajectoire T1).

Définition: Pour chaque trajectoire, T_k, il existe un flux entre deux zones (i,j), noté F(i,j) si les conditions suivantes sont respectées :

$$p_1 \subset i \ \textit{OU} \ (p_2 \subset i \ \textit{ET} \ p_1 \in V\textit{BTS} \text{ du point } p_2) \tag{2}$$

ET

$$p_n \subset j \ \textit{OU} \ (p_{n-1} \subset j \text{ ET } p_n \in V\textit{BTS} \text{ du point } p_{n-1}) \tag{3}$$

La Figure 1 illustre cette procédure d'estimation des flux à partir de l'exemple de deux trajectoires individuelles (T1 et T2) générées par deux mobiles, donc deux utilisateurs différents. Compte tenu de la définition précédente, la trajectoire T1 peut être appréhendée comme un flux car celle-ci s'inscrit d'une part, dans l'espace défini par la zone d'étude (les zones en grisé) ; elle comporte, d'autre part, au moins deux points qui sont situés dans des zones différentes, la différence entre ces zones étant liée à l'existence d'une limite qu'il convient de franchir pour lier une origine (i) à une destination (j). En revanche, les données de trajectoire T2 ne sauraient être considérées comme un flux, la condition (3) n'étant pas respectée : un seul des points qui composent sa trajectoire es inclue dans la zone d'étude.

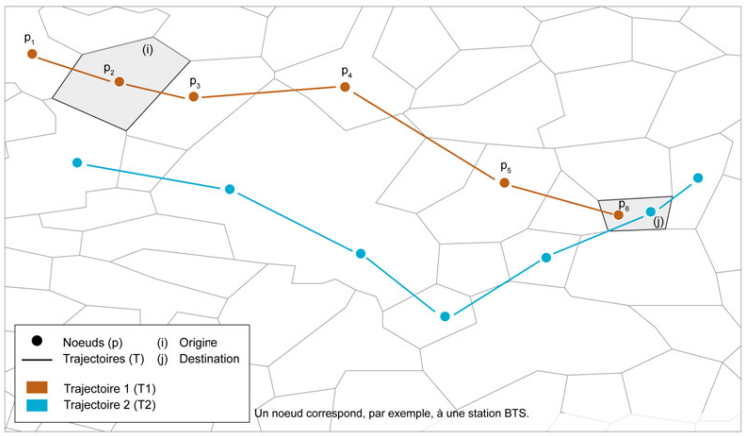


Figure 1. La définition spatiale des flux OD

La distinction entre les T1 et T2 est liée au filtrage spatiotemporel appliqué sur la mesure du nombre de déplacements entre les couples de lieux. Elle dépend du filtre temporel mais également de la résolution spatiale. Sur la Figure 1, si l'ensemble des zones (en blanc) avait été considéré comme faisant partie de la zone d'étude, les informations de T2 auraient participé à la génération des flux de zones à zones. Si la prise en compte du filtre territorial – qui intervient dans un second temps de la

procédure – ne pose pas de problèmes particuliers, du moins autre que des effets de zonages, il n'en est pas de même pour le filtre temporel.

L'une des difficultés rencontrées dans la procédure de discrétisation temporelle de trajectoires issues de traces de téléphones mobiles réside, en effet, dans la définition de sous-trajectoires spatiotemporelles qui décrivent le phénomène de mobilité, tout en tenant compte de l'usage qui est fait du téléphone portable, de l'activité téléphonique de l'individu en déplacement. Cette étape est rendue primordiale en raison de l'hétérogénéité de taille et de forme des cellules de la couverture du réseau de télécommunication, mais aussi de l'absence de continuité spatiale du maillage. La probabilité qu'un point p_i soit exclusivement inclus dans une zone donnée, à un moment donné, est de ce fait, très faible. De la même manière, la probabilité que plusieurs points soient inclus dans la même zone – ce qui revient à une mesure de flux intra-zonal – est très élevée dans les espaces intra-urbains ce qui ne conduit toutefois pas nécessairement à une meilleure définition des flux interzones.

4. Application

Nous testons notre approche de constitution de matrices de flux sur des données de téléphonie mobile (CRA) collectées sur la Région de Picardie. La méthode que nous présentons dans un second temps consiste à identifier des filtres temporels, ce qui revient à segmenter la série en sous-populations à l'aide de l'algorithme EMD. Nous analysons, enfin, les différentes matrices temporelles obtenues.

4.1. Présentation des données utilisées

Les données utilisées dans cette étude proviennent de trois sources : les CRA, le découpage territorial en pôles urbains et les données de recensement de l'INSEE.

Les données issues des CRA ont été enregistrées du 1er septembre au 15 octobre 2007 en Picardie, et dans quatre villes qui n'appartiennent pas à la région mais qui présentent une interaction importante avec celle-ci : Paris, Reims, Lille et Rouen. Afin de réduire le volume de données, ne sont sélectionnés que les individus dont la durée de présence dans la zone d'étude est d'au moins trois jours. Nous considérons qu'au delà de trois jours, la présence dans la zone d'étude n'est que ponctuelle, elle constitue de ce fait un bruit. L'échantillon ainsi obtenu porte sur 10 145 916 mobiles (utilisateurs). Ayant observé qu'en moyenne, 17% des individus ont au moins un événement par jour et que 50% en ont plus de quatre, nous faisons l'hypothèse que la pertinence du filtrage dépend du nombre d'événements localisés, l'événement étant défini comme le SMS ou l'appel, tous deux entrants ou sortants. Nous choisissons alors de ne retenir que les individus qui présentent au moins quatre positions temporelles, ce qui constitue un premier filtre des données CRA, selon le nombre d'événements de l'individu et par jour. La Figure 2 illustre la répartition temporelle des événements agrégés par heure, par jour et par semaine.

8

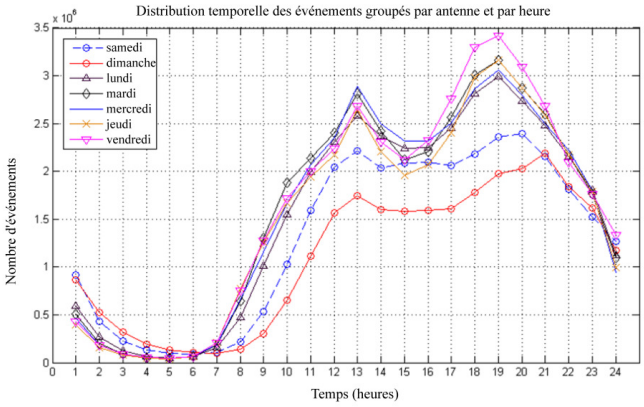


Figure 2. *Distribution temporelle des événements*

Il convient de noter les pics d'activité observés à 13h00 et à 19h00 en semaine et à 13h00, à 20h00 et à 21h00 les week-ends. Ils semblent liés soit à la pause-déjeuner, soit aux déplacements quotidiens des individus sélectionnés et que l'on suppose réguliers. Notons également que la distribution de ces événements est moins importante le week-end, sauf la nuit, entre 01h00 et 05h00 du matin.

Le deuxième jeu de données représente les pôles urbains de la Région de Picardie et ceux des quatre villes sélectionnées (voir Figure 3).

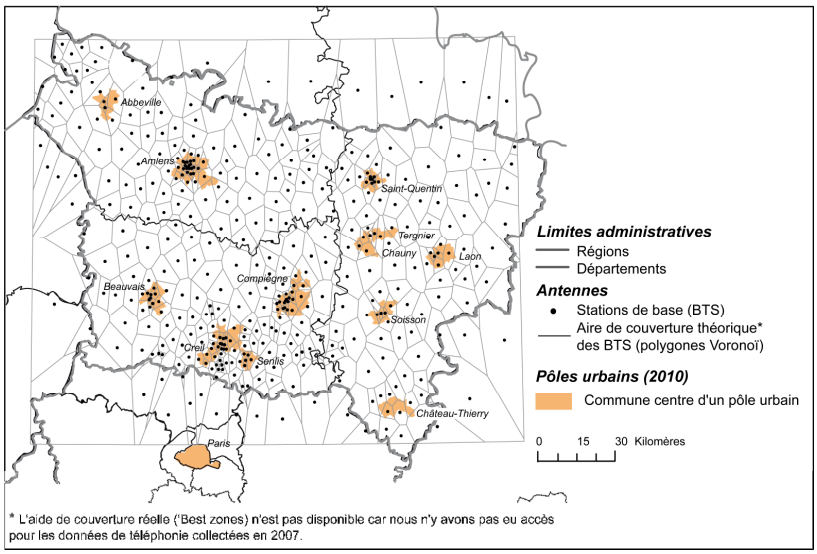


Figure 3. *Localisation des pôles et des stations de base (BTS)*

D'après l'INSEE (2007), le pôle urbain est une unité urbaine qui offre au moins 10 000 emplois et qui n'est pas située dans la couronne d'un autre pôle urbain.

Nous avons eu recours à deux types de données afin de valider notre analyse : le recensement de population et le fichier MOBPRO 2007 de l'INSEE. Seuls les déplacements domicile-travail effectués par les individus de 15 ans ou plus ayant un emploi ainsi que les déplacements domicile-étude[3] effectués par les individus de 2 ans ou plus, inscrits dans un établissement d'enseignement, ont été retenus.

4.2. L'estimation d'une matrice de flux OD issue de données des CRA

La validation de notre approche implique de comparer nos résultats avec des données de références telles que celles de l'INSEE (voir *supra*), afin de réduire les erreurs dues à l'hétérogénéité des informations disponibles dans les CRA. Nous avons ainsi limité le périmètre de la zone d'étude aux déplacements des individus qui résident en Picardie et qui travaillent dans l'un des seize pôles urbains sélectionnés dans le cadre de l'étude. Sur le plan thématique, ces déplacements correspondent à des navettes que nous avons tenté de détecter en identifiant dans un premier temps, les communes mises en relations dans le cadre de déplacements de type domicile-travail puis en calculant dans un second temps, la valeur du flux communes*communes de mobiles correspondant.

4.2.1 Détection de la commune domicile-travail

Il convient de garder en mémoire que les informations disponibles dans les CRA ne concernent que l'activité du téléphone, aucune caractéristique de l'individu porteur du mobile (de nature démographique, socio-économique ou professionnelle) n'est directement disponible. C'est la raison pour laquelle nous avons développé une méthode de détection des communes de domicile et de travail / étude pour chacun des individus. La méthode, appliquée à l'échelle nationale des communes[4] et pas uniquement pour la région de Picardie, consiste :

– pour la commune de résidence, à déterminer la station de base qui présente le nombre de connexions le plus important pendant la nuit : de 20h00 à 08h00 du matin, en semaine et le week-end ;

– pour la commune de travail, à déterminer la station de base qui présente le nombre de connexions le plus important pendant les heures de travail en journée : de 8h00 à 20h00 en semaine.

---

[3] Les déplacements domicile-étude ont été pris en compte car les porteurs de téléphones mobiles peuvent être des individus de tout âge, qui exercent ou non une activité professionnelle.

[4] Pour des raisons de confidentialité des données, les lieux de domicile et de travail ont été attribués à la commune.

	Communes de résidence	Communes de travail	Mobiles ayant la commune de résidence détectée	Mobiles ayant la commune de travail détectée
Effectifs	9054	9060	13 401 044	13 401 544

Tableau 1. *Effectifs des communes de résidences/travail et de mobiles impliquées dans l'analyse*

D'après le Tab. 1, notre méthode a permis de détecter la commune de résidence et de travail pour quelques 13 401 000 téléphones mobiles. Parmi les 36 700 communes françaises, seules 9054 communes de résidence et 9060 communes de travail sont concernées par nos données. Ce biais s'explique, en partie, par le fait qu'en zone rurale, l'aire de couverture des BTS englobe plusieurs communes tandis qu'en zone urbaine, la couverture des BTS est plus dense (voir Fig. 2).

Nous avons étudié la corrélation entre les fichiers 'communes de résidence' et 'communes de travail' ainsi générés, avec les données de recensement de l'INSEE. Il convient de noter que l'analyse de corrélation n'a été effectuée que sur les 9054 communes de résidence et les 9060 communes de travail ; les autres communes étant exclues. La forte corrélation obtenue, respectivement à 0,91 et de 0,9 nous a ensuite autorisés à extraire les données relatives à la Picardie, puis à identifier les individus qui travaillent dans l'un des pôles urbains sélectionnés. Nous avons ensuite estimé les flux domicile-travail entre ces pôles.

4.2.2 Estimation des flux domicile-travail

Etant donné que les données de flux de référence de l'INSEE ne représentent que les navettes domicile-travail / étude, nous avons appliqué, dans un premier temps, deux filtres temporels et cela, afin d'obtenir une matrice domicile-travail / étude. Ces deux filtres concernent les 'heures de pointe' en semaine, [06h00-10h00] pour le 'matin' et [16h00-20h00] pour le 'soir', ce qui nous a conduits à générer deux jeux de données pour chacun des jours de la semaine, soit quatorze fichiers.

Dans un second temps, la méthode de construction de matrice de flux OD présentée en section 3.2 a été appliquée à chacun de ces jeux de données pour générer deux matrices par jour, de flux inter et intra zonaux. Nous avons ensuite comparé les matrices de flux obtenues à partir de données CRA avec les données de référence de l'INSEE. Les coefficients de corrélation obtenus pour les matrices 'matin' et 'soir' sont très homogènes et ce, quel que soit le jour de la semaine : ils varient respectivement entre 0,92 et 0,93 et entre 0,93 et 0,94. Ces très fortes corrélations soulignent l'intérêt de l'utilisation de telles données issues de la téléphonie mobile pour reconstituer des matrices de flux. Elles montrent aussi l'adéquation des filtres temporels 'matin' et 'soir' retenus pour générer des flux domicile-travail à partir de données individuelles de déplacement. Nous avons

souhaité aller plus loin, en filtrant davantage les données afin de générer des matrices qui reflètent plus finement l'activité des individus.

4.2.3 Identification et analyse des effets des filtres temporels

Ce troisième temps de l'analyse porte sur la procédure de filtrage temporel et sur ses effets sur les données de flux disponibles dans les matrices. Nous avons vu, en section 4.2.2, qu'une matrice de flux est nécessairement construite sur une fenêtre temporelle donnée. Ces intervalles de temps sont généralement définis au regard des objectifs de l'analyse de flux particuliers tels que ceux qui se sont produits, par exemple, en heure de pointe ou en heure creuse ; les jours ouvrés ou le week-end. Cependant, il n'est pas toujours aisé d'identifier le filtre à retenir, celui des heures de pointes pour les week-ends, par exemple. L'enregistrement des données est, de plus, aléatoire ce qui nécessite une étape préalable de discrétisation des trajectoires. Pour cela, nous proposons l'identification de filtres temporels de manière automatique et sans *a priori*.

La méthode mise en œuvre est fondée sur le traitement du signal émis par l'activité des téléphones mobiles. Ce signal correspond à un échantillonnage par minute (et par jour) des événements des CRA mobile auquel est appliqué l'algorithme EMD (Huang *et al.*, 1998). L'EMD est une méthode de décomposition du signal optimisée pour l'analyse des signaux naturels, qui sont majoritairement non linéaires et non stationnaires. La méthode génère une somme de signaux sinusoïdaux organisée en composantes principales, comme la méthode d'Analyse en Composantes Principales (ACP) qui renseignent sur la dynamique et la stationnarité du signal global étudié. Chaque composante représente un phénomène distinct des autres, dont les conséquences peuvent n'apparaître qu'à un moment donné (une portion) du signal global; à noter que cette information de variabilité locale sur chaque composante indépendante n'apparait pas dans une transformée de fourrier ou dans une ACP. Les composantes peuvent ensuite être utilisées pour définir les fenêtres temporelles spécifiques aux données dont il est question.

La Figure 4 illustre les quatre signaux primitifs (ou composantes temporelles) générés par l'EMD sur nos données. Chaque composante explique une partie de la dynamique et de la stationnarité du signal ; la somme de toutes les composantes correspond au signal d'origine. De ce fait, plus l'amplitude d'une composante est élevée et plus sa contribution au signal principal est forte, à l'heure donnée.

Il est important de noter que les quatre composantes obtenues ne correspondent pas à des catégories de comportements individuels. Elles indiquent des variations de l'activité liée à l'utilisation du mobile au cours de la journée. L'interprétation de la Figure 4 s'inscrit, en effet, dans le cadre d'un raisonnement mathématique fondé sur le principe de décomposition d'un signal qui n'augure en rien des comportements individuels, l'objectif étant ici de segmenter les données puis de les regrouper. Ainsi, une composante qui présente six amplitudes maximales indique l'existence de six variations locales du signal d'origine ; ces variations ne correspondent donc pas à six événements de même type qui se seraient produits dans la journée.

12

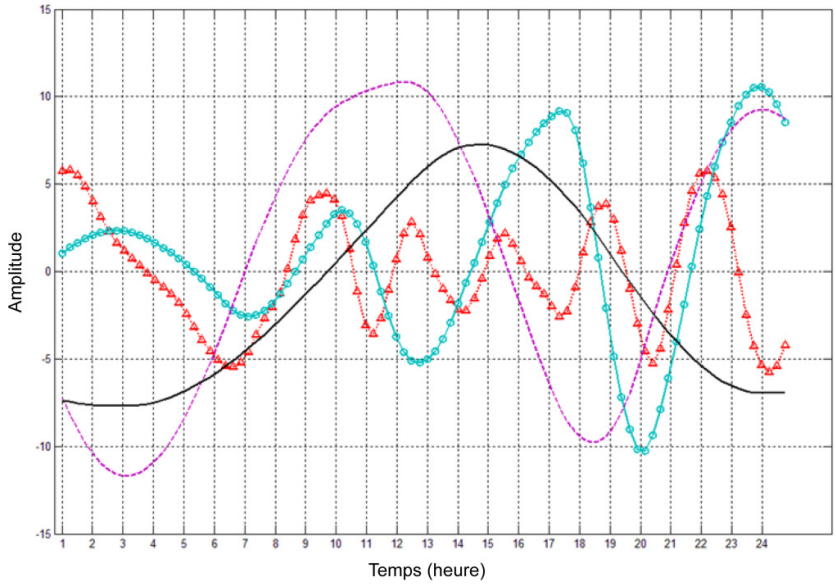


Figure 4. *Composantes primitives du signal représentant l'activité téléphonique*

En faisant l'hypothèse que chaque signal obtenu décrit un type de variation, nous avons sélectionné les points d'inflexions correspondant à une amplitude nulle. Notons que seuls les filtres temporels correspondant aux amplitudes positives et pendant la journée ont été utilisés dans notre analyse. Pour les quatre composantes nous avons obtenu les filtres temporels suivants :

– la composante n°1 (ligne continue-noir) présente une seule amplitude maximale, centrée sur 14h00 et correspondant à une distribution gaussienne ;

– la composante n°2 (ligne pointillée-magenta) est composée de deux amplitudes maximales, nous donnant un filtre pendant la journée : [07h00 – 15h30] ;

– la composante n°3 (ligne composée de cercles-cyan) présente quatre amplitudes maximales ; elle identifie les filtres : [08h30-11h30], [14h30-18h30] ;

– la composante n°4 (ligne composée de triangles - rouge) comporte pendant la journée, quant à elle, quatre amplitudes maximales à 10h00, 12h30, 15h30 et 18h30.

Le test du khi-deux appliqué aux huit matrices à des fins de comparaison, nous a conduits aux trois résultats suivants : (*i*) l'application d'un filtre similaire sur les cinq journées de travail, quelles qu'elles soient, ne conduit pas à une variation sensible du nombre[5] de flux inter pôles. Ce premier résultat souligne la robustesse

[5] C'est-à-dire du nombre de liens entre les communes, sachant qu'un lien correspond à un individu.

des fenêtres temporelles identifiées ; il indique aussi une certaine constance dans l'activité téléphonique spécifique aux jours de la semaine ; (*ii)* la variation des flux est, en revanche, sensible lorsqu'un filtre identique est appliqué sur les jours de la semaine et du week-end. Ce résultat suggère une variation de l'activité téléphonique spécifique à la fin de semaine ou, du moins, différente de celle qui est réalisée en semaine ; (*iii)* le nombre de flux dépend du filtre temporel utilisé pour sélectionner les trajectoires ; plus l'amplitude est grande, plus ce dernier augmente.

Ces trois résultats nous ont conduits à poser l'hypothèse selon laquelle il existe des profils d'activité téléphonique pendant les déplacements inter pôles et en fonction des jours (semaine, week-ends). Notre hypothèse semble être soutenue par les résultats de Couronné *et al.* (2011), qui ont montré qu'il y a une forte corrélation entre l'activité téléphonique (appel et SMS) et les déplacements des individus, ce qui signifie que plus les individus se déplacent, plus ils communiquent. Pour la valider, nous avons réalisé une classification des différentes matrices de flux temporelles, en utilisant la méthode des K-means. La typologie obtenue nous permet d'identifier les filtres temporels qui permettent d'obtenir des flux similaires et de regrouper les différentes matrices en sept classes :

– la classe n° 1 regroupe deux fenêtres temporelles : [8h30-10h30] et [14h30-18h30] définies du lundi au jeudi ; elle correspondrait à des flux réguliers domicile-travail / étude (les flux du vendredi, supposés atypiques, n'étant pas inclus) ;

– la classe n°2 correspond aux deux fenêtres suivantes : [10h00-18h00] et [7h00-15h00] du lundi au jeudi ;

– la classe n°3 regroupe les flux de [14h30-16h30] et [18h00-19h30] du lundi au mercredi, ainsi que le vendredi ; elle définirait soit les périodes creuses, soit les flux retours vers le domicile qui font l'objet d'une activité moindre de fin de la journée ;

– la classe n°4 identifie des flux qui diffèrent de tous les autres, sur la fenêtre temporelle [08h30-11h30], pour tous les jours de la semaine ; elle correspondrait aux flux en période de pointe, dans la matinée ;

– la classe n°5 identifie les flux de la fenêtre temporelle : [10h00-18h00] mais pour la seule journée de vendredi, étant complémentaire à ceux de la classe n°1;

– la classe n°6 regroupe les flux pour les fenêtres temporelles : [08h30-10h30] et [14h30-18h30] pour le vendredi, étant complémentaire à ceux de la classe n°2 ;

– la classe n°7 regroupe les matrices de flux pour les fenêtres temporelles : [11h30-13h30] pour la semaine et [14h30-16h30] pour la seule journée du jeudi ; cette classe correspondrait aux flux générés pendant la pause-déjeuner.

La représentation cartographique de ces sept classes de flux inter-pôles (voir la Figure 5) met en évidence deux axes de communications locales : Saint-Quentin, Tergnier, Chaumont, Soisson et Château-Thierry pour le premier ; entre Compiègne, Senlis et Paris pour le second. Les communications apparaissent, dans l'ensemble, plus nombreuses du lundi au jeudi, lors des déplacements réguliers domicile-travail que les autres jours.

14

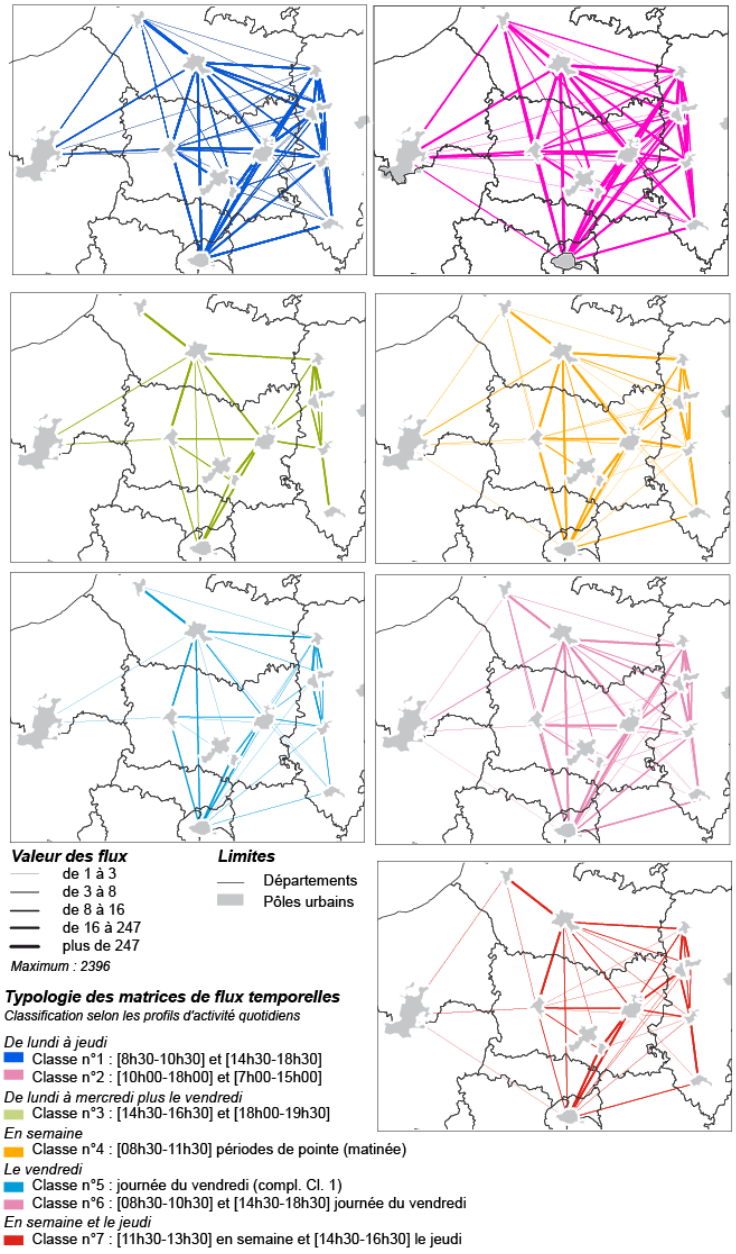


Figure 5. *Cartographie de la typologie temporelle des flux issue de l'activité quotidienne*

5. Conclusion

Cet article propose une nouvelle méthode d'estimation des matrices de flux OD temporelles adaptée aux données de téléphones mobiles. Cette matrice temporelle peut être considérée comme dynamique puisque nous pouvons la générer, à partir des données brutes, en agissant sur le filtre temporel.

L'objectif de notre proposition est double : de proposer une méthode de génération de matrice de flux OD dynamique qui permet, de prendre en compte les incomplétudes et les imprécisions des données de téléphonie et d'identifier les filtres temporels utilisés pour échantillonner les données de téléphonie mobile.

La méthode proposée s'appuie sur la relation topologique d'inclusion et sur la notion de voisinage. C'est, en effet, la notion de voisinage qui nous permet de prendre en compte l'imprécision de la localisation et la cadence aléatoire des enregistrements liées à l'utilisation du téléphone portable.

Pour ce qui est du filtrage temporel, nous proposons une méthode fondée sur une segmentation temporelle du signal. Elle est utile lorsque des informations – qui portent sur le choix des filtres temporels utiles pour la génération de matrices de flux OD dynamiques – ne sont pas disponibles *a priori* mais aussi pour affiner les filtres temporels, lorsque ces derniers existent. Par exemple, l'analyse du signal temporel nous a permis de définir les heures de pointes, selon l'activité téléphonique légèrement différentes de celles généralement utilisés dans la littérature pour les flux domicile-travail. Cette méthode basée sur une analyse EMD permet d'identifier les filtres temporels en fonction de l'activité téléphoniques des utilisateurs. Cette méthode a l'avantage de conserver le domaine temporel, c'est-à-dire les informations qui portent sur l'évolution temporelle du signal au cours de la journée, contrairement à une transformée de Fourrier où toute notion d'information temporelle est perdue.

L'approche proposée a été testée sur des données générées par des résidents de la Région de Picardie. Les résultats obtenus sont encourageants, ce d'autant plus qu'ils ont pu être comparés avec des données issues des recensements réalisés par l'INSEE. Cependant, l'approche proposée dans le cadre de cet article confirme aussi les limitations de l'inférence de flux OD, à partir des données issues de téléphones portables. En effet, la résolution de ces données dépend de la fréquence d'utilisation de la carte SIM, mais aussi de la structure du réseau de télécommunications. Notons aussi que si certaines de ces limites sont prises en compte dans notre méthode, d'autres n'ont pas pu l'être : c'est le cas des procédures de filtrages agrégatifs qui mériteraient d'être davantage analysées. Malgré ces limites, l'utilisation de la trace circulant sur un réseau radio-mobile sera très certainement de plus en plus exploitée, et considérée comme source complémentaire pour l'étude du comportement humain, de la mobilité individuelle et des flux agrégés OD.

Pour conclure, il est important de noter que la méthode proposée est généralisable. Elle peut être appliquée à tout type de données issues de téléphones mobiles : des flux directs ou de transit, de personnes (mobilités domicile-travail) ou de marchandises. L'intérêt des données de téléphonie mobile réside, en effet, dans

les possibilités d'analyses spatio-temporelles des flux à des niveaux de résolution fins, en choisissant différentes fenêtres temporelles (les périodes creuses, les périodes de pointe, l'heure par heure, pendant la semaine ou pendant les vacances). Elles permettent d'étudier des flux de mobilité de week-end, qui génèrent des trajectoires différentes de celles des jours ouvrables, ce qui n'est pas sans importance dans un contexte où les enquêtes classiques ne saisissent que rarement les flux réalisés pendant les week-ends.

Disposer d'une matrice de flux OD qui reflète raisonnablement la répartition temporelle des échanges est indispensable dans diverses applications de gestion des transports, depuis la planification à court terme des déplacements jusqu'au contrôle et à la gestion du trafic quotidien et cela, avec un coût d'accès peu élevé.

Remerciements:

Nous tenons à remercier Cezary Ziemlicki pour sa contribution aux prétraitements des données et les discussions riches que nous avons eues.

Bibliographie

Ahas R., Aasa A., Roose A., Mark, Silm S. (2008) Evaluating passive mobile positioning data for tourism surveys: An Estonian case study, *Tourism Management*, vol. 29, p. 469-486.

Asakura Y., Takamasa I. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice,* vol. 41, n° 7, p. 684-690.

Bahoken, F., Olteanu-Raimond, A-M., 2013, Designing Origin-Destination Flow Matrices from Individual Mobile Phone Paths. The effect of spatiotemporal filtering on flow measurement, *Proceeding of the International Cartographic Association Conference ICA'2013*, Dresdes, 15 p.

Bavoux J.-J., Beaucire F., Chapelon L., Zembri P. (2005), *Géographie des transports*, Paris, Armand Colin.

Blondel V., Guillaume JL., Lambiotte R., Lefebvre E. (2008) Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment,* vol. 2008, n°10, P10008.

Byeong-Seok, Y., Kyungsoo, C. (2005) Origin-destination estimation using cellular phone as information. *Journal of the Eastern Asia Society for Transportation Studies,* vol. 6, p. 2574–2588.

Calabrese F., Di Lorenzo G., Liu L., Ratti C. (2011) Estimating origin-destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area, *IEEE Pervasive Computing*, vol.10, n°4, p.36-44.

Caceres N., Wideberg J., Benitez F. (2007) Deriving origin-destination data from a mobile phone network, *Intelligent Transport Systems*, IET vol.1, n° 1, p. 15-26.

Couronné T., Smoreda Z., Olteanu-Raimond A.-M., (2011) Chatty mobiles: Individual mobility and communication pattern, *Actes du colloque Analysis of Mobile Phone Datasets and Networks*, MIT, Cambridge, MA

Frias-Martinez V., Soguero C., Frias-Martinez E. (2012) Estimation of Urban Commuting Patterns Using Cellphone Network Data, *Actes du colloque ACM SIGKDD Workshop on Urban Computing,* Beijing, China.

Giannotti F., Nanni M., Pedreschi D., Pinelli F., Renso C., Rinzivillo S., Trasarti R. (2011) Unveiling the complexity of human mobility by querying and mining massive trajectory data, *International Journal on Very Large Data Bases*, vol. 20, n° 5, p.695-719.

González MC., Hidalgo CA., Barabási AL. (2008) Understanding individual human mobility patterns, *Nature,* vol. 453, p. 779–782.

Huang N.E., Shen Z., Long S.R., Wu M.L., Shih H.H., Zheng Q., Yen N.C., Tung C.C., Liu H.H. (1998) The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, vol.454, n° 1971, p. 903-995.

Insee (2010) *Base communale du zonage en aires urbaine* 2010, http://www.insee.fr/fr/methodes/default.asp?page=zonages/aires_urbaines.htm

Insee (2010) *Fichiers de détail : Mobilités et migrations,* issus du recensement de population de 2009 : MOBPRO - Mobilités professionnelles des individus : déplacements commune de résidence / commune de travail. http://www.recensement.insee.fr/fichiersDetailTheme.action?codeTheme=MOB-MIG

Olteanu-Raimond AM., Couronne T., Fen-Chong J., Smoreda Z. (2012) Le Paris des visiteurs, qu'en disent les téléphones mobiles ?, *Revue Internationale de la Géomatique*, vol. 22, n°3, p.413-437.

Phithakkitnukoon S., Horanont T., Di Lorenzo G., Shibasaki R., Ratti C. (2010) Activity-Aware Map: Identifying human daily activity pattern using mobile phone data. HBU'10 *Proceedings of the First international conference on Human behavior understanding*, p. 14-25.

Reades J., Calabrese F., Sevtsuk A., Ratti C. (2007) Cellular Census: Explorations in Urban Data Collection, *IEEE Pervasive Computing,* vol. 6, p.30-38.

Schneider C.M., Belik, V., Couronné, T., Smoreda, Z., González, M.C. (2013) Unravelling daily human mobility motifs, *Journal of the Royal Society Interface*, vol. 10, n° 84: 20130246.

Sevtsuk A., Ratti C. (2010). Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks, *Journal of Urban Technology*, vol. 17, n° 1, p. 41–60.

Smith B, Pack M., Lovell D., Sermons M. (2001). Transportation management applications of anonymous mobile call sampling. *Actes du colloque 11th Annual Meeting of ITS America,* Miami, USA.

Smoreda Z., Olteanu-Raimond AM., Couronné T. (2013) Spatiotemporal data from mobile phones for personal mobility assessment, In: J. Zmud, M. Lee-Gosselin, M. Munizaga, JA. Carrasco (eds.) *Transport Survey Methods: Best Practice for Decision Making*. Emerald Group Publishing, London, p.745-767.

Song C., Qu Z., Blumm N., Barabasi A-L. (2010). Limits of Predictability in Human Mobility. *Science*, vol. 327, n° 5968, p. 1018-1021.

Zhang Y., Qin X. Dong S., Ran B. (2010). Daily O-D matrix estimation using cellular probe data. *Transportation Research Board 89th Annual Meeting,* Washington, USA.

Simulation des effets de structure des réseaux sur les polluants émis par l'automobile personnelle

Julie Prud'homme[1], Cyrille Genre-Grandpierre[1], Didier Josselin[1,2], Rémi Prud'homme[3]

1. UMR ESPACE 7300 CNRS
74, rue Louis pasteur, (F) 84029 AVIGNON cedex 1, France
julie.prud-homme@etd.univ-avignon.fr

2. Laboratoire d'Informatique d'Avignon (LIA), Université d'Avignon et des Pays de Vaucluse
339, chemin des Meinajaries, Agroparc, BP 1228, 84911 Avignon Cedex 9, France

3. TVC-LISP
36, avenue La Bruyère, 38100 Grenoble, France

RÉSUMÉ.
Les transports contribuent aux émissions de polluants atmosphériques, pour environ la moitié du CO_2 et des NOx émis. Cela génère beaucoup d'impacts directs et indirects sur la santé publique et sur l'effet de serre. Dans une large mesure, les études du domaine ont montré le rôle des modes de transport sur les émissions de polluants, en fonction des types de véhicules utilisés et des pratiques de conduite favorisant la consommation d'énergie à cause des changements de vitesse des véhicules (accélération, décélération). Toutefois, nous manquons de connaissances précises sur la localisation des polluants émis par les véhicules se déplaçant sur le réseau. Cet article propose une méthodologie pour estimer les émissions de polluants atmosphériques à l'échelle du tronçon de route et évaluer l'impact de la structure du réseau sur ces émissions. Nous simulons l'usage de la voiture personnelle selon différents scénarios, avec plusieurs types de réseaux routiers qui induisent divers plus courts chemins en temps. Les changements de vitesse et de forme du trajet des véhicules apparaissent avoir un effet non négligeable sur les émissions de polluants et sur leur distribution spatiale.

ABSTRACT.
Transportation strongly contributes to pollutant emissions in atmosphere (about 50% of the total CO_2 and NOx) with a lot of direct and indirect effects on public health and greenhouse effect. On a large scale, research has studied the role on pollutant emissions of transport modes, types of vehicle and driving practices which favour (or not) energy consumption due

to vehicle speed and acceleration/deceleration. However, we miss an accurate knowledge about the location of pollutant emission due to vehicle shortest paths induced by the road network structure. In this research, we propose a methodology which enables to estimate pollutant emissions at the road section scale. We simulate private car use according to several scenarios with different kinds of road networks. Tuning speed and shape influences both the level of emissions and on their spatial distribution.

MOTS-CLÉS : émission et cartographie de polluants atmosphériques, structure de réseaux routiers, usage de l'automobile personnelle, simulations spatiales.

KEYWORDS: Pollutant emission estimation and mapping, road network structure, private car use, spatial simulations.

1. Introduction

Dans la perspective du développement durable, la maîtrise voire la réduction des émissions de polluants ainsi que le contrôle de leur concentration spatiale dans les zones urbaines sont de première importance. Toutefois, il reste délicat de comprendre et de diagnostiquer les causes profondes de la pollution atmosphérique due à l'activité des transports, de nombreuses échelles devant être considérées.

Plusieurs programmes de recherche sur le transport et l'environnement ont abouti à une estimation correcte du rôle respectif des modes de transport, des types de véhicules, des températures des moteurs, des vitesses des véhicules et des pratiques de conduite (accélération, freinage) sur la consommation d'énergie et l'émission de polluants (ARTEMIS: Assessment and Reliability of Transport Emission Models and Inventory System, Boulter & McCrae, 2007 ; COPERT: Computer Program to Calculate Emissions from Road Transport program, Ntziachristos & Samaras, 2009). Ces éléments sont utiles et intéressants si l'on considère de grands territoires où l'information est agrégée. Toutefois, ils ne permettent pas de localiser les pollutions à une échelle fine (tronçon), échelle à laquelle les individus sont effectivement sensibles.

Il existe par ailleurs des modèles météorologiques basés sur des réseaux de capteurs distribués dans l'espace géographique. Ils sont utilisés notamment pour informer les conducteurs sur les vitesses appropriées (par exemple sur autoroute) afin de réduire les émissions de polluants lors des pics de chaleur estivale (Malherbe & Rouïl, 2003). Ces modèles interviennent à une échelle intermédiaire puisqu'ils sont basés sur les conditions météorologiques locales et de trafic. Ils interpolent les niveaux de pollution observés. Toutefois, la faible densité de mesures conjuguée à des méthodes d'interpolation qui peinent à prendre en compte les facteurs locaux (morphologie urbaine) peuvent générer des erreurs de prédiction.

Ainsi, un des enjeux majeurs est d'établir une relation fiable entre les polluants effectivement présents dans l'air et les sources qui les produisent. Bien qu'intéressante et prometteuse, la combinaison des deux types de modèles précédents ne permet pas d'estimer les émissions de polluants à une résolution spatiale suffisamment fine. Au-delà du diagnostic de pollution, il convient de rechercher les causalités, plus particulièrement l'impact de la structure des réseaux. De fait, il reste très difficile d'évaluer l'impact des déplacements réels, en incluant les vitesses, les accélérations et les décélérations des véhicules, dont on connaît l'impact déterminant sur les émissions (Joumard *et al.*, 1995a, Jensen, 1995; Hickman *et al.*, 1999).

Notre modèle, basé sur l'approche la plus usitée, implémentée dans COPERT 4 (Ntziachristos & Samaras, 2009), vise à estimer les niveaux d'émission de polluants à l'échelle du tronçon de route en intégrant les effets de la structure des réseaux sur les vitesses des véhicules. Appliqué à l'agglomération d'Avignon, le modèle intègre deux types de paramètres variables :

– des flux origine-destination qui simulent (ou non) une certaine attractivité du centre-ville (par exemple pour des motifs d'achat ou de travail) ;

– des structures de réseaux routiers, plus précisément la topologie et/ou la vitesse allouée aux différentes sections considérées.

Cette approche nous permet d'aborder les questions suivantes. Pouvons-nous identifier une structure de réseau (en termes de vitesse et de forme) favorable à la réduction des émissions de polluants ? Comment se distribuent spatialement les émissions produites par les véhicules personnels ? Y a-t-il des points de concentration des émissions ?

Suite à cette introduction du contexte et des objectifs, nous décrivons le cadre méthodologique pour estimer les émissions, les paramètres des scénarios simulés et la chaîne de traitement géomatique de l'information. Nous présentons ensuite des résultats synthétiques avant de conclure.

2. Méthodologie

2.1. Facteurs d'émission de polluants atmosphériques

Notre modèle, GREEN (Geographical Reasoning on Emission Estimations based on road Network shape), dérive du projet MEET (Methodologies for Estimation of Emissions from Transport, Hickman *et al.*, 2003). Il a déjà été utilisé avec succès pour évaluer les émissions des véhicules d'un transport à la demande (GREEN-DRT, Prud'homme *et al.*, 2011).

La nature et les quantités de polluants produits par les véhicules dépendent de plusieurs facteurs :

– le *type et la cylindrée des véhicules* (citadine, berline, monospace, minibus, bus, autobus) ; dans notre cas nous ne considérons que des véhicules de type monospace ou berline ;

– la *vétusté du véhicule et sa capacité à réduire les émissions*, sur la base des normes fixées par l'Union Européenne (cf. Tableau 1) ; dans notre cas le véhicule standard est neuf et émet des valeurs moyennes de polluants qui se rapprochent des niveaux théoriques estimés par les constructeurs (par exemple, pour le CO_2, 170 g/km en essence et 140 g/km en gasoil) ;

– le *type de carburant* ; essence et gasoil sont comparés dans notre étude ;

– la *température ambiante*, fixée ici à 15°C, est celle du moteur ; lorsqu'un moteur est froid, il surémet ; au-delà de 6 km le véhicule est considéré comme chaud et fonctionne de manière optimale (Joumard *et al.*, 1995b) ;

– la *vitesse du véhicule* qui dépend du réseau ; la relation entre la vitesse et l'émission n'est pas linéaire ; plus précisément la consommation est plus importante dans les faibles vitesses et dans les grandes vitesses; dans l'intervalle 50-70 km/h ; les émissions rapportées au kilomètre sont minimales ; nous considérons que les véhicules roulent toujours à la vitesse maximale autorisée sur le tronçon ;

– les *accélérations et les décélérations du véhicule* (Int Panis *et al*., 2006) sous l'effet des intersections, des stops et des feux (Brundell-Freij & Ericsson, 2005) ; dans notre cas, nous ne disposons pas des informations sur les feux ni les stops ou les céder-le-passage, mais nous tenons compte de tous les changements de vitesse ; l'intensité d'accélération ou de décélération des voitures est fixée à 0.5 m/s; ainsi, plus les itinéraires des véhicules sont hétérogènes en termes de vitesse, plus les émissions de polluants sont importantes ;

– le *rythme de conduite associé au conducteur* ; nous ne considérons qu'un seul comportement moyen qui atteint la vitesse maximum autorisée sans à-coups et pratique un freinage maîtrisé et adapté (André *et al.*, 2006) ;

– les *distances parcourues* ; les véhicules sont censés parcourir les plus courts chemins en temps de leur origine à leur destination ; en effet, nous postulons que le temps est le critère référent pour le choix de l'auto-mobilité personnelle ;

Le tableau 1 précise les niveaux d'émission de gaz (hors micro-particules) selon le carburant, la température et l'action du moteur.

Moteur	Essence		Diesel	
	CO_2 en g/km	**NOx** en g/km	**CO_2** en g/km	**NOx** en g/km
Chaud	175.972	0.0385	141.656	0.472
Froid	308.963	0.68	400.085	0.203
Accélérant	+20.889	+0.009	+108.994	+0.134

Tableau 1. Emission et surémission (+) de polluants selon le fonctionnement du moteur et le type de carburant

2.2. Modèle global d'évaluation des émissions de polluants

L'originalité de GREEN réside dans la prise en compte, autant que faire se peut, de la structure fine du réseau routier. Une fois que nous connaissons les plus courts chemins pratiqués par les véhicules et la séquence des tronçons de route, nous appliquons un calcul d'émission (en grammes) pour chaque section E_{sec}. L'émission globale sur l'ensemble du territoire est la somme E_{tot} des émissions locales : $E_{tot}=\sum E_{sec}$ (1). Pour chaque section, l'émission locale d'un polluant k est constituée de l'émission à chaud E, à froid E_{cold}, et en accélération E_{acc} : $E_{sec}=\sum_{k} E+E_{cold}+E_{acc}$ (2).

L'émission à chaud (Ntziachristos & Samaras, 2009) est calculée comme suit : $E=\sum D_{V,j}\times e_{V,j,k}$ (3) où $D_{V,j}$ est la longueur de la section j parcourue par le véhicule V, $e_{V,j,k}$ est le facteur d'émission du polluant k correspondant à la vitesse autorisée sur la section j pour le véhicule V.

L'émission à froid est estimée (Hickman *et al.*, 2003) par : $E_{cold}=\omega\left[f(V)+g(T)-1\right]$ (4) où ω est la référence d'émissions en excès (à 20°C et 20 km/h), $f(V)$ est la correction de la valeur de référence en fonction de la vitesse, $g(T)$ est la correction selon la température ambiante.

L'émission en accélération (Int Panis *et al.*, 2006) est évaluée par la fonction : $E_{acc}=max\left[E_0, f1+f2.v+f3.v^2+f4.a+f5.a^2+f6.v.a\right]$ (5) où f sont des coefficients spécifiques à chaque véhicule et à chaque polluant, v est la vitesse instantanée en km/h, a est l'accélération instantanée, en m.s^{-2}.

Parmi les facteurs d'émission identifiés précédemment, la vitesse, le rythme de conduite et les distances parcourues déterminent l'émission à chaud, en lien avec les caractéristiques du véhicule et le carburant utilisé. Les changements de vitesse et la température ambiante modifient les valeurs de référence, notamment de l'émission à froid et en accélération.

2.3. Scénarios

Les scénarios concernent l'agglomération d'Avignon regroupant plus de 450 000 habitants dans un rayon de 20 km autour du centre ville (figure 1). Les paramètres variables des scénarios sont les flux en origines destinations des véhicules privés, ainsi que les différentes structures de réseaux testées. Certains paramètres ont été fixés pour l'ensemble des simulations. Le réseau provient de la BD Géoroute de l'IGN composée de 27 700 arcs, incluant les sens de circulation et les carrefours. 1000 véhicules privés simulés se déplacent sur le réseau pendant une journée. La simulation des origines-destinations a été réalisée à partir de 59 points virtuels localisés sur le territoire, plus concentrés en centre-ville.

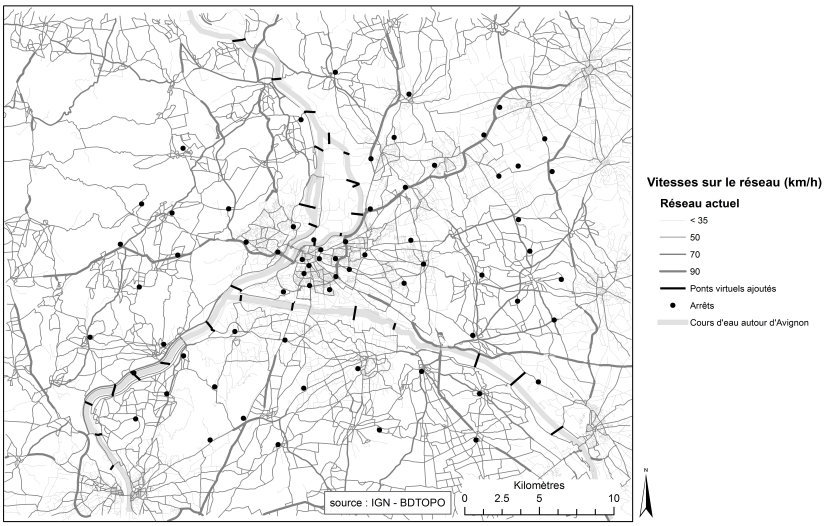


Figure 1. Le réseau actuel de l'agglomération d'Avignon utilisé pour les simulations, avec les vitesses autorisées, les arrêts et (dans certains scénarios) les ponts virtuels ajoutés

Deux types de structure de flux sont considérés. Le premier est aléatoire et pourrait correspondre à des mobilités de loisir pendant le week-end. Dans ce cas, les flux (généralement faibles dans la pratique) ne sont pas prédictibles comme ceux des

navetteurs. Ce cas de figure correspond également à un « tube témoin » théorique car le réseau peut être potentiellement exploré dans son entièreté. Chaque point a ainsi la même probabilité d'être sélectionné en origine ou en destination. Le second type simule des flux convergeant des 59 points possibles vers une sélection de 17 autres concentrés en centre-ville. Ce scénario mime les mouvements des navetteurs dans la semaine pour des motifs scolaires et professionnels. Des corridors de mobilité et des points de polarisation apparaissent tout au long de la journée. Dans cet article, nous ne prenons pas en compte les pics de mobilités du matin ou du soir, ni les congestions sur le trafic. Pour une simulation donnée, les émissions sont sommées sur la journée.

12 % des routes du réseau *actuel* de l'agglomération d'Avignon permettent de rouler à plus de 50 km/h, notamment celles qui convergent vers le centre-ville. Les vitesses autorisées sur les différentes sections ont été fixées conformément à la législation française, même si elles peuvent différer des pratiques de conduite en France. Ne connaissant pas dans le détail les écarts aux limites de vitesse, nous gardons à l'esprit que les simulations du réseau « actuel » sous-estiment dans leur globalité les émissions. Par ailleurs nous ne disposons pas des informations sur les congestions dues au trafic. Sur la base du réseau actuel, nous avons construit trois nouveaux réseaux théoriques avec des vitesses différentes selon les types de sections (cf. Tableau 2). L'analyse s'appuie sur les différences de vitesses modélisées.

Actuel	**Rapide**	**Lent**	**Homogène**
90	110	80	80
70	90	65	75
50	60	45	60
35	45	35	45
25	35	25	30

Tableau 2. Vitesses (km/h) associées aux sections de 4 types de réseaux selon différents scénarios

Le réseau *rapide* simule une augmentation de toutes les vitesses. Cela pourrait correspondre au cas où tous les conducteurs dépasseraient systématiquement les vitesses autorisées. À l'opposé, le réseau *lent* modélise des vitesses systématiquement plus faibles qu'actuellement. Cette tendance est d'actualité dans la planification urbaine, car elle permet d'améliorer la sécurité et la compétitivité des modes de transport publics (bus sur voie réservée) ou alternatifs (vélo). Le cas du réseau *homogène* est particulier puisqu'il réduit globalement la variabilité de la vitesse entre les différents types de tronçons. De ce fait, il tend à diminuer les concentrations spatiales de véhicules dans des sections à grande vitesse, les congestions, les concentrations de bruit et de pollution (De Vlieger *et al.*, 2000).

Par ailleurs, seuls 5 ponts permettent de traverser le Rhône ou la Durance pour accéder aux différentes parties de l'agglomération, ce qui contribue à concentrer les flux. Nous avons donc créé un ensemble de ponts virtuels qui augmentent sensiblement la connectivité du réseau et l'accessibilité, et permettent de fluidifier virtuellement le trafic. Chaque scénario de vitesse est simulé avec ou sans ajout des nouveaux ponts. Au total, nous étudions 8 réseaux croisés avec 2 configurations de flux (aléatoires ou en convergence), pour 2 types de véhicules (essence ou diesel) en calculant 3 types d'émission (CO_2, NOx et micro-particules pour le diesel notamment).

2.4. Chaîne de traitement de l'information géographique

L'ensemble des informations est traité au sein d'un système d'information géographique (ArcGIS) combiné à différents programmes dédiés tels que Common LISP (gestion des échanges de fichiers et calculs des émissions), LispStat (création des fichiers de mobilité en origines et destinations) et Access (système de gestion des bases de données). La figure 2 décrit les différentes fonctions et le cheminement de l'information. Le système d'information géographique permet de traiter les plus courts chemins sur la base du réseau calibré et de cartographier les émissions calculées à l'échelle du tronçon de route.

Nous avons donc simulé 16 scénarios d'usage de véhicules personnels. Nous disposons en sortie des informations suivantes concernant les émissions : (i) des estimations, prenant en compte les démarrages à froid, les accélérations et décélérations des véhicules, calculés soit sur l'ensemble du territoire, soit à l'échelle des tronçons, (ii) d'autres critères complémentaires, tels que la somme des distances parcourues, les vitesses moyennes, ainsi qu'un indice d'homogénéité local H_n, calculé sur le graphe.

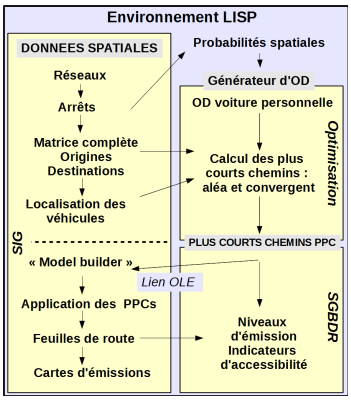


Figure 2. Chaîne de traitement de l'information géographique du modèle GREEN

L'*indice d'homogénéité* H_n prend en compte la structure topologique locale des plus courts chemins en temps pratiqués en théorie par les véhicules. Il fournit une estimation de la distribution spatiale de la vitesse pour chaque itinéraire n d'un véhicule. Il s'agit du ratio entre le nombre de carrefours avec rupture de vitesse j_{break} et le nombre total de carrefours j_{total} . Plus cet indice est élevé, plus le réseau est hétérogène en termes de vitesse. Il est calculé comme suit : $H_n = \frac{j_{break}}{j_{total}} \times 100$ (6).

3. Résultats

La méthode de simulation a été calibrée statistiquement. Nous avons en effet calculé un intervalle de confiance à partir de 50 simulations aléatoires par scénario. Les intervalles de confiance sont restés suffisamment serrés et discriminants entre les scénarios pour que nous puissions fournir les résultats sans incertitude. Dans cet article, nous ne présentons qu'une partie des résultats, chaque simulation comportant 1000 véhicules et représentant assez fidèlement le scénario simulé. Pour plus de détails sur les résultats comparatifs des émissions, voir Prud'homme *et al.*, 2013, dont cet article s'est largement inspiré.

3.1. Le réseau homogène contribue de façon significative à la réduction des émissions

Les valeurs des émissions du réseau *homogène* sont les plus faibles. Voici la position de chaque type de réseau par ordre décroissant d'émission, quel que soit le polluant considéré : *rapide>actuel>lent>homogène.* On constate par ailleurs que les réseaux *lent* et *homogène* sont relativement similaires, alors que les autres réseaux montrent des comportements particuliers.

Pour ce qui est du CO_2, indicateur privilégié de l'effet de serre, on constate des écarts de valeurs différents selon que les véhicules utilisent l'essence ou le gasoil. Pour l'essence, le réseau *homogène* permet de sauver un peu moins d'une demi-tonne de CO_2, soit environ 4 % des émissions totales du réseau *rapide*. Pour un véhicule privé familial émettant 150 g/km, cela constitue près de 2400 km parcourus, représentant 14 % de l'ensemble de la distance parcourue par les véhicules. Pour les véhicules diesel, ce constat est similaire mais encore plus marqué, puisque la différence entre le réseau *homogène* et le réseau *rapide* (plus d'une demi-tonne de gaz) permet potentiellement d'économiser 8000 kilomètres parcourus. Globalement, l'usage du gasoil n'est pas favorable à la réduction des émissions de polluants atmosphériques. Les résultats observés vont dans le même sens pour les émissions de NOx.

Y a-t-il un effet des patrons spatiaux de flux (aléatoire *vs* convergent) sur les émissions ? Les tendances observées précédemment se confirment dans une moindre mesure, les flux aléatoires exprimant davantage de différences (entre 2 et 11%) que les flux convergents (entre 0,1 et 6,9%). Tout comme l'ajout de ponts virtuels, qui aurait dû augmenter significativement la connectivité du réseau, l'attractivité du centre a tendance à réduire les écarts entre les différents scénarios simulés. En effet, beaucoup d'itinéraires des véhicules se ressemblent, car ces derniers utilisent les mêmes types de voiries pour rejoindre le centre.

3.2. Le degré d'homogénéité de la vitesse explique le niveau d'émission

Sur un réseau ayant des vitesses plus hétérogènes, les automobilistes ont tendance à réaliser des trajets plus longs en faisant des détours pour accéder le plus vite possible aux voies les plus rapides. De ce fait, le réseau homogène permet de réduire les distances parcourues et donc les émissions. Mais cela n'est pas la seule explication de la réduction des émissions. En effet, on peut voir sur la figure 3, le rôle de la distribution (statistique et spatiale) des changements de vitesse le long des plus courts chemins grâce à l'indice *H*, particulièrement élevé sur les réseaux *rapide* et *actuel* (figure 3).

L'effet de l'hétérogénéité des vitesses est encore plus visible par l'observation des surémissions liées aux accélérations et aux décélérations des véhicules. Pour les véhicules diesel, qui émettent davantage aux changements de vitesse, ces émissions sont pratiquement doublées entre un réseau *homogène* et un réseau *rapide* ! Par rapport au réseau *homogène*, elles représentent une augmentation de 40 % pour le réseau *actuel*, et même 22 % pour le réseau *lent.* De plus, les véhicules diesel émettent des particules dangereuses dans l'atmosphère (PM), dont l'impact sur certaines maladies comme le cancer est aujourd'hui avéré.

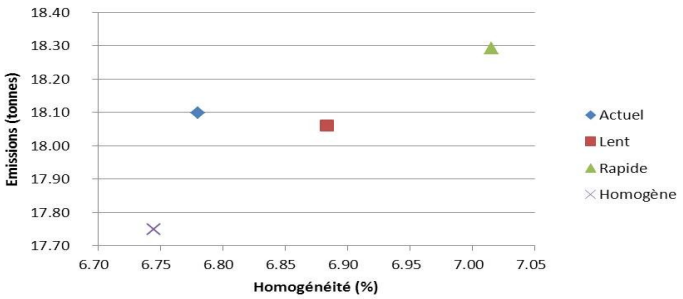


Figure 3. Position des différents types de réseaux selon leur niveau d'émission globale et l'homogénéité des vitesses (H)

Toutefois, même si les réseaux rapides sont défavorables en termes d'émission de polluants atmosphériques, ils apportent d'autres avantages aux ménages. En effet, ils

réduisent de 13 à 35 % les temps de parcours des véhicules personnels par rapport aux autres types de réseaux. Ce gain de temps est largement dû à une augmentation moyenne de la vitesse sur les réseaux plus rapides (autoroutes ou doubles voies par exemple). Plus exactement, la structure des réseaux rapides engendre une meilleure accessibilité, grâce à une forte hiérarchie des vitesses (Foltête *et al.*, 2012). La hiérarchie des vitesses est en effet observée le long des réseaux par une augmentation graduelle des vitesses autorisées et pratiquées, en partant des centres urbains où les vitesses sont lentes, vers la périphérie où les boulevards -voire les voies rapides- facilitent les déplacements à plus grande vitesse. Dans cette perspective, certaines politiques de planification urbaine s'orientent d'ailleurs vers une réduction de la hiérarchie des vitesses (Dupuy, 1999 ; Genre-Grandpierre, 2011). Dans une certaine mesure, les niveaux d'émission de polluants peuvent résulter d'un arbitrage entre les temps de déplacement des ménages et la réduction, voire l'homogénéisation généralisée, des vitesses sur le réseau.

3.3. Répartition spatiale et points chauds des émissions de polluants

Une analyse globale des polluants émis par les voitures personnelles reste insuffisante pour évaluer l'impact local des pratiques de mobilité par les ménages utilisant leur(s) voiture(s) personnelle(s). Notre méthodologie permet de répondre à ce besoin. Comme on pouvait s'y attendre, plus la hiérarchie des vitesses est élevée, plus il apparaît des points de concentration des émissions, à cause de la centralité d'intermédiarité (Freeman, 1977). En d'autres termes, la hiérarchie des vitesses n'est pas favorable à une dispersion des flux, comme l'explique la théorie du mouvement naturel (Penn *et al.*, 1998). En conséquence, des vitesses élevées induisent des concentrations périphériques de pollution (cf. figure 4), notamment sur des corridors ou des “points chauds” (cas pour les flux aléatoires).

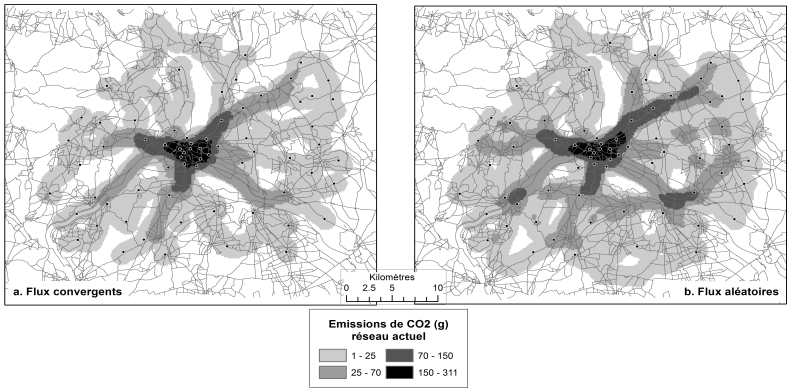


Figure 4. Emissions de CO_2 (en grammes) sur le réseau actuel de l'agglomération d'Avignon (a. flux convergents; b. flux aléatoires)

L'ajout de ponts contribue également, dans une certaine mesure, à réduire la concentration locale et la variance des émissions de polluants (pour environ 10 %), comme l'illustre la figure 5 où les différences entre les réseaux *rapide* et *homogène* se sont estompées dans le cas d'un scénario de flux aléatoires. Par ailleurs, même si elle ne correspond pas à des volumes très importants de gaz émis localement, la différence d'émission de CO_2 entre ces deux réseaux est très marquée. Son empreinte dessine des canyons ou des poches concentrant les polluants atmosphériques, au centre-ville comme en périphérie.

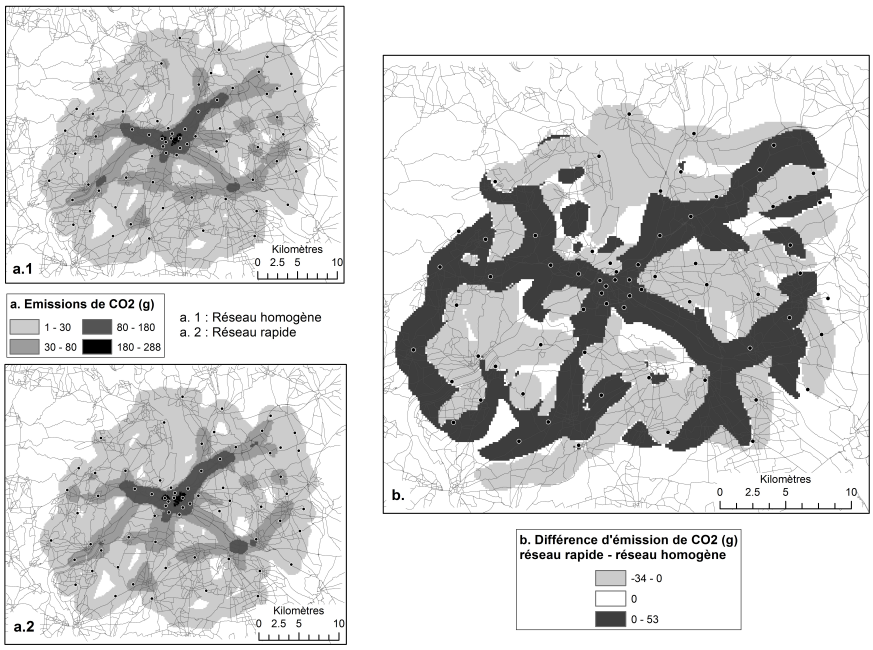


Figure 5. Emission de CO_2 (g) - a.1 réseau homogène; a.2 réseau rapide; b. différence d'émission de CO_2 (rapide – homogène)

Conclusion

L'approche proposée permet d'estimer les émissions de polluants atmosphériques à une résolution spatiale fine. Qui plus est, notre modèle intègre une grande partie des connaissances récentes des modes d'émission liés au fonctionnement des véhicules. À ce titre, cette recherche participe à l'établissement du lien entre l'impact des structures de réseaux et leurs capacités intrinsèques d'émission, avec la localisation observable des polluants atmosphériques débouchant sur des préconisations liées à la qualité de l'air.

Les résultats obtenus militent en faveur d'une homogénéisation des vitesses sur les réseaux, dont les effets s'avèrent non seulement bénéfiques sur la qualité de l'air respiré, mais également sur la sécurité, la gêne due au bruit. Par ailleurs, l'homogénéisation des vitesses tend à réduire les effets de coupures spatiales (Héran, 2011), favorisant ainsi le développement des modes doux (vélo, marche à pied).

Toutefois, prôner une réduction et une homogénéisation des vitesses aura un impact (certes modéré) sur l'accessibilité du territoire en temps. C'est la contrepartie à accepter pour le développement de réseaux plus favorables à la qualité de vie urbaine. Tout l'enjeu consiste à montrer que le gain, pour la santé notamment, est supérieur à la contrainte imposée par ce type de réseau, avec la difficulté que les réductions de polluants atmosphériques générées, sans être négligeables, ne sont pour autant pas extrêmement importantes à l'échelle d'une agglomération moyenne.

Ces résultats pourront être avantageusement complétés, par exemple en utilisant des modélisations multi-agents (Banos & Genre-Grandpierre, 2012), par des simulations intégrant l'interaction et les comportements des conducteurs, ainsi que l'occurrence de congestions du trafic générées par les pics de flux à certains moments de la journée. Sur d'autres sites, ils pourront être comparés à des flux et des trafics observés, voire à des pollutions mesurées, dès lors que les données existent et peuvent être traitées à l'échelle du tronçon.

Références

André M., Rapone M., & Joumard R. (2006). Analysis of the cars pollutant emissions as regards driving cycles and kinematic parameters. *INRETS-LTE*.

Banos A., & Genre-Grandpierre C. (2012). Towards New Metrics for Urban Road Networks: Some Preliminary Evidence from Agent-Based Simulations. In *Agent-Based Models of Geographical Systems*. Springer.

Boulter P., & McCrae I. (2007). ARTEMIS: Assessment and Reliability of Transport Emission Models and Inventory Systems-final report. *TRL Published Project Report.*

Brundell-Freij K., & Ericsson E. (2005). Influence of street characteristics, driver category and car performance on urban driving patterns. *Transportation Research Part D: Transport and Environment*, 10(3), 213-229.

De Vlieger I., De Keukeleere D., & Kretzschmar J. G. (2000). Environmental effects of driving behaviour and congestion related to passenger cars. *Atmospheric Environment*, 34(27), 4649–4655.

Dupuy G. (1999). *La dépendance automobile: symptômes, analyses, diagnostic, traitements.* Paris: Anthropos.

Foltête J.-C., Genre-Grandpierre C., & Josselin D. (2011). Impacts of Road Networks on urban mobility. In M. Thériault & F. D. Rosiers (Éd.), *Modeling urban dynamics: mobility, accessibility and real estate value. London* : John Wiley and Sons: ISTE Ltd., ; Hoboken, NJ.

Freeman L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1), 35-41.

Genre-Grandpierre C. (2011). L'efficacité des déplacements automobiles en milieu urbain. In *Données urbaines.* Economica.

Héran F. (2011). *La ville morcelée: effets de coupure en milieu urbain.* Paris: Economica.

Hickman J., Hassel D., Joumard R., Samaras Z., & Sorenson S. (2003). Environnement, energy and transport. Written material extended version by EU-PORTAL, transport teaching material, based on MEET and COST program.

Int Panis L., Broekx S., & Liu R. (2006). Modelling instantaneous traffic emission and the influence of traffic speed limits. *Science of The Total Environment*, 371(1–3), 270-285. doi:10.1016/j.scitotenv.2006.08.017

Jensen S. S. (1995). Driving patterns and emissions from different types of roads. *Science of The Total Environment*, 169(1–3), 123-128.

Joumard R., Jost P., Hickman J., & Hassel D. (1995 - a). Hot passenger car emissions modelling as a function of instantaneous speed and acceleration. *Science of The Total Environment*, 169(1-3), 167-174.

Joumard R., Vidon R., Pruvost C., Tassel P., & De Soete G. (1995 - b). Evolution des émissions de polluants des voitures particulières lors du départ à froid. *Science of The Total Environment*, 169(1-3), 185-193.

Malherbe L., & Rouil L. (2003). Méthodes de représentation de la qualité de l'air Guide d'utilisation des méthodes de la géostatistique linéaire (p. 70). Laboratoire Central de Surveillance de la Qualité de l'Air - INERIS.

Ntziachristos L., & Samaras Z. (2009). Exhaust emissions from road transport. EMEP/EEA emission inventory guidebook - COPERT.

Penn A., Hillier B., Banister D., & Xu J. (1998). Configurational modelling of urban movement networks. *Environment and Planning B: Planning and Design*, 25(1), 59-84.

Prud'homme J., Josselin D., & Aryal J. (2011). Quantitative Analysis of Pollutant Emissions in the Context of Demand Responsive Transport. In B. Murgante, O. Gervasi, A. Iglesias, D. Taniar, & B. O. Apduhan (Éd.). *Computational Science and Its Applications* - ICCSA 2011 (Vol. 6782, p. 439-453). Berlin, Heidelberg: Springer Berlin Heidelberg.

Prud'homme J., Genre-Grandpierre C., Josselin D., Prud'homme R. (2013). A methodological framework to simulate the effects of the network structure on pollutant emissions of private cars, soumis à IJAEIS.

Analyse de la mobilité d'un champ de Barkhanes (Mauritanie) à partir des images SAR

Kamel Hachemi[1] et Yves-François Thomas[1]

1. Laboratoire de Géographie Physique, UMR 8591 du CNRS, Université Paris-1 et Université Paris-Est
1 place Aristide Briand, 92195 Meudon Cedex.
hachemi_kamel@yahoo.fr et yves-francois.thomas@cnrs-bellevue.fr

RESUME.
Dunes épousant la forme d'un croissant allongé dans le sens du vent, les barkhanes sont présentes là où l'apport de sable est faible. Elles se déplacent le plus souvent sur un substrat « dur » dans un contexte de vents sinon unidirectionnels, du moins saisonnièrement unidirectionnels. Le vent faisant rouler le sable qui remonte la pente jusqu'au sommet de la dune et vient avalancher sur le versant aval, phénomène fait provoquer le déplacement de la dune. La dynamique des barkhanes menace localement l'existence de certaines infrastructures: routes, domaines aéroportuaires, villages. L'objectif de cette étude est de produire une analyse quantitative de la mobilité d'un champ de barkhanes situé au nord-est de la ville de Nouadhibou (Mauritanie). L'exploitation diachronique d'images SAR, acquises par les satellites ERS-1/2 et ENVISAT sur une période plus de 15 ans entre mai 1995 et octobre 2010, a permis de répertorier plus de 900 barkhanes et de suivre l'évolution de la morphologie du champ de dune. Cette approche a été couplée avec l'analyse des données de vent issues de la base de données ECMWF Interim Reanalysis pour la période 1983-2012. Les barkhanes sont parfaitement identifiables: elles apparaissent comme des secteurs de faible rugosité et rétrodiffusion (zones sombres) sur un plancher de plus forte rétrodiffusion et rugosité (zones claires). Le repérage des barkhanes est obtenu par filtrage puis segmentation d'image, le contour de leur morphologie –en 2D– est acquis par contourage, enfin leur caractérisation morphométrique connue par analyse de contours. Une suite d'indicateurs a été dérivée : longueur, envergure, périmètre, aire, centre de gravité, indice de compacité, vitesse de déplacement, etc. Diverses relations déjà observées in situ ont été retrouvées et confirmées : Longueur = f (Envergure) ; Aire = f (Longueur) ; Vitesse de déplacement = f (Longueur) ; etc. Le déplacement du centre de gravité des barkhanes vers le SSW (moyenne = 190.4°) sous l'effet des vents alizés (moyenne = 189.5°) est clairement mis en évidence. Le temps mis par une barkhane à parcourir sa propre longueur a pu être calculé, avec la distance parcourue et la vitesse de déplacement. Il varie entre 3 mois et plus de 18 ans selon la taille des édifices dunaires avec une valeur modale de 6 mois. L'imagerie SAR permet de détecter et de localiser de faibles mouvements de dunes sur un laps de temps court. L'analyse multi-temporelle conduite par emploi d'images d'amplitude a montré la possibilité : (i) de cartographier les distributions des sables mobiles et des formes dunaires ; (ii) de déterminer la vitesse et la direction de la remise en mouvement du sable ; (iii) de localiser les zones les plus touchées et les plus menacées par le

processus d'ensablement. Elle fournit une méthode utile pour répondre à la problématique de l'analyse du processus d'ensablement, de la remise en mouvement du sable dunaire, elle permet d'appréhender l'évolution et la dégradation du milieu sur différentes périodes.

ABSTRACT.

Dunes have a shape of a crescent lying in the direction of the wind, barchans are presented where the sand supply is limited the most often. They move on a hard substrate in the context of unidirectional winds, they at least seasonally unidirectional. The wind rolls the sand up the slope to the top of the dune and slides it on the downstream side. This mechanism causes the movement of the dune. Barchans Dynamic threats the existence of certain local infrastructure: roads, airport areas, villages. The objective of this study is to produce a quantitative analysis on the mobility of field barchans located northeast of the city of Nouadhibou (Mauritania). The diachronic operating SAR images which were acquired by the ERS-1/2 and ENVISAT over a 15 year period (between May 1995 and October 2010), helped to identify more than 900 barchans and to monitor the morphology of the dune fields. This approach has been coupled with the analysis of wind data coming from the database of ECMWF Interim Reanalysis data for the period of 1983-2012. Barchans are clearly identifiable: they appear as areas of low roughness and backscatter (dark areas) on a floor of higher roughness and backscatter (bright areas). The location of barchans is obtained by filtering and image segmentation, the contour of their 2D morphology is acquired by contouring their morphometric characteristics known by contour analysis. A series of indicators has been derived: length, size, perimeter, area, center of gravity, index of compactness, speed, etc. Various relationships already observed in situ were found and confirmed: Length = f (Span) Area = f (length) Travel speed = f (length) and so on. The displacement of the center of gravity towards the SSW barchans (mean = 190.4 °) due to trade winds (mean = 189.5 °) is clearly highlighted. The time needed to traverse by a barchans its own length could be known calculations of the distance and speed. This time varies between three months and 18 years depending on the size of dune building with a modal value of six months. The SAR imaging can detect and locate small movements of dunes on a short time span. The multi-temporal analysis using of amplitude images showed the ability: (i) to map the distribution of mobile sand dune forms and (ii) to determine the speed and direction of remobilization of sand, (iii) to locate the areas most affected and threatened by the silting process. It provides a useful method to address the problem of the analysis on of the silting of the remobilization of sand dune, it helps to understand the evolution and environmental degradation over different periods.

MOTS-CLES : Barkhane, SAR, ERS-1/2, ENVISAT, dynamique, Mauritanie.

KEYWORDS: Barchane, SAR, ERS-1/2, ENVISAT, dynamic, Mauritania.

1. Introduction

Au nord-est de la ville de Nouadhibou, à cheval sur la frontière entre la Mauritanie et le Sahara Occidental les paysages sont caractérisés par la présence de vastes espaces de sédiments sableux localement organisés en champs de barkhanes. Les dunes réactivées menacent l'existence de terrains agricoles, de villages ainsi que de routes. Quantifier la dynamique des champs de barkhanes fournit une information précieuse pour la gestion des ressources locales.

Les capteurs SAR (Synthetic Aperture Radar), capteurs actifs embarqués sur des satellites en orbites polaires mesurent les deux composantes du signal rétrodiffusé (amplitude et phase). L'amplitude de signal est fonction des caractéristiques physiques et diélectriques de l'élément d'image. La phase dépend des caractéristiques géométriques contient l'information de distance entre le radar et la cible, elle permet donc de mesurer les déplacements verticaux du sol de l'ordre d'1/2 longueur d'onde. Les images radar ont l'avantage de pouvoir être prises par tout temps, de jour comme de nuit, quelle que soit la couverture nuageuse, à la différence des plateformes qui opèrent dans le domaine du visible. Elles (images d'amplitude) ont été utilisées dans plusieurs domaines, comme la cartographie du dommage urbain du aux catastrophes naturelles ou industrielles, les changements des limites côtières (Ba *et al.,* 2007), le suivi de la déforestation, la surveillance de la fonte des glaces, la lithologie dans l'environnement polaire (Budkewitsch *et al.,* 1996a ; 1996b ; 1997), la cartographie géologique (Singhroy et Saint-Jean, 1999), la cartographie structurale (Wade *et al.,* 2001), la classification des cultures (Bruniquel et Lopes, 1994 ; Lopes et Sery, 1997), la reconnaissance des formes, l'aménagement du territoire (Rudant *et al.,* 1996), la détection et la caractérisation des nappes d'hydrocarbure dans les mers et les océans (Mercier *et al.,* 2004), la détermination de l'humidité de surface du sol en zones arides et semi arides (Troufleau *et al.,* 1994), la détection de phénomènes méso-échelle en océanographie (Laborde et Deveaux, 1996), la détection et l'analyse des changements du paysage dans la région subcarpatique de Buzau, Roumanie (Hachemi *et al.,* 2009 ; 2010), le suivi des phénomènes hydrologiques (les crues, les inondations, etc.), la cartographie de l'extension de l'inondation (novembre 1994) dans la région de Piémont, Italie (Brivio *et al.,* 2002), l'identification des caractéristiques optimales pour la surveillance de l'inondation dans des zones urbaines (McMillan *et al.,* 2006). Dans l'exploration géologique en zones arides, Baghdadi *et al.* (2005) ont démontré que l'imagerie radar SAR, en bande «L» ou «C», montre un plus fort potentiel de cartographie que les images optiques ASTER, surtout pour les structures linéaires comme les dykes ainsi que les structures circulaires.

Les environnements désertiques et subarides fournissent un contexte particulièrement favorable aux études historiques de suivi des évolutions de la radiométrie et de la géométrie des paysages grâce aux archives des données SAR des satellites ERS-1/2 (pour les deux dernières décennies). Les évolutions actuelles pouvant être décrites par emploi des données ENVISAT.

L'objectif de ce travail est : (i) de produire, sur une durée d'observation de plus de quinze ans (1995-2010), une analyse quantitative de la mobilité des champs de

dunes de type barkhanes situé au nord-est de la ville de Nouadhibou ; (ii) de confronter la mobilité observée avec la mobilité attendue à partir d'observations climatologiques ; (iii) de produire une cartographie digitale de la géomorphométrie du champ de barkhanes.

2. Localisation et caractéristiques de la zone d'étude

La zone d'étude est située au nord-est de la ville de Nouadhibou et centré par 21.4° N – 16.8° W sur la frontière entre le Sahara Occidental et la Mauritanie (Figure 1).

Figure 1. Carte géographique de la région (zone d'étude en rouge).

Le champ de barkhanes étudié (Figure 2), situé au nord-est de la ville de Nouadhibou, est un territoire désertique dont le paysage est caractérisé par la présence de vastes espaces de sédiments sableux localement organisés en champs de barkhanes. Dune épousant la forme d'un croissant allongé dans le sens du vent. Elle naît là où l'apport de sable est faible et se déplace le plus souvent sur un substrat « dur » dans un contexte de vents sinon unidirectionnels, du moins saisonnièrement unidirectionnels. Le vent faisant rouler le sable qui remonte la pente jusqu'au sommet de la dune et vient avalancher sur le versant aval, phénomène fait provoquer le déplacement de la dune. La figure 3 présente le schéma de la morphologie et du déplacement d'une barkhane.

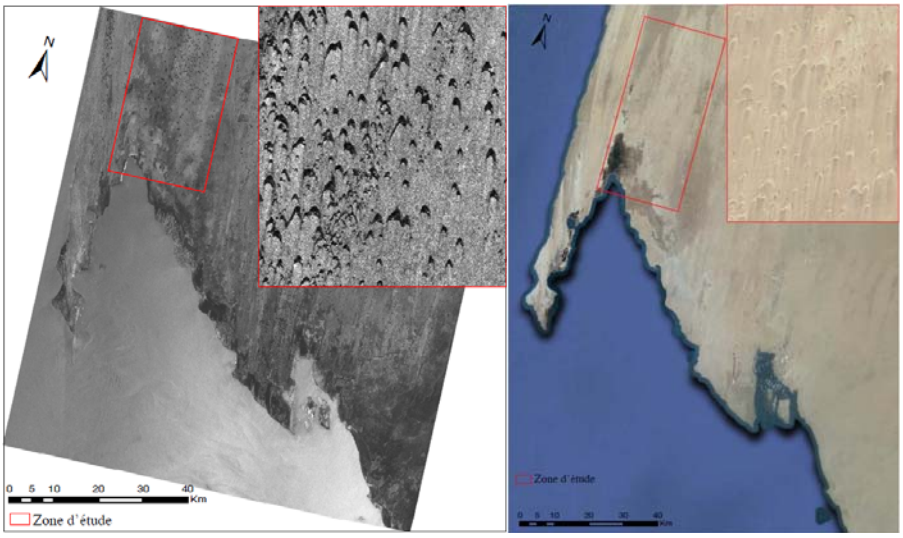


Figure 2. a) Image amplitude Envisat, 2010 (à droit) et b) Image en pseudo couleurs de la zone d'étude (Image Digital Globe, 2013)(à gauche).

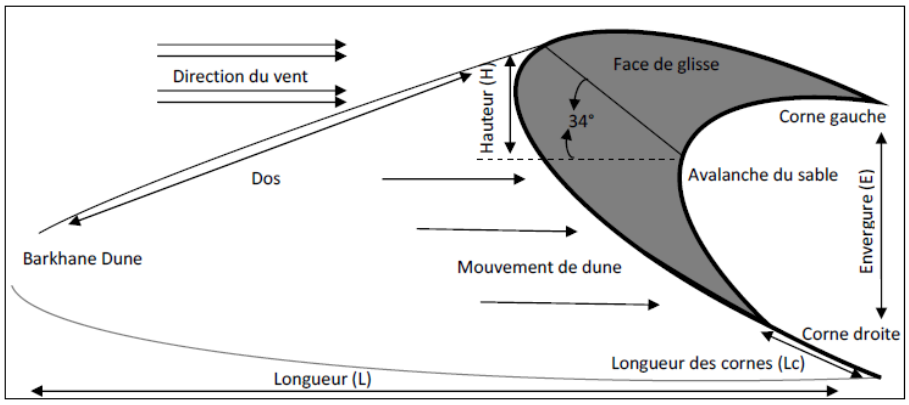


Figure 3. Schéma de la morphologie et du déplacement d'une barkhane.

6

3. Acquisition et traitement des données

L'acquisition des images SAR a été fournie par l'ESA (Agence Spatiale Européenne) dans le cadre d'un projet sur l'étude du phénomène d'ensablement dans la région Mauritanie – Sahara Occidental.

3.1. Données radar SAR utilisées

Les données SLC (Single Look Complex) utilisées, ont été traitées par compression d'impulsion en direction radiale en distance et par la synthèse d'ouverture SAR en direction azimutale à partir du même centre de traitement et d'archivage (UK-PAF). Dans ce type de produit, chaque pixel est représenté par une valeur complexe (partie réelle et partie imaginaire) d'où l'amplitude et la phase du signal sont extraites.

Pour cette étude, nous avons utilisé 9 images de passage descendant : 2 images mono vues complexes (SLC) des satellites ERS-1 et 7 images stellite ENVISAT. Ces images ont respectivement 26460 lignes et 4900 colonnes pour ERS et 26890 lignes et 5178 colonnes pour ENVISAT. Elles couvrent une surface de (100km x 100km) avec une résolution de 4m (en azimut) et 20m (en distance) et polarisation verticale/verticale (V/V). Le tableau (1) synthétise les caractéristiques de ces images.

Tableau 1. Données utilisées ERS-1 et ENVISAT

N°	Date	Heure	Satellite	Orbite	Track	Centre (Lat°/Lon°)
1	05/06/1995	11:38:24.78	ERS-1	20337	309	21,05/-16,66
2	27/11/1995	11:38:23.43	ERS-1	22842	309	21,05/-16,66
3	08/06/2004	11:08:29.97	ENVISAT	11886	309	21,04/-16,59
4	11/09//2007	11:08:37.67	ENVISAT	28920	309	21,04/-16,59
5	13/05//2008	11:08:33.00	ENVISAT	32427	309	21,04/-16,59
6	17/02/2009	11:08:29.40	ENVISAT	36435	309	21,04/-16,59
7	24/03/2009	11:08:30.15	ENVISAT	36936	309	20,99/-16,60
8	31/08//2010	11:08:25.37	ENVISAT	44451	309	21,04/-16,59
9	05/10/2010	11:08:21.04	ENVISAT	44952	309	21,04/-16,59

3.2. Traitement effectués

A l'aide des images radar SLC obtenues, nous avons pu extraire 9 images d'amplitudes. Nous avons aussi calibré et géo-référencié ces images. Le calibrage radiométrique consiste à mettre l'image produite (intensité) à sa représentation réelle

(terrain). Cette étape (radiométrique absolue) a été assurée en englobant : correction de la perte de gain en distance radiale ; correction de gain d'antenne ; correction des constantes de calibration… recommandées par l'ESA (Laur, 1992 ; Laur, *et al.,* 1993 ; 1998 ; 2004 ; Rosich et Meadows, 2004). Aussi, le géo-référencement permet d'avoir les images produites dans un même système de projection cartographique (UTM zone 28-N, WGS84). Cette étape a été déroulée automatiquement en attribuant les caractéristiques géographiques de chaque image. Ces deux étapes sont indispensables afin de procéder à une analyse diachronique. Les barkhanes sont parfaitement identifiables (Figure 4). Elles apparaissent comme des secteurs de faible rugosité et rétrodiffusion (zones sombres) sur un plancher de plus forte rétrodiffusion et rugosité (zones claires).

Le repérage des barkhanes est obtenu par filtrage puis segmentation d'image et le contour de leur morphologie –en 2D– est acquis par contourage. Cette dernière étape a été soigneusement effectuée à partir deux méthodes combinées ; (i) manuellement, par digitalisation des barkhanes une après l'autre ; (ii) la deuxième a été déroulée automatiquement par discrimination à partir d'un seuil représentant la surface de glisse (couleur noire) des zones des barkhanes.

Enfin leur caractérisation morphométrique est connue par analyse de contours, dont une suite d'indicateurs a été dérivée : longueur, envergure, périmètre, aire, centre de gravité, indice de compacité, vitesse de déplacement, etc. Diverses relations déjà observées in situ ont été retrouvées et confirmées : Longueur = f (Envergure) ; Aire = f (Longueur) ; Vitesse de déplacement = f (Longueur) ; etc.

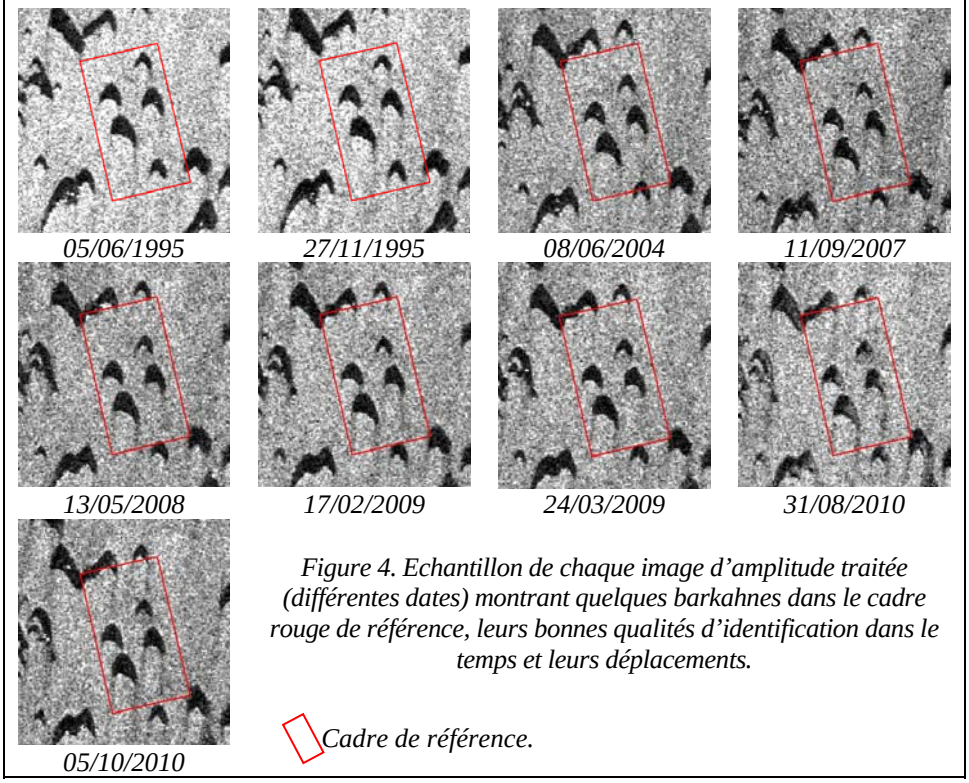


Figure 4. Echantillon de chaque image d'amplitude traitée (différentes dates) montrant quelques barkahnes dans le cadre rouge de référence, leurs bonnes qualités d'identification dans le temps et leurs déplacements.

4. Résultats

Nous avons répertorié et suivi l'évolution de la morphologie de plus de 900 barkhanes dans cette zone. Grace à ce suivi temporel d'une période de plus de 15 ans (1995/2010), les premiers résultats obtenus se résument ci-dessous :

(1) Les dimensions des barkhanes du champ de dunes étudié (longueur, envergure) ont pu être précisées, puis par analyse diachronique leur vitesse de déplacement (Figure 5).

(2) Les relations affines entre longueur et envergure des barkhanes ont pu être précisées (Figure 6-a).

(3) De même, les relations entre l'envergure en fonction de la vitesse de déplacement des barkhanes entre le 08/06/2004 et 05/10/2010 (Figure 6-b) et le turnover des barkhanes en fonction de leur taille entre 08/06/2004 et 05/10/2010 (Figure 6-c) ont pu être décrits. Le Turnover est le temps que met une barkhane à parcourir sa propre longueur ($T_{turnover} = \frac{L}{v}$) calculé avec la distance parcourue (L) et la vitesse de déplacement (v). Il varie entre 3 mois et plus de 18 ans selon la taille

des édifices dunaires avec une valeur modale de 6 mois. La fonction de densité de probabilité caractérisant le $T_{turnover}$ semble pouvoir être utilisé pour caractériser la dynamique d'un champ de dunes.

(4) Enfin, les direction de déplacements des barkhanes entre le 08/06/2004 et le 05/10/2010 exprimées en m/an (Figure 7-a) et la direction des vents issues de la base de données ECMWF Interim Reanalysis pour la période 1983-2012 exprimées en en m/s (Figure 7-b) ont été comparées. Le déplacement du centre de gravité des barkhanes vers le SSW (moyenne = 190.4°) sous l'effet des vents alizés (moyenne = 189.5°) est clairement mis en évidence.

Le résultat global de la vitesse du champ de barkhane étudié dans cette période de plus de 15 ans se résume dans le tableau 2. Il faut aussi noter que les résultats représentés dans les graphes ci-dessous représentent les images ENVISAT (2004/2010).

Tableau 2. Déplacement et vitesse globale du champ de barkhane étudié

Barkhane 05061995		Barkhane 05102010		Surfce (S) (m^2)	Déplacement (m)	Vitesse (m/an)
nombre	%	nombre	%			
528	57,70	532	58,65	S < 20000	571,62	38,10
290	31,69	292	32,19	20000 < S < 60000	409,58	27,30
95	10,38	81	8,93	S > 60000	351,87	23,45
Champ de barkhane			100%	Tout	444,35	29,62

10

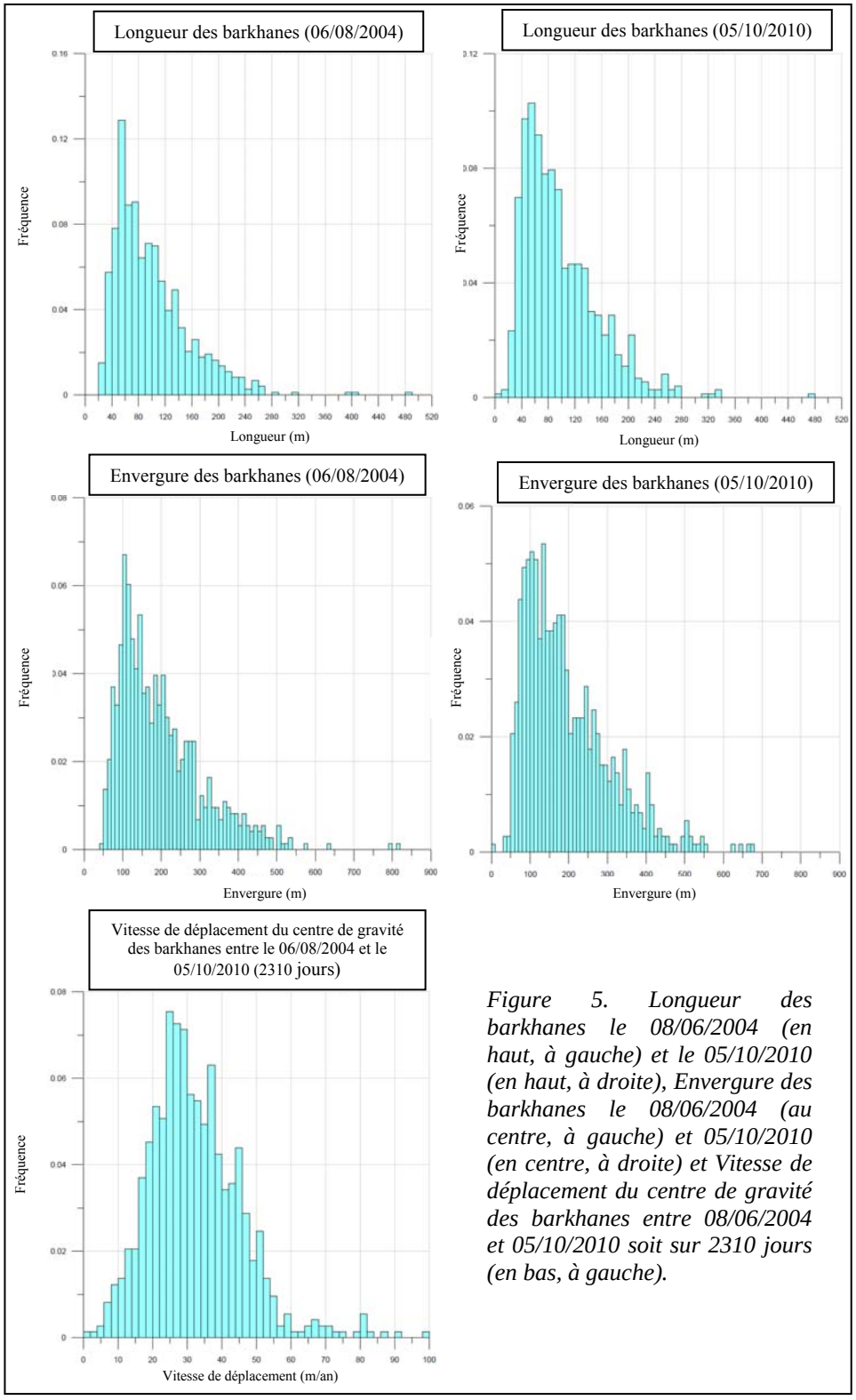


Figure 5. Longueur des barkhanes le 08/06/2004 (en haut, à gauche) et le 05/10/2010 (en haut, à droite), Envergure des barkhanes le 08/06/2004 (au centre, à gauche) et 05/10/2010 (en centre, à droite) et Vitesse de déplacement du centre de gravité des barkhanes entre 08/06/2004 et 05/10/2010 soit sur 2310 jours (en bas, à gauche).

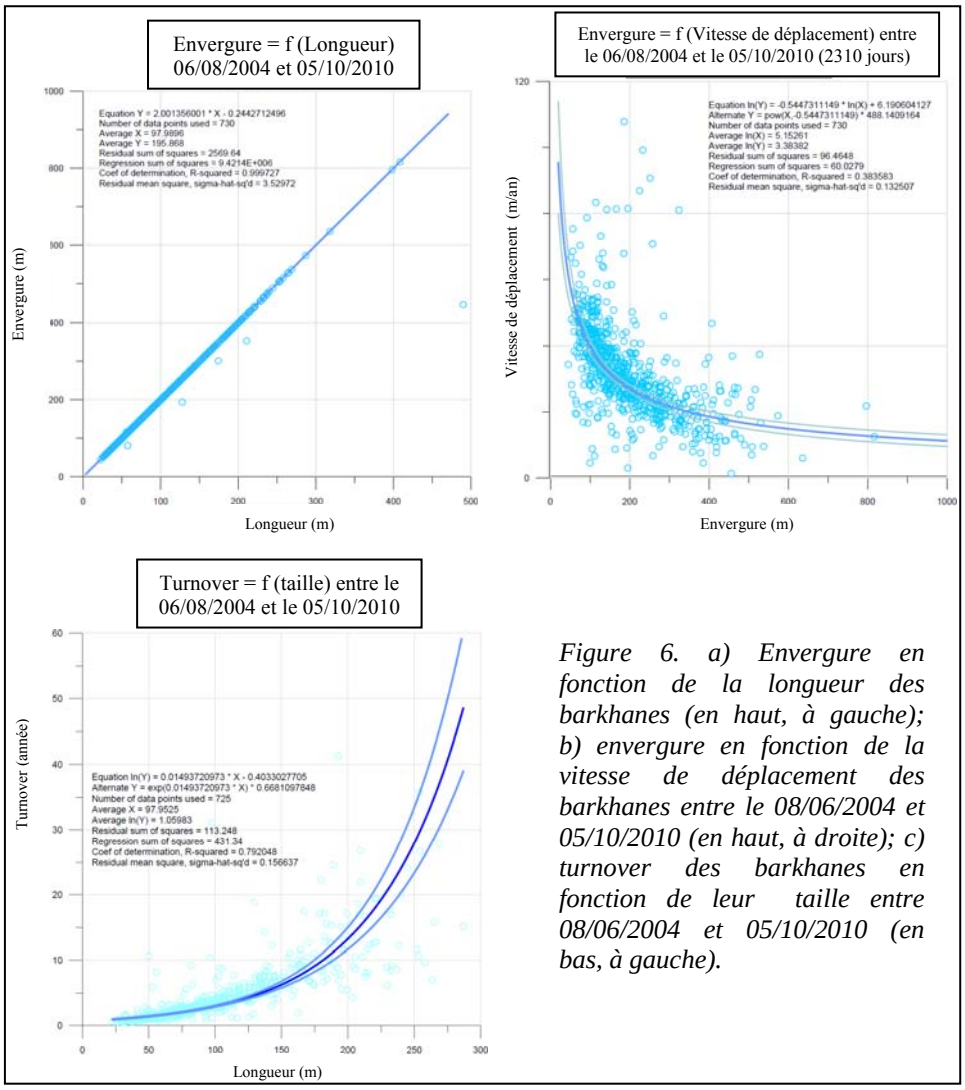


Figure 6. a) Envergure en fonction de la longueur des barkhanes (en haut, à gauche); b) envergure en fonction de la vitesse de déplacement des barkhanes entre le 08/06/2004 et 05/10/2010 (en haut, à droite); c) turnover des barkhanes en fonction de leur taille entre 08/06/2004 et 05/10/2010 (en bas, à gauche).

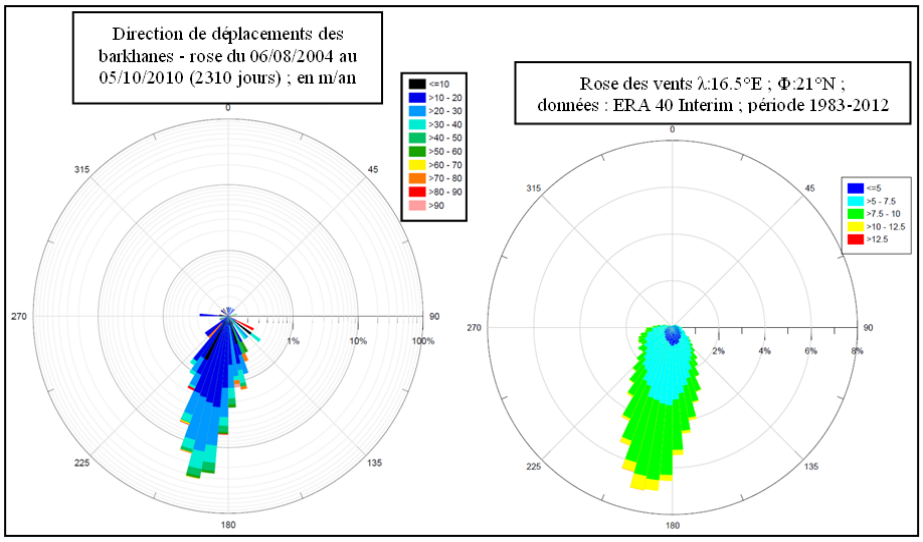


Figure 7. a) Direction de déplacements des barkhanes entre le 06/08/2004 et le 05/10/2010 en m/an (à gauche) ; b) Direction des vents issues de la base de données ECMWF Interim Reanalysis pour la période 1983-2012 en m/s (à droite).

4. Conclusion

Ce travail, nous a permis de répertorier plus de 900 barkhanes dans la région. Nous avons aussi suivi leur évolution sur une période de plus de 15 ans (mai 1995/octobre 2010). Nous avons mis en évidence le déplacement des barkhanes sous l'effet des vents alizés du nord-est. Grâce à l'analyse de changements, nous avons pu classer les barkhanes en trois catégories selon leur vitesse par rapport à leur dimension. Les barkhanes à petites dimensions, surface de glisse inférieure à 20000 m^2 (représentent 58 % de la totalité des barkhanes répertoriés) leurs vitesses ne dépassent pas les 38 m/an ; les barkhanes à moyennes dimensions, surface de glisse comprise entre 20000 et 60000 m^2 (32 %) ont des vitesses de 27 m/an ; par contre, les barkhanes à grandes dimensions, surface de glisse supérieure à 60000 m^2 (10 %) ont des vitesses de 23 m/an. L'imagerie SAR permet de détecter et de localiser de faibles mouvements de dunes sur un laps de temps court. L'analyse multi-temporelle conduite par emploi d'images d'amplitude a montré la possibilité : (i) de cartographier les distributions des formes dunaires ; (ii) de déterminer vitesse et direction de la remise en mouvement du sable ; (iii) de localiser les zones les plus touchées et les plus menacées par le processus d'ensablement. Elle fournit une méthode utile pour répondre à la problématique de l'analyse du processus d'ensablement, de la remise en mouvement du sable dunaire, elle permet d'appréhender l'évolution et la dégradation du milieu sur différentes périodes.

5. Remerciements

Nous remercions l'Agence Spatiale Européenne (ESA) de nous avoir fourni les images SAR des satellites ERS et ENVISAT dans le cadre d'un projet de recherche (n° 11148) sur l'étude des Barkhanes dans la région de la Mauritanie.

Bibliographie et références

Ba K., Wade S., Niang I., Trebossen H., Rudant J.-P. (2007). Cartographie radar en zone côtière à l'aide d'images multidates RSO d'ERS-2 : Application au suivi environnemental de la langue de barbarie et de l'estuaire du fleuve Sénégal. *Revue Télédétection*, vol. 7, n° 1-2-3-4, p. 129-141.

Baghdadi N., Grandjean G., Lahondere D., Paillou P., Lasne Y. (2005). Apport de l'imagerie satellitaire radar pour l'exploration géologique en zones arides. *Comptes Rendus (CR) Geosciences*, vol. 337, lssue 8, June, p. 719-728.

Brivio A., Colombo R., Maggi M., Tomasoni R. (2002). Integration of remote sensing data and GIS for accurate mapping of flooded areas, *Published in: International Journal of Remote Sensing* , vol. 23, Issue 3 February 2002, p. 429-441.

Bruniquel J., et Lopes A. (1994). Analysis and Enhancement of multitemporal SAR data. *Image and Signal Processing of Remote Sensing*, 2315: p. 342-353.

Budkewitsch P., D'Iorio M. A., Harisson J. C. (1996a). SAR Expressions of Geology in the Canadian Arctic. *Proceedings for the 26th International Symposium on Remote Sensing of Environment / 18th Symposium of the Canadian Remote Sensing Society,* du 25 au 29 mars 1996, Vancouver, p. 88-91.

Budkewitsch P., D'Iorio M.A., Harrison J.C. (1996b). C-band radar signatures of lithology in arctic environments: preliminary results from Bathurst Island, Nunavut. *Current Research 1996-B, Geological Survey of Canada*, p. 67-72.

Budkewitsch P., D'Iorio M.A. (1997). Contributions Toward Understanding C-band SAR Data for Lithological Discrimination and Structural Mapping in the Canadian Arctic. *Proceedings of the 12th International Coference on Applied Geologic Remote Sensing*, Denver, Colorado (USA), 17-19 November, p. I-38 - I-41.

Hachemi K., Abdellaoui A., Ozer A., Grecu F., Visan G. (2009). Apport de l'imagerie radar SAR (Images d'Amplitude) pour l'analyse du changement dans la région de Buzau (Roumanie). *Revista de Geomorfologie*, vol., 11, p. 63-72.

Hachemi K., Abdellaoui A., Grecu F., Ozer A., Visan M., (2010). Association d'images diachroniques avec un MNA pour une meilleure interprétation des changements de paysage dans la région de Buzau (Roumanie). *Revista de Geomorfologie*, vol. 12, p.53-65.

Laborde P., et Deveaux M. (1996). Etude des apports et limites du Radar à Ouverture Synthétique dans l'observation de l'océan à la Méso-échelle à partir des images SAR ERS-1 recueillies lors de la compagne Sémaphore, *Rapport d'étude, série Hydrographie, Service Hydrographique et Océanographique de la Marine (SHOM)*, Février 1996, Centre militaire d'Océanographie, Bureau de recherche et d'étude SHOM-Météo, Rapport d'étude n° 001/96, p. 40.

Laur H. (1992). ERS-1 SAR Calibration: Derivation of Backscattering Coefficient σo in ERS-1 SAR PRI Products. *ESA/ESRIN, Issue 1, Rev. 0,* 17 October 1992.

Laur H., Meadows, P.J., Sanchez, J.I., Dwyer E. (1993). ERS-1 SAR Radiometric Calibration. *Proceedings of the CEOS SAR Calibration Workshop, ESTEC,* Noordwijk, The Netherlands, ESA Publication WPP-048, September 1993, p. 257-281.

Laur H., Bally P., Meadows P., Sanchez J., Schaettler B., Lopinto E., Esteban D. (1998). ERS SAR calibration derivation of the backscattering coefficient σo in ESA ERS SAR PRI products. *Document No: ES-TN-RS-PM-HL09, Issue 2, Rev. 5b,* 7 september 1998.

Laur H., Bally P., Meadows P., Sánchez J., Schättler B., Lopinto E., Esteban D. (2004). ERS SAR Calibration: Derivation of σ0 in ESA ERS SAR PRI Products. *ESA/ESRIN, ES-TN-RS-PM-HL09*, Issue 2, Rev. 5f, November 2004.

Lopes A., et Sery F. (1997). Optimal Speackle reduction for the production model in Multilook Polarimetric data and the complex Wishart distribution. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5): p. 632-647.

McMillan A., Morley J. G., Adams B. J., Chesworth S. (2006). Identifying Optimal SAR Imagery Specifications for Urban Flood Monitoring: A Hurricane Katrina Case Study. *4th International Workshop on Remote Sensing for Disaster Response,* Magdalene College, Cambridge.

Mercier G., Derrodey S., Pieczynski W. (2004). Segmentation Multi-échelle de nappes d'hydrocarbure. *Traitement du Signal*, vol. 21(4), p. 329-346.

Rosich B., et Meadows P. (2004). Absolute calibration of ASAR Level 1 products. *ESA/ESRIN, ENVI-CLVL-EOPG-TN-03-0010*, Issue 1, Revision 5, October 2004, p. 26.

Rudant J.P., Baltzer F., Lointier M., Maitre H., Mougin E., Penicand C., Prost M. (1996). Apport des images radar satellitaires ERS-1 et JERS-1 dans le domaine de la cartographie générale et thématique en contexte tropical humide. *SFPT société Française de Photogrammétrie et Télédétection*, (142): p.15-31 et 81-82.

Singhroy V., et Saint-Jean R. (1999). Effects of relief on the selection of RADARSAT-1 incidence angle for geological applications. *Canadian Journal of Remote Sensing*, vol. 25, n°3, p. 211-217.

Troufleau D., Audoin A., Vidal A., Moran M.S., Goodrich D.C., Washburn J. (1994). Importance des corrections radiométriques dues au relief pour les données SAR du satellite ERS-1 : Applications à l'hydrologie. *Xème journées hydrologiques* -Orstom, Septembre 1994, p. 635-650.

Wade S., Ndoye A., Mbaye M. (2001). Fusion d'images optique et radar : application à la cartographie du massif granitique de Bambadji (Falémé, Sénégal oriental). *Teledetection*, 2001, vol. 2, n° 2, p. 119-127.

Utiliser les Smartphones et les outils de géo-localisation pour améliorer les connaissances et les prévisions des inondations rapides : enjeux, contraintes et perspectives

Jules Sekeouia Kouadio[1], Johnny Douvinet[1], Loic Grasland[1]

1. UMR CNRS 7300 ESPACE, Université d'Avignon et des Pays de Vaucluse
74 rue Louis Pasteur – 84029 Avignon Cedex 1 – France
jules.kouadio@alumni.univ-avignon.fr, *johnny.douvinet@univ-avignon.fr*, loic.grasland@univ-avignon.fr

RESUMÉ. *Cet article propose de discuter des enjeux, contraintes et possibilités offertes par et pour le développement d'une application Smartphone qui pourraient améliorer les prévisions et les connaissances actuelles sur le mode de fonctionnement des inondations rapides. Il est aussi question d'évaluer l'efficacité des systèmes pervasifs en matière d'alerte en temps réel. Le caractère participatif et volontariste de ce travail soulève des questions liées à l'usage et à la gestion de données personnelles. L'identification des zones à risque, la géolocalisation des personnes en situation de danger ou la réception d'un message d'alerte posent des problèmes liés à la vie privée des individus, qui pourraient pourtant devenir des « citoyens-capteurs ».*

ABSTRACT. *This article focuses on problems and constraints and opportunities regarding the development of one specific application to better know or forecast flash floods. We also want to assess the effectiveness of pervasive systems in terms of real-time alert. The participatory and proactive nature of our approach raises problems related to the use and management of personal data. Indeed, the location of risk areas, the geo-localization of persons endangered and the reception of one alert message faces to the population liberties. But how can we make a return on that experience without dropping anonymity?*

MOTS-CLES *: inondations rapides, système d'alerte, application, Smartphones, France*

KEYWORDS*: flash flood, alert system, applications, Smartphone, France.*

1. Contexte de recherche

Les crues rapides constituent la forme la plus originale et la plus dangereuse des inondations observées en France. Elles se caractérisent par un écoulement torrentiel et par leur soudaine apparition et entraînent des dommages plus ou moins importants selon l'importance des enjeux et la vulnérabilité des secteurs géographiques affectés (Vinet., 2010). Si le SCHAPI, Service Central d'Hydro-Météorologie en Appui à la Prévision des Inondations) surveillent près de 21 100 kilomètres de linéaire de cours

d'eau (en juin 2013), les 103 000 km restant (près de 45% sont associés à des petits tronçons susceptibles de générer des crues rapides) ne font toujours pas l'objet d'une vigilance particulière. Ce constat peut s'expliquer par plusieurs raisons : i) l'absence de stations hydrométriques sur des tronçons en aval desquels les enjeux ne sont pas élevés ; ii) une méconnaissance des seuils de déclenchement des crues sur un grand nombre de bassins (PUB, 2006) ; iii) les incertitudes des radars pluviométriques qui rendent les modélisations hydrologiques et hydrauliques délicates (Lobligeois *et al.,* 2012 ; Douvinet *et al*., 2012). Pour pallier ces insuffisances et suite à la survenue des inondations dramatiques de 2010 à La-Faute-sur-Mer (28 février, 53 morts) et dans la Dracénie (15 juin, 27 morts ; plus de 700 millions d'euros de dommages), l'Etat a toutefois engagé des réflexions et différents projets rassemblés dans le PSR – Plan Submersions Rapides (DGPR, 2011).

Dans ce contexte, notre travail a pour objectif de contribuer à la réduction de la vulnérabilité des personnes mobiles lors de la survenue de tels phénomènes à travers l'usage des technologies Smartphones. Il s'agit plus spécifiquement d'introduire une nouvelle approche en matière d'alerte à travers l'usage d'une application qui se veut volontariste et gratuite. Ce sujet est vite opportun tant les Smartphones sont devenus populaires et bien adoptés en France (24 millions d'utilisateurs en 2012 ; 4 millions de Smartphones vendus depuis le 1er janvier 2013). La mise en œuvre de nouveaux systèmes d'alerte pourrait apporter de nouvelles connaissances sur les crues rapides (modes de fonctionnement, seuils de déclenchement, pluie critique) et contribuer au remplacement des systèmes actuellement utilisés, qui datent de la seconde guerre mondiale et qui apparaissent désuets (qui est capable de reconnaître les sirènes et le type d'alerte généré aujourd'hui parmi l'ensemble du grand public ?)

Avec l'entrée, dans le Web social, des outils de géo-localisation, le partage de la localisation (coordonnées GPS) et des déplacements (méthode du tracking), allant même parfois jusqu'à la diffusion de ces informations, individuelles ou collectives, sur des réseaux sociaux plus ou moins ouverts (amis ou professionnels) par le canal de ces appareils, demeurent des pratiques courantes. En contrepartie, des demandes de plus en plus fortes sont formulées pour accéder à des réseaux urgentistes et/ou d'appels de secours à partir des technologies Smartphone. Le succès d'applications comme *La Croix Rouge* (aider une personne démunie), *!ncident* (informer d'un incident en temps réel) ou *Quake-Up* (alerte séisme) l'illustre parfaitement.

Mais avant de développer une application, il faut en mesurer les fondements, les enjeux et les limites ; ainsi, il s'agira dans cet article : 1) de savoir si des applications ont déjà été développées sur ces questions (et si oui, sur quelles critères ont-elles été créées et sont-elles opérationnelles en France ?) ; 2) d'identifier les contraintes qui peuvent limiter l'utilisation des outils Smartphone (peut-on créer un système d'alerte en utilisant les données de géo-localisation par exemple ?) ; 3) de voir si la création d'une nouvelle application (mono-risque) complique encore plus l'information et/ou la gestion d'un risque (autrement dit, contribue-t-on à une meilleure alerte auprès de la population exposée et/ou à une meilleure responsabilisation face aux gestes à tenir en cas de crise ?)

2. Etat des lieux des applications développées pour une mise en alerte face aux risques inondation ou inondation rapide : quelle est la position de la France ?

La téléphonie mobile a été bien accueillie grâce à sa fonctionnalité qui permet aux personnes « mobiles et/ou en déplacement » d'être joignable, « partout et à tout moment », contrairement à la téléphonie fixe. Depuis 1992, une nouvelle génération de téléphones portables est apparue : les Smartphones. Ces appareils (*Ordiphones* en français) disposent, au-delà de leur fonction primaire, d'un nombre très important de fonctionnalités (connectique WIFI, géo-localisation par GPS, cartographie couplée au calcul d'itinéraire, etc.). Il est possible d'installer des applications additionnelles (plus de 330 000 existent sur *Apple Store* en 2012), voire de supprimer celles livrées lors de l'acquisition de l'appareil, ce qui permet à l'utilisateur de configurer son téléphone selon ses propres usages et envies.

Une étude effectuée sur 17 applications nous a amené à les dissocier en 2 grands groupes : 1) des applications à caractère consultatif, qui permettent à l'utilisateur de visualiser des informations et des instructions relatives à un danger (sans être associé à la survenue d'un épisode en temps réel). C'est le cas pour *FloodWatch* (développé sous iOS6) ou *FloodMap* (sous Symbian OS) par exemple ; 2) des applications à caractère participatif qui offrent, en plus, la possibilité d'une localisation du danger tout en contribuant à l'amélioration de sa description (ex : l'application de la ville de Nice). Avec elles, la notion de « citoyen-capteur » prend tout son sens : l'utilisateur reçoit des informations mais il peut aussi, en retour, améliorer les connaissances sur les phénomènes inondant (il déclenche l'alerte ou peut en être averti).

En regardant la nature des territoires couverts par les applications étudiées, deux oppositions sont mises à jour : 1) les applications couvrent d'importantes surfaces (Philippines, Etats-Unis ou Thaïlande) et leur mise en place répond à des demandes étatiques (en lien avec la récurrence et la sinistralité des inondations) ; 2) les autres applications sont spécifiques à des entités communales (à Nice ou à Athis-Mons en France par exemple) et les surfaces couvertes sont bien plus petites (< à 20 km2). Les enjeux sont alors très différents et l'opérationnalité n'est pas non plus du même ordre. L'application *River an Stream Pro* permet, par exemple, de surveiller plus de 9000 rivières composant le réseau hydrographique des Etats Unis (ainsi, pourquoi ce système n'est pas développé par le SCHAPI pour la France ?) alors que l'application de Nice concerne 4 petites rivières. Les applications locales dépendent en réalité des initiatives politiques émanant des communautés de communes ou des villes qui ont décidées d'inclure ces outils dans une politique globale de prévention des risques. Développer sa propre application nécessite la mobilisation de ressources (humaines, matérielles, budgétaires) et des autorisations spéciales. De plus, ces systèmes ont été conçus dans des contextes variés, avec des objectifs et des usages différents, n'étant alors pas fonctionnels en dehors des territoires d'investigation. On peut par ailleurs trouver des petites « briques » qui permettent de déclarer un incident ou de connaître l'intensité des pluies (*via* Météo France), mais ces données contribuent à une information générale sur les risques sans inciter à tel ou tel comportement.

Des différences de pro-activité sont aussi observées : soit l'utilisateur est obligé d'ouvrir l'application pour savoir si il y a ou non une alerte (on parle de pro-activité

passive) ; soit le système expert est capable de déclencher tout seul l'alerte à partir du moment où certains seuils sont franchis au niveau des capteurs. Dans ce cas, on parle de pro-activité active, qui peut dépendre du centre de gestion si ce dernier se base sur les avis / photographies / informations envoyées de la part de la population ; la population participe alors elle-même à cette pro-activité.

On remarque par ailleurs que ces applications sont mono-risques ou multirisques (interface unique ou multiple de contribution ou de consultation mais pour différents types de phénomènes). L'usage du GPS se fait à travers des modes de cartographie intégrée une fois que l'utilisateur a accepté de transmettre ses coordonnées, et des photos relatives aux phénomènes peuvent être jointes dans le message émis. La taille de l'espace mémoire et le prix des applications ont aussi servi à l'étude comparative.

Si provisoire soit cet inventaire, aucune application ne porte exclusivement sur les inondations rapides qui pourraient survenir en France. Dès lors, quelles sont les raisons expliquant cette absence ? Les lacunes dans les connaissances actuelles sur leur mode de fonctionnement et sur la localisation précise des pluies intenses en sont sans doute à l'origine (prévoir une crue rapide nécessite d'avoir un certain nombre de certitudes sur l'occurrence du phénomène, et pas seulement en étudiant l'intensité des pluies), mais nous pensons surtout que l'absence d'applications est dépendante de contraintes juridiques et réglementaires fortes en France.

3. Des contraintes techniques et réglementaires importantes en France

Bien qu'alléchant, le développement d'applications basées sur la géolocalisation (via le système GPS) et sur les risques (naturel ou technologique) demeure limité en France à cause de contraintes omniprésentes. La protection des personnes physiques à l'égard de l'utilisation et des traitements des données de géolocalisation demeure régie en France par les lois 2004-669 du 9 juillet 2004 et 2004-801/182 du 6 août 2004 (portant sur les protections individuelles au sens large). Afin d'éviter que des personnes non autorisées n'accèdent aux données, il est obligatoire de prendre des mesures de sécurité (accès à un site avec un identifiant et un mot de passe), selon la norme 51 de la CNIL. Cet usage est également protégé par un ensemble de lois associées au droit pénal (article 226-1 et suivants sur la protection de la vie privée ; article 226-16 relatif aux droits des personnes résultant de traitements informatiques ; loi du 6 janvier 1978). Toutes ces contraintes constituent bien un frein face à notre problématique, mais ce ne sont pas les seules car des contraintes réglementaires sont aussi observées, notamment sur la gestion de l'alerte et les responsabilités associées. En France, le décret n°2005-1269 (12 octobre 2005) relatif au code d'alerte national décrit toutes les mesures destinées à alerter et informer « en toutes circonstances la population, soit d'une menace ou d'une agression au sens des articles L. 1111-1 et L. 1111-2 du Code de la défense, soit d'un accident, d'un sinistre ou d'une catastrophe au sens de la loi du 13 août 2004 ». Ces mesures sont toujours mises en œuvre par les « détenteurs » des moyens de communication vers le grand public, ce qui indique qu'aucune personne ne peut, à titre individuel, véhiculer une alerte *via* les canaux officiels de communication. Le déclenchement de l'alerte est une tâche qui incombe aux mairies (associés à son pouvoir de police), aux préfets ou aux sous-préfets, voire

à l'Etat si l'alerte dépasse les limites départementales. Une telle procédure nécessite une validation et des contrôles à différents niveaux. En se basant sur leurs capteurs hydrométriques, pluviométriques et/ou des modèles de prévisions, le SCHAPI (avec Météo France) alerte en premier lieu l'un des 22 Services de Prévisions des Crues (SPC) en cas de forte probabilité d'occurrence d'une crue (associée à un degré de dommages faible, moyen ou élevé). Les informations sont relayées aux services de l'Etat (mairies, préfectures) et à la population (via le site www.vigicrues.fr et lors des bulletins de prévisions météorologiques sur les chaînes hertziennes). Le portail Internet est mis à jour à 8h et 14hUTC. Quand le niveau de vigilance passe du jaune au orange ou rouge, le Centre Opérationnel de Gestion Interministériel de Crise (COGIC) est directement informé pour faciliter ensuite les relais avec les acteurs locaux (SDIS, mairie, DDT, police, association de riverains). Dans tous les cas, il s'agit pourtant d'une alerte classique partant de l'administration vers les administrés, donc l'administration détient exclusivement le pouvoir d'information en cas de risques majeurs (qu'ils soient naturels ou technologiques – selon le Code Général des Collectivités Territoriales) et le maire est de toute façon le seul responsable de la sécurité des biens et des personnes sur son territoire en cas de sinistre (Denolle et Douvinet, 2010). Le dispositif actuel apparaît alors fortement technocratique. Il peut même paraître désuet (il date au passage de la seconde guerre mondiale) et bien loin des nouvelles techniques d'information et/ou de communication. Peu de citoyens savent relier le son des sirènes aux comportements à adopter en période de crise. Ce constat s'explique par l'absence d'exercices de simulation réguliers et par l'excès d'informations, où le citoyen place le risque loin de ses préoccupations centrales.

4. Perspectives et conclusions

Après l'étude des applications existantes et des contraintes liées, nous cherchons actuellement à développer un système opérationnel, et réfléchissons à la meilleure forme, structure et mode de fonctionnement de l'application que nous pourrions créer. Il s'agirait d'une application mono-risque (face aux crues rapides) à caractère contributif qui intègre les données de géolocalisation, qui fait appel à l'appareil photographique et qui laisse le libre choix à l'utilisateur pour l'installation (sur le téléphone ou vers une version en ligne via Q-R-Code). Une fois ouverte, une interface simple et intuitive propose à l'utilisateur un pré-formulaire (ce qui permet de le guider dans les différents choix). A ce stade, d'autres technologies seront sans doute employées telles que la Réalité Augmentée ou la Synthèse vocale. Après remplissage le formulaire serait validé et le contenu transmis à un centre de décision (préconçu en amont ou localisé dans un service opérationnel), qui en retour pourrait diffuser l'alerte dans un rayon de 5 kilomètres autour du point émetteur. La diffusion des messages se ferait ensuite par un mode « pair-à-pair » permettent une connexion entre les téléphones sans passer par les opérateurs téléphoniques via le système DTN (Delay Tolerent Network), procédé déjà utilisé par GeoCam de Cylab par exemple.

La réussite de l'application implique une très bonne (parfaite ?) connaissance de l'environnement dans lequel elle sera déployée, et de ses aptitudes à répondre aux besoins exprimés par les futurs utilisateurs. Notre expérimentation portera sur deux

départements sensibles aux crues rapides : le Vaucluse et le Var. Afin d'identifier les territoires potentiellement concernés par l'application, nous avons sélectionné puis écarté tous les tronçons disposant de capteurs hydrométriques intégrés à la vigilance actuelle du SCHAPI ; occulté toutes les communes situées à proximité de linéaires surveillés ; conservé les communes qui sont traversées par plusieurs tronçons ; puis, pour finir, écarté les rivières et canaux circulant en plaine. Parmi les 205 communes sur 315 au départ (71 %), il s'agira : 1) de mener des enquêtes par questionnaires administrés auprès d'un échantillon de population situé au plus près des rivières ou des cours d'eau (ayant des densités de population variées) afin de mieux estimer le pourcentage d'utilisateurs Smartphone dans ces zones ; 2) évaluer la faisabilité du système avec les prévisionnistes et les gestionnaires des risques aux échelles locales (depuis les maires, les préfectures, le Service de Prévision des Crues SPC-Grand Delta jusqu'au SCHAPI), tout en mesurant la perception des élus et de la population face à la mise en place d'une telle application. D'ici là, des capteurs seront placés sur certains cours d'eau (en amont des communes enquêtées) pour tester la chaine d'information (de l'émission à la diffusion des messages d'alerte) en mode « pair à pair » (sans passer par les opérateurs). Il faudra définir un système de traitement des données émises par les utilisateurs, tout en protégeant le système. En complément, les leviers techniques, règlementaires ou juridiques devront être levés.

Remerciements

Les auteurs tiennent à remercier pour leur soutien financier la Région PACA, le bureau d'études Risques et Développement, la Structure Fédérative de Recherche Agor@ntic de l'Université d'Avignon et des Pays de Vaucluse.

Bibliographie

Direction Générale de la Prévention des Risques (DGPR) (2011). *Plan de Submersion Rapide*, février 2011, 234 p.

Denolle A.-S., Douvinet J. (2010) – Les marges de manœuvre des maires face à l'application des Plans de Prévention du Risque Inondation (PPRI). Analyse d'un point de vue juridique. *RISEO, Risques, études et observations*, 2010-1, 38-51.

Douvinet J., Mallet F., Escudier A. (2012) – *Localisation des bassins présentant la plus forte prédisposition au risque de ruissellement et de crues rapides à l'échelle de la Seine-Maritime.* Rapport interne UAPV-SCHAPI, 125 p.

Lobligeois F., Andreassian V., Perrin C., Loumagne C. (2012) – Investigating the influence of rainfall spatial variability on flow simulation. Proceedings of the seventh European conférence on radar in meteorology and hydrology. 4p.

Vinet F. (2010) - *Le risque inondation*. *Diagnostic et gestion*, Éditions Tec & Doc, 245 p.

ARCAMA-3D: Une application mobile à base de réalité augmentée connectée au Web des Données

Betül Aydin[1], Jérôme Gensel[1], Philippe Genoud[1], Sylvie Calabretto[2], Bruno Tellez[3]

1. *Laboratoire d'Informatique de Grenoble, équipe STEAMER. 681, rue de la Passerelle, 38402 Saint Martin d'Hères*
   Betul.Aydin@imag.fr, Jerome.Gensel@imag.fr, Philippe.Genoud@imag.fr

2. *INSA de Lyon, 20, avenue Albert Einstein, 69621 Villeurbanne Cedex*
   Sylvie Calabretto@liris.cnrs.fr

3. *Université Claude Bernard Lyon 1, 43 boulevard du 11 Novembre 1918 69622 Villeurbanne Cedex*
   Bruno.Tellez@liris.cnrs.fr

RESUME. ARCAMA-3D (Réalité Augmentée pour les applications mobiles sensibles au contexte avec 3D) est une application qui propose de superposer à la vue de l'environnement dans lequel l'utilisateur évolue une représentation 3D de celui-ci. Elle agit comme une interface en réalité augmentée, qui en situation de mobilité, permet l'exploration de l'espace environnant de manière interactive et adaptée aux besoins de l'utilisateur. Dans cet article, nous montrons comment les objets de la réalité augmentée qui se superposent aux objets du monde réel de la scène capturée par la caméra du smartphone sont couplés avec le nuage du Web des Données ouvertes et liées (Linked Open Data (LOD) cloud). D'une part, cela offre la possibilité d'enrichir les modèles 3D avec des informations structurées provenant du Web des Données, auxquelles les artefacts de la réalité augmentée donneront accès de manière interactive. D'autre part, cela donne la possibilité aux personnes qui publient des données sur le nuage du Web des données, de lier ces données avec des modèles 3D. A ces fins, nous proposons un modèle de données 3D extensible qui prend en compte l'évolution temporelle des entités du monde réel (comme les bâtiments, les monuments, etc.).

ABSTRACT. ARCAMA-3D (Augmented Reality for Context Aware Mobile Applications with 3D) is a application that offers to overlay a 3D representation of the surroundings that acts as an augmented reality interface. While the user visits a place, it allows the exploration of her environment in an interactive manner adapted to her needs. In this paper, we show how the augmented reality objects, which are overlaid on the real world objects that are captured by the camera of the mobile device, are coupled with the Linked Open Data (LOD) cloud. On the one hand, it offers the possibility of enriching the 3D models with structured information from the LOD, to which augmented reality objects provide access interactively. On the other hand, it provides opportunity for people who publish information on the LOD cloud to link their

data with the 3D models. To this end, we propose an extensible data model that takes into account the temporal evolution of the real world entities (such as buildings, monuments, etc.).

MOTS-CLES : *mobilité et ubiquité, modélisation spatio-temporelle, web sémantique, web des données, réalité augmentée.*

KEYWORDS: *mobile and ubiquitous computing, spatiotemporal modeling, semantic web, linked open data, augmented reality*

1. Introduction

Les applications mobiles ubiquitaires s'appuient sur des approches centrées utilisateur dites sensibles au contexte dans lesquelles le système prend en compte l'évolution du contexte de l'utilisateur pour fournir dynamiquement des informations. Toutefois, les contraintes liées à la taille réduite de l'écran de la plupart des dispositifs mobiles et le mode de consultation en situation de mobilité qui nécessite souvent une réponse immédiate, imposent de fournir une information la plus pertinente possible. Par exemple, dans une application de découverte d'un territoire, la présentation systématique et exhaustive de tous les points d'intérêt qu'il comporte non seulement nuira à la lisibilité de cet ensemble d'informations mais risquera également de détourner l'utilisateur de l'observation de l'environnement.

Dans cet article, nous présentons ARCAMA-3D (Réalité Augmentée pour les applications mobiles sensibles au contexte avec 3D) (Aydin *et al.*, 2012a ; 2012b), une application mobile de Réalité Augmentée (RA) (Azuma, 1997) sensible au contexte qui minimise l'effort de l'utilisateur pour interagir avec son environnement et accéder à des informations susceptibles de l'intéresser. ARCAMA-3D superpose à la vue captée au travers de la caméra du dispositif mobile (de type *smartphone)* des représentations 3D des objets du monde réel. Ces objets servent de support aux interactions de l'utilisateur pouvant pour accéder à des données disponibles sur son environnement. Pour cela, nous interconnectons notre application au Web des Données (Heath et Bizer, 2011) qui, en structurant les connaissances disponibles sur le Web, permet une approche sémantique dans la recherche d'informations. En enrichissant les modèles 3D de métadonnées *thématiques* et *temporelles* publiées et liées sur le Web des Données, nous permettons la mise en place de scenarios applicatifs dans lesquels l'interface de réalité augmentée peut assister efficacement l'utilisateur dans son exploration du monde réel en fonction de ses choix et de ses préférences.

Le papier est organisé de la manière suivante. Dans la section 2, nous présentons quelques-uns des travaux liés à la réalité augmentée qui intègrent l'information géo-référencée, ainsi que les aspects qui différencient notre système de l'existant. La section 3 décrit nos travaux précédents, la plate-forme ARCAMA-3D et introduit le modèle de données (l'ontologie) que nous avons définie pour interconnecter les objets 3D avec le Web des Données liées. Nous montrons également dans cette section comment cette interconnexion peut, d'une part étendre l'utilisation d'ARCAMA-3D, et, d'autre part, contribuer à l'extension du nuage du Web des Données avec des informations structurées actuellement non représentées dans ce

nuage. Dans la section 4, nous décrivons les expérimentations réalisées, et, dans la section 5, nous tirons divers enseignements de ce travail et présentons les poursuites envisagées.

2. Contexte et travaux connexes

Pour une personne qui visite un lieu, rechercher et fournir *toutes* les informations à partir du seul critère de sa géolocalisation, ne peut pas être considéré comme entièrement *satisfaisant*. En effet, les plates-formes de réalité augmentée mobiles sensibles au contexte devraient également envisager de filtrer les informations afin de ne retenir que les informations utiles en phase avec les attentes ou les centres d'intérêt de l'utilisateur. Dans ce but, il nous semble que les technologies du Web sémantique peuvent être utilisées afin de permettre une exploration pertinente de l'information disponible sur le Web grâce à l'expression de connaissances générales et/ou spécialisées de manière formelle et structurée.

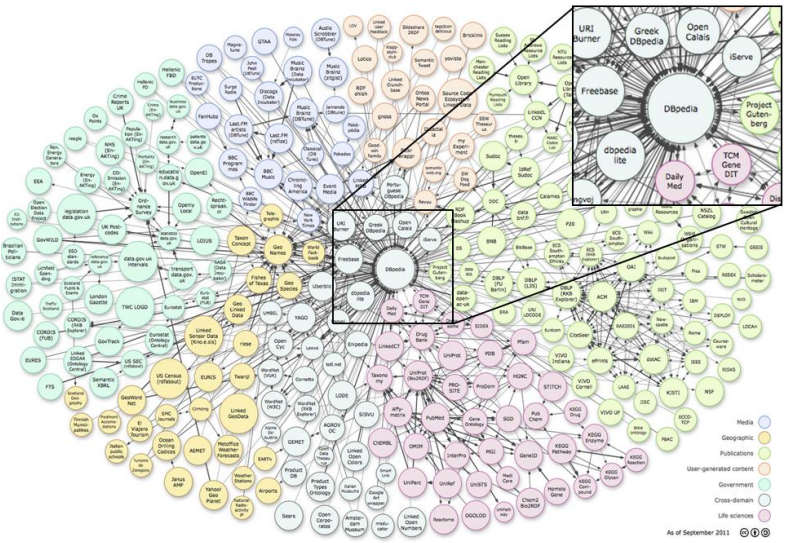


Figure 1. Le nuage du Web des Données.

Les données liées (*linked data*) est un terme utilisé pour désigner une bonne pratique recommandée pour représenter, exposer, partager et interconnecter des informations en utilisant les technologies du Web sémantique : IRIs[1], RDF,

[1] Internationalized Resource Identifier : généralisation des URI (Uniform Resource Identifiers) pouvant contenir des caractères Unicode non ASCII.

SPARQL[2]. Les données ouvertes et liées (*Linked Open Data* ou LOD) constituent un effort collectif pour publier des jeux de données disponibles sous licence ouverte en respectant les principes du *linked data* (Heath et Bizer, 2011). Parmi ceux-ci, Dbpedia, extrait à partir des contenus de l'encyclopédie Wikipédia, joue un rôle central en reliant entre eux de nombreux autres jeux de données, tels que Flickr, Revyu, ou encore des profils FOAF (voir Figure 1). Une part de ces données contient des informations géo-localisées qui peuvent être exploitées par des applications, en particulier en situation de mobilité. Cependant, peu d'applications mobiles de réalité augmentée ont exploré cette possibilité.

Par rapport à la problématique abordée de découverte d'un territoire en situation de mobilité, mSpace mobiles (Wilson *et al.*, 2005) est l'une des premières applications mobiles qui utilisent le Web sémantique pour soutenir l'exploration de ressources d'information. Il s'agit d'un guide de la ville de Londres offrant des connexions à différents domaines ou sources d'intérêts (cinéma, musique, *etc.*). En utilisant la localisation de l'utilisateur explorant la ville, l'application interroge différentes bases de connaissances (IMDB, les archives vidéo de la BBC, *etc.*). Bien que cette application ne s'appuie aucunement sur la réalité augmentée et que son interface utilisateur ne permette pas une exploration rapide et réellement efficace, elle demeure intéressante par l'utilisation et l'exploration de plusieurs bases de connaissances reliées entre elles.

DBpedia Mobile, quant à lui, est un navigateur sensible au contexte fonctionnant sur les appareils mobiles (Becker et Bizer, 2009). Conçu pour l'exploration d'un lieu touristique, il indique sur une carte les emplacements situés à proximité de l'utilisateur pour lesquels des données DBpedia sont disponibles. L'utilisateur peut alors explorer les informations associées à ces emplacements grâce aux liaisons entre DBpedia et d'autres sources de données (par exemple les pages Wikipédia associées). Il permet également aux utilisateurs de publier leur localisation, photos et commentaires sur le Web des Données. On remarque ici que c'est à l'utilisateur de choisir l'information et que l'interface n'exploite pas la réalité augmentée.

En revanche, dans (Reynolds *et al.*, 2010), sont abordés plusieurs cas d'utilisation possibles du Web des Données dans le cadre d'applications mobiles à base de réalité augmentée. (VanAart *et al.*, 2010) explore l'utilisation d'ensembles de données liées afin de rechercher et de consulter des informations sur le patrimoine culturel. A partir de la localisation et de la direction vers laquelle se dirige l'utilisateur, les points d'intérêts situés à proximité sont identifiés à partir du nuage du Web des Données liées. Les auteurs se sont concentrés ici sur la fusion et l'alignement de plusieurs ressources associées à un même point d'intérêt, ainsi qu'à la correction par enrichissement sémantique des données erronées (erreurs dans les géo-références de données, les annotations, etc.).

En conclusion, ces différents projets de recherche montrent le potentiel de l'utilisation du Web des Données dans des applications mobiles où l'accès à

[2] T. Berners-Lee. Linked data - design issues. Accédé à http://www.w3.org/DesignIssues/LinkedData.html, 2013-04-22, 2006.

l'information repose sur la localisation de l'utilisateur. Cependant, le risque d'afficher *toutes* les informations géo-localisées en particulier dans les applications de réalité augmentée est d'exposer rapidement l'utilisateur nomade (Olsson *et al.*, 2012) à une surcharge cognitive.

3. Contribution

Dans nos précédents travaux, nous avons proposé une plate-forme de réalité augmentée mobile, ARCAMA-3D, qui s'appuie sur l'utilisation d'un modèle 3D léger (Aydin *et al.*, 2012a). Les objets du monde réel sont représentés par un modèle géométrique 3D semi-transparent superposé à la vue réelle captée par la caméra du dispositif mobile (voir Figure 2). Ainsi, le monde réel, représenté par ces objets 3D, devient l'interface utilisateur, et en interagissant avec ces objets, l'utilisateur peut poser des questions comme "*de quoi s'agit-il ?*", "*Y-a-t-il quelque chose d'intéressant (pour moi) autour de moi ?*", *etc.*

Figure 2. Une représentation de l'application ARCAMA-3D. (a) sur la gauche: vue réelle (b) sur la droite: un modèle 3D de l'environnement est superposé à la vue réelle avec des « objets rouges » capables d'interaction qui donnent accès à l'information.

L'application ARCAMA-3D acquiert continuellement les données de géolocalisation ainsi que d'orientation à l'aide des capteurs embarqués (GPS, accéléromètre, gyroscope, etc.) de l'appareil mobile (smartphone, tablette PC, *etc.)* sur lequel elle s'exécute. Nous assurons l'alignement du modèle 3D dans la scène réelle en affinant les données captées par utilisation d'un filtre de Kalman et en fusionnant ces données (Kalman, 1960). La fusion des données captées permet une estimation précise de la position, et, par conséquent, de superposer le modèle 3D à la vue réelle, et de maintenir cette superposition tout au long du déplacement de l'utilisateur (Aydin *et al.*, 2012b). ARCAMA-3D considère la plupart des objets de la réalité augmentée comme étant transparents, au sens propre, et ne les affiche que lorsque des informations significatives sont disponibles pour ces objets, suivant en cela la recommandation de (Olsson *et al.*, 2012). Les objets affichés ont différentes

couleurs pour indiquer à la fois qu'ils sont interactifs et qu'ils contiennent des informations intéressantes pour l'utilisateur relativement à un thème d'intérêt donné.

Les modèles 3D exploités par ARCAMA-3D jouent donc ici un rôle clef. Ils sont en effet le support des objets de réalité augmentée qui donnent accès à l'information. Nous proposons de représenter ces modèles 3D par une ontologie qui rende compte des caractéristiques propres aux modèles 3D (niveaux de détail, période couverte, *etc.*) et permette de les lier aux descriptions ontologiques des objets que ces modèles 3D représentent, établissant par là même un lien précieux entre des objets de la réalité décrits sur le Web des Données et un ou plusieurs modèles 3D où ces objets apparaissent. Pour atteindre cet objectif, nous intégrons la description géométrique (modèle 3D) d'une entité du monde réel correspondant à un objet d'intérêt à proximité (OIP) (Aydin *et al.*, 2012a) d'ARCAMA-3D dans un graphe RDF. Ce graphe utilise un vocabulaire contrôlé défini par une ontologie OWL (arcama-owl). Cette ontologie définit un modèle générique pour les entités ARCAMA-3D. Le graphe RDF correspondant à une entité peut être considéré comme une instance de ce modèle générique. Rappelons que la représentation des modèles 3D utilisés par ARCAMA-3D présente deux avantages : 1) elle permet de relier - et donc d'enrichir - ces modèles 3D avec d'autres ensembles de données structurées (en particulier DBpedia); 2) symétriquement, elle offre la possibilité à d'autres ensembles de données du nuage du Web des Données, d'associer des modèles 3D aux informations qu'ils contiennent.

3.1. Modèle 3D de données

Nous avons créé un modèle de données 3D dans lequel nous stockons les représentations géométriques des entités du monde réel et les informations qui leur sont associées. Lors de la modélisation, nous avons fait en sorte que le modèle proposé soit extensible, et que la représentation d'un OIP puisse être associée à des informations temporelles et thématiques. Aussi, dans la conception d'un modèle de données 3D, nous pensons nécessaire d'intégrer des caractéristiques *spatiales (quelle(s) portion(s) de l'espace recouvre ce modèle 3D), temporelles (à quelle(s) époques(s) correspond ce modèle 3D ?)*, et *thématiques (quel(s) objet(s) du monde réel décrit ce modèle 3D ?)* dans le modèle. Nous adoptons une approche semblable à celle décrite dans (Parent *et al.*, 1997) qui propose un modèle conceptuel spatio-temporel mettant l'accent sur l'orthogonalité des instances *spatiale*, *temporelle* et *structurelle*. Dans ce travail, les auteurs défendent cette approche pour des objets géographiques non urbains (rivières, lacs, cantons, *etc.*) ou pour des données spatiales 2D relatives à des objets urbains (la surface d'un château, *etc.*). Nous prolongeons cette approche dans notre modèle aux données 3D afin de permettre l'exploitation d'informations spatiales, temporelles et thématiques liées aux OIP pendant l'expérience de RA de l'utilisateur.

Pour illustrer ce modèle de données 3D à l'aide d'un exemple concret, nous choisissons un objet bien connu du monde réel : l'édifice *Sainte-Sophie*, à Istanbul en Turquie. Le choix de cette entité nous paraît tout indiqué, l'architecture mais

également le rôle (la fonction) de ce bâtiment ayant toutes deux évolué au cours du temps.

La ressource DBpedia représentant *Sainte-Sophie*[3] lui attribue les rôles successifs de *cathédrale orthodoxe*, *mosquée* et *musée* (voir Figure 3). Cependant, aucune indication liée à la période de ces changements fonctionnels, ni à ses modifications architecturales, n'est présente dans cette description. Le but de notre modèle de données est de combler en partie ces lacunes. Ainsi, le modèle proposé prend en compte à la fois l'évolution temporelle dans l'espace (géométrie) et les données thématiques (attributs) de tout OIP décrit. Par conséquent, pour un OIP donné, nous avons besoin de représentations multiples de l'objet, chacune d'entre elles correspondant à un changement selon la dimension *thématique* et/ou selon la dimension *géométrique*. La Figure 3 montre cette évolution pour *Sainte-Sophie*, et les différentes représentations (R_1, R_2, R_3, R_4) que le modèle devra supporter pour cet OIP, chaque représentation correspondant à un changement architectural (géométrique) et/ou fonctionnel (rôle) de l'OIP.

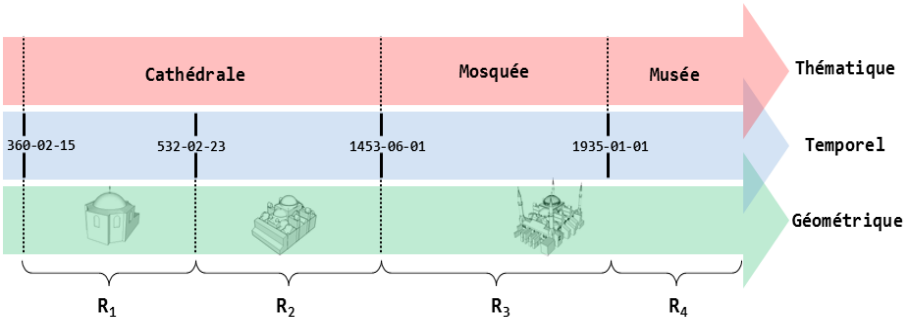


Figure 3. Les représentations de Sainte-Sophie, selon les changements thématiques et géométriques.

3.3. L'ontologie arcama-owl

Afin de construire ces représentations multiples, nous avons défini une ontologie OWL qui décrit un modèle générique pour un OIP. Le diagramme de classes UML de la Figure 4 montre les différents concepts (classes) et les relations entre ces concepts (propriétés) définis dans cette ontologie.

Ce modèle de classes (et ses instanciations) joue un rôle pivot : il permet la liaison entre les modèles 3D associés aux OIP et les informations relatives à ces OIP que l'on trouve sur le Web des Données. Les classes qui le constituent sont donc :

– La classe `Entity` correspond à un OIP (l'entité du monde réel Sainte-Sophie, par exemple). Ce modèle permet de regrouper plusieurs représentations temporelles

[3] http://dbpedia.org/resource/Hagia_Sophia/

de la même entité (voir classe TemporalRepresentation). L'association `represents` permet d'interconnecter ce modèle avec d'autres ressources décrites dans d'autres *datasets* publiés sur le web des données (par exemple, la ressource Dbpedia pour *Sainte-Sophie*).

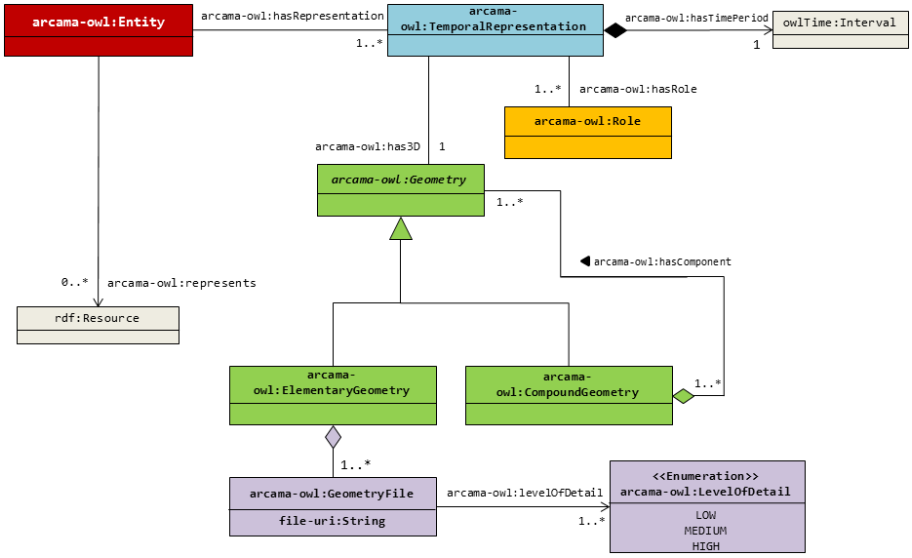


Figure 4. Diagramme de classes UML de l'ontologie arcama-owl.

– La classe `TemporalRepresentation` est au cœur de notre modèle. Elle correspond à une représentation de l'entité pour une période donnée. Cette période est décrite par un intervalle OWL-Time[4] dans l'association `hasTimePeriod`. Une occurrence temporelle regroupe les attributs spatiaux et thématiques décrivant l'entité au cours de cet intervalle de temps. Ces attributs sont décrits respectivement par des classes `Geometry` et `Role`. Toute modification des informations `Geometry` et/ou `Role` nécessite la création d'une autre `TemporalRepresentation` associée à un nouvel intervalle de temps.

– L'association `hasRole` permet de définir le(s) rôle(s) d'un objet `TemporalRepresentation` au travers de la classe `Role` qui détient des propriétés sur la fonction de l'objet. Cette information est utilisée pour classer les représentations temporelles selon les critères thématiques sélectionnés par l'utilisateur. En fonction de ceux-ci, elle permet de filtrer les représentations temporelles qui seront affichées dans l'interface de réalité augmentée et d'autoriser ou non l'accès aux informations supplémentaires liées à l'entité modélisée. Tous les rôles autorisés ici sont définis par une autre ontologie que nous avons créée et qui compte actuellement 125 classes correspondant aux différentes catégories de

[4] http://www.w3.org/TR/owl-time/

structures architecturales, par exemple, les édifices religieux (*cathédrale*, *mosquée*, *temple*, *etc.*), les sites historiques (*château*, *monument*, *pyramide*, *etc.*). Afin de faciliter la liaison avec les ressources issues d'autres jeux de données définis à l'aide de leur propre ontologie, une mise en correspondance entre les rôle et les classes de ces différentes ontologies est définie. Initialement, nous avions envisagé de nous baser sur les rôles utilisés dans l'ontologie de DBpedia[5], mais il s'est avéré, d'une part que la structuration proposée par DBpedia n'était pas satisfaisante, et, d'autre part, que le typage des ressources pouvait être de qualité très variable (selon la qualité des informations fournies par les contributeurs Wikipédia). L'utilisation de notre propre ontologie de domaine et la mise en place de mécanismes de médiation nous a permis de nous affranchir de ces contingences.

– Pour décrire la géométrie associée à une instance temporelle, nous utilisons le patron de conception composite (Gamma *et al.*, 1995) qui permet de combiner des géométries en une structure arborescente qui représente une hiérarchie partie-tout (*whole-part*). `Geometry` est donc une classe abstraite avec deux sous-classes `ElementaryGeometry` et `CompoundGeometry`.

– La classe `CompoundGeometry` permet de définir une géométrie complexe composée de plusieurs géométries différentes qui peuvent être des géométries complexes ou des géométries élémentaires.

– La classe `ElementaryGeometry` décrit une feuille dans la structure arborescente des géométries. Étant donné que pour un objet précis, nous voulons être en mesure de proposer différents modèles 3D à différents niveaux de détail, la classe `ElementaryGeometry` regroupe un ou plusieurs objets `GeometryFile`. Chaque objet `GeometryFile` est décrit par une URI qui permet d'accéder à un fichier physique contenant une description de l'objet 3D dans un format usuel (KML, CityGML, VRML, COLLADA *etc.*).

– La classe `LevelOfDetail` décrit le niveau de détail d'un `GeometryFile`. Nous avons défini trois niveaux : *LOW* correspond à un prisme rectangulaire (équivalent au niveau de détail LOD1 de CityGML), *MEDIUM* correspond à une géométrie 3D plus détaillée sans texture, *HIGH* correspond aux objets 3D texturés (il peut même représenter un objet scanné au laser).

La Figure 5 présente une partie du graphe RDF correspondant à une instanciation de ce modèle pour notre exemple précédent de *Sainte-Sophie*. Afin d'en faciliter l'interprétation en relation avec le modèle présenté précédemment, les nœuds ressources sont colorés avec une couleur identique à celle utilisée pour la classe arcama-owl à laquelle ils appartiennent (Figure 4).

Dans ce graphe RDF (voir Figure 5), *Sainte-Sophie* est représentée par deux instances temporelles: model:HagiaSophia#TI1 et model:HagiaSophia# TI2. La première est l'instanciation du modèle pour l'intervalle R_2 de la Figure 5, et correspond à un changement de géométrie. La seconde représente le modèle pour l'intervalle R_3 et correspond à la fois à un changement de rôle et à un changement de géométrie.

[5] http://mappings.dbpedia.org/server/ontology/classes/

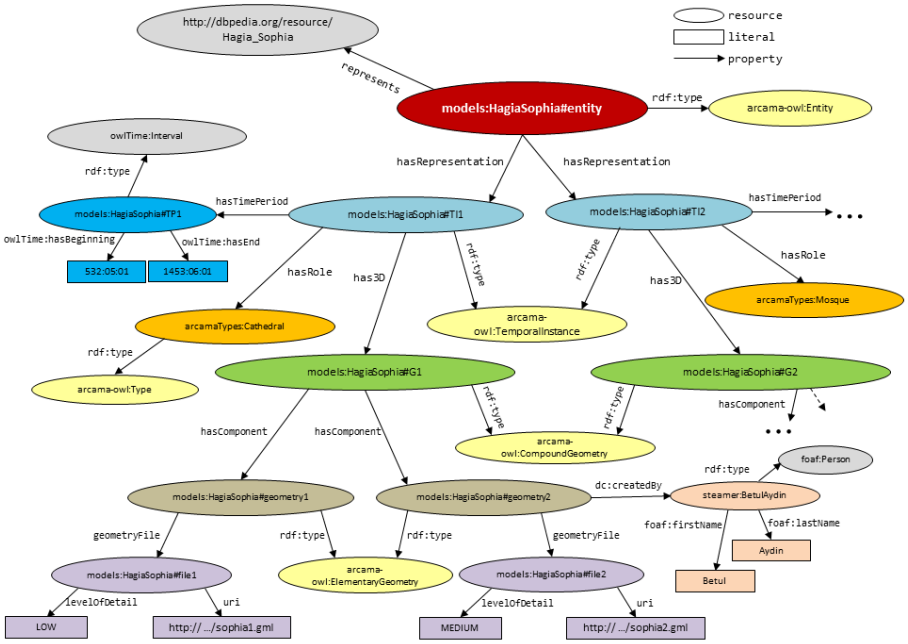


Figure 5. Graphe RDF représentant Sainte-Sophie avec différentes instances temporelles.

4. Mise en Oeuvre

La description RDF des entités ARCAMA-3D, ainsi que les fichiers correspondant aux géométries 3D sont stockés sur un serveur (ARCAMA-3D serveur). Pour l'enregistrement des graphes RDF, nous utilisons le triplestore Parliament[6] (il peut être interrogé via des requêtes SPARQL ou accédé directement via l'API Java Jena[7]). Dans ces graphes RDF, les URIs correspondant aux objets géométriques désignent des fichiers (au format .obj) accessibles via un serveur http (Apache). Une application destinée aux développeurs de nouveaux modèles 3D permet leur dépôt sur le serveur ARCAMA-3D : téléversement des fichiers géométrie sur le serveur http, insertion du modèle arcama-owl dans le triplestore Parliament, liaison avec d'autres jeux de données (DBpedia en particulier).

L'application ARCAMA-3D se présente alors comme une application client/serveur dont le fonctionnement est schématisé par la Figure 6.

[6] http://parliament.semwebcentral.org/, Parliament propose en particulier une implémentation expérimentale de GeoSPARQL qui propose des extensions geospatiales à SPARQL
[7] http://jena.apache.org/

Premièrement, le dispositif mobile (client ARCAMA-3D), se connecte au serveur ARCAMA-3D (1). En utilisant les informations de géolocalisation et les préférences de l'utilisateur (les Rôles qui l'intéressent), le serveur interroge (via des requêtes SPARQL) les données RDF annotant les modèles 3D (2) et retourne aux clients les entités susceptibles de véhiculer des informations intéressantes (3). Les géométries 3D correspondantes sont alors superposées à la vue réelle sous une forme semi-transparente avec une couleur différente selon les informations qui leur sont associées. Ensuite, lorsque l'utilisateur sélectionne l'un des objets 3D de la réalité augmentée (technique de picking 3D), l'identification de celui-ci est transmise au serveur (4) qui, grâce aux liens établis dans le modèle arcama-owl, recherche dans le web des données les informations pertinentes (requêtes SPARQL) (5). Ces informations sont transmises en retour à l'application cliente (6) qui peut les proposer alors à l'utilisateur.

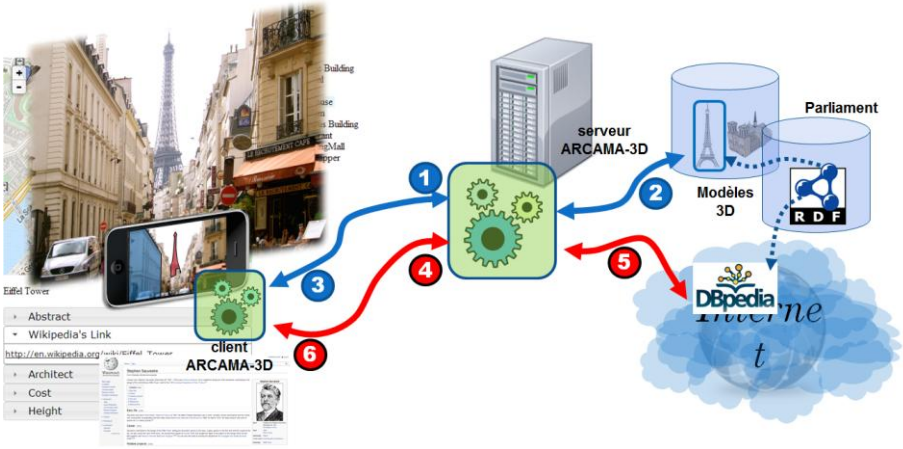


Figure 6. L'architecture de l'application ARCAMA-3D

5. Conclusion et travaux futurs

De nouveaux jeux de données sont continuellement ajoutés au nuage du Web des Données qui constitue une base d'informations sémantiquement liées. Nous décrivons dans cet article un système dédié à des applications mobiles à base de réalité augmentée, qui inclut une représentation spatiale, temporelle et thématique permettant de lier les modèles 3D aux informations disponibles sur le nuage du Web des Données et relatives aux entités du monde réel que ces modèles 3D décrivent. Cette interconnexion avec le Web des Données ne contribue pas seulement à représenter des ensembles de données 3D sur le nuage, mais améliore également le processus d'intégration de données du système de réalité augmentée.

Les aspects temporels supportés par le modèle proposé rendent compte de l'évolution des objets du monde réel dans leur forme, fonction, localisation, etc. Dans la continuité de notre projet, nous comptons améliorer la représentation de

l'identité des objets du monde réel et la caractérisation de leur évolution dans le temps, en adaptant des travaux précédents travaux précédents (Plumejeaud, 2010).

Remerciements : Cette recherche est financée par la Région Rhône-Alpes.

Bibliographie

Aydin B., Gensel J., Calabretto S., Tellez B., (2012). *Arcama-3d - a context-aware augmented reality mobile platform for environmental discovery*. In Proceedings of the 11th International Symposium on Web and Wireless Geographical Information Systems: W2GIS'12, pages 17-26

Aydin B., Gensel J., Calabretto S., Tellez B., (2012). *Arcama-3d: Une application mêlant réalité augmentée et 3d pour la présentation d'informations contextuelles*. In 8èmes journées francophones Mobilité et Ubiquité (UBIMOB), Anglet, France, pages 34-41.

Azuma R. (1997) *A survey of augmented reality. Presence: Teleoperators and Virtual Environments*, 6(4):355-385.

Becker C., Bizer C., (2009). *Exploring the geospatial semantic web with dbpedia mobile*. *Web Semant.* 7(4):278-286.

Gamma E., Helm R., Johnson R., Vlissides J., (1995). *Design patterns: elements of reusable object-oriented software*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Heath T., Bizer C., (2011). *Linked Data: Evolving the Web into a Global Data Space*. 1st edition, Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

Kalman R. E. (1960). *A New Approach to Linear Filtering and Prediction Problems*. Transactions of the ASME - Journal of Basic Engineering, (82 (Series D)):35-45.

Olsson T., Kärkkäinena T., Lagerstama E., Ventä-Olkkonen L., (2012). *User evaluation of mobile augmented reality scenarios*. JAISE, 4(1):29-47.

Parent C., Spaccapietra S., Zimanyi E., Donini P., Plazanet C., Vangenot C., Rognon N., Crausaz P. A., (1997). *MADS, modèle conceptuel spatio-temporel*. Revue Internationale de Géomatique, 7(3-4):317-352.

Plumejeaud C. (2011). *Modèles et méthodes pour l'information spatio-temporelle évolutive*. Thèse en informatique, Université de Grenoble.

Reynolds V., Hausenblas M., Polleres A., Hauswirth M., Hegde V., (2010). *Exploiting Linked Open Data for Mobile Augmented Reality*. In W3C Workshop: Augmented Reality on the Web, Barcelona, Spain.

Wilson M., Russell A., Smith D. A., Owens A., Schraefel M. C., (2005). *mSpace mobile: A mobile application for the semantic web*. In Proceedings of the End User Semantic Web Workshop at ISWC'05.

SESSION 5 :

Modélisation et gestion des données spatiales

Optimisation des performances des opérations d'agrégation au sein des entrepôts de grilles spatialisées

Myoung-Ah Kang[1], Mehdi Zaamoune[2], François Pinet[2], Sandro Bimonte[2], Philippe Beaune[3]

1. LIMOS-UMR CNRS 6158, ISIMA, Université Blaise Pascal, Clermont-Ferrand
2. Irstea, UR TSCF, Clermont-Ferrand
3. Agaetis, Clermont-Ferrand

RÉSUMÉ.
Le problème du stockage et du requêtage de gros volumes de grilles spatialisées est un verrou à traiter. Dans cet article, nous proposons une méthode pour optimiser les requêtes visant à agréger des grilles matricielles stockées dans des bases de données. Dans notre approche, nous proposons d'estimer le résultat réel plutôt que calculer le résultat exact. Ceci permet de gagner en temps d'exécution. Un avantage de notre méthode est qu'elle ne nécessite pas de re-implémenter une partie du système de gestion de base de données. Notre approche se base sur une nouvelle structure de données et sur un modèle spécifique de requêtes SQL. Nous appliquons ici nos travaux aux entrepôts de données relationnels.

MOTS-CLES : Données spatiales, Base de données.

CES TRAVAUX SONT FINANCES PAR LA REGION AUVERGNE, LE FEDER, LA SOCIETE AGAETIS ET IRSTEA.

1. Introduction

Les grilles (régulières ou non) constituent un format des données spatiales classiques (Laurini et Thompson, 1992). Dans cette représentation, chaque cellule est associée à une mesure (habituellement une valeur numérique). Par exemple, dans les cas les plus simples, les cellules d'une grille matricielles peuvent être représentées par des pixels où les couleurs correspondent à différentes valeurs (ou intervalles de valeurs) d'une mesure ; par exemple des températures, des données sur la qualité atmosphérique, etc. Aujourd'hui, dans de nombreux domaines d'activité, de plus en plus de grilles sont produites. La quantité de données disponibles ne fait qu'augmenter. Généralement, ces données peuvent être mesurées par des moyens de télédétection ou des capteurs terrestres, produites grâce à des

relevés manuels ou des enquêtes, etc. ; il s'agit alors le plus souvent de mesures de phénomènes observés. Ces données peuvent être aussi produites par simulation ; elles peuvent alors porter sur des phénomènes passés, présents (mais non mesurés ou mesurables en pratique) ou des prévisions. L'accroissement des performances de calcul des ordinateurs a permis de considérablement raffiner les échelles spatiales et temporelles des résultats de simulation. Dans de nombreux domaines, les données simulées ont des pas de temps et des échelles spatiales toujours plus fines. Le problème du stockage et de l'analyse des gros volumes de grilles est un défi à relever. Dans le présent article, nous contribuons à ce domaine, en proposant une méthode d'optimisation des requêtes portant sur les grilles (matricielles) dans les bases de données (BD). Nous appliquons ici nos travaux aux entrepôts de données (relationnels) (Pinet et Schneider, 2010; Thomsen, 1997) ; notre contribution pourrait néanmoins être aussi utilisée dans les BD « classiques ». A noter aussi que l'objectif de l'article est d'optimiser un certain type d'agrégation portant sur une structure de données bien particulière ; ces points sont décrits à la section 2. La section 3 décrit notre méthode d'optimisation. Les sections 4 et 5 proposent une évaluation de cette méthode. La section 6 conclut en donnant des perspectives de recherche.

2. Problématique

2.1 Problème de performance

Les entrepôts de données (ED) constituent un moyen moderne pour stocker et agréger les informations (Pinet et Schneider, 2010; Thomsen, 1997). Cette technologie offre tous les avantages des bases de données (sécurisation, sauvegarde, contrôle d'intégrité, etc.), tout en facilitant et optimisant les opérations d'agrégations sur les données (typiquement l'équivalent des opérations de GROUP BY en SQL). Un exemple d'opérations d'agrégation peut être le calcul de la somme des ventes en Euros de plusieurs magasins, agrégées par mois et/ou par région de France. Dans l'esprit des ED et des systèmes OLAP associés (« On-Line Analytical Processing ») (Thomsen, 1997), l'objectif est de fournir très rapidement les données agrégées à l'utilisateur. Avec des outils OLAP, l'utilisateur visualise et « explore » les données de manière itérative ; à noter que l'affichage des données spatiales sur des cartes est rendu possible grâce aux outils « Spatial OLAP » (Rosemarie McHugh *et al.*, 2009; Pestana *et al.*, 2005). Avec ces outils l'utilisateur combine différents critères d'agrégation (par ex. agrégations des données par jour, semaine, mois, année, ville, département, région, pays, type de produits, etc.), puis change ses critères d'agrégation pour obtenir une autre vue sur les données. Dans ce type d'outils centrés sur l'interaction avec un utilisateur, les performances en termes de temps de réponse aux requêtes sont essentielles. Actuellement, les performances des ED ne sont pas adaptées pour agréger les données de type grille spatiale. Nous allons illustrer les problèmes que nous abordons sur ce point dans cet article.

Supposons que l'on souhaite agréger p grilles matricielles (de taille $m*n$) en utilisant la fonction de somme. Chaque grille couvre exactement la même région

spatiale. Le résultat de l'agrégation sera une seule grille de taille *m*n*. Une méthode d'agrégation simple est de faire la somme de chaque cellule (*x*,*y*) :

$$Grille_résultat(x,y) = \sum_{i=1}^{p} Grille_i(x,y)$$

Ceci reviens à une somme de matrices (chaque grille étant une matrice). Si on prend maintenant la fonction moyenne, la formule sera :

$$Grille_résultat(x,y) = (\sum_{i=1}^{p} Grille_i(x,y))/p$$

La figure 1 illustre l'agrégation de 3 petites grilles avec la fonction moyenne.

Moyenne (

1	2	2
1	1	3
1	7	1

;

3	2	4
1	1	6
1	3	1

;

2	2	6
1	1	6
1	2	1

)

=

2	2	4
1	1	5
1	4	1

Figure 1. Moyenne de 3 grilles

Dans les ED classiques, les *p* objets à agréger sont habituellement simples (par exemple on agrège *p* montants de ventes). Dans le cas des ED de grilles, contrairement aux ED classiques, les *p* objets à agréger sont composites ; ce sont des grilles composées de plusieurs cellules. Ceci pose très souvent des problèmes de performance. Il faut en effet agréger *m*n* valeurs pour obtenir une grille résultat. Supposons que l'on ait comme données de base, une grille de taille 500*500 par jour (donc des grilles de petite taille). Si l'on veut fournir à l'utilisateur un ensemble de grilles agrégées par mois sur les derniers 10 ans (donc 120 grilles de taille 500*500), il faudra au final agréger 30 millions de valeurs (=500*500*12*10). On voit bien les problèmes de performance introduits par la spatialité des données à agréger ; les ED actuels ne sont pas prévus pour l'agrégation de grilles.

Nous proposons donc dans cet article une nouvelle méthode pour optimiser le temps des agrégations de grilles matricielles (et de même taille) dans les entrepôts. Notre méthode se base sur le calcul d'une estimation de la grille agrégée, estimation moins coûteuse à réaliser en temps que le calcul du résultat exact. Avec notre approche, il est possible de régler la précision des résultats en fonction des performances souhaitées ; plus les résultats produits par notre système sont proches du résultat réel et plus le gain en temps d'exécution des requêtes diminuera (et

inversement). Nous traitons dans cet article les fonctions d'agrégation de somme et de moyenne.

2.2 Travaux existants portant sur l'optimisation de l'agrégation des grilles spatiales dans les bases de données

Les opérations sur les grilles spatialisées (somme, moyenne, etc.) sont habituellement formalisées dans une « map algebra ». L'utilisation des grilles (ou des champs continus) dans les ED ou les base de données est un sujet souvent abordé – voir par exemple (Câmara *et al.*, 2005; Gómez *et al.*, 2012; R. McHugh, 2008; R. McHugh *et al.*, 2007). Néanmoins très peu de travaux ont porté sur l'optimisation du processus d'agrégation des grilles. Dans (Guitiéreez et Baumann, 2008), il est question d'agrégation de grilles dans les bases de données en considérant une pré-agrégation des données (c'est-à-dire un pré-calcul et pré-stockage de résultats agrégés). Les problèmes évoqués sont liés à la découverte des parties des données pré-agrégées qui peuvent être utiles à une requête. La limite de l'approche réside dans la nécessité de stocker des données supplémentaires (pré-agrégées) ; surtout sachant que les volumes de stockage des données grilles (même non pré-agrégées) posent déjà des problèmes. Dans (Zhang *et al.*, 2004), les auteurs considèrent une agrégation de plusieurs cellules appartenant à une même grille – ce qui constitue un cas différent de celui traité dans notre article.

3. Notre méthode d'optimisation pour l'agrégation des grilles matricielles

L'agrégation, cellule par cellule (avec les fonctions de somme ou de moyenne) de p grilles matricielles de taille $m*n$ demande au moins $m*n*p$ opérations arithmétiques (additions). Notre approche consiste à réduire la valeur de p en n'utilisant que certaines grilles dans les calculs. Nous détaillons ci-dessous les différentes étapes de notre approche.

3.1 1er étape : La clusterisation

Dans la première étape de notre approche, nous regroupons les différentes grilles de l'ED en clusters selon leur similarité. Chaque cluster contiendra ainsi des grilles qui sont similaires (c'est-à-dire qui se « ressemblent »). Une grille ne doit appartenir qu'à un et un seul cluster. Dans le présent article, nous ne proposons pas de fonction de similarité particulière. Le choix de la fonction de similarité dépendra du type de grille considérée (de la nature des données). Ce choix sera effectué par un expert des données. Le résultat du regroupement en clusters dépend entièrement de la fonction de similarité utilisée. Les paramètres de la clusterisation peuvent par exemple être le nombre de clusters souhaité ou bien un pourcentage de similarité souhaité entre les grilles au sein de chaque cluster.

L'appartenance d'une grille à un cluster sera une information stockée de manière pérenne dans l'ED. Le modèle type de schéma logique d'ED que nous proposons est

présenté en Figure 2. La table Grilles stocke les identifiants des grilles ainsi que le n° cluster associé. Les cellules ainsi que leur *x,y* et leur valeur sont stockés dans la table Cellules. Le reste du schéma reprend la structuration en dimension classique des ED. Plusieurs tables de dimensions peuvent être définies. Il s'agit des différentes dimensions d'analyse. Par exemple, en pratique, elles peuvent correspondre à une table Temps, Espace, Type, etc. Plus de détails généraux sur la structure des ED peuvent être trouvés dans (Pinet et Schneider, 2010; Thomsen, 1997).

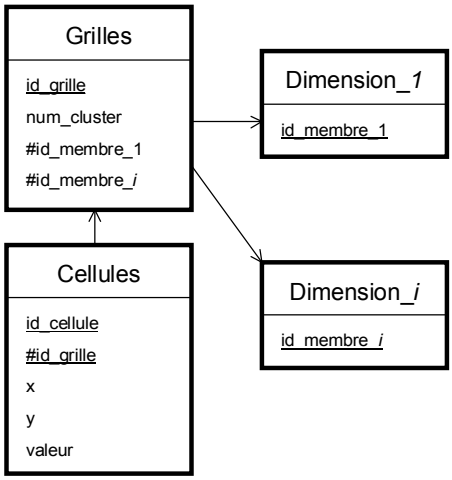


Figure 2. Schéma type d'ED pour le stockage des grilles

3.2 2ème étape : La requêtage

Dans les ED relationnels, les requêtes classiques d'agrégation peuvent s'exprimer en SQL, comme présenté dans la requête A ci-dessous ; on considère ici que l'on agrège les grilles de l'ED présenté en Figure 3, selon le jour (dimension temporelle) et la source d'acquisition de données (dimension « mode d'acquisition »).

Requête A :

```
SELECT jour, source_acquisition, id_cellule , sum(valeur)  FROM cellules C, grilles G
WHERE  G.id_grille  = C.id_grille  GROUP BY jour, source_acquisition, id_cellule
```

Cette requête fait la somme de toutes les grilles pour tout couple (jour, source_acquisition) pour chaque cellule (id_cellule).

6

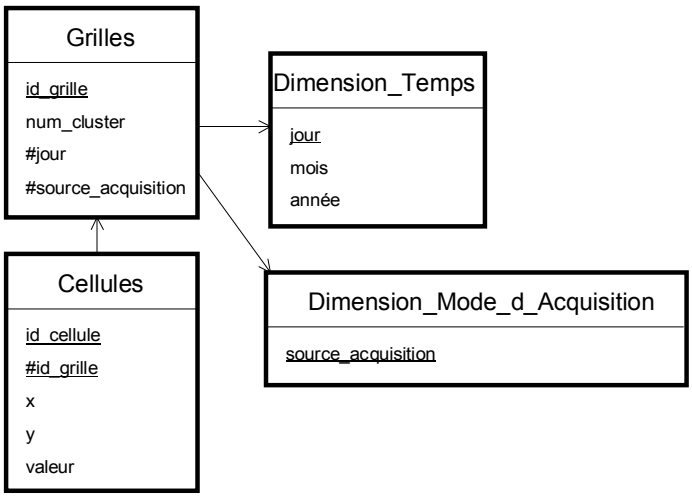


Figure 3. Un exemple d'ED de grilles

Nous proposons un autre modèle de requête qui va, cette fois, calculer une estimation du résultat final afin de réduire les temps de calcul. Ce modèle de requête est illustré avec la requête B ci-dessous qui utilise la clusterisation proposée dans la sous-section précédente. Cette requête va produire une estimation de la grille réelle calculée dans la requête A.

Requête B :

```
SELECT jour, source_acquisition, id_cellule, sum(valeur * nombre_cluster)
FROM cellules C, grilles G,
        ( SELECT max(id_grille) id_grille_rep, numero_cluster, count(*)  nombre_cluster
        FROM    grilles G
        GROUP BY  jour, source_acquisition, num_cluster ) as TempCluster
WHERE  G.id_grille = C.id_grille and G.id_grille = TempCluster.id_grille_rep
GROUP BY jour, source_acquisition, id_cellule
```

Dans cette requête, seules certaines grilles vont être utilisées pour l'agrégation. En fait, la requête B implémente l'algorithme présenté à la Figure 4. Intuitivement, pour calculer la grille associée au couple (jour, source d'acquisition), l'approche consiste à prendre une seule grille dans chaque cluster (plutôt que toutes), puis à

multiplier les valeurs de ses cellules par le nombre de grilles du cluster pour l'occurrence (jour, source d'acquisition).

Pour mieux comprendre la différence entre les requêtes A et B, prenons la clusterisation d'un ED présenté à la Figure 5. Les 30 grilles de l'ED ont été regroupées en 3 clusters (en fonction de leur similarité) : les clusters 1 et 2 contiennent 9 grilles chacun, le cluster 3 contient 12 grilles. Les identifiants des grilles sont indiqués sur la figure (de g01 à g30). Nous avons aussi indiqué les valeurs des attributs « jour » et « source d'acquisition » pour chacune des grilles.

Pour chaque couple (*jour*, *source d'acquisition*) noté *co* :

- $R_{co} = \{\}$
- Pour chaque cluster *i* :
  - S = l'ensemble des grilles associées à *co* dans le cluster *i*
  - *no* = le nombre de grilles contenues dans *S*
  - Si $S > 0$:
    - Prendre la dernière grille *g* dans *S*
    - Calculer une nouvelle grille rg_i contenant les valeurs des cellules de *g* multipliées par le nombre *no*.
    - $R_{co} = R_{co} \cup \{rg_i\}$
- Si $R_{co} = \{\}$ alors il n'y a pas de résultat pour le couple *co*
- Si $R_{co} <> \{\}$ alors la grille résultat pour le couple *co* sera égale à la somme de toutes les grilles rg_i contenues dans R_{co}

Figure 4. Algorithme qu'implémente la requête B

Pour le couple (jour 2, source A) par exemple, la requête A (qui n'utilise pas la clusterisation) fait la somme des grilles g25 à g30 du cluster 3. Dans la requête B, seule la grille g30 sera accédée et les valeurs de ses cellules seront multipliées par 6 pour obtenir la grille résultat ; car il y a 6 grilles associées au couple (jour, source d'acquisition) dans le cluster. Cette approche considère que puisque toutes les grilles d'un cluster sont similaires, il est possible de n'en prendre qu'une et de pondérer ses valeurs, plutôt que de faire la somme de toutes les grilles. On obtient donc une estimation du résultat réel mais on réduit le nombre d'opérations arithmétiques effectuées.

Etudions maintenant plus en détail l'écriture SQL de la requête B. La sous-requête TempCluster est utilisée pour connaître le nombre de grilles de chaque cluster pour chaque couple (jour, source d'acquisition). La colonne nombre_cluster du résultat de cette sous-requête correspond alors au nombre de grilles appartenant à chaque cluster pour un couple (jour, source d'acquisition) donné. Nous avons utilisé la grille possédant l'identifiant maximal comme grille représentative d'un cluster pour le couple (jour, source d'acquisition) ; la colonne id_grille_rep résulte de max(id_grille). Le choix de la grille représentative peut être facilement adapté selon le type ou la nature des données.

8

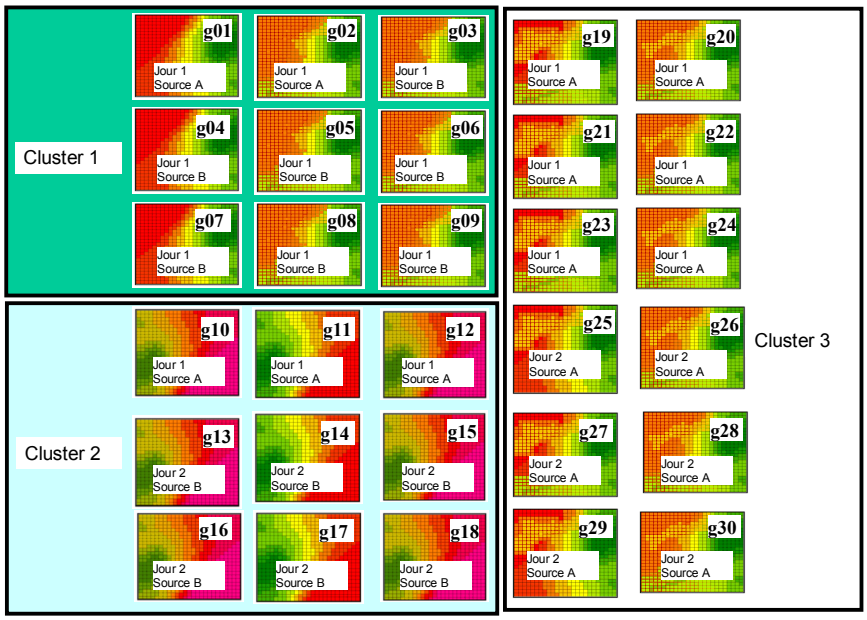


Figure 5. Un exemple d'ED de grilles clusterisées

Par exemple, voici le résultat de la sous-requête TempCluster sur les données de la Figure 5 :

Jour	Source d'acquisition	Numero_cluster	Id_grille_rep	Nombre_cluster
Jour1	SourceA	C1	g02	2
Jour1	SourceA	C2	g12	3
Jour1	SourceA	C3	g24	6
Jour1	SourecB	C1	g09	7
Jour2	SourceA	C3	g30	6
Jour2	SourceB	C2	g18	6

Dans ce tableau, afin de faciliter la compréhension de la requête, nous avons ajouté les colonnes « jour » et « source d'acquisition », qui ne font pas partie du résultat de la sous-requête Tempcluster.

Les grilles de chaque couple (jour, source d'acquisition) sont donc regroupées selon leur num_cluster. Les clusters appartenant au même couple (jour, source d'acquisition) sont coloriés avec la même couleur. Donc le premier couple bleu (Jour1, SourceA) possède trois clusters (C1, C2 et C3) et les autres couples possèdent un cluster chacun. Le même cluster peut exister dans plusieurs couples de

la requête, comme présenté dans le tableau, par ex. le cluster C1 existe dans le coupe (Jour1, SourceA) et le coupe (Jour1, SourceB).

Au total, la requête A réalisera les calculs ci-dessous ; comme indiqué plus haut, chaque somme de deux grilles implique en fait $m*n$ additions de cellules.

Somme des grilles pour le couple (jour 1, source A) :

g01+g02+g10+g11+g12+g19+g20+g21+g22+g23+g24, soit $10*m*n$ additions de cellules

Somme des grilles pour le couple (jour 1, source B) :

g03+g04+g05+g06+g07+g08+g09, soit $6*m*n$ additions de cellules

Somme des grilles pour le couple (jour 2, source A) :

g25+g26+g27+g28+g29+g30, soit $5*m*n$ additions de cellules

Somme des grilles pour le couple (jour 2, source B) :

g13+g14+g15+g16+g17+g18, soit $5*m*n$ additions de cellules

Donc un total de $26*m*n$ opérations arithmétiques.

La requête B réalisera les calculs ci-dessous ; ici, certaines valeurs de cellule sont multipliées par une constante, conformément à l'algorithme présenté plus haut.

Somme des grilles pour le couple (jour 1, source A) :

(g02*2)+(g12*3)+(g24*6), soit $2*m*n$ additions de cellules et $3*m*n$ multiplications.

Somme des grilles pour le couple (jour 1, source B) :

g09*7, soit $m*n$ multiplications.

Somme des grilles pour le couple (jour 2, source A) :

g30*6, soit $m*n$ multiplications.

Somme des grilles pour le couple (jour 2, source B) :

g18*6, soit $m*n$ multiplications.

Donc un total de $8*m*n$ opérations arithmétiques.

La requête B permet donc de réduire le nombre d'opérations arithmétiques effectuées. Comme nous le montrerons dans notre partie expérimentation, le gain obtenu en pratique peut être important.

Les exemples donnés ici portent sur l'utilisation de la fonction somme. L'optimisation que nous proposons peut aussi s'appliquer de manière directe avec la fonction de moyenne. Il suffit d'effectuer des opérations de division supplémentaires.

10

A noter, enfin, que notre technique d'optimisation fonctionne aussi lorsqu l'on utilise un critère d'analyse « plus élevé » au sein de la dimension ; par exemple, si on agrège par année et par source d'acquisition (plutôt que par jour et source) – voir la requête C. Ici, pour montrer comment adapter notre requête avec de nouvelle condition, nous avons aussi rajouté une condition supplémentaire dans le WHERE (nous sélectionnons les années supérieures à 2010).

Requête C :

```
SELECT annee, source_acquisition, id_cellule, sum(valeur * nombre_cluster)
FROM cellules C, grilles G, dimension_temps DT ,
        (SELECT max(id_grille) id_grille_rep, numero_cluster, count(*)  nombre_cluster
        FROM grilles G, dimension_temps DT
        WHERE annne > 2010 and G.jour = DT .jour
        GROUP BY annee, source_acquisition, num_cluster ) as TempCluster
WHERE   annne > 2010 and G.jour = DT .jour and
        G.id_grille = C.id_grille and G.id_grille = TempCluster.id_grille_rep
GROUP BY annee, source_acquisition, id_cellule
```

4. Analyse de la précision des grilles résultats

La précision des grilles résultats dépend des paramètres utilisés pour le processus de clusterisation. Généralement, plus le taux de similarité sur lequel se base la clusterisation est élevé, plus la précision du résultat est élevée (mais aussi moins le gain de performance est élevé). Ce réglage « performance vs. précision » devra être ajusté en fonction des applications. Nous avons commencé une étude théorique sur l'analyse de la précision des grilles résultats, étude que nous résumons ici.

Considérons qu'une clusterisation a produit C clusters contenant chacun p grilles, avec $C > 1$. Les grilles ont toutes $m*n$ cellules. Dans chaque cluster, t % des cellules des grilles sont toutes dans un intervalle de valeur $[i-s, i+s]$. La valeur de i peut varier d'un cluster à un autre. Les valeurs de m, n, t, s et p sont constantes d'un cluster à l'autre. Par contre, l'ensemble des cellules composant les t % peut varier d'un cluster à un autre. Avec cette configuration, nous avons identifié le « pire des cas ». Nous avons en effet déterminé l'écart maximal possible entre une grille résultat réelle et une grille résultat estimée, dans le cas de l'agrégation de la totalité des grilles. Nous traiterons ici d'une agrégation de type moyenne.

Nous avons déterminé qu'au pire des cas, nous sommes sûrs que $(C - (1/(1-t)))*m*n$ cellules du résultat final auront un écart de plus ou moins s avec le résultat réel, pour les cas où $C >= 1/(1-t)$. En effet, intuitivement, le pire des cas apparaît lorsque les $(1-t)$ % des cellules non similaire changent au maximum d'un cluster à l'autre. Néanmoins, la probabilité d'apparition de ce pire des cas est habituellement extrêmement faible. Nous continuerons notre étude théorique pour, entre autres, être

capable de calculer cette probabilité. Les expérimentations réelles présentées dans la section suivante ont montré une très bonne qualité des résultats en pratique.

5. Expérimentation

Nous avons réalisé une expérimentation des performances avec une machine de 4Go de RAM et un CPU i5 2,60 GHz. Nous avons utilisé 2411 grilles à la résolution 100 par 100. Une grille est donc composée de 10000 points. Deux grilles sont dans le même cluster si elles ont un certain % de cellules similaires. L'expérimentation a été réalisée sous Postgresql sur des données de qualité de l'air. Avant de lancer chaque type de requête, nous avons redémarré systématiquement la machine pour éviter « l'effet de cache » du système de gestion de BD ainsi que celui du système d'exploitation. Le temps d'exécution est le temps moyen de 5 exécutions de chaque requête. L'exécution de notre requête type sans utilisation des clusters a été faite (en moyenne) en 88515 ms.

Voici les temps d'exécution pour les cas où l'on utilise la clusterisation. Nous avons expérimenté quatre clusterisation, avec pour chacune différents nombres de clusters.

	Nombre moyen de grilles dans un cluster	Nombre de clusters	Temps d'exécutions
Cas 1	2,30	1048	53243 ms
Cas 2	2,79	862	43259 ms
Cas 3	5,67	425	27596 ms
Cas 4	10,045	240	17799 ms

Avec ce tableau, nous pouvons constater que l'augmentation du nombre moyen de grilles dans un cluster permet de réduire le temps d'exécution. Le gain en temps est important (voir le cas 4) même avec des clusters contenant très peu de grilles. Ces clusters restent donc petits et, par conséquent, ils ont donné des résultats proches du résultat réel. Le gain en temps devrait être encore meilleur si on augmente la taille des clusters et que l'on se permet des résultats plus éloignés des résultats réels.

6. Bilan et perspectives

Nous avons ici proposé un moyen d'optimiser les performances de certaines requêtes d'agrégation (moyenne et somme) sur les grilles, au sein des ED relationnels. Pour cela nous avons proposé d'estimer les résultats réels. A noter que l'utilisateur, après avoir vu une grille résultat estimée, peut souhaiter visualiser la grille réelle correspondante. Dans ce cas l'ED pourra toujours facilement la fournir (voir requête A). Avec notre méthode, l'utilisateur pourra en effet explorer rapidement les grilles estimées, puis s'arrêter sur certaines et demander les résultats exacts correspondants. Un avantage de notre technique est en effet que les résultats exacts peuvent toujours être calculés si on le souhaite (puisque les données de base

sont conservées). Un autre avantage de notre approche est que la clusterisation ne nécessite pas de modifier la structure des données (hormis l'ajout d'un numéro de cluster dans une table). Enfin, notre méthode ne nécessite pas de re-implémenter une partie du système de gestion de BD – notre structure de clusters est directement utilisable en SQL.

Nous travaillons actuellement sur les points suivants : **(1) Proposer une technique de réécriture de requêtes.** La requête B (utilisant la clusterisation) est plus compliquée à écrire que la requête A. L'idée serait de permettre de convertir automatiquement une requête sans clusterisation en une requête utilisant la clusterisation. Ceci permettrait de simplifier la tâche des utilisateurs et permettrait aussi d'assurer une compatibilité avec les systèmes OLAP relationnels existants (qui utilisent des modèles de requête n'utilisant pas la clusterisation). **(2) Adapter notre technique à d'autres fonctions d'agrégation. (3) Donner la possibilité de regrouper dans un même cluster des parties de grilles** (plutôt que des grilles entières). Afin d'améliorer la qualité des résultats estimés, différentes parties d'une même grille pourraient ainsi se retrouver dans plusieurs clusters. **(4) Permettre l'agrégation de grilles qui sont à des résolutions différentes**. **(5) Continuer les expérimentations** - en considérant différents types de clusterisation et différents types de données. **(6) Raffiner les études théoriques sur l'estimation des erreurs dans les résultats.**

Bibliographie

Câmara, G., Palomo, D., De Souza, R. C. M., De Oliveira, O. R. F. (2005). *Towards a generalized map algebra: Principles and data types.* Paper presented at the 7th Brazilian Symposium on GeoInformatics, GEOINFO 2005.

Gómez, L. I., Gómez, S. A., Vaisman, A. A. (2012). *A generic data model and query language for spatiotemporal OLAP cube analysis*.

Guitiéreez, A., Baumann, P. (2008). *Computing Aggregate Queries in Raster Image Databases Using Pre-Aggregated Data.* Paper presented at the WCECS 2008, October 22 - 24, 2008, San Francisco, USA.

Laurini, R., Thompson, D. (1992). *Fundamentals of spatial information systems*.

McHugh, R. (2008). *Intégration de la structure matricielle dans les cubes spatiaux, Université Laval, 128 p.*

McHugh, R., Bédard, Y., Hubert, F. (2007). *Potentiel de la structure matricielle pour optimiser l'analyse des données géodécisionnelles*. Paper presented at the Géocongrès Québec 2007.

McHugh, R., Roche, S., Bédard, Y. (2009). Towards a SOLAP-based public participation GIS. *Journal of Environmental Management, 90*(6), 2041-2054.

Pestana, G., da Silva, M. M., Bedard, Y. (2005). *Spatial OLAP modeling: an overview base on spatial objects changing over time.* Paper presented at the Computational Cybernetics, 2005. ICCC 2005. IEEE 3rd International Conference on.

Pinet, F., Schneider, M. (2010). Precise Design of Environmental Data Warehouses. *Operational Research, 10*(3), 349-369.

Thomsen, E. (1997). *OLAP Solutions: Building Multidimensional Information Systems*: John Wiley & Sons.

Zhang, J., Gertz, M., Aksoy, D. (2004). *Spatio-temporal aggregates over raster image data.*

Interfaçage de bases de données photographiques et géographiques par appariement de lignes

Youssef ATTIA[1], Thierry JOLIVEAU[1], †Eric FAVIER[2]

1. ISTHME-EVS, Université Jean Monnet, CNRS
6 rue Basse des Rives, 42023 F-Saint-Etienne cedex 02
youssef.attia@univ-st-etienne.fr

2. ENISE
58, rue Jean Parot
42023 F-Saint-Etienne cedex 02, France

RESUME. *Cet article propose une méthodologie pour assigner une position géographique à une photographie urbaine à partir de son seul contenu. Le principe est de comparer cette photographie avec des images 3D générées automatiquement à partir d'un modèle virtuel des bâtiments et du terrain dont on connaît la localisation. La correspondance se fait au moyen d'un processus d'appariement de lignes détectées dans la photographie et de celles extraites des images. Cette extraction est fondée sur l'usage d'une transformée de Hough. L'appariement est suivi d'une analyse statistique qui propose une estimation de localisation avec une précision géographique. Les résultats à partir d'une base de données de test constituée sur la ville de Saint-Etienne sont présentés et évalués et des pistes d'amélioration et de poursuite sont évoquées.*

ABSTRACT. *The aim of this paper is to describe a new method for assigning a geographical position to an urban picture based only on the content of the picture. The photograph is compared to a sample of geolocated 3D images generated automatically from a virtual model of the terrain and the buildings. The relation between the picture and the images is built through the matching between detected lines in the photograph and in the images. The lines extraction is based on Hough transform. This matching is followed by a statistical analysis to propose a probable location of the picture with an estimation of accuracy. The paper presents and discusses the results of an experiment with data about Saint-Etienne, France and ends with leads for improving and extending the method.*

MOTS-CLES : *SIG, photo, appariement de lignes, lignes de Hough, reconstitution 3D, étiquettes, ville*

KEYWORDS:*GIS, photo, matching lines, Hough lines, 3D reconstruction, labels, city*

2

1. Introduction

La démocratisation des appareils photos numériques et l'apparition du Web a contribué à la mise en ligne de centaines de millions de photographies. Ces dernières se sont propagées d'abord sur des sites personnels puis, plus récemment, sur des plates-formes contributives et sur les réseaux sociaux. De nos jours, les utilisateurs publient, partagent et commentent leurs photographies sur le Web. Ils annotent ces photographies avec des informations sur les circonstances, le lieu, le contenu ou tout autre élément décrivant ce qu'ils prennent en photo.

La plupart des plates-formes Web de partage des photographies, comme par exemple Panoramio, Zooomr, Lo.ca.lise.us ou Flickr, attachent une position géographique à chaque photographie (Torniai et al., 2007). Les logiciels grand public de retouche et classement de photographies tels Picasa de Google ont aussi rapidement proposé aux utilisateurs de localiser manuellement leurs photographies sur les globes virtuels de Google ou Yahoo par exemple. Le succès de ces solutions auprès des internautes, malgré le caractère fastidieux et approximatif des techniques de localisation manuelle, est bien le signe de l'intérêt que suscite actuellement la géolocalisation des photographies.

Les techniques actuelles de positionnement automatique lors de la prise de vue ont certaines limites liées aux erreurs de localisation du GPS en milieu urbain ou de précision dans la triangulation à partir des antennes GSM ou des bornes Wifi. Mais surtout, il reste un nombre important de photographies non localisées ou localisées approximativement. Dans de nombreux domaines, qu'il s'agisse des utilisations personnelles ou de bases de données professionnelles, le besoin d'outils de localisation automatique ou semi-automatique des photographies est patent. Nous avons proposé dans Attia et al. (2009a) et Attia (2012) une approche de géolocalisation des photographies par comparaison interactive de celles-ci avec les vues obtenues en naviguant dans un modèle 2,5D du paysage urbain, construit à partir d'un Modèle Numérique de Terrain auquel sont ajoutées des extrusions de bâtiment en fonction de leur hauteur. Un démonstrateur a été mis en œuvre qui permet de se déplacer dans l'environnement de synthèse jusqu'à obtenir un rendu analogue à celui d'une photo déterminée et donc de localiser celle-ci en lui affectant la position et les paramètres de la caméra virtuelle (Attia et al., 2009b). Cette méthode nécessite l'intervention de l'utilisateur et qu'il sache approximativement où les photographies ont été prises.

Nous proposons ici comme dans le travail de thèse une méthode pour parvenir automatiquement à une localisation de la photographie en la comparant avec un échantillon de vues de synthèse extrait automatiquement de cet environnement virtuel. Le principe est d'extraire des indicateurs comparables entre les photographies et les vues de synthèse de manière à les apparier puis de les localiser en se fondant sur les coordonnées des vues de synthèse les mieux appariées (Attia 2012). Le critère choisi est le nombre de lignes communes à la photographie et à la vue de synthèse.

Nous revenons d'abord sur les méthodes de géolocalisation automatique de photographies à partir de leur contenu qui existent dans la littérature, puis nous exposons notre nouvelle approche et notre processus de validation. Nous donnons enfin les résultats et les statistiques qui permettent de fournir une évaluation de l'efficacité de notre proposition et ses limites, avant de proposer des pistes d'amélioration.

2. Méthodes de géolocalisation automatiques

Pour qu'elle soit automatique, une méthode de géolocalisation ne doit nécessiter aucune intervention humaine dans l'ajout d'une information géographique à une photographie à partir de son contenu. Les méthodes proposées dans la bibliographie restent encore assez peu nombreuses.

Une équipe de chercheurs à l'Université de Maryland a développé une méthode semi-automatique d'annotation d'images (Suh et Bederson, 2007). Elle profite d'une photo géolocalisée pour retrouver les personnes de cette photographie dans d'autres photographies. Elle affecte à toute personne portant les mêmes vêtements dans d'autres photos un même emplacement géographique.

James Hays et Alexei A. Efros tentent de trouver l'information sur la position géographique d'une photographie en la comparant avec tous les éléments d'une base de données d'images géolocalisées construite à cet effet (Hays et Efros, 2008). Les photographies touristiques sont les seuls cas où cette technique est efficace, car elle nécessite la présence dans les photographies d'éléments emblématiques du paysage. Dans un sens inverse, Keita Yaegashi et Keiji Yanai utilisent la position géographique de la prise de la photographie pour aider à détecter le contenu (Yaegashi et Yanai, 2009).

Dans sa thèse, Moslah (2011) a produit un modèle urbain capable de détecter les différents éléments des façades comme les fenêtres, les corniches et les balcons. Ce travail a déployé une grammaire descriptive des éléments constituant un paysage urbain.

3. Approche proposée

Notre proposition est de réaliser un appariement entre la photographie dont on cherche la localisation et un échantillon de vues de synthèse tirées d'un modèle 3D de la ville ou du quartier concerné. Nous cherchons à retrouver la position et l'orientation exacte de la photographie en nous positionnant dans le monde virtuel de telle façon qu'on ait un contenu d'image analogue. C'est donc le contenu de l'image qui permet de déduire la localisation. Les valeurs de qualité d'appariement vont ensuite permettre de calculer une position à partir des coordonnées géographiques des photographies appariées et d'évaluer son erreur.

4

3.1. Principe de l'appariement

On utilise une base de données SIG qui contient les bâtiments du milieu urbain, l'élévation du sol au pied de chaque bâtiment et sa hauteur. Toutes ces données permettent de créer un monde virtuel représentatif du monde réel. Il est possible de positionner une caméra dans un emplacement quelconque de ce monde virtuel généré par la synthèse d'image.

Un contenu identique entre une photographie et une photo de synthèse n'est ordinairement possible que si les deux images partagent la même position, la même focale et la même direction de visée. Nous proposons de retrouver l'image de synthèse qui correspond à une photo réelle au moyen d'un appariement de lignes entre la photographie et une collection d'images de synthèse.

Notre procédure d'appariement comporte cinq phases :

1. Création du monde virtuel en se fondant sur une couche de données géographiques contenant les bâtiments.

2. Sélection d'images de synthèse aux alentours de la position présumée de la photographie réelle objet de l'appariement. Les images de synthèse sont réalisées idéalement avec un intervalle régulier de 2 mètres. 8 images sont prises dans chaque position avec une rotation de 45° entre elles. Le choix de cet intervalle est le résultat d'une expérimentation par la pratique. Il tient compte de la nature du tissu urbain.

3. Préparation de la photographie en éliminant le sol.

4. Essai d'appariement de la photographie traitée avec les différentes images de synthèse sélectionnées à la phase 2.

5. Affectation à la photographie des paramètres de positionnement des images de synthèse en appliquant une pondération fonction de la valeur de l'appariement .

Notre approche d'appariement pour fusionner les données géographiques avec les photographies est présentée en détail dans la figure 1.

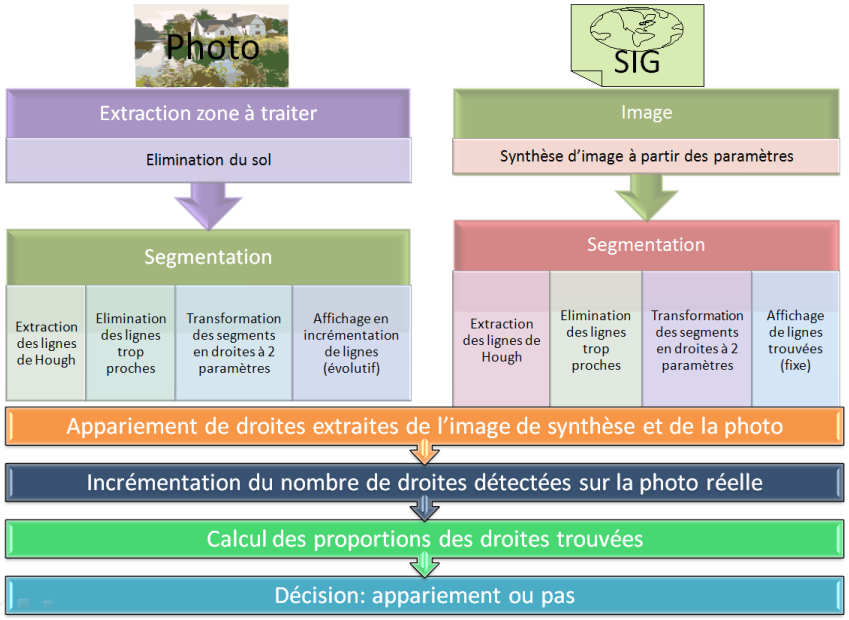


Figure 1. La démarche d'interfaçage par appariement de lignes

L'appariement des images consiste à passer d'une comparaison entre deux images à une comparaison entre deux listes de lignes. C'est une correspondance (« matching ») entre les lignes détectées dans une image de synthèse et celles détectées dans une photographie.

Nous avons choisi d'utiliser les lignes qui définissent assez bien les bâtiments dans une ville et de les apparier avec celles qui apparaissent sur les images de synthèse.

3.2. Méthode d'extraction de lignes

La méthodologie d'extraction de ligne choisie est la transformation de Hough. Paul V. C. Hough a proposé en 1962 une méthode qui permet de détecter dans une image les formes paramétriques comme les droites et les cercles (Hough, 1962 ; Maître 1985). Cette technique (Valentin, 2009) a été revue par d'autres chercheurs (Duda et Hart, 1972 ; Rosenfeldn, 1969) dix ans plus tard pour passer de la détection d'alignements sur un oscilloscope à une méthode de détection de lignes. La transformation de Hough permet de passer d'un point sur une image à une ligne droite dans un espace de paramètres (Duda et Hart, 1972).

Selon Cohen et Toussaint (1977), les expérimentations montrent que la transformée de Hough est une des meilleures méthodes de détection de lignes pour

6

une image bruitée. Pour Crowley (2010), c'est même une technique "optimale" pour détecter les lignes dans les images très bruitées. Elle permet par exemple de trouver les bords de routes avec un éclairage minimum (Kneepkens, 2005). Cependant, les éléments détectés par cette transformation ne sont pas des segments mais des droites, qui n'ont pas besoin d'être continues pour être détectées. En effet, cette méthode permet de retrouver les points alignés. En manipulant un seuil d'extraction, il est possible de détecter des lignes constituées de quelques points alignés ou de dizaines de points orientés selon la même direction. Nous désignons ce seuil par SP dans l'algorithme de la section suivante.

Un autre argument en faveur de l''utilisation des droites de Hough est lié à la représentation des lignes. En effet chaque ligne détectée est caractérisée parfaitement par seulement deux paramètres numériques (Figure 2) :

- Une distance ρ exprimée en pixels qui sépare la ligne de l'origine du repère. La valeur maximale que peut prendre ρ est celle de la diagonale de l'image.
- Une inclinaison θ, exprimée en radians et qui varie de $+\infty$ à $-\infty$.

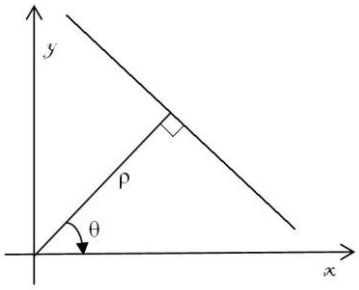


Figure 2. Représentation d'une ligne.

Cette légèreté de représentation est importante pour notre application, vu le nombre important d'images et de lignes à traiter et enregistrer.

Nous illustrons dans la figure 3 les lignes extraites dans une photographie et dans l'image de synthèse correspondante. Les lignes en communs sont en rouge.

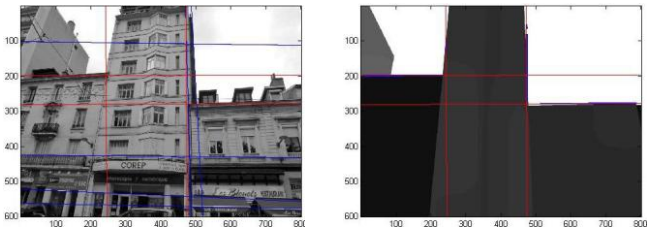


Figure 3. Exemple d'extraction de ligne

3.3. Algorithme

L'algorithme 1 est celui utilisé pour l'appariement de lignes. Il permet de calculer le nombre de lignes en commun entre une photographie et une image de synthèse.

Algorithme 1. Démarche d'appariement.

Appariement_de_lignes*(Image S ,image P) {*
Transformation de l'image S et P en niveaux de gris
Détecter LS lignes de l'image S
Filtrer les lignes trop rapprochées parmi LS
C ← faux
Tant que *C est faux* **faire**
    Détecter LP lignes de l'image P avec le seuil SP
    Filtrer les lignes trop rapprochées parmi LP
    similitude ← 0
    Pour tout *ligneP ∈ LP* **faire**
        Pour tout *ligneS ∈ LS* **faire**
            Si *(ρP ≈ ρS et θP ≈ θS)* **alors**
            similitude ← similitude +1
            fin condition
        fin boucle
    fin boucle
    Si *(similitude ≥* $\frac{4}{5}$*LP et LP ≤ 3 LS)* **alors**
        C ← vrai
    fin condition
    Si *(LP ≥ 3 × LS)* **alors**
    C ← vrai
    fin condition
    Seuil SP ← SP - 0.01
fin boucle
}

Si $\frac{4}{5}$ des lignes de la photographie sont trouvées dans une image de synthèse, nous considérons, d'après les expérimentations, que les deux images sont appariées. Nous arrêtons de diminuer le seuil d'extraction dès lors que nous détectons trois fois plus de lignes dans l'image de synthèse qu'il en existe dans la photographie.

4. Localisation de la photographie

L'utilisateur peut déclarer un quartier ou une position approximative d'une photographie ancienne scannée ou récente et demander au système de lui proposer une localisation probable de la photographie en s'appuyant sur les images de synthèse du voisinage.

8

L'appariement est lancé pour les images de synthèse se situant dans cette zone. La validation du positionnement géographique de la photographie se fait en utilisant deux indicateurs : le taux d'appariement (voir paragraphe 6 ci-dessous) et la distance séparant les images avec les meilleurs taux et position de la photographie. En combinant les deux, on peut donner un indicateur de la qualité de localisation de la photographie.

L'estimation du calcul de l'erreur de localisation se fait par le calcul du barycentre des localisations des images de synthèse appariées à la photographie, pondérées par le taux d'appariement. La position du barycentre pondéré (X_t, Y_t) est calculée selon la formule (1).

$$X_t = \frac{\sum_{t=1}^{n} t_{ix_i}}{\sum_{t=1}^{n} t_i} \quad Y_t = \frac{\sum_{t=1}^{n} t_{iy_i}}{\sum_{t=1}^{n} t_i} \qquad (1)$$

Les (x_i, y_i) correspondent aux coordonnées de la position des images de synthèse qui s'apparient avec un taux t_i à la photographie.

5. Validation du concept

5.1. Choix techniques

Nous avons utilisé l'outil ArcScene d'ESRI pour construire notre modèle 3D et les images de synthèse. Notons qu'il est possible de préciser les coordonnées X, Y et Z de l'observateur, les coordonnées X, Y et Z de la cible, l'angle de vision qui correspond à la focale et l'angle d'inclinaison verticale et horizontale (figure 4).

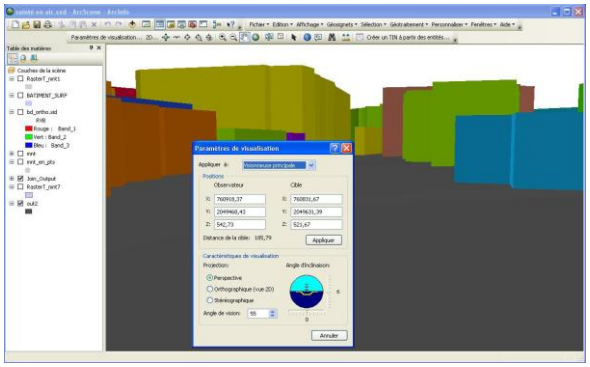


Figure 4. Paramètres de visualisation dans ArcScene. (Source : capture d'écran d'ArcScene. ESRI)

Le traitement des tableaux de lignes est effectué avec l'outil Matlab.

5.2 Traitement d'image sur les images de synthèse

Les images de synthèse subissent quatre traitements spécifiques.

1. *Transformation des images en couleur en niveaux de gris*. Cette étape est obligatoire pour permettre la détection de lignes par transformation de Hough.

2. *Augmentation du seuil pour la transformation de Hough*, pour moins détecter les marches d'escalier.

3. *Filtrage des lignes*. Nous avons créé un petit module permettant d'éliminer les lignes trop proches les unes des autres. Dans un milieu urbain les lignes correspondant aux bâtiments sont habituellement bien espacées. Ce module traite les lignes par ordre et cherche les lignes très voisines. Il les élimine en ne gardant que la première ligne détectée de chaque groupe de lignes voisines.

Le voisinage est défini par expérimentation. Deux lignes parallèles ayant un même degré d'inclinaison θ et dont la différence entre leurs deux distances ρ ne dépasse pas 10 pixels, sont considérées comme voisines.

4. *Augmentation du contraste*. Les couleurs des bâtiments peuvent être proches entre elles ou voisines de la couleur du sol surtout en transformant les images du mode couleur au mode niveaux de gris. L'augmentation du contraste permet une détection plus facile des contours et par conséquence des lignes.

5.3 Traitement d'image sur les photographies

Les photographies dans un environnement urbain ne contiennent pas seulement des bâtiments. On y trouve des arbres, des pancartes publicitaires, des signalisations, des voitures, des piétons, des cyclistes... Afin d'éliminer ces éléments qui se situent pour la plupart des cas au pied des bâtiments, nous avons utilisé la méthode de Derek Hoiem, Alexei A. Efros et Martial Hebert (Hoiem et al., 2005). Ceux-ci proposent une heuristique qui permet de segmenter l'image en trois parties (surface horizontale ou base, zones verticales et ciel). Quatre critères servent à déterminer l'appartenance d'une zone à une de ces trois parties: la localisation de la zone traitée dans l'image, la couleur, la texture et le point de vue (Hoiem et al., 2007).

5.4 Jeu de données

Nous avons constitué un jeu de données d'apprentissage qui regroupe 20 photographies prises dans le milieu urbain stéphanois composé de bâtiments anciens du XIXe siècle ou plus récents du XXe siècle et du XXIe siècle. Une fois l'analyse de ces photographies réalisée et l'arbre de décision créé, nous avons effectué la phase d'évaluation de notre démarche d'interfaçage par appariement de lignes avec 13 autres photographies. Cela fait donc un total de 33 photographies prises à Saint-Étienne.

10

Nous avons créé 46 images de synthèse prises aux environs de ces photographies. En moyenne nous avons pris 2.3 images de synthèse pour chacune des positions des prises de vue photographiques. Parmi ces 46 images de synthèse, 20 sont très proches de l'une des photographies. Les 26 autres en différent par l'angle de visée, la focale ou l'orientation.

La base de données d'images de synthèse que nous utilisons pour valider notre approche est composée de 83 images de synthèse prises dans 29 positions géographiques différentes et correspondent aux 33 photographies des deux phases : apprentissage et évaluation.

Un échantillon de ces photographies ainsi que les images de synthèse correspondantes sont présentés dans la figure 5.

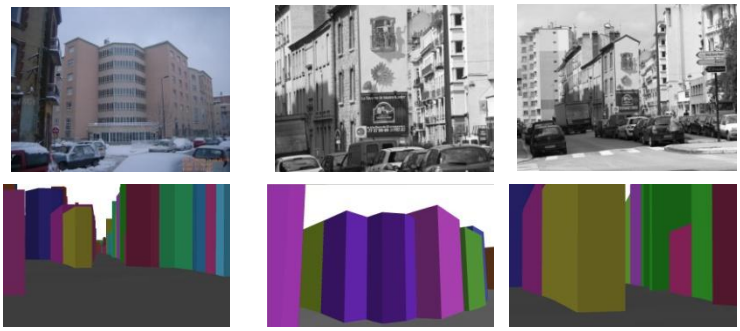


Figure 5. Echantillon de photographies et d'images de synthèse

Les pointeurs sur la carte de la figure 6 indiquent les positions des appareils photographiques lors des prises de vue.

Figure 6. Positions des images de synthèse

Les photographies ont été prises en utilisant cinq appareils différents : deux appareils photographiques équipés d'une puce GPS intégrée, deux Smartphones, un réflex classique couplé avec un GPS pour récupérer la position géographique et avec une boussole analogique pour préciser la direction de visée pendant la prise de la photographie.

La couche de données utilisée est la version 2 de la BD TOPO dans le système de coordonnées RGF93. Les données que nous utilisons pour déterminer la hauteur du sol sont fournies par le modèle numérique de terrain (MNT) au format raster de l'IGN.

6. Résultats

L'indicateur choisi pour évaluer la relation entre la photographie et l'image est le taux d'appariement calculé par la formule (2).

$$Taux = \frac{Nombre\ \ de\ lignes\ \ communes}{X} \times 100 \qquad (2)$$

12

X est le nombre minimal de lignes détectées sur la photographie permettant d'avoir le maximum possible de lignes en commun.

Nous appelons taux d'auto-reconnaissance le taux de correspondance entre une image de synthèse et la photographie qui lui correspond véritablement. Le taux minimal constaté dans notre base d'apprentissage est de 21.14 %. Le taux maximal est de 73.3 %, la valeur médiane est de 38 % et l'écart type est de 13.56 %. C'est la base que nous utilisons pour juger de l'appariement d'une photographie à une image de synthèse.

L'analyse des différents taux d'appariements et du taux d'auto-reconnaissance, nous permet de dégager l'arbre de décision de la figure 7.

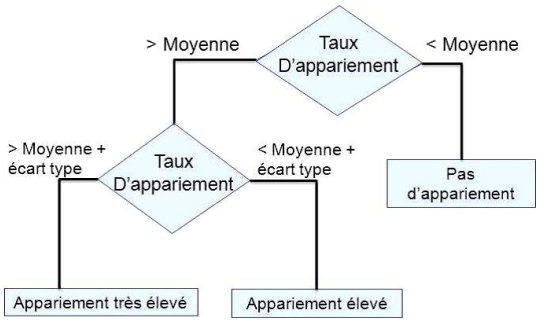


Figure 7. Arbre de décision.

Nous étudions trois niveaux d'imprécision de la localisation d'une photographie, à 200 mètres, 100 mètres et 50 mètres de la position supposée. A chaque fois ne sont prises en compte que les images de synthèse localisées dans le périmètre correspondant.

Avec une incertitude de 50 mètres autour d'une position, en appliquant notre arbre de décision, nous avons une distance moyenne de 12 mètres entre la bonne position de la photographie et l'image de synthèse signalée comme la correspondante à cette photo. Une incertitude de 100 mètres engendre en moyenne 29 mètres séparant la bonne position et celle donnée par notre approche. Dans un cercle de rayon 200 mètres, nous pouvons trouver en moyenne la position d'une photographie à 55 mètres de sa position réelle.

Si nous n'avons pas de localisation supposée *a priori*, nous prenons toutes les images de synthèse ayant un taux de reconnaissance supérieur à la moyenne. Le décalage moyen entre la position réelle de la photographie et sa position calculée est alors de 218 mètres.

7. Limites de la méthode

Les bases de données géographiques sont la première cause d'erreur dans l'appariement. Dans la base de données utilisée, des bâtiments différents sont très souvent regroupés, ce qui réduit le nombre de lignes dans l'image de synthèse et perturbe l'appariement avec la photographie où tous les bâtiments apparaissent.

Le deuxième élément perturbateur est lié aux caractéristiques des prises de vue. La méthode fonctionne bien pour les photographies qui sont prises depuis un point de vue au sol et avec un axe de prise de vue proche de l'horizontal. En se plaçant dans des situations spéciales : prise de photographie d'un étage supérieur d'un bâtiment ou orientation de l'appareil photographique par rapport à l'axe horizontal par exemple, l'appariement est rendu difficile et les performances baissent sensiblement ...

La présence de plusieurs obstacles ou écrans pénalise aussi notre approche. Les décorations festives et la neige sur les arbres accentuent les perturbations et par conséquence le nombre de lignes parasites, ce qui fait baisser le score de l'appariement.

8. Conclusion et perspectives

Nous avons démontré que les lignes peuvent être suffisamment représentatives d'un paysage urbain pour permettre d'apparier de manière satisfaisante une photographie urbaine avec une image de synthèse représentant la vue du même point dans une même orientation. Il devient dès lors possible de localiser la photographie à partir des coordonnées géographiques des images de synthèse appariées avec la photographie.

La principale limite de cette recherche est liée au nombre encore trop réduit d'images de synthèse de référence dans notre base de test et surtout à leur caractère irrégulier. Pour étalonner les taux d'erreur et valider complètement la démarche, il faudrait disposer d'une base de référence plus complète et plus systématique d'images de synthèse. Il faudrait aussi multiplier les tests dans des paysages urbains différents de manière à vérifier la performance de la méthode dans des zones où les bâtiments sont plus uniformes que dans le centre de ville de Saint-Etienne.

Plusieurs pistes d'amélioration sont envisageables. La première concerne l'élimination des obstacles entre l'appareil photographique et les bâtiments afin de favoriser la ressemblance entre le monde réel et la modélisation 3D. Les techniques d'élimination des objets situés au-devant de la scène se multiplient dans le monde du traitement de l'image et deviennent de plus en plus performantes. Cela devrait permettre d'augmenter le taux d'appariement. Une deuxième piste envisageable est de travailler avec une modélisation 3D de type fil de fer, qui permettrait de produire directement les lignes des bâtiments. Il faudrait tester aussi l'apport de modèles 3D avec des niveaux de détails plus élevés qui produiraient des images avec un plus grand nombre de lignes. Enfin, il est envisageable, vu la montée en puissance des

processeurs et des mémoires internes des téléphones, de porter les traitements sur ces appareils pour procéder au traitement juste après la prise de vue.

Bibliographie

Attia Y., Joliveau T., Favier E. et Delaup B. (2009a). *Extraction de contenu photographique en utilisant les SIG. MajecSTIC 2009, Novembre 2009, Université d'Avignon, 5 p.*

Attia Y., Joliveau T., Favier E., Delaup B. et Rana S. (2009b). *SIG et extraction de contenu photographique. Une proposition méthodologique, SAGEO 2009, Paris, 25-27 Novembre 2009, 15 p.*

Attia Y. (2012). *Interfaçage de bases de données photographiques et géographiques par appariement de ligne, Thèse en information géographique et application, Université Jean Monnet,Saint-Etienne, 181 p.*

Cohen M. et Toussaint G. (1977), *On the detection of structures in noisy pictures. Pattern Recognition, vol. 9, no. 2, pages 95-98, 1977.*

Crowley (2010), *Crowley J. L. Détection et Description de contraste, cours ENSIMAG, décembre 2010, 15 p., http://www-prima.inrialpes.fr/jlc/Courses/2010/ENSI3.FAI/ENSI3.FAI.S6.pdf*

Duda R. et Hart P. (1972), *Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM, vol. 15, no. 1, pages 11-15, janvier 1972.*

Hays J. et Efros A. A. (2008), *IM2GPS : estimating geographic information from a single imag, Computer Vision and Pattern Recognition, 2008.CVPR 2008.Conference IEEE, pages 1-8, 2008.*

Hoiem D. ,Efros A. A. et Hebert M. (2005), *Geometric context from a single image, IEEE Computer Society, volume 1, pages 654-661, 2005.*

Hoiem D., Efros A. A. et Hebert M. (2007), *Recovering Surface Layout from an Image. Article, International Journal of Computer Vision, vol. 75, no. 1, pages 151_172, 2007.*

Hough P. V. C. (1962), *Method and Means for Recognizing Complex Patterns. Rapport technique 3,069,654, Décembre 1962, 19 p*

Kneepkens *R.E.J.* (2005), *Hough based road detection. Thèse Technische Universiteit Eindhoven, avril 2005, 81 p.*

Maître (1985), *Maître H. Un panorama de la transformation de Hough. Cours traitement de signal, vol. 2, no. 4, pages 305-317, 1985. http://documents.irevues.inist.fr/bitstream/handle/2042/2334/02.PDF+TEXTE.pdf?sequence=3*

Moslah O (2011),Toward large scale urban environment modeling from images, Thèse de doctorat en Sciences de l'information et de la communication , Université Cergy Pontoise, Juillet 2011, 107 p.

Rosenfeld A. (1969*), Picture processing by computer. ACM Computing Surveys, volume 1, pages 147-176, New York, septembre 1969.*

Suh B.et Bederson B. B (2007), *Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition. Interacting with Computers, vol. 19, no. 4, pages 524-544, Juillet 2007.*

Torniai C., Battle S. et Cayzer S. (2007), *Sharing, Discovering and Browsing Geotagged Pictures on the Web. Rapport technique, Digital Media Systems Laboratory, HP Laboratories, pages 159-170, Bristol, 2007.*

Valentin R. (2009), *Reconnaissance de formes - Transformée de Hough. Rapport technique, ENSEIRB-MATMECA à l'Institut Polytechnique de Bordeaux, Bordeaux, 2009, 14 p.*

Yaegashi K. et Yanai K, (2009*), Can Geotags Help Image Recognition? Advances in Image and VideoTechnology, volume 5414, pages 361-373; Berlin, Heidelberg, 2009.*

S2PViewer : un prototype de visualisation de motifs spatio-temporels

Hugo Alatrista-Salas[1,3], Sandra Bringay[2], Frédéric Flouvat[3], Nazha Selmaoui-Folcher[3], Maguelonne Teisseire[1]

1. TETIS - Irstea
500, rue J. F. Breton, 34093 Montpellier, Cedex 5, FRANCE
prenom.nom@teledetection.fr

2. Lirmm
161, rue Ada, 34392 Montpellier, Cedex 5, FRANCE
prenom.nom@lirmm.fr

3. PPME – Université de la Nouvelle Calédonie
98851 Nouméa, Cedex, NOUVELLE CALEDONIE
prenom.nom@univ-nc.nc

RESUME.

Le volume des données collectées et stockées dans des bases de données spatio-temporelles augmente. Il devient donc crucial de fournir des synthèses permettant aux experts de mieux appréhender ces données afin de prendre des décisions pour agir immédiatement. Dans ce contexte, la fouille de données spatiales, qui permet d'identifier de nouveaux motifs dans ces bases, couplée à des méthodes de visualisation, facilite le travail des experts. La visualisation apporte notamment une forte valeur ajoutée pour la compréhension de dynamiques spatio-temporelles de motifs extraits. Dans cet article, nous présentons une nouvelle approche de visualisation dédiée aux motifs spatio-temporels. Notre approche a été validée sur une base de données contenant des informations sur le suivi épidémiologique de la dengue en Nouvelle Calédonie.

ABSTRACT.

The volume of stored data in spatiotemporal databases increases rapiditly. It becomes crucial to provide summaries from these data allow experts to make decisions and to act immediately on the results. In this context, spatial data mining, which helps to identify new patterns in these databases coupled with visualization methods, facilitates the work of experts. Visualization derived from data mining can provide great add value, often crucial for the understanding of some spatiotemporal dynamics. In this paper, we present a new visualization approach of spatiotemporal patterns. Our proposition has been tested on a database containing information about epidemiological surveillance of dengue in New Caledonia.

MOTS-CLES : Base de données spatio-temporelle, épidémies, fouille de motifs séquentiels.

KEYWORDS: Spatiotemporal databases, epidemics, sequential patterns mining.

1. Introduction

De nombreux phénomènes évoluent dans l'espace et le temps. La modélisation de ces phénomènes est souvent complexe, non seulement en raison de leur nature spatiale et temporelle, mais aussi du fait des interactions possibles entre les événements participant aux phénomènes. L'étude de la propagation d'une maladie vectorielle (e.g. la dengue) dans une ville en est un exemple typique. Les experts en santé publique savent que l'évolution de l'épidémie dépend de facteurs environnementaux (e.g. climat, proximité de zones infectées en points d'eau, mangroves...) et d'interactions entre les humains et le vecteur de transmission (e.g. le moustique qui transporte la maladie). Toutefois, l'impact de ces facteurs environnementaux et de ces interactions reste encore mal connu. Dans ce contexte, les méthodes d'Extraction de Connaissances à partir des Données (ECD) apportent des solutions via l'identification sans hypothèse *a priori* de relations entre variables et événements, caractérisées dans l'espace et dans le temps. Malheureusement, l'exploitation des connaissances extraites par l'expert est souvent limitée, car il lui est difficile de s'approprier ces nouvelles connaissances qui sont parfois tout aussi complexes à interpréter que les données initiales, notamment, lorsque l'on cherche à représenter des dynamiques spatiales et temporelles (Cao *et al.*, 2011). La mise en place de méthodes et d'outils permettant de mieux restituer ces connaissances est donc un enjeu majeur.

Dans cet article, nous nous focalisons sur la fouille de données spatio-temporelles, et plus particulièrement, sur la restitution des connaissances obtenues aux experts. Pour cela, nous nous appuyons sur un nouveau type de motifs appelés *motifs spatio-séquentiels*, définis dans Alatrista-Salas *et al.* (2012). Ces motifs sont basés sur une extension des motifs séquentiels (Agrawal et Srikant, 1995), de façon à considérer ensemble les dimensions spatiale et temporelle. Nous proposons une approche de visualisation permettant aux experts de mieux appréhender les interactions spatiales et temporelles entre les différents facteurs représentés par ces motifs. À la différence des approches classiques, la méthode de visualisation retenue souligne les dynamiques spatio-temporelles, tout en prenant en compte l'environnement proche. Notre méthode de visualisation peut être appliquée à d'autres types de motifs, où les dimensions spatiales et/ou temporelles sont présentes, telles que les co-localisations (Shekhar et Huang, 2001) et les séquences temporelles de Tsoukatos et Gunopoulos (2001).

Nous ne proposons pas uniquement une méthode de visualisation de motifs spatio-temporels, mais tout un environnement permettant de faire une analyse détaillée de ces motifs à différentes échelles (des motifs globaux aux objets spatiaux locaux). Plus précisément, notre environnement de visualisation offre les avantages suivants :

- un affichage synthétique et schématique des motifs sous forme de graphes colorés (avec possibilité d'associer des icones aux nœuds). Trois visualisations sont proposées en fonction des besoins des experts et du type de motifs ;
- un affichage détaillé des zones et des dates où sont apparus les événements (i.e. des occurrences des motifs). À partir d'un motif, il est possible d'identifier les zones impactées sur une carte (et inversement). Une frise chronologique permet de visualiser les dates des évènements représentés par les motifs (avec deux niveaux de détails) ;
- des statistiques détaillées sur les zones (e.g. nombre d'habitants) et les caractéristiques temporelles des motifs (e.g. durée moyenne).

À notre connaissance, il n'existe pas d'approches de visualisation qui propose ce type de fonctionnalités associées à des motifs spatio-temporels. Notre proposition a été appliquée au suivi épidémiologique de la dengue dans la ville de Nouméa. Les premiers retours des experts professionnels de santé confirment l'intérêt d'une telle plateforme.

L'organisation de cet article est la suivante : un état de l'art complet est présenté dans la Section 2. Ensuite, dans la Section 3, nous introduisons les motifs spatio-séquentiels ainsi que la problématique soulevée. La Section 4 détaille notre approche de visualisation. Nous présentons le prototype de visualisation et son application aux données de la dengue dans la Section 5, pour conclure et présenter les perspectives en Section 6.

2. Etat de l'art

La visualisation d'informations est un thème très étudié ces dernières années dans diverses disciplines informatiques, et en particulier pour les résultats obtenus après l'étape de fouille de données.

Dans un contexte général, les techniques de visualisation ont été largement discutées dans la littérature (voir par exemple, Tufte, 1983 ; Peuquet, 1994 ; Bertin, 2010). Ces auteurs insistent, entre autres, sur l'importance de l'affichage visuel des informations de façon à rendre plus facile l'interprétation des résultats obtenus. Un graphisme visuellement attrayant, montrant l'information à interpréter, est plus intéressant et souvent plus efficace en terme d'interprétation qu'un affichage immédiat de quelques chiffres ou une représentation purement textuelle (Tufte, 1983). Ces auteurs donnent également des indications sur les spécifications techniques à prendre en compte lors de la représentation visuelle des données. Ils arrivent à la conclusion que la visualisation graphique d'informations doit être précise, claire et efficace. Ces trois caractéristiques doivent être accompagnées de conventions d'ordre technique comprenant : la sélection des couleurs, le choix des formes, les polices, la forme et le remplissage des lignes, le rangement des espaces de design et bien d'autres. Par ailleurs, Keim (2002) propose une classification des techniques de visualisation, selon le type de données à visualiser, e.g. données unidimensionnelles, bidimensionnelles, hiérarchiques, etc.

Plus spécifiquement, les systèmes de visualisation de motifs sont souvent associés à de nouvelles techniques d'extraction de connaissances appliquées à différents domaines. En effet, il est très difficile d'aborder le problème de la visualisation des connaissances extraites par la fouille de données sans aborder la méthode qui précède le système de visualisation. Plusieurs approches ont été proposées pour visualiser des motifs séquentiels. Par exemple, Pak Chung Wong *et al.* (2000) ont appliqué une technique d'extraction de motifs séquentiels aux données textuelles et l'ont accompagnée d'un prototype de visualisation permettant l'analyse des motifs obtenus sur des grands corpus. Subasic et Berendt (2008) ont proposé une méthode et un outil de visualisation pour cartographier et interagir avec les publications scientifiques postées sur le Web en utilisant des méthodes de fouille de textes. Plus récemment, Sallaberry *et al.* (2011) ont présenté un cadre pour la modélisation et la visualisation de motifs séquentiels permettant d'identifier les associations et les relations hiérarchiques entre des données associées à des gènes humains.

Dans le cas des données spatio-temporelles, un état de l'art très complet a été fait par Adrienko *et al.* (2003). Ils présentent un inventaire des techniques d'exploration visuelles existantes en fonction du type de données et des méthodes de fouille de données utilisées. Trois types de données spatio-temporelles ont été étudiés : (1) les données représentant des changements existentiels des objets spatiaux, e.g. occurrence, (2) les données reflétant des changements dans les propriétés spatiales des objets, et ; (3) les données représentant des variations temporelles des différents attributs thématiques.

Par ailleurs, Bertini et Lalanne (2010) étudient le rôle de la visualisation et des techniques de fouille de données sur le processus d'extraction de connaissances à partir des données (ECD). Ils distinguent trois catégories de techniques : (1) les techniques fondamentalement visuelles mais qui exigent l'exécution d'un processus de calcul avant la visualisation des résultats ; (2) les techniques où la fouille de données est l'étape prédominante et les résultats sont montrés en s'appuyant sur un système de visualisation, et ; (3) les techniques où la fouille de données et la visualisation sont totalement intégrées (il est impossible de distinguer lequel des deux processus joue un rôle prédominant). Ils ont aussi proposé des extensions possibles à ces travaux.

Du coté applicatif, Ping *et al.* (2008) ont visualisé des motifs représentant des changements au niveau d'objets géographiques (e.g. expansion de villes). Les motifs extraits étaient des règles d'association spatio-temporelles. Ils les ont représenté en prenant en compte quatre caractéristiques : date, durée, ordre et fréquence. Plus récemment, Burauskaite-Harju *et al.* (2012) ont proposé un ensemble de méthodes permettant, entre autres, la visualisation des dépendances spatiales entre des épisodes des précipitations temporellement synchronisés.

Les travaux étudiant le problème de la visualisation dans le processus ECD sont donc nombreux. Toutefois, contrairement à notre approche, ils ne sont pas adaptés à des motifs spatio-temporels complexes faisant à la fois intervenir les dynamiques spatiale et temporelle et un environnement proche. Un motif tel que *« la présence de cas de dengue dans une région est souvent précédée de températures élevées dans une zone située près de réservoirs d'eau »* pourrait difficilement être observé via ces méthodes.

3. Motifs spatio-séquentiels : concepts et définitions

Dans cette section, nous présentons la méthode de fouille de données permettant d'extraire des *motifs spatio-séquentiels* (cf. Alatrista-Salas *et al.*, 2012).

3.1. Définitions préliminaires

Une base de données spatio-temporelle $stDB$ est un ensemble d'informations structurées contenant des composantes géographiques (e.g. quartiers, rivières, etc.), des composantes temporelles (e.g. des dates) et des données décrivant les composantes géographiques à un temps donné. Plus formellement, une base de données spatio-temporelle est définie comme un triplet $stDB = (D_T, D_S, D_A)$ où D_T est la dimension temporelle, D_S est la dimension spatiale and $D_A = \{D_{A_1}, D_{A_2}, \dots, D_{A_p}\}$ est un ensemble des dimensions d'analyse associées aux attributs. Le Tableau 1 montre un exemple de base de données spatio-temporelle et sera utilisée dans le reste de cette section.

Tableau 1. Exemple d'une base de données spatio-temporelle (où b=bas, m=moyen, h=haut)

Zone	*Date*	*Température*	*Précipitations*	*Vent*	*Nombre de cas de dengue*
Z_1	12/03/2013	T_b	P_m	V_m	--
Z_1	13/03/2013	T_m	P_m	V_b	3
Z_1	14/03/2013	T_b	P_m	V_m	11
Z_2	12/03/2013	T_m	P_m	V_m	4
Z_2	13/03/2013	T_b	P_m	V_b	2
Z_2	14/03/2013	T_b	P_b	V_m	--

La dimension temporelle est associée à un domaine de valeurs noté $dom(D_T) = \{T_1, T_2, \dots, T_t\}$ où $\forall i \in [1..t]$, T_i est souvent appelé estampille temporelle et $T_1 < T_2 < \dots < T_t$.

Chaque dimension D_{A_i} ($\forall i \in [1..p]$) appartenant à la dimension d'analyse est associée à un domaine de valeurs noté par $dom(A_i)$. Dans ce domaine, les valeurs peuvent être ordonnées ou non.

La dimension spatiale est associée à un domaine de valeurs noté $dom(\{D_S) = \{Z_1, Z_2, \dots, Z_l\}$ où $\forall i \in [1..l], Z_i$ est une zone.

Définition 1. Item et Itemset

Soit un événement *I*, traditionnellement appelé *item*, une valeur littérale pour la dimension D_{A_i}, $I \in dom(D_{A_i})$. Un itemset, $IS = (I_1\ I_2 \dots I_n)$ avec $n \leq p$ est un ensemble non vide d'items tel que $\forall\ i, j \in [1..n], \exists k, k' \in [1..p], I_i \in dom(D_{A_k}), I_j \in dom(D_{A_{k'}})$ et $k \neq k'$.

Définition 2. Séquence d'itemsets

Une séquence d'itemsets S est une liste ordonnée, non vide, d'itemsets notée $< s_1\ s_2 \dots s_q >$ où s_j est un itemset.

Tous les événements produits dans une même zone sont regroupés et triés par date. Par exemple, considérons les événements produits dans la zone *Z1* du *12/03/2013* au *14/03/2013* selon la séquence $S = <(T_b\ P_m\ V_m)(T_m\ P_m\ V_b 3)(T_b\ P_m\ V_m\ 11)>$ indiquée dans le Tableau 1. Ceci signifie que hormis les événements T_b, P_m, et V_m qui se sont produits au même temps dans la zone *Z1*, i.e. lors de la même transaction, les autres événements de la séquence se sont produits dans deux autres dates dans la même zone. Une *n*-séquence est une séquence composée de *n* items, dans notre exemple, S est une *10*-séquence.

Une séquence $< s_1\ s_2 \dots s_q >$ est une sous-séquence d'une autre séquence $< s'_1\ s'_2 \dots s'_m >$ s'il existe des entiers $< i_1\ i_2 \dots i_j \dots i_q >$ tels que $1 \leq i_1 \leq i_2 \leq \dots \leq i_p \leq \dots \leq i_m$ Par exemple, la séquence $<(AB)(C)>$ est inclue dans la séquence $<(AB)(CD)>$ car $(AB) \subseteq (AB)$ et $(C) \subseteq (CD)$.

3.2. Motifs spatio-séquentiels

Nous allons étendre les définitions présentées dans la Section 3.1 de façon à prendre en compte la dynamique spatiale des données. Pour cela, nous définissons la relation *dans* entre une zone Z et un itemsets IS comme l'occurrence de l'itemset IS dans la zone Z au temps t dans la base de données $stDB$. Plus formellement :

$$\begin{cases} dans(IS, Z, t) = vrai\ si\ IS\ apparait\ dans\ stDB\ pour\ la\ zone\ Z\ au\ temps\ t \\ dans(IS, Z, t) = faux\ sinon \end{cases}$$

Ensuite, nous définissons la notion de voisinage entre zones qui peut être de différente nature, par exemple, les zones qui se trouvent à une certaine distance ou celles qui partagent une frontière, etc. Dans ce travail, deux zones Z_i et Z_j sont voisines si :

$$\begin{cases} voisin(Z_i, Z_j) = vrai\ si\ Z_i\ et\ Z_j\ partagent\ une\ frontière \\ voisin(Z_i, Z_j) = faux\ sinon \end{cases}$$

Définition 3. Itemset spatial

Soient IS_i and IS_j deux itemsets, IS_i et IS_j sont spatialement proches si $\exists\ Z_f, Z_g \in dom(D_S)$ et $\exists\ t \in dom(D_T)$ tel que $dans(IS_i, Z_f, t) \wedge dans(IS_j, Z_g, t) \wedge voisin(Z_f, Z_g)$ est vrai. Deux itemsets IS_i et IS_j qui sont spatialement proches, forment un itemset spatial noté $I_{ST} = IS_i \cdot IS_j$. Les itemsets spatiaux représentent un profil (non-exhaustif par rapport à l'ensemble des attributs disponibles) d'une zone et de son environnement proche.

Pour alléger les notations, nous introduisons un opérateur de groupement d'itemsets associés à l'opérateur $\cdot$ (voisin) et noté []. Le symbole θ représente l'absence d'itemsets dans une zone. Par exemple, l'itemset spatial $(\theta \cdot [B; C])$ peut être interprété comme deux événements différents apparus dans deux zones proches à une zone où aucun événement ne s'est produit à un moment donné.

Définition 4. Association entre zone, itemset spatial et le temps

Soit $I_{ST} = IS_i \cdot IS_i$ un itemset spatial, $Z \in dom(D_S)$ une zone et $t \in dom(D_T)$ une estampille temporelle, nous définissons la relation *vérifier* qui représente la présence de l'itemset spatial I_{ST} dans Z au temps t comme suit :

$$\begin{cases} vérifier(I_{ST}, Z, t) = vrai\ si\ dans(IS_i, Z, t) = vrai\ et\ Z' \in dom(D_S)\ tel\ que \\ \qquad voisin(Z, Z') = vrai\ et\ dans(IS_j, Z', t) = vrai \\ vérifier(I_{ST}, Z, t) = false\ sinon \end{cases}$$

Définition 5. Inclusion d'itemsets spatiaux

Un itemset spatial $I_{ST} = IS_i \cdot IS_i$ est inclus, noté par $\subseteq$, dans autre itemset spatial $I'_{ST} = IS'_k \cdot IS'_l$, si et seulement si $IS_i \subseteq IS'_k$ et $IS_j \subseteq IS'_l$.

Puis, nous modélisons la notion d'évolution d'événements dans les zones en prenant en compte leur relation de voisinage via la notion de séquence spatiale.

Définition 6. Séquence spatiale

Une séquence spatiale ou simplement *2S* est une liste ordonnée d'itemsets spatiaux notée par $s = < I_{ST_1}\ I_{ST_2} \ldots I_{ST_m} >$ où $I_{ST_i}, I_{ST_{i+1}}$ satisfaisant la contrainte de séquentialité temporelle, i.e. $i \in [1\,..\,m-1]$.

Ensuite, nous définissons une mesure d'élagage appelée *support absolu* pour des séquences spatiales définie comme le nombre de zones contenant la séquence étudiée et satisfaisant les contraintes de proximité des *itemsets spatiaux*. Plus formellement, cette mesure peut être définie comme suit :

Définition 7. Support absolu d'une séquence spatiale

Soit la *2S* $s = < I_{ST_1}\ I_{ST_2} \ldots I_{ST_m} >$, le support absolu de s représenté par $supp_{abs}(s,\ stDB)$ est défini comme le nombre de zones de la base de données *stBD* qui vérifient s, autrement dit :

$$supp_{abs}(s, stBD) = \left| \begin{matrix} Z \in dom(D_S)\ tel\ que\ \forall i \in [1..m], \exists t_i \in dom(D_T) \\ et\ vérifier(I_{ST_i}, Z, t_i) = vrai \end{matrix} \right|$$

De la même manière, nous définissons le support relatif dénoté par $supp_{rel}(s,\ stDB)$ pour une *2S* s comme le ratio entre le support absolu et le nombre de zones total.

Définition 8. Support relatif d'une séquence spatiale

Soient la *2S* $s = < I_{ST_1}\ I_{ST_2} \ldots I_{ST_m} >$, $supp_{abs}(s,\ stDB)$ le support absolu de s et $|dom(D_S)|$ le nombre total de zones de la base de données spatio-temporelle *stDB*, le support relatif de s est défini par :

$$supp_{rel}(s, stBD) = \left| \frac{supp_{abs}(s, stBD)}{|dom(D_S)|} \right|$$

Définition 9. Motif spatio-séquentiel

Soient la *2S* s et σ le support minimum spécifié par l'utilisateur, si $supp_{rel}(s, stDB) \geq \sigma$ alors s est une *2S* fréquente appelée motif spatio-séquentiel (S2P) où $supp_{rel}$ est le support relatif d'une séquence spatiale. Si la valeur de σ est proche de *1*, des motifs dit « rares » seront extraits. Contrairement, si σ est proche de *0*, des motifs « triviaux » seront restitués.

4. S2Pviewer : un environnement pour visualiser des motifs spatio-temporels

Dans cette section, nous présentons une nouvelle approche de visualisation des motifs spatio-séquentiels précédemment définis. Cette méthode a été intégrée dans un prototype de visualisation appelé S2PViewer qui permet de visualiser la dynamique spatiale (voisinage proche) ainsi que la dynamique temporelle. S2PViewer est basé sur Javascript, JQuery et D3 et le prototype est accessible sur le site http://datamining.univ-nc.nc/.

4.1. Une visualisation globale des motifs sous forme de graphes colorés

Comme indiqué dans la section précédente, un motif spatio-séquentiel (S2P) est une séquence d'itemsets spatiaux. Notre approche de visualisation doit permettre, entre autres, d'avoir un aperçu général de ces motifs, de donner un sens à ce qu'ils représentent et de trouver un motif pertinent parmi les nombreux motifs inintéressants (Bertini et Lalanne, 2010). Pour représenter visuellement de tels motifs, il faut prendre en compte deux dynamiques : la dynamique spatiale, représentée par les itemsets spatiaux (l'opérateur spatial) et la dynamique temporelle, représentée par l'aspect séquentiel.

Un itemset spatial représente l'état courant d'une zone (ses évènements ou caractéristiques) ainsi que celui de ses voisins proches, à un instant donné. Trois types d'itemsets spatiaux peuvent être représentés à l'aide des opérateurs tels que · (voisin), *[]* (groupement) et le symbole θ (absence). La Figure 1 illustre ces trois cas en utilisant une représentation à base de graphe.

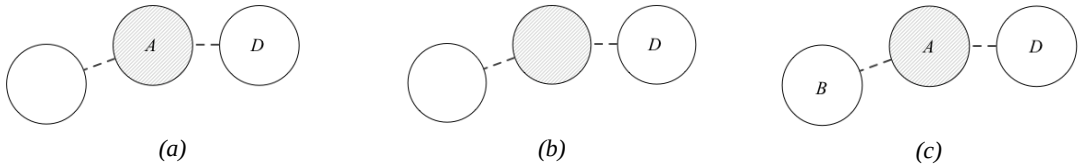


Figure 1. Représentation graphique des itemsets spatiaux (a) A·D (b) θ ·D (c) A· [B ; D]

Dans la Figure 1, chaque cercle représente une zone, et les étiquettes représentent les évènements associés à la zone (i.e., les itemsets). Dans les trois figures, la zone centrale colorée est la zone étudiée. Les lignes pointillées représentent le voisinage spatial. Cette représentation du voisinage spatial a été définie par Peuquet D.J., (1994) comme une représentation de la « contiguïté spatiale ». La longueur et l'angle des lignes pointillées n'ont pas de signification particulière.

Les motifs spatio-séquentiels décrivent l'évolution temporelle d'une zone et de son environnement proche à différentes estampilles temporelles. Cette évolution temporelle est représentée par une succession d'itemsets spatiaux. Le S2P illustré dans la Figure 2 est composé de trois estampilles temporelles. La première est composée de deux événements *A* et *B* apparus dans la zone étudiée. Après, il n'existe aucun événement dans la zone étudiée mais *B* et *C* sont apparus dans des zones voisines. Enfin, *P* est apparu dans la zone étudiée et *Q* et *R* sont apparus autour. Les flèches pleines représentent la dynamique temporelle.

Comme on peut le constater, cette représentation respecte le postulat énoncé par Tufte (1983) : la quantité d'informations (dimensions) affichées ne dépasse pas le nombre de dimensions présente dans les données. En effet, la temporalité est représentée par un arc orienté, qui symbolise une succession d'événements, et l'aspect spatial (voisinage) est représenté par des arcs en pointillés.

Afin d'enrichir cette représentation, nous utilisons la taille et la couleur des nœuds pour donner visuellement des informations supplémentaires à l'expert. La taille du nœud est directement proportionnelle au nombre d'items (i.e., d'évènements) qu'il représente. La couleur est utilisée pour

identifier visuellement certains évènements d'intérêt particulier pour l'expert (e.g., la présence de dengue). Une solution efficace pour représenter les types d'entités (e.g., présence de dengue ou absence de dengue) est l'utilisation des régions fermées colorées (Ware, 2004). Nous avons choisi des cercles de deux couleurs différentes : le vert et le rouge car ce sont des couleurs opposées dans le modèle de couleurs appelé *color opponent-process model*. En effet, un œil sain les distingue parfaitement, contrairement à d'autres paires de couleurs (Ware, 2004). Aussi, la couleur rouge est souvent employée, par convention, pour représenter les maladies, le danger, etc., contrairement à la couleur verte qui est souvent utilisée pour représenter des personnes saines, des zones non polluées, etc.

Un point reste à préciser par rapport à cette représentation : comment positionner les éléments du graphe dans l'espace ? En effet, un même graphe peut être affiché de différentes façons en fonction de la position des nœuds et des arcs. La Figure 3 présente trois exemples de configuration. Une technique classique pour afficher des graphes est d'utiliser les algorithmes basés sur les forces (Kaufmann et Wagner, 2001). Cette approche positionne les nœuds du graphe de façon à limiter les croisements entre arêtes. Un des avantages de cette approche est de permettre à l'utilisateur de déplacer les nœuds du graphe. Lorsqu'un nœud est déplacé, le graphe est alors réorganisé automatiquement à la volée (c.f. Figure 3a).

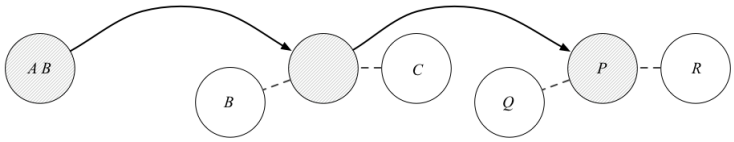


Figure 2. Représentation graphique du S2P <(AB) (θ · [B ; C]) (P · [Q ; R])>

Notre objectif premier est d'avoir un affichage qui soit le plus intuitif possible pour les utilisateurs, i.e., un affichage facilitant la compréhension du motif. Pour un motif spatio-séquentiel représentant l'évolution dans le temps d'une zone (et de son environnement), un affichage sous forme de frise chronologique (ou ligne du temps) semble être un affichage naturel. Cet affichage revient simplement à positionner les nœuds (de gauche à droite) sur un arc orienté de rayon fixe. Les nœuds représentant l'environnement proche sont répartis « autour » de cette ligne de temps reliés par des lignes pointillées. Cette représentation est illustrée par la Figure 3b. Une représentation plus classique du temps sous la forme d'une ligne droite n'a été pas retenue car elle ne permet pas de positionner autant d'éléments dans une même fenêtre.

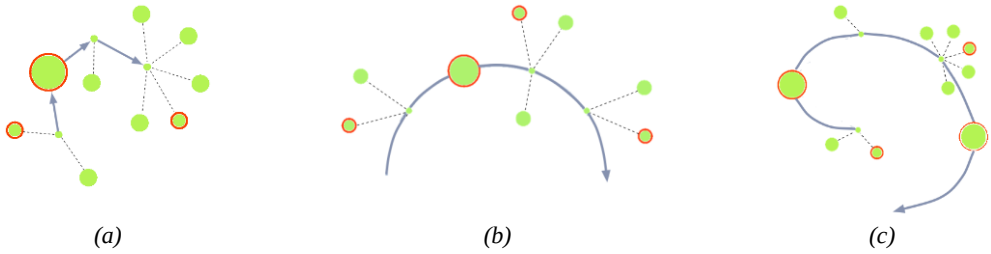


Figure 3. Trois types de visualisation (a) graphe, (b) arc, (c) spirale

Par ailleurs, lorsque le motif à afficher est très long et contient de nombreux itemsets spatiaux, cet affichage sous forme d'arc peut ne pas suffire. Dans ce contexte, nous avons privilégié une visualisation basée sur des coordonnées polaires – au lieu des coordonnées cartésiennes. En effet, le ratio de l'arc peut être incrémental et dépend de la taille du motif (Bertin, 2010). L'angle thêta représente la succession ou le temps. Le ratio de l'arc étant incrémental, cette approche permet de visualiser de longs motifs sous la forme d'une spirale d'Archimède (cf. Figure 3c).

Les conventions citées précédemment sont intégrées dans notre prototype. Nous laissons à l'utilisateur le choix du type de positionnement (graphe « dynamique », arc et spirale). En effet, chaque approche a ses avantages et inconvénients. La représentation en arc est la plus naturelle, mais elle n'est pas adaptée si les motifs sont longs. La représentation en spirale est le plus adaptée à longs motifs, mais elle est moins intuitive. La représentation en graphe « dynamique » (basée sur les algorithmes de forces) est probablement la moins adaptée à notre problème. Par contre, elle pourra être particulièrement intéressante pour afficher des motifs à n dimensions, par exemple, pour représenter la propagation d'une épidémie.

La représentation d'un motif donne une vision globale de celui-ci (disposition des événements au cours du temps et localisation dans l'espace). Elle peut être complétée par une visualisation à une granularité plus fine (celle des zones et des temps où est apparu le motif), comme nous le décrivons dans la sous-section suivante.

4.2. Vers une visualisation d'information à différents niveaux

Notre approche de visualisation des motifs S2P se veut une approche globale avec des possibilités de visualisation d'informations détaillées localement sur les éléments d'un motif (e.g. où et quand apparaît ce motif ?).

La visualisation de la dynamique temporelle a pour objectif de valoriser l'information concernant la durée de l'occurrence d'un S2P et où il est apparu. Deux points ont été pris en compte au moment de la représentation de la dynamique temporelle : (1) pour chaque motif, une synthèse contenant les périodes, dates du début et de fin de l'apparition du motif dans les zones concernées sera représentée en utilisant des blocs de couleurs différentes. Actuellement, la visualisation des zones impactées pour le motif sélectionné est présentée à l'aide d'une liste ; (2) des statistiques concernant la durée maximale, minimale et moyenne d'un motif, par rapport aux zones où le motif est impacté, seront présentées également. Ces valeurs calculées « à la volé » permettront à l'expert d'avoir une référence des caractéristiques concernant la durée d'apparition des motifs sur les zones associées.

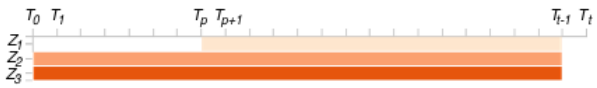


Figure 4. Exemple de représentation de la dynamique temporelle d'une séquence spatiale

Pour présenter la dynamique temporelle (voir Figure 4), les conventions à prendre en compte pour sa conception sont différentes de celles considérées pour la représentation de la dynamique spatiale. En effet, si l'on utilise des coordonnées polaires, alors une déformation des motifs les plus éloignés sera perçue (ils seront plus longs). De plus, comme cette visualisation n'est pas centrale mais indicative, une représentation en utilisant des coordonnées cartésiennes est privilégiée. Donc, nous allons représenter l'apparition d'un motif avec un bloc représentant les dates du début et fin d'un motif dans une zone.

Par exemple, soit le S2P $s =< (\boldsymbol{A})(\boldsymbol{B}) >$ et la séquence caractérisant une zone $S =< (\boldsymbol{A}C)(D)(D)(\boldsymbol{B}C) >$, nous représentons l'apparition de la séquence spatiale s dans S avec un bloc qui commence au temps t_0 et finis au temps t_3. Si un S2P apparaît de multiples fois dans une même zone, il sera tracé plusieurs fois avec la même couleur. De même, si le motif apparaît dans plus d'une zone, de multiples blocs seront tracés en utilisant des différentes couleurs.

4.3. Le processus d'analyse spatio-temporelle avec S2PViewer

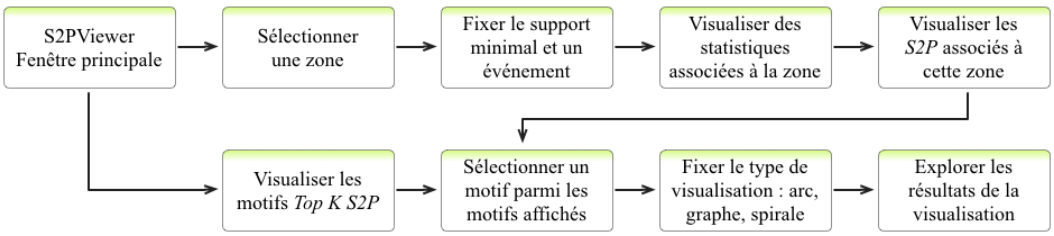


Figure 5. Diagramme de flux du processus d'analyse de S2P

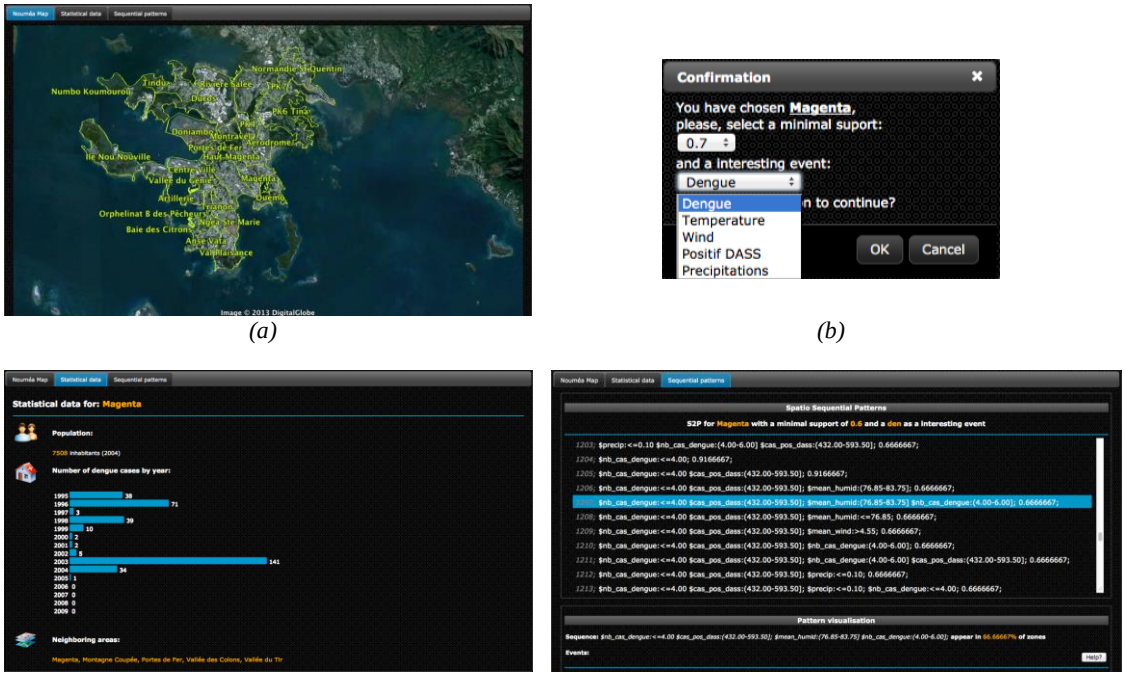
Notre prototype de visualisation est basé sur la technique décrite par Andrienko *et al.*, (2003) appelée « interaction de cartes » dans laquelle, des événements spatio-temporels sont « mappés » en utilisant une carte. La Figure 5 montre le diagramme de flux du processus d'analyse spatio-temporelle des S2P. Le prototype de visualisation est composé de trois étapes qui se traduisent par trois vues interactives dans notre interface :

1. La première vue – fenêtre principale – permet de sélectionner une zone. La Figure 6a montre l'exemple des données épidémiologiques (données dengue en Nouvelle Calédonie) avec comme zones spatiales les quartiers de Nouméa. Dans cette vue, une seule zone pourra être

sélectionnée à chaque fois que l'on désire afficher des motifs la décrivant. Une fois la zone sélectionnée, un support minimal doit être choisi (le support représente la probabilité d'apparition de ces motifs dans la base de données). Enfin dans cette vue, l'utilisateur devra sélectionner un événement considéré comme plus intéressant pour lui, sinon un événement (par exemple *dengue*) est sélectionné par défaut (voir Figure 6d).

2. La deuxième vue montre des informations statistiques décrivant la zone choisie dans la vue précédente (voir Figure 6b). Notamment, nous avons ajouté de l'information concernant la population de la zone (recensement de l'année 2004), le nombre de cas de dengue par année, et des informations sur les zones voisines, i.e. ceux qui partagent une frontière commune avec la zone sélectionnée auparavant.

3. En fin (Figure 6c), la troisième vue montre les S2P appartenant à la zone sélectionnée ayant un support supérieur ou égal au support fixé dans la première vue. Cette vue interactive permet, la sélection d'un motif, la visualisation graphique de ce motif ainsi que la visualisation de la dynamique temporelle associée (voir Figure 7 et 8). Si aucune zone n'a été sélectionnée, les vingt plus longs motifs seront automatiquement visualisés.



(a) *(b)*

(c) *(d)*

Figure 6. Les 4 vues du prototype de visualisation : (a) sélection d'un quartier, (b) sélection du support minimal et d'un événement intéressant, (c) visualisation des informations associées au quartier sélectionné, et (d) visualisation des 20 plus longs motifs ou des S2P associés au quartier sélectionné

5. Cas d'étude : analyse d'une épidémie de dengue

5.1. Contexte

Notre approche est générique et peut être appliquée à différents problèmes comme l'érosion des sols, de pollution des rivières, etc. Dans cet article, nous avons testé notre prototype sur une base de données spatio-temporelle contenant des données épidémiologiques de suivi de la dengue. Ces données ont été collectées dans la ville de Nouméa (Nouvelle Calédonie) sur un territoire divisée en *32* quartiers couvrant *45,7 km²*. Cette division spatiale a été proposée par la *Direction des Affaires Sanitaires et Sociales en Nouvelle Calédonie* (DASS). Cet ensemble de données contient des informations associées à des données démographiques, entomologiques, météorologiques, ainsi que des données de planification urbaine et des données médicales.

Afin d'obtenir des données catégorielles (données séparables dans des catégories qui s'excluent mutuellement), une discrétisation a été faite pour transformer les données continues en données nominales à l'aide de la méthode de discrétisation des *fréquences égales*. Les données sont stockées

selon trois classes de valeurs dans lesquelles ils sont classées : basse, moyenne et haute. Ces classes ont été discutées avec les experts. Le Tableau 2 montre des exemples de motifs extraits par notre méthode. Par exemple, le dernier motif peut être interprété par : dans *60%* des zones, les événements *haute température* et *haute humidité* apparaissent dans une zone voisine avant la présence de *peu de poubelles* et de l'apparition d'un certain nombre de *cas de dengue* (inferieur à 6) dans la zone d'étude.

Tableau 2. Exemple de S2P

Motif spatio-séquentiel	*Support*
< (mean_temper:>23.55 ihre_index:>34.82)(nb_cas_dengue:<=6.00 mean_temper:<=23.55)(mean_humid:(76.85-83.30]) >	0,7
< (waste_container:<=39.00 nb_cas_dengue:<=6.00)(θ·[comunity_gather:<=20.00 ; nb_cas_dengue:<=6.00 ; graveyard:<=2.00]) >	0,6
< (θ·[mean_temper:>23.55 ; mean_humid:>83.30])(waste_container:<=39.00 nb_cas_dengue:<=6.00) >	0,6

Sur les données de la dengue, nous avons souligné dans Alatrista-Salas *et al.*, (2012), que notre approche permet de trouver des motifs tels que : *peu de dépôts d'eau, peu de précipitations et faible présence de locaux publics suivi par quelques cas de dengue, peu de précipitations et du vent*. Cependant, notre approche trouve également des motifs plus complexes tels *que peu de dépôts d'eau, peu de précipitations et faible présence de locaux publics suivi par peu de dépôts d'eau et beaucoup de précipitations dans deux zones voisines, suivi par présence de dengue dans une zone voisine à la zone d'étude*. L'exemple souligne de la richesse des motifs extraits par notre approche en mettant en évidence l'influence des zones voisines.

Notre outil de visualisation permet de visualiser tout type de séquences d'itemsets (ensemble d'événements) décrivant une zone (objet géo-référencé) et son environnement proche (des zones voisines) quelque soit l'application, e.g. érosion des sols, pollution des rivières, épidémies, etc.

5.2. Analyse sémantique des motifs obtenus

La Figure 7 représente visuellement un S2P. Comme a été décrit dans la Section 4.1, chaque cercle symbolise un ensemble d'événements apparaissant à un moment donné. La taille des cercles reflète le nombre d'évènements caractérisant la zone. L'absence d'événement dans la zone choisie (i.e. θ) est représentée par un cercle de petite taille. L'arc de cercle représente le temps et les lignes pointillées représentent la dimension spatiale (à côté de).

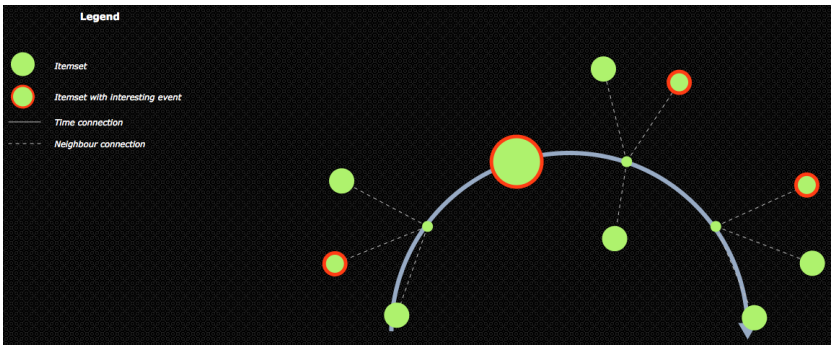


Figure 7. Représentation d'un S2P à l'aide du S2PViewer

La couleur rouge sert à identifier les itemsets qui contiennent l'événement *dengue* ou tout autre événement choisi lors de la première étape (voir Figure 6d). Cette caractéristique permet d'identifier facilement, d'un côté, la position (i.e. l'estampille temporelle) dans le motif S2P où se trouve l'événement intéressant, et d'un autre coté, les autres événements appartenant au même itemset.

Comment l'ont mentionné Andrienko *et al.*, (2003), il est nécessaire de différencier les événements temporaires des événements durables. Les changements de caractéristiques décrivant une zone n'ont pas la même interprétation s'ils apparaissent dans des périodes courtes ou longues. Dans ce contexte, la représentation de la dynamique temporelle de notre approche permet d'identifier aussi bien la durée d'un motif que la périodicité de son apparition dans toutes les zones. Par exemple, la Figure 8 illustre un S2P qui apparaît périodiquement dans les quartiers *PK7, PK6 Tina, Ducos et Rivière Salée*. Cette information peut devenir cruciale pour les experts quand une épidémie est sur le point de se déclencher.

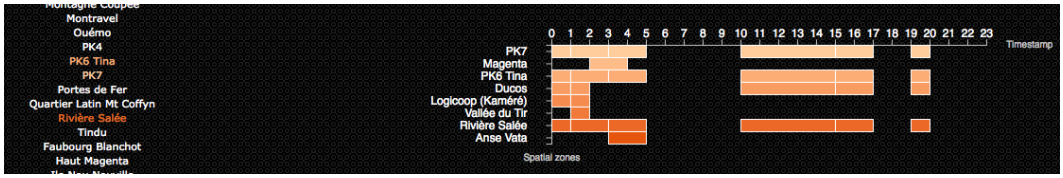


Figure 8. Représentation de la dynamique temporelle

5.3. Temps de réponse et validation pour les experts

Notre prototype de visualisation prend comme fichier d'entrée les motifs obtenus lors de l'étape de fouille de données et transforme cette information en objets à visualiser. Cette transformation ne consomme pas beaucoup de ressources. Au cours de nos expérimentations, nous avons affiché facilement plus de 3000 motifs en quelques secondes. Ce temps démontre la réactivité de notre prototype. Cette expérimentation met également en avant la nécessité de développer des fonctionnalités pour filtrer et/ou classer les motifs. En effet, il est difficilement envisageable de faire interpréter autant de motifs à un expert. Ce problème a motivé l'intégration dans notre prototype d'une fonctionnalité permettant de visualiser les « top-k » motifs (c.à.d. les *k* meilleurs motifs).

Concernant la validation de notre prototype, nous avons travaillé en partenariat avec des experts en santé public de l'Institut Louis Pasteur, de l'IRD et de la Direction des Affaires Sanitaires et Sociales de la Nouvelle Calédonie (DAAS) afin de recueillir des expertises, d'analyser les besoins et ainsi de résoudre les problèmes de lisibilité et d'utilisabilité de notre prototype. Lors de ces collaborations, nous nous sommes intéressés aux données à fouiller et aux besoins des utilisateurs finaux de S2PViewer pour réunir les informations concernant : (1) l'utilité, i.e. le prototype doit permettre d'aboutir à un résultat et ce résultat doit être pertinent pour ses objectifs, et ; (2) l'utilisabilité, i.e. le prototype doit permettre de réaliser une action rapidement et efficacement. Nous avons ainsi conçu trois types de visualisation, soit en arc, en graphe ou en spirale, lesquelles ont été développées suite aux remarques des experts.

6. Conclusion et perspectives

Dans cet article, nous avons proposé une nouvelle approche de visualisation permettant la restitution de motifs spatio-séquentiels (S2P). Ces motifs décrivent l'évolution d'un ensemble de caractéristiques "spatialisées" dans le temps en prenant en compte l'environnement voisin. Cet outil de visualisation permet de souligner la dynamique spatiale et temporelle des motifs extraits. Nous avons testé notre méthode d'extraction et de visualisation des S2P sur des jeux de données réelles dont celles associées au suivi épidémiologique de la dengue en Nouvelle Calédonie. Les résultats montrent l'intérêt de l'approche pour extraire et visualiser efficacement des S2P très riches.

Les perspectives associées à ce travail sont nombreuses. À terme, nous utiliserons des icones avec différentes couleurs pour représenter et identifier des événements. En effet, la représentation des événements par des icones avec des couleurs variant selon l'intensité (e.g. haute → rouge, basse → bleu) semble tout à fait pertinente. Cette idée a été soutenue par les experts en épidémiologie impliqués dans le projet. Nous voudrions aussi réduire le nombre de motifs à visualiser. Pour cela, nous sommes à la recherche de méthodes permettant, par exemple, de montrer les plus longues séquences ou celles contentant un événement spécifique à une position donnée (e.g. l'apparition de la dengue dans le dernier itemset). Une autre perspective consiste à étendre au spatial des mesures spatialisées comme la moindre contradiction temporelle (cf. Alatrista-Salas *et al.*, 2011). Par exemple, pour trouver des motifs qui se contredisent ou non selon les zones ou des périodes temporelles ou des événements. Par ailleurs, des améliorations sur la visualisation de motifs sont en cours de développement, notamment la représentation des motifs en utilisant une vue plus abstraite permettant à l'utilisateur de s'orienter dans les données. Nous voudrions également proposer le prototype en application mobile pour « smartphones ». En effet, il sera possible de récupérer la position géo-référencée de l'utilisateur, identifier la zone et afficher les informations et les S2P appartenant à cette zone automatiquement.

Remerciements :

Nous tenons à remercier la Direction des Affaires Sanitaires et Sociales de la Nouvelle Calédonie, l'Institut Pasteur, l'IRD et l'Université de Nouvelle Calédonie (Convention 2010) pour les données de suivi

épidémiologique de la dengue, ainsi qu'Arnaud Salaberry et Pierre Accorsi pour leur contribution dans la conception et le développement de l'outil de visualisation.

Bibliographie

Agrawal R., et Srikant R., (1995). *Mining sequential patterns*. Data Engineering, International Conference on, 0:3, 1995.

Andrienko N., Andrienko G., Gatalsky P., (2003). *Exploratory spatio-temporal visualization: an analytical review*. Journal of Visual Languages & Computing, Volume 14, Issue 6, p. 503-541, 2003.

Alatrista-Salas H., Bringay S., Flouvat F. Selmaoui-Folcher N., et Teisseire M., (2012). *The Pattern Next Door: Towards Spatio-sequential Pattern Discovery*. Advances in Knowledge Discovery and Data Mining - 16th Pacific-Asia Conference PAKDD'12, p. 157-168, 2012.

Alatrista-Salas H., Cernesson F., Bringay S., Azé J., Flouvat F. Selmaoui-Folcher N., et Teisseire M., (2011). *Recherche de séquences spatio-temporelles peu contredites dans des données hydrologiques*. Revue des Nouvelles Technologies de l'Information, RNTI-E-22, p. 165-188, 2011

Bertini E. et Lalanne D., (2010). *Investigating and reflecting on the integration of automatic data analysis and visualization in knowledge discovery*. ACM SIGKDD Explor. Newsletter, p. 9-18, 2010.

Bertin, J., (2010). *Semiology of Graphics: Diagrams, Networks, Maps*. Economic & Social Research Institute, 2010.

Burauskaite-Harju A., Grimvall A., Walther A., Achberger C. et Chen D., (2012). *Characterizing and visualizing spatio-temporal patterns in hourly precipitation records*. Journal of Theoretical and Applied Climatology, Vol. 109, p. 333-343, 2012.

Cao L., Zhang H., Zhao Y., Luo D. et Zhang C., (2011). *Combined mining : Discovering informative knowledge in complex data*. Systems, Man, and Cybernetics, IEEE Transactions on, 41(3), p. 699-712, 2011.

Kaufmann M. et Wagner D., (2001). *Drawing Graphs: Methods and Models*. Lecture Notes in Computer Science, Springer, 2001.

Keim D.A., (2002). *Information Visualisation and Visual Data Mining*. Visualisation and computer graphics, IEEE Transaction on, 7(1), p. 100-107, 2002

Pak Chung Wong, Cowley, W., Foote, H., Jurrus, E., Thomas, J., (2000). *Visualizing sequential patterns for text mining*. InfoVis 2000. IEEE Symposium on, p.105-111, 2000.

Peuquet D.J., (1994). *It's about Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems*. Annals of the Association of American Geographers , Vol. 84, No. 3, pp. 441-461, 1994.

Ping Y., Xinming T., et Shengxiao W., (2008). *Dynamic cartographic representation of spatio-temporal data*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B2, p. 7-12, 2008

Sallaberry A., Pecheur N., Bringay S., Roche M., Teisseire M., (2011). *Sequential patterns mining and gene sequence visualization to discover novelty from microarray data*. Journal of Biomedical Informatics, Volume 44, Issue 5, p. 760-774, 2011.

Shekhar S. et Huang Y., (2001). *Discovering spatial co-location patterns a summary of results*. Advances in Spatial and Temporal Databases, p. 236-256, 2001.

Subasic, I. et Berendt, B., (2008). *Web Mining for Understanding Stories through Graph Visualisation*. ICDM '08. Eighth IEEE International Conference on, p. 570-579, 2008.

Tsoukatos I.I. et Gunopulos D., (2001). *Efficient mining of spatiotemporal patterns*. Advances in Spatial and Temporal Databases, p. 425-442, 2001.

Tufte, E.R., (1983). *The visual display of quantitative information*. The Visual Display of Quantitative Information, Graphics Press, 1983.

Ware, C., (2004). *Information Visualization: Perception for Design*. Interactive Technologies, Elsevier Science, 2004.

Exploitation du vague spatial dans le SOLAP : vers une approche de conception prenant en compte les risques d'usage

Elodie Edoh-Alove[1], Sandro Bimonte[1], François Pinet[1], Yvan Bédard [2]

1. Irstea Clermont-Ferrand,
24 avenue des Landais, 63170 Aubière, France
prenom.nom@irstea.fr

2. Centre de Recherche en Géomatique, Département des Sciences Géomatiques
Pavillon Louis-Jacques Casault, Université Laval, Québec, QC, Canada
yvan.bedard@scg.ulaval.ca

RÉSUMÉ. Les systèmes « Spatial OLAP » (SOLAP) permettent l'analyse multidimensionnelle de grands volumes de données spatiales. Le vague spatial est une imperfection courante des données. De nombreux travaux proposent de nouveaux modèles pour gérer le vague spatial. Néanmoins, l'implémentation de ces modèles dans les cubes de données et leur utilisation avec des outils SOLAP sont encore à l'état embryonnaire. Aussi, nous présentons dans cet article une nouvelle approche pour concevoir des cubes de données spatiales prenant en compte la tolérance des utilisateurs aux risques de mauvais usages des cubes.

ABSTRACT. Spatial OLAP (SOLAP) systems allow multidimensional analysis of huge volume of spatial data. Spatial vagueness is a usual spatial data imperfection. Several works propose new models for handling spatial vagueness. However, the implementation of those models in data cubes and their use with SOLAP tools are still in an embryonic state. Thus, we present in this paper a new approach for designing spatial data cubes based on users tolerance to the risks of data cubes misuses.

MOTS-CLES : vague spatial, conception de cube, Spatial OLAP, risque d'usage

KEYWORDS: Spatial Vagueness, Risk of use, Spatial OLAP, Data cube design

1. Introduction

Le couplage des technologies OLAP[1] avec certaines fonctionnalités des SIG[2] a donné naissance à un nouveau système d'analyse nommé Spatial OLAP (SOLAP). Il est défini comme « *Une plate-forme visuelle supportant l'exploration et l'analyse spatio-temporelle faciles et rapides des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation via un affichage cartographique, tabulaire ou en diagramme statistique* » (Bédard, 1997).

Un « cube de données spatiales » est l'implémentation du modèle spatio-multidimensionnel souvent exploité dans ces technologies. Il contient l'ensemble des dimensions (ou axes) d'analyse, comme le temps ou l'espace, et les faits (c'est-à-dire les sujets d'analyse). Les dimensions peuvent être composées de hiérarchies comprenant plusieurs niveaux (Ex : niveaux « jour », « mois », « années » pour une dimension temporelle). Les instances des niveaux sont appelées des membres. Les faits quant à eux sont des combinaisons des mesures (valeurs analysées habituellement numériques) et des croisements de dimensions correspondants. Avec une telle structure de données, l'utilisateur d'une application SOLAP peut facilement et de manière interactive agréger les mesures le long des niveaux des dimensions.

Comme pour tout système, il convient de prêter une attention particulière à l'incertitude relative aux données spatiales exploitées avec les outils SOLAP et à son impact sur la fiabilité des analyses. En effet, les données spatiales sont souvent des « représentations fausses » mais utiles de la réalité (Devillers et Jeansoulin, 2005). Par exemple, dans le cas des données spatiales incluant une représentation géométrique, il n'est pas toujours possible de définir exactement où un objet commence et finit (Ex : forêt, rivière, habitat). Cette situation donne lieu à une incertitude souvent appelée « vague spatial ». Dans la communauté informatique, le terme « vague spatial » est utilisé pour décrire soit (1) des frontières floues ou (2) une incertitude de position ou de forme (Schneider, 1999). La notion de frontières floues a été généralisée par la notion de « vague de forme » incluant tous les problèmes de frontières larges (Bejaoui, 2009). Le terme « vague de localisation » est aussi utilisé dans la communauté pour décrire les problèmes de manque de connaissance sur la forme ou la position d'un objet avec des frontières réelles existantes (incertitude de position), ou l'inaptitude à mesurer un tel objet précisément (incertitude de mesure) (Hazarika et Cohn, 2001;Schneider, 1999). Dans notre travail, nous nous concentrons sur les données empreintes de vague de forme et de vague de localisation ainsi que les données spatiales dérivées (Ex : le périmètre ou la surface des objets spatiaux ou de leur agrégat).

Plusieurs travaux proposent l'utilisation de modes de représentation spécifiques (objets vagues en opposition aux types d'objet classiques point, ligne, polygone) pour mieux modéliser le vague spatial : il s'agit des modèles exacts (Bejaoui, 2009;Pauly et Schneider, 2010), flous (Dilo *et al.*, 2007), probabilistes (Burrough et

1. On Line Analytical Processing
2. Système d'Information Géographique

Frank, 1996) ou grossiers (Worboys, 1998). Néanmoins, peu de travaux (Siqueira *et al.*, 2012) s'intéressent aux problématiques inhérentes à l'intégration des objets vagues dans les cubes de données spatiales (Ex : l'agrégation des objets, les relations hiérarchiques et spatiales entre les niveaux d'une dimension spatiale, la visualisation et l'exploration dans le système SOLAP).

Dans nos travaux, nous recherchons plutôt comment exploiter les données empreintes de vague spatial avec les systèmes SOLAP classiques. Dans ce papier, nous présentons avant tout un état de l'art des travaux existants (Section 2) puis un cas d'étude fictif qui nous permettra d'illustrer nos contributions (Section 3). Ensuite nous proposons les grandes lignes d'une approche de conception de cubes de données spatiales qui prendrait en compte les risques liés à l'exploitation des sources de données vagues (Section 4). Nous parlons également de l'influence du vague spatial sur les éléments du cube de données spatiales et des risques qui peuvent en découler (Section 5) et pour terminer nous offrons un aperçu sur nos travaux futurs (Section 6).

2. Etat de l'art

Très peu de travaux portent sur l'intégration des objets vagues dans les cubes de données spatiales. Les auteurs de (Siqueira, Aguiar Ciferri, Times et Ciferri, 2012) ont proposé très récemment une extension du modèle spatio-multidimensionnel pour la prise en compte des modèles exacts de (Bejaoui, 2009;Pauly et Schneider, 2010) avec des techniques particulières pour le stockage et l'interrogation des cubes de données (par exemple un stockage séparé des régions sures et des régions incertaines d'un attribut géométrique vague). Ils ne proposent pas encore d'outil implémentant leurs nouvelles définitions et leurs techniques. Les auteurs de (Jadidi *et al.*, 2012) proposent une approche algorithmique basée sur la théorie des ensembles flous pour gérer, entre autres, les frontières flous des zones de risque d'érosion dans les cubes.

Ainsi, bien qu'il soit intéressant d'intégrer les objets vagues dans les cubes de données spatiales, il reste encore beaucoup à faire avant de pouvoir concevoir, implémenter et exploiter ces cubes de données spatiales dans la pratique. En effet, les outils SOLAP existants ainsi que les bases de données (Pauly et Schneider, 2010) ne supportent que le stockage et l'interrogation de données spatiales vues comme des entités discrètes certaines (point, ligne, polygone). Dans une telle situation, il est impératif de gérer les risques encourus relativement à l'usage des cubes produits.

La gestion du risque comme axe de recherche est l'objet d'un intérêt grandissant de la part de la communauté Géomatique. Le risque est défini dans les travaux de (Gervais et al., 2009;Lévesque, 2008) comme la probabilité d'occurrence d'un usage inapproprié du cube de données spatiales combinée à la gravité de cet usage inapproprié (ISO/IEC 51, 1999). Il est représenté par une valeur qualitative (Ex : faible, moyen, élevé) qui varie selon le contexte d'usage du cube. Une méthode de gestion de risque basée sur celle définie par (ISO/IEC 51, 1999) est également proposée dans lesdits travaux. Elle se déroule en quatre phases: identification des usages inappropriés possibles du cube spatial (identification des risques) ;

identification et établissement de stratégies pour prévenir ces mauvais usages (analyse et évaluation du risque, réponse au risque) ; surveillance du risque ; et enfin documentation du processus de gestion du risque.

Ainsi, pour pallier le fait que les systèmes SOLAP existants ne gèrent pas les objets vagues, nous pensons qu'il faut développer une nouvelle approche de conception de cubes qui intègre une gestion des risques d'usage relatifs au vague spatial dans la définition même du schéma spatio-multidimensionnel. Notons que dans cet article, nous définissons un risque d'usage comme étant un danger, un inconvénient plus ou moins probable auquel l'utilisateur est exposé lorsqu'il exploite ou interprète les données du cube dans son processus décisionnel (adaptation de la définition du Larousse Français). Il s'agit par exemple des risques de surestimation, sous-estimation, perte d'argent.

Aucun outil pour aider à produire (ou à déduire) un cube en se basant sur les risques acceptables par l'utilisateur, n'a été proposé.

3. Cas d'étude

Dans cette section nous présentons un exemple fictif d'application SOLAP agro-environnementale. Cet exemple sera utilisé tout le long de l'article pour illustrer nos contributions. Cette application est destinée à l'analyse des activités d'épandage de pesticides pour une culture de type blé et la campagne 2012 de traitement des cultures. Les données sont analysées dans le cadre du contrôle de la contamination des eaux de surface par du pesticide. Le cube de données présenté à la Figure 1 a été défini en utilisant le profil UML présenté dans (Boulil *et al.*, 2011a) ; ce profil est destiné à la représentation des cubes de données spatiales. Le cube est composé : (1) d'une dimension spatiale « Zones » contenant les zones (représentées par des polygones) sur lesquelles on observe la quantité de pesticide appliquée, organisées suivant la hiérarchie « Parcelles culturales →Exploitations agricoles →Zones inondables » ; (2) d'une dimension temporelle « Date » avec la hiérarchie « Jour→Mois » et d'une dimension thématique « Matières Actives» (des pesticides) avec comme unique niveau « Matières Actives ». Le fait (Pesticides) est décrit par les mesures : *SurfaceEpandue* qui représente la surface où le pesticide a été appliqué, *QuantiteMa* qui représente la quantité en Kg de matière active épandue, *QuantiteMaParHa* qui représente la quantité en Kg de matière active épandue par Ha (*QuantiteMa*/*SurfaceEpandue*). *QuantiteMa* et *SurfaceEpandue* sont agrégées le long des hiérarchies avec l'opération Somme puis *QuantiteMaParHa* est calculée à partir des valeurs agrégées.

Avec ce cube, il est possible de répondre aux requêtes SOLAP comme : « *Quelle est la quantité totale de pesticide (par Ha) appliquée en janvier 2012 pour toutes les exploitations agricoles et toutes les matières actives ?* » ou « *Quelle est la quantité totale de pesticide (par Ha) appliquée au cours de la campagne de 2012 dans les zones inondables ?* ». Son implémentation directe cache des risques d'usage liés au fait que le caractère vague des niveaux zones inondables (frontières larges) et

5

parcelles culturales (frontières bien définies mais localisées de façon incertaine) de la dimension spatiale ait été négligé (Cf. Figure 2).

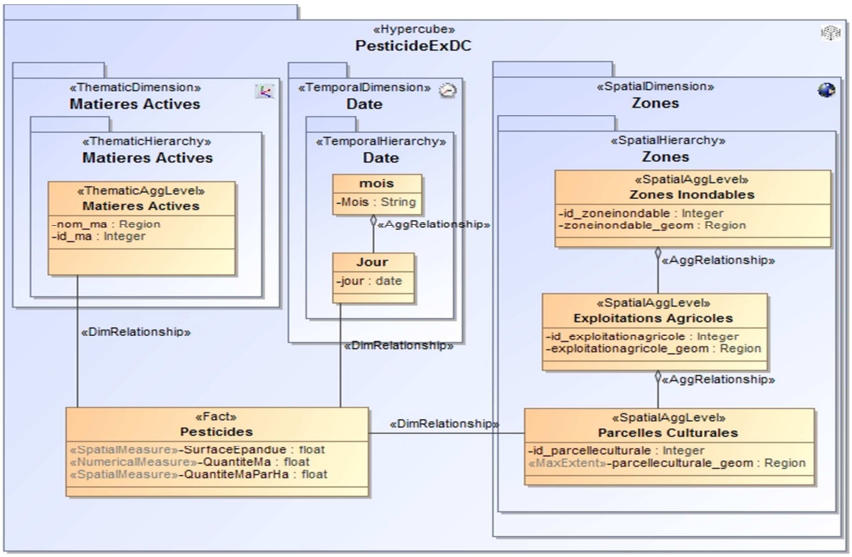


Figure 1 : Schéma initial du cube PesticideExDC

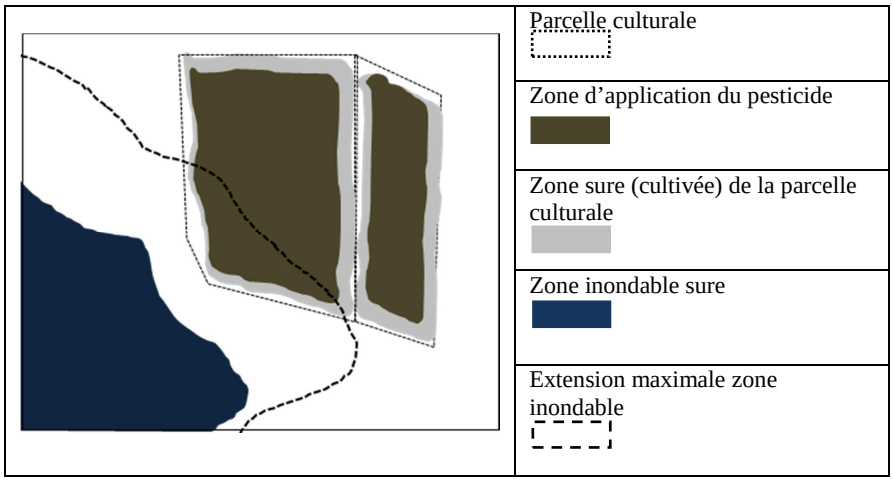


Figure 2 : Schéma représentant le vague sur les zones inondables et les parcelles culturales

En effet, comme montré sur la Figure 2, la parcelle culturale a une zone cultivée réelle (zone sure) qui ne couvre pas toute la parcelle culturale; ainsi l'utilisateur pourrait croire que la mesure *SurfaceEpandue* (obtenue en calculant la surface de la parcelle) est exacte alors qu'elle serait surestimée. La densité (Quantité/Surface), qui hérite de l'incertitude sur la surface serait quant à elle sous-estimée. D'un autre côté, les zones inondables aussi amènent une incertitude supplémentaire sur la mesure *QuantiteMaParHa* car d'une part la mesure *QuantiteMa* obtenue par agrégation (somme des quantités de matière active correspondantes aux zones d'intersection entre les parcelles culturales et les zones inondables) est incertaine et la *SurfaceEpandue* correspondante est empreinte de vague spatial (vu que la zone inondable l'est). Les limites des exploitations agricoles correspondent quant à elles aux limites cadastrales.

Un tel cube ne serait pas de bonne qualité pour tous les usages. En particulier, il présente des risques pour l'utilisateur qui souhaite étudier les contaminations des eaux de surface par du pesticide (du fait du vague des zones inondables et des parcelles culturales).

4. Approche pour la conception de cubes de données spatiales prenant en compte les risques d'usage

L'objectif principal de notre travail est d'offrir aux concepteurs des outils aidant à l'élaboration de schémas conceptuels et mettant en œuvre une approche qui prendrait en compte la tolérance des utilisateurs aux risques d'usage liés à l'exploitation des sources de données vagues.

Afin de répondre à notre objectif principal, nous avons tout d'abord étendu l'approche de conception de cube spatial classique (Malinowski et Zimányi, 2008) avec la méthode de gestion de risque (Lévesque, 2008) (Cf. Figure 3). Ainsi nous y avons rajouté (en rose) des étapes d'identification: des données sources vagues ; de l'influence de ce vague sur les éléments du cube ; des risques d'usage encourus relatifs au vague. Nous avons aussi intégré (toujours en rose) une étape pour le choix et la validation de stratégies pour gérer les risques « non tolérables ». Lors de la phase de spécification des besoins, nous préconisons une mise à contribution des utilisateurs potentiels identifiés, non plus seulement dans la détermination des besoins en analyses, mais aussi dans une première étape d'identification des risques encourus (risques potentiels identifiés) relatives aux analyses prévues (Ex : risque de mauvaise interprétation des mesures associées aux zones inondables). En parallèle, une spécification de besoins orientée-sources est effectuée, à partir des données sources identifiées, selon les étapes de l'approche classique. Par la suite, lors de la phase de conception, un schéma initial est déduit des spécifications résultantes de la démarche orientée-utilisateurs. Il sera fusionné avec celui déduit dans la démarche orienté-sources après marquage du vague spatial (« Ajouter support vague spatial ») existant dans les sources, par exemple, les attributs contenant l'extension géométrique minimale et celle maximale des zones inondables sont marqués si existants. Par la suite, l'approche préconise de déterminer l'impact des attributs marqués de vague sur les membres et mesures agrégées du cube (« Identifier impacts

du vague »). Cette étape ainsi accomplie, il est conseillé de mettre à jour les risques identifiés (Ex : Risque de surestimation ou sous-estimation de la *QuantitéMa* dans les zones inondables) avant de passer à une étape de catégorisation (« Ajouter degrés de tolérance aux risques ») de ces risques selon leurs degrés de tolérance (Ex : Risque de sous-estimation inacceptable- degré 0). L'étape suivante dans l'approche est l'application des stratégies (« Appliquer stratégies contrôle risques ») dans le but de déduire un schéma final approprié, par exemple, remplacer les géométries des zones inondables par leur extension maximale.

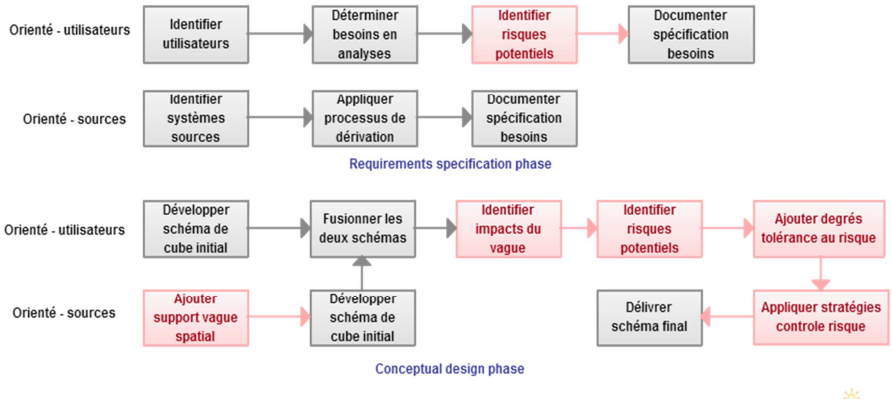


Figure 3 : Approche hybride (orienté-utilisateur/source) prenant en compte les risques d'usage

5. Le vague, source de risques d'usage

Dans cette partie, nous présentons nos premières réflexions sur le vague spatial et son impact sur les éléments du cube de données spatiales (notamment les mesures). Nous présentons tout d'abord la typologie des géométries (exploitées dans le cube) que nous considérons dans notre étude (Section 5.1) puis nous présentons notre étude d'impact (Section 5.2) en faisant ressortir le vague et les risques associés. Pour finir, nous présenterons quelques exemples de stratégies pour gérer les risques.

5.1. Typologie des géométries

Nous nous sommes concentrés dans un premier temps sur les géométries polygonales. Les géométries exploitées dans le cube peuvent correspondre à différentes réalités du phénomène spatial observé et donc être de différents types (Cf. Figure 4) :

- *L'extension minimale (selon le modèle exact de (Bejaoui, 2009))*. Elle correspond à la partie où la présence du phénomène est sure. Nous l'appellerons MinExtent dans la suite de l'article.
- *L'extension maximale selon le modèle exact de (Bejaoui, 2009)*. Elle inclut la partie sure mais également les zones où le phénomène est probablement présent. Nous l'appellerons MaxExtent.
- Géométrie discrète classique : géométrie choisie pour représenter le phénomène sans tenir compte, le cas échéant, de ses frontières larges ou de sa localisation vague.

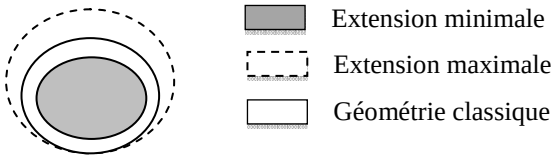


Figure 4 : Représentation formelle des types de géométries

Cette typologie est utile tant pour le marquage du vague comme préconisé par notre approche globale, que pour l'analyse de l'impact du vague sur les éléments du cube puis la déduction des risques encourus.

5.2. Etude de l'impact du vague sur les mesures

Un membre ou une mesure spatiale géométrique, qu'il soit de type MinExtent, MaxExtent ou classique, crée de l'incertitude sur les niveaux et les mesures agrégées (en fonction de l'opération d'agrégation mise en œuvre). Il est important d'identifier comment cette incertitude se décline et s'accumule sur les mesures afin de spécifier plus tard des règles permettant de déduire le risque connaissant les besoins en analyses de l'utilisateur (cube initial + agrégations). Afin de respecter les limites de pages, nous nous limiterons dans cet article au cas de membres spatiaux géométriques MaxExtent.

5.2.1. Dans un niveau

L'introduction dans le cube d'un niveau spatial géométrique MaxExtent (tous les membres sont MaxExtent) crée une incertitude de type « mesure maximum » sur les mesures détaillées et sur les mesures agrégées pour une opération d'agrégation additive (somme). Il en résulte des risques de type « risque niveau » (Cf. Tableau 1).

En effet, supposons que les zones d'application du pesticide soient prises dans leur extension maximale (c'est-à-dire dans les limites de la parcelle culturale) :

- Les valeurs de la mesure *SurfaceEpandue* associée à ces zones seront empreintes de « vague » que nous typons « mesure maximum ». Elle revêt un risque de surestimation par l'utilisateur.
- Ce vague de type « mesure maximum » se retrouvera également sur les valeurs de *SurfaceEpandue* associées à l'union de toutes les parcelles culturales (qui sont disjointes). Elle revêt également d'un risque de surestimation car il s'agit d'une erreur systématique qui se propage à l'agrégat de façon additive.
- La densité calculée *QuantiteMaParHa*, que ce soit pour chaque membre ou pour tous les membres du niveau revêt un risque de sous-estimation. Cela peut conduire ultimement à croire qu'un épandage est règlementaire car la densité tomberait sous la norme maximale, alors qu'en réalité elle dépasse cette norme et a un effet inacceptable sur l'eau avoisinante.

5.2.2. Dans une hiérarchie

Il est possible d'avoir différentes combinaisons de niveaux spatiaux dans une même hiérarchie (MaxExtent → Classique ; MaxExtent → MaxExtent ; MaxExtent → MinExtent, etc.). Pour chacune de ces combinaisons, la mesure agrégée récupère l'incertitude introduite par les membres en fonction des relations topologiques entre les géométries en jeu ainsi que l'opération d'agrégation utilisée (ici la somme) le cas échéant. Il en découle des risques de type « risque hiérarchie ». Le Tableau 1 ci-après présente des exemples relatifs à notre cas d'étude pour la mesure *QuantiteMa*. Aussi, nous avons étudié le cas où les parcelles culturales étaient les enfants directs des zones inondables afin d'illustrer les relations MaxExtent → MaxExtent et MaxExtent → MinExtent.

5.3. Stratégies de gestion des risques

Dans cette section nous montrons quelques exemples de stratégies applicables au schéma du cube pour gérer les risques identifiés. Ces stratégies sont définies en fonction des degrés de tolérance aux risques des utilisateurs (0 – inacceptable, 1 – indésirable, 2 – acceptable, 3 – négligeable).

Si l'utilisateur juge le risque sur *QuantiteMaParHa* (Cf. section 5.2.1) pour les parcelles culturales comme étant négligeable (tolérance 3), le cube initial lui sera fourni. Par contre, s'il le juge inacceptable (tolérance 0), le risque est évité en lui proposant par exemple d'enlever le niveau « Parcelles Culturales » pour ne garder que les exploitations agricoles et les zones inondables (Cf. *Figure **5***). Si le risque est indésirable (tolérance 1), il est possible de lui proposer des politiques de contrôle comme par exemple le remplacement de la géométrie par l'extension minimale grâce à l'utilisation du GPS sur les tracteurs ou avions d'épandage (au lieu de prendre les limites cadastrales). Enfin, si le risque est acceptable (tolérance 2), le cube initial lui est proposé avec une politique de communication (il pourrait s'agir par exemple d'état d'alerte associé aux résultats des analyses comme proposé par (Lévesque, 2008)).

10

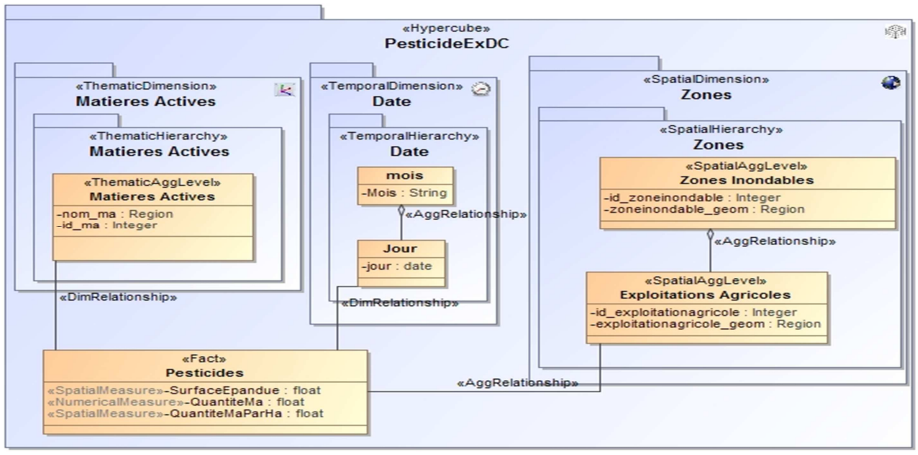
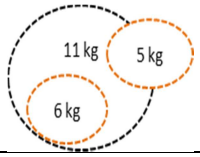
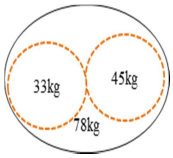
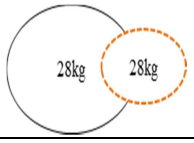
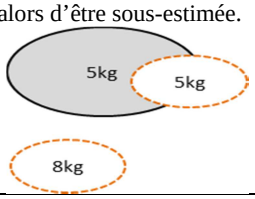


Figure 5 : Cube final avec tolérance 0 au risque de sous-estimation de la QuantiteMaParHa pour les parcelles culturales

Tableau 1 : Impacts du vague (MaxExtent) sur les mesures agrégées des parents directs et exemples de risques associés

	Relations hiérarchiques	**Impact**	**Exemples de Risques**
Hiérarchie CAS1	«maxExtent » Zones Inondables «maxExtent » Parcelles Culturales	Mesure maximum	***« Risque niveau »*** Ex : Risque de surestimation de la quantité de pesticide pour les zones inondables.
			« Risque hiérarchie » Ex : La répartition de la quantité de pesticide dans la parcelle culturale n'étant pas connue (sure), les valeurs calculées pour les zones inondables peuvent être surestimées dans les cas d'intersections (Risque Hiérarchie). 

11

Hiérarchie CAS2	«classique » Exploitations Agricoles «maxExtent » Parcelles Culturales	Mesure maximum restreinte classique	Le fait d'agréger vers un niveau classique non vague permet d'éliminer le risque de surestimation induit par le niveau inférieur **(Pas de risque niveau)**. 
			 ***« Risque hiérarchie »*** Par contre dans les cas d'intersection, ce risque reste présent.
Hiérarchie CAS3	«minExtent » Zones Inondables «maxExtent » Parcelles Culturales	Mesure maximum restreinte minExtent	***« Risque niveau »*** Ex : Risque de sous-estimation de la quantité de pesticide pour les zones inondables.
			« Risque hiérarchie » Ex : La zone inondable prise dans son extension minimale n'intersecte/contient pas certaines parcelles qui sont alors exclues de l'agrégation. La quantité de pesticide risque alors d'être sous-estimée. 

6. Conclusion et travaux futurs

Dans cet article nous présentons une revue des travaux existants sur la gestion du vague spatial dans les systèmes conventionnels et dans les systèmes SOLAP. Nous avons remarqué que l'angle d'approche relatif à la gestion du risque d'usage est une façon plus pertinente de garantir la fiabilité des analyses sans demander trop d'efforts aux utilisateurs. Aussi, nous avons proposé les étapes formelles d'une approche de conception de cube prenant en compte les risques d'usage. Nous avons également amorcé l'étude des impacts de l'introduction des données vagues dans le cube sur les mesures ainsi que les risques qui en découlent. Nous avons finalement présenté quelques stratégies de gestion des risques à appliquer sur le schéma multidimensionnel. Dans nos travaux futurs, nous généraliserons ces réflexions (typologie des risques, typologie des opérations à appliquer au schéma et conséquences des modifications dynamiques sur la structure du cube final) afin d'en

exploiter les résultats dans la mise en œuvre d'un outil pour déduire les risques d'usage connaissant les attributs géométriques vagues introduits dans le cube. Nous travaillerons également sur la définition de stratégies (règles) à mettre en œuvre pour gérer les risques et produire un schéma répondant à la tolérance de l'utilisateur aux risques encourus.

Le but final de ces travaux est de proposer un outil de conception (étendant des outils existants (Boulil *et al.*, 2011b;Lévesque, 2008;Romero et Abelló, 2009)) qui prendra en entrée les sources de données et un schéma initial de cube issu de l'analyse des besoins des utilisateurs. Cet outil, basé sur le formalisme standard UML, sera conçu pour : (1) exploiter les données discrètes fournies par les sources de données traditionnelles ; (2) produire de façon semi-automatique, non seulement les schémas des cubes de données spatiales, mais aussi les agrégations pertinentes et autorisées ainsi que des éléments de visualisation, contrairement aux méthodes semi-automatiques existantes. Le concepteur peut interagir avec l'outil pour valider les schémas proposés.

Bibliographie

Bédard Y. (1997). Spatial OLAP. *Forum annuel sur la R-D, Géomatique VI: Un monde accessible*.

Bejaoui L. (2009). *Qualitative topological relationships for objects with possibly vague shapes: implications on the specification of topological integrity constraints in transactional spatial databases and in spatial data warehouses*. Université Blaise Pascal.

Boulil K., Bimonte S. and Pinet F. (2011a). Un modèle UML et des contraintes OCL pour les entrepôts de données spatiales. De la représentation conceptuelle à l'implémentation. *Ingénierie des Systèmes d'Information*, vol. 16, Rapport technique n° 6, p. 11-39.

Boulil K., Bimonte S., Pinet F. and Chanet J.-P. (2011b). A UML Profile and OCL-based Constraints for Spatial Data Cubes. *Information Systems*, vol. Rapport technique n° p. 45.

Burrough P.A. and Frank A.U. (1996). *Geographic Objects with Indeterminate Boundaries* Taylor & Francis, London.

Devillers R. and Jeansoulin R. (2005). *Qualité de l'information géographique*, Hermès science publications, Paris.

Dilo A., de By R.A. and Stein A. (2007). A system of types and operators for handling vague spatial objects. *International Journal of Geographical Information Science*, vol. 21, Rapport technique n° 4, p. 397-426.

Gervais M., Bédard Y., Levesque M.-A., Bernier E. and Devillers R. (2009). Data Quality Issues and Geographic Knowledge Discovery. *Geographic Data Mining and Knowledge Discovery*, vol. Rapport technique n° p. 99-115.

Hazarika S. and Cohn A. (2001). Qualitative Spatio-Temporal Continuity Spatial Information Theory. Springer Berlin / Heidelberg, p. 92-107.

ISO/IEC 51 (1999). ISO/IEC Guide 51 Aspects liés à la sécurité - Principes directeurs pour les inclure dans les normes.

Jadidi A., Mostafavi M.A., Bédard Y. and Long B. (2012). Towards an Integrated Spatial Decision Support System to Improve Coastal Erosion Risk Assessment: Modeling and Representation of Risk Zones. *FIG Working Week 2012*, Rome, Italy.

Lévesque M.-A. (2008). *Formal Approach for a better identification and management of risks of inappropriate use of geodecisional data*. Laval University.

Malinowski E. and Zimányi E. (2008). Designing Spatial and Temporal Data Warehouses. *Advanced Data Warehouse Design*, Berlin, Springer - Verlag Berlin Heidelberg, p. 307-343.

Pauly A. and Schneider M. (2010). VASA: An algebra for vague spatial data in databases. *Information Systems*, vol. 35, Rapport technique n° 1, p. 111-138.

Schneider M. (1999). Uncertainty Management for Spatial Datain Databases: Fuzzy Spatial Data Types Advances in Spatial Databases. Springer Berlin / Heidelberg, p. 330-351.

Siqueira T.L.L., Aguiar Ciferri C.D., Times V.C. and Ciferri R.R. (2012). Towards Vague Geographic Data Warehouses. *Geographic Information Science*, Springer Berlin Heidelberg, p. 173-186.

Worboys M. (1998). Computation with imprecise geospatial data. *Computers, Environment and Urban Systems*, vol. 22, Rapport technique n° 2, p. 85-106.

WPS Bassin Versant

Un outil de modélisation hydrologique intégré à une Infrastructure de Données Spatiales

R. Béra[1,2,3], H. Squividant[1,2,3] , P. Aurousseau[1,2,3,4]

1Agrocampus Ouest, , F-35000 Rennes, France
2INRA, UMR1069 Sol Agro et hydrosystème Spatialisation,Rennes, France
3Université européenne de Bretagne, France
4Conseil Scientifique de l' environnement de Bretagne

RÉSUMÉ. *Nous proposons dans le présent article de rendre les outils d'analyse hydrologique développés dans notre UMR disponibles au travers d'une Infrastructure de Données Spatiales web interopérable. L'approche utilisée est celle du développement d'addons pour la plateforme OGC geOrchestra. Ces addons permettent le déchenchement d'algorithmes et la récupération en temps réel des résultats via la norme OGC WPS. Nous présentons ensuite l'exemple du WPS Bassin Versant et son mode de fonctionnement. Nous mettons en particulier en évidence que l'usage des normes OGC permet simplement à partir d'une application web (accédée par un simple navigateur) de lancer le traitement distant (sur un serveur de processus) de données distantes (sur un serveur cartographique) pour en afficher le résultat, et le cas échéant de le déposer sur un serveur cartographique tiers.*

ABSTRACT. *The proposal in the present paper is to make accessible the hydrology analysis tools that were developed by our rersearch team in the past years through an interoperable Spatial Data Infrastructure. To this aim we chose to develop addons for the geOrchestra OGC-compliant platform. Such addons trigger algorithms and retrieve their output in real time through OGC standard WPS. We then introduce a watershed WPS addon and its functioning modes. In so doing we exemplify the fact that the use of OGC standards make it straightforward (and transparent to the user operating a common web browser) to remotely trigger a process on a distant server, then applied to distant data present on a remote cartographic server, and drop the outcome on a third-part cartographic server while visualising it all on a browser.*

MOTS-CLÉS : *normes OGC, Web Processing Service, addon, Infrastructure de Données Spatiales, geOrchestra, MNTSurf, modélisation hydrologique, bassin versant.*

KEYWORDS: *OGC standards, Web Processing Service, addon, Spatial Data Infrastructure, geOrchestra, MNTSurf, hydrological modelling, watershed.*

1. Introduction et enjeux

La connaissance et la diffusion des données et outils d'analyse, de diagnostic, et de modélisation environnementales est plus que jamais un enjeu crucial pour la définition de politiques adaptées d'aménagement du territoire et de gestion des ressources environnementales. Ce constat est particulièrement pertinent s'agissant des problématiques de l'eau, les besoins y étant multiples et les publics intéressés variés : la complexification des enjeux de gestion de l'eau et la multiplication des acteurs dans les territoires (institutions publiques nationales, collectivités locales, commissions locales de l'eau) rend indispensable la mise à disposition en temps réel des données publiques disponibles mais aussi d'outils d'analyse et de modélisation de qualité.

Dans ce contexte il est légitime de répondre à ce besoin par la mise en œuvre d'une infrastructure de données spatiales (IDS) permettant l'accès aux données sur le territoire et mettant un certain nombre d'outils d'analyse et de modèles à la porté d'un public large et varié.

En particulier la détermination des limites des bassins versants est un exercice fortement dépendant de la qualité des données topographiques de base : cartes topographiques à différentes échelles, modèles numériques de terrain (MNT) de résolutions différentes et de qualités variables, eux-même tributaires de la correction des erreurs et artefacts, etc., l'impact d'une variation infime de ces paramètres étant potentiellement important en terme de délineation du bassin versant et conséquemment de gestion de la ressource.

Nous proposons donc ici d'intégrer à une IDS basée sur l'architecture geOrchestra[(a)] (2009) un certain nombre de fonctions, dont le but est de faciliter la gestion de la ressource en eau, l'exemple choisi étant celui du calcul de bassins versants.

Le développement des solutions de géomatique web permet la diffusion de ces données et outils en temps réels à l'ensemble des publics concernés. De nombreux outils conceptuels et logiciel ont été récemment développés dans ce domaine, y compris dans le domaine de la modélisation hydrologique (Feng *et al.*, 2011 ; Castronova *et al.*, 2012). Or trop souvent encore ceux-ci s'intègrent imparfaitement les uns aux autres ou font un usage limité ou partiel des possibilités offertes par les standards actuels. Il apparaît donc primordial de favoriser autant que possible l'interopérabilité des systèmes et outils, afin de permettre à tous les utilisateurs de bénéficier des avancées les plus récentes.

Les publics intéressés par la mise à disposition d'outils hydrologiques sont variés:

[(a)]geOrchestra est une architecture de données spatiales initialement développée par la société Camp To Camp suivant le cahier des charges de Geobretagne (l'IDS régionale opérée conjointement par la DREAL Bretagne et la région Bretagne), lequel imposait comme bases l'utilisation de logiciels Open Source (Geoserver, Geonetwork, Openlayers, etc.) et le versement du code développé spécifiquement (en particulier le visualiseur mapfishapp) à la communauté Open Source, dans le but d'en assurer la perennité.

– les chercheurs : modélisation hydrologique et des transfert des polluants (pesticides, nitrates, matières organiques en suspension) pour une meilleure compréhension du cycle de l'eau ;

– les gestionnaires de l'eau : institutions publiques nationales, comme l'Office National de l'Eau et des Milieux Aquatiques (ONEMA), ou les Schémas Directeurs d'Aménagement et de Gestion des Eaux (SDAGE) ; collectivités locales et commissions locales de l'eau, etc., dont les missions recouvrent la qualité et la surveillance du bon état environnemental des ressource en eau, leur utilisation responsable dans un but de durabilité ;

– le grand public : en dernier ressort une diffusion des connaissances sur l'eau et la mise à disposition d'outils adaptés favorise un contrôle citoyen de la qualité et de la disponibilité des ressources. Outre l'intérêt en terme de vulgarisation et de partage des connaissances, l'approche ainsi engagée répond aux impératifs de l'Union européenne en matière environnementale (Aarhus, 1998), transcrits en droit constitutionnel français (CE, 2004 ; intégré au bloc constitutionnel en 2005). De même l'ONU considère désormais la sécurité de l'eau (accès et qualité de la ressource) comme un enjeu majeur pour les années à venir (ONU, 2010).

La conjonction des contextes technique, législatif, voire philosophique et moral plaide pour le développement du partage non seulement des données mais aussi des outils permettant leur manipulation au sens large :

– volonté de l'Union Européenne (directive INSPIRE, 2007) d'assurer la disponibilité de la donnée publique et son accessibilité (obligation de faire publicité des données disponibles et, au travers des métadonnées, de donner toute information pertinente à leur sujet et en dernier ressort d'y accéder) ;

– mouvement Open Data et mise en place d'un service public de la donnée, au niveau national (DGF, 2011), mais aussi international et local (DC, 2013) ;

– volonté des acteurs de l'information spatiale (communautés de développeurs, universités et entreprises), regroupés notamment dans l'Open Geospatial Consortium de favoriser l'interopérabilité des systèmes et l'échange des flux (OGC, 1994).

Notre démarche fait suite à la nécessité de valoriser l'expertise accumulée à l'UMR SAS[(b)] en recherche et modélisation hydrographique, laquelle s'est traduite par la production d'outils largement utilisés par les gestionnaires de bassins et que nous souhaitons partager de la manière la plus large et transparente possible. Dans cet article, nous nous situons dans la perspective de l'intégration à l'IDS geOrchestra de fonctions dont le but est de faciliter la gestion de la ressource en eau, l'exemple choisi étant celui du calcul des contours topographiques des bassins versants.

Afin d'assurer une utilisation/réutilisation/diffusion maximale des systèmes, données, modèles et outils, notre démarche implique de répondre à l'ensemble des exigences liées à l'utilisation d'une IDS : (1) ouverture (logiciel libre, ouverture et accessibilité de la donnée, des outils d'analyse et modèles), (2) interopérabilité

[(b)] Unité de Recherche Mixte (INRA, Agrocampus Ouest) 1069 : Sol Agrohydrosystèmes et Spatialisation (http://www.rennes.inra.fr/umrsas) ; INRA : Institut National de la Recherche Agronomique.

(respect des normes OGC; modèle de développement inclusif), (3) modularité (conception par briques logicielles substituables, déploiement à la carte), (4) évolutivité (intégration facile des innovations et mises à jour: par exemple services de processus pour la publication élargie d'outils, méthodes et modèles).

Après avoir détaillé les options technologiques retenues (section 2), nous présentons le fonctionnement du WPS Bassin Versant (WPS BV) et de l'addon associé, en interaction avec les autres services web utilisés (section 3) avant de conclure par les développements et perspectives en cours et à venir.

2. Technologie

geOrchestra. Nous utilisons l'Architecture de Données Spatiales (ADS) Open Source geOrchestra (2009) composée des modules principaux suivant :

– Geonetwork (catalogue) ;

– Geoserver (serveur cartographique) ;

– mapfishapp (visualiseur intégrant OpenLayers, GeoExt, ExtJS).

Cette ADS est instanciée sous la forme de l'Infrastructure de Données Spatiales (IDS) GéoSAS (2009) pour la mise en œuvre.

Web Processing Service. La norme WPS (OGC WPS standard, 2007) édicte les modalités des entrées/sorties entre clients et serveurs pour les algorithmes et modèles (processus) appliqués aux données géoréférencées. Les données en entrée peuvent être locales au serveur ou distantes (client), utiliser des formats image (comme GeoTIFF) ou normes d'échange de données (SHAPE Zippé, GML, etc.). Les processus peuvent avoir tout degré de complexité. La norme WPS permet l'interopérabilité des services WPS, donc l'interaction de ces services et leur réutilisation (et la réutilisation du code associé).

Un service OGC doit pouvoir exécuter (côté serveur) trois types d'opérations à la requête du client.

– *GetCapabilities* permet au client d'effectuer une requête et d'obtenir les métadonnées de service (Capabilities) en retour . Les métadonnées de service décrivent les aptitudes du serveur WPS concerné : nom et description de chaque processus hébergé sur le serveur WPS, version de la norme utilisée pour les échanges client-serveur.

– *DescribeProcess* renvoie sur requête du client les informations détaillées sur le fonctionnement des différents processus offerts en service : entrées attendues, sorties possibles, ainsi que leurs formats respectifs.

– *Execute* permet l'exécution par le client du processus avec les paramètres choisis et renvoie le résultat produit.

MNTSurf. L'ensemble des services détaillés plus bas repose sur le logiciel MNTSurf (Squividant, 1994), développé à l'UMR SAS, dédié aux traitements des

modèles numériques de terrain (MNT) et à la modélisation hydrologique. Cet outil propose une méthode originale de correction des MNT par importation d'une couche vecteur représentant le réseau hydrographique d'une part et par la suppression des anomalies de drainage s'appuyant sur des structures de données de type graphe (arbres de drainage), d'autre part (Aurousseau et Squividant, 1997). MNTSurf implémente une représentation de l'écoulement de surface par une description du chemin de l'eau entre les mailles du MNT sous la forme d'un modèle de drainage monodirectionnel ou d'un modèle de drainage multidirectionnel (Aurousseau et Squividant 1996).

MNTSurf est à la base de nombreux travaux de recherche : modélisation des zones humides potentielles (Mérot, 2002), modélisation des transfert et transformation d'azote (Beaujouan *et al.*, 2001, 2002), modélisation de transfert de pesticides (Tortrat *et al.*, 2004), système d'aide à la décision pour l'amélioration de la qualité de l'eau, l'expertise sur les bassins versants en contentieux (et le plan algues vertes), suivi environnemental et synthèse sur l'évolution des concentrations et flux d'azote en Bretagne du Conseil Scientifique à l'Environnement de Bretagne (CSEB, 2008).

Les fonctions principales de MNTSurf permettent en particulier :

– le calcul du contour du bassin versant à l'amont de n'importe quel point sélectionné ; le point considéré est l'exutoire du bassin versant : tout écoulement de surface dans ce bassin versant converge vers cet exutoire.

– le calcul du réseau hydrographique modélisé sous forme d'arbre ;

– la localisation des stations de mesure présentes à proximité de tout point du réseau hydrographique.

3. MNTSurf dans geOrchestra via un WPS

L'originalité de l'outil proposé consiste en l'intégration de traitements WPS à geOrchestra. À Chaque fonction de MNTSurf correspond un WPS. Le fonctionnement du WPS BV présenté ici pourra être généralisé à l'ensemble des fonctions MNTSurf.

Outre le fait de simplifier l'accès au calcul des bassins versants (simple clic dans un navigateur), l'intégration du WPS BV à une IDS permet de rendre les données en entrée et sortie interopérables.

La figure 1 résume les différents types de flux entre les modules, ainsi que les solutions logicielles utilisées.

Paramétrage. Le WPS BV est paramétrable par le menu *Hydrologie / Bassin Versant / Paramètres*. Il est ainsi possible de sélectionner le MNT utilisé en entrée du calcul, la superficie minimale des BV calculés, le mode de lissage des contours de la couche produite, le zoom automatique ou non sur la couche produite.

Exécution. Le WPS BV se lance à partir du menu *Hydrologie / Bassin Versant* accessible via la barre d'outils du visualiseur de geOrchestra. Il accepte 4 types d'entrées distinctes, soit au choix :

a) un clic souris en un point de la carte transmettant un couple de coordonnées (x,y) ;

b) un fichier (liste de points) au format GML importé dans le visualiseur à partir du poste de l'utilisateur ;

c) la référence à une couche WMS complète de points (dans ce cas, les données transitent uniquement entre le serveur cartographique et le serveur de WPS) ;

d) une sélection d'objets (de type exutoires, donc de points) dans une couche WMS de points diffusée par un serveur OGC tiers ; une variante consiste à passer non plus la sélection mais la référence à cette sélection ainsi que la référence à la couche WMS (cas e).

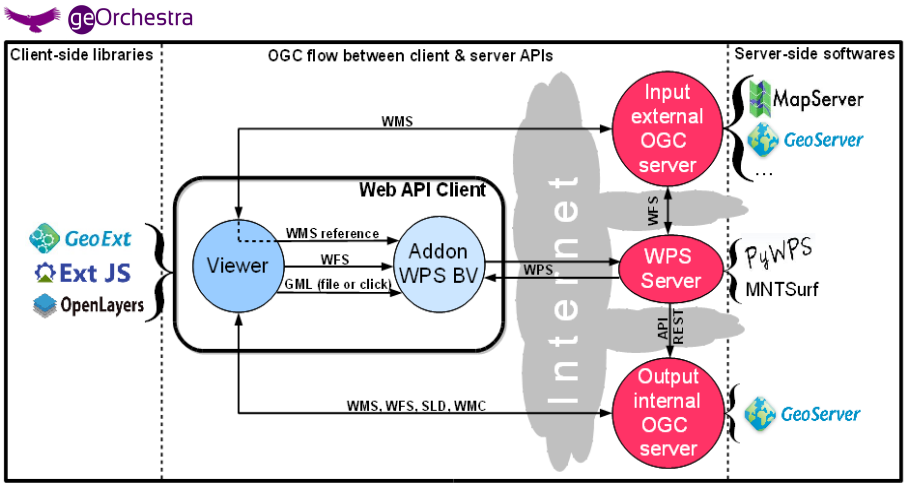


Figure 1. Schéma fonctionnel synthétique de l'architecture intégrant l'addon et le WPS BV.

Résultat. Une fois calculée, la couche produite par le WPS BV est automatiquement déposée sur un serveur OGC dédié puis est ajoutée comme une couche WMS standard à la liste des couches disponibles. De ce fait, elle bénéficie des fonctionnalités OGC du visualiseur (client web) comme : l'affichage (WMS[c]), le téléchargement au format SHAPE (WFS[c]), les requêtes spatiales ou attributaires (WMS/WFS), la création de style (SLD[c]), la sauvegarde de contexte (WMC[c]), etc.

[c] respectivement Web Map Service, Web Feature Service, Styled Layer Descriptor, et Web Map Context sont des normes OGC. *cf.* http://www.opengeospatial.org/standards/{wms,wfs,sld,wmc} pour un descriptif précis.

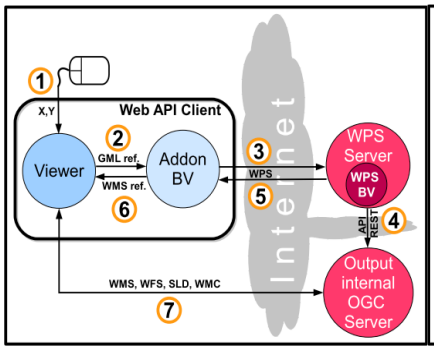


Fig. 2a.

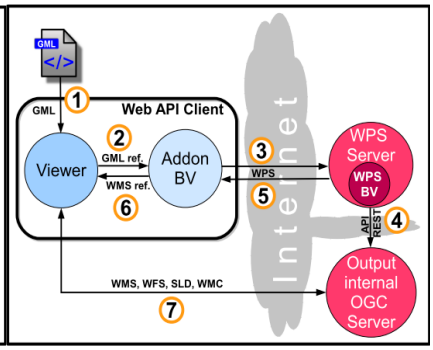


Fig. 2b.

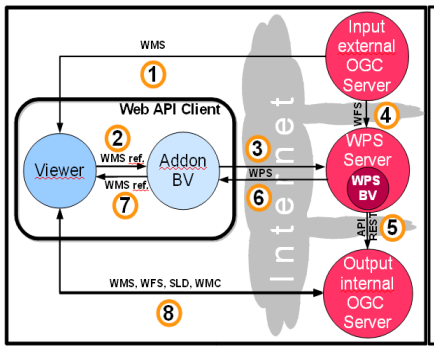


Fig. 2c.

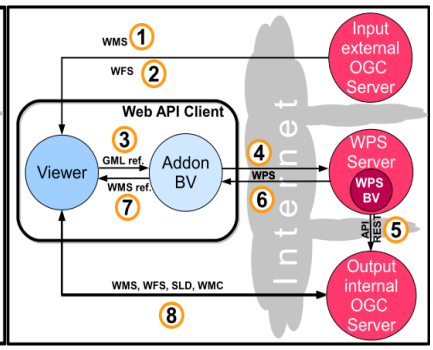
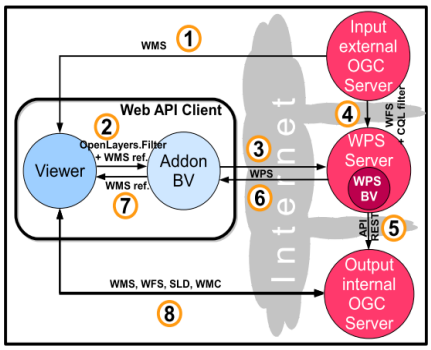


Fig. 2d.



2a. *Calcul à partir d'un clic souris.*
2b. *Calcul à partir d'une couverture spatiale (fichier de points) au format GML.*
2c. *Calcul à partir de la référence à une couche OGC complète (points).*
2d. *Calcul à partir d'une sélection dans une couche de points au format OGC réalisée dans le visualiseur.*
2e. *Calcul à partir de la référence à une couche de points OGC ainsi qu'à la description d'une sélection sur cette couche.*

Figure 2e.

Figure 2. Flux entre client et serveurs en fonction du type d'entrée pour l'addon BV.

La figure 2 détaille la chronologie des différents flux circulant entre les modules de l'IDS à l'exécution de l'addon BV en fonction de l'option choisie, laquelle dépend du type d'entrée : clic souris (fig. 2a), fichier GML (fig. 2b), référence à une couche

OGC (fig. 2c), sélection dans une couche OGC (fig. 2d et fig. 2e : dans le premier cas la sélection est transmise, tandis que dans le second seule la référence ou description de la sélection est transmise). Le tableau 1 détaille les différents flux associés à chaque option et leur enchaînement chronologique.

Nature du flux	2a	2b	2c	2d	2e
Transmission de coordonnées (x,y) déclenchée par un clic de souris dans le visualisateur	1				
Flux entrant sous la forme d'un fichier GML (liste de points) en direction du visualiseur		1			
Flux WMS en provenance du serveur cartographique vers le visualiseur (client)			1	1	1
Flux WFS en provenance du serveur cartographique vers le visualiseur (client) après sélection				2	
Le visualiseur transmet la réference du GML à l'addon	2	2		3	
Flux transmettant la réference du WMS			2		
Tranmission des références du WMS et du filtre (Openlayer.Filter) traduisant la requête					2
Flux WPS déclenché par l'addon vers le serveur WPS	3	3	3	4	3
Flux WFS du serveur cartographique vers le serveur WPS			4		
Flux WFS du serveur cartographique vers le serveur WPS, La sélection est transmise par filtre CQL					4
API REST pour dépôt du résultat sur serveur cartographique tiers	4	4	5	5	5
Flux WPS en sortie (du serveur WPS vers l'addon)	5	5	6	6	6
L'addon transmet la réference du résultat (adresse du WMS) au visulaliseur	6	6	7	7	7
Requêtes du client (visualiseur) et réponses du serveur tiers pour transmission du résultat (Flux OGC : WMS,WFS,SLD,WMC)	7	7	8	8	8

Tableau 1. Comparatif des différents flux en fonction de l'entrée de l'addon BV. Les numéros de la première ligne « nature du flux » permettent la mise en correspondance avec la figure 2 (fig.2a à 2e).

La figure 3 illustre l'utilisation de l'addon et du WPS BV dans l'interface mapfishapp de geOrchestra, disponible sur le site GéoSAS (BV-WPS, 2013). Un certain nombre de couches de données y sont chargées en WMS et affichées dans le visualiseur : le fond de carte OpenStreetMap, les stations hydrographiques de Bretagne. D'autres fonds de cartes sont chargés mais non affichés (scan de l'IGN, orthophoto composite). Tous ces fonds de carte sont disponibles en flux OGC sur le serveur cartographique GéoBretagne. Parmi les stations hydrologiques une sélection a été réalisée. La table attributaire des stations sélectionnées est visible en bas d'écran, et la sélection est matérialisée par un surlignage orange sur la carte. Dans l'exemple de la figure 3 la sélection concerne toutes les stations hydrographiques qui sont les exutoires de bassins versants d'une superficie supérieure à 500 hectares

(champ « bv » dans la table attributaire associée, visible dans la dernière colonne affichée). L'addon BV est ensuite sélectionné dans le menu déroulant *outils*, avec l'option *calcul BV depuis sélection*, déclenchant la requête WPS. Le résultat, c'est-à-dire les bassins versants ayant pour exutoire les stations de la sélection, est ensuite ajouté sous la forme d'une couche OGC apparaissant en bleu sur la carte et ayant un nom de la forme «bv_date_heure».

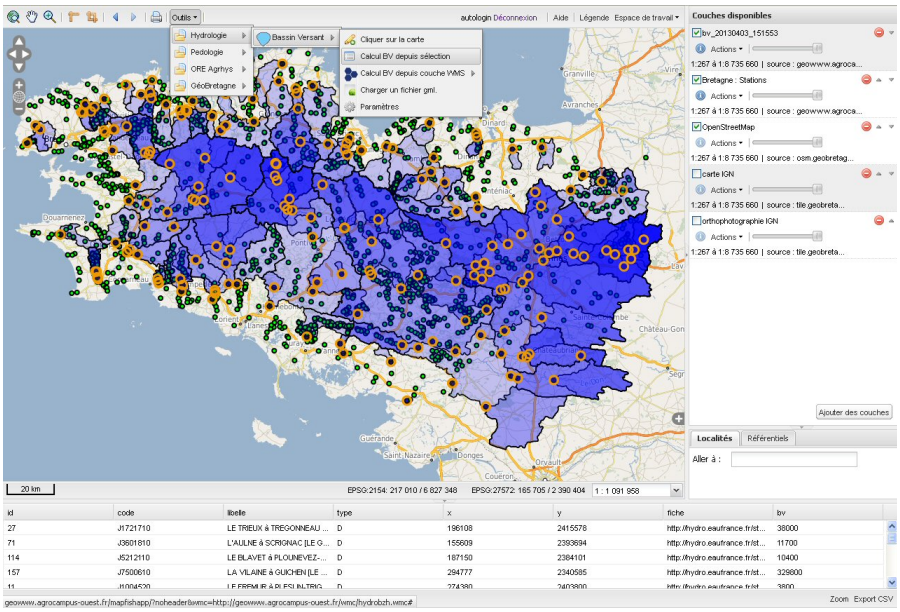


fig. 3. L'addon BV et le résultat du WPS BV dans le visualiseur.

5. Conclusion et Perspectives

Nous avons démontré l'utilisation dynamique (en temps réel) d'une fonction d'analyse hydrologique via un navigateur web faisant un usage intensif des standards OGC, le tout par une une utilisation exclusive de logiciels libres, et tirant avantage d'IDS ouvertes conformément au mouvement Open Data et à la directive européenne INSPIRE. Ce schéma de fonctionnement, les normes OGC (assurant l'interopérabilité), et la solution logicielle (modulaire) retenues permettent la transposition de l'approche à d'autres fonctions et outils d'analyse, suivant le principe suivant, dont le WPS BV détaillé plus haut est un exemple : l'utilisation concomitante de données distantes disponibles sur un serveur cartographique aux normes OGC, d'une part, et de processus d'analyse et/ou de modélisation spatiale accessibles sur un serveur de processus distant lui aussi OGC-compatible d'autre part, afin d'exécuter un traitement spatial (bénéficiant au passage des performances propres du serveur) dont le résultat peut être déposé sur un serveur cartographique OGC tiers et visualisé en temps réel dans le navigateur web client. Il est ainsi

possible d'une façon générale d'effectuer un traitement spatial sur des données délocalisées à l'aide de processus eux aussi délocalisés.

Suivant cette logique nous travaillons à la mise à disposition sous la forme de WPS d'un ensemble de fonctions issues de MNTSurf, avec pour effet une accessibilité maximale, en particulier par l'intermédiaire d'addons geOrchestra, mais aussi via tous clients WPS (comme par exemple celui du système d'information géographique QGIS). Il s'agit en particulier de proposer en WPS des outils faisant le lien entre modélisation et mesures en temps réel comme la Macro Flux (Vinson, 2003).

D'autre part, un certain nombre d'addons à geOrchestra déjà opérationnels, comme celui développé conjointement avec l'UMR LISAH[(d)] dans le Projet de Visualisation des Données AgroEnvironnementales (Vidae, 2009, 2013) sont aussi sur la liste des WPS en cours de finalisation. Cet addon permet l'interrogation des bases de données issues des chroniques de mesures en continu ou intermitentes de la physique et de la chimie de l'eau en chacun des points de mesure disséminés sur les sites des Observatoires de Recherche en Environnement (ORE) AgrHyS[(e)] et OMERE[(f)] situé respectivement dans le massif armoricain et le bassin Méditerranéen.

Enfin le développement de services de processus Web ne se résume pas pour nous au domaine hydrographique : nous proposons ainsi un WPS d'interrogation de bases de données sur la composition du sol pour le projet Websol (Chafchafi et Bargeot, 2013) et sa déclinaison régionale pour les sols de Bretagne (SdB, 2013), les possibilités dans le domaine étant elles aussi nombreuses.

D'autres projets travaillent à l'étoffement de l'offre de processus web, avec en particulier la mise sous WPS de l'ensemble des fonctions GRASS (52north.org) ou encore de l'encapsulage WPS d'un certain nombre de fonctions ArcGIS. Cette tendance, associée à l'accessibilité accrue des données est appelée à démultiplier et transformer les usages en géomatique.

Nous nous inscrivons complètement dans cette démarche et en proposons de plus un usage simplifié ne nécessitant pas même l'installation d'un client SIG (par exemple QGIS et ses différents plugins OGC) : le seul prérequis est un navigateur web. La large diffusion des smartphones et tablettes, qui en sont munis est donc en passe de rendre la géomatique (données et outils) véritablement pervasive et ubiquitaire.

(d) Unité de Recherche Mixte (INRA, AgroM, IRD) : Laboratoire d'étude des Interactions Sol - Agrosysteme – Hydrosystème). AgroM : École Nationale Supérieure Agronomique de Montpellier ; IRD : Institut de Recherche pour le Développement. http://www.umr-lisah.fr/

(e) Observatoire de Recherche en Environnment sur les temps de réponse dans les AgroHydroSystèmes. http://www.inra.fr/ore_agrhys

(f) Observatoire Méditerranéen de l'Environnement Rural et de l'Eau. http://www.umr-lisah.fr/omere/

6. Bibliographie

Aarhus, 1998 : Convention d'Aarhus. http://www.unece.org/fileadmin/DAM/env/pp/documents/cep43f.pdf

Aurousseau, P., Squividant, H., 1996 : Use of tree and graph structures for computation modelling of derived variables in digital elevation model software. http://viviane.roazhon.inra.fr/spanum/publica/dem/dem2.htm

Aurousseau, P., Squividant, H., 1997 : Correction of Digital Elevation Models using drainage pattern constraints. http://viviane.roazhon.inra.fr/spanum/publica/contrain/contrain.htm

Beaujouan, V.,Durand., P, Ruiz, L.,2001 : Modelling the effect of the spatial distribution of agricultural practices on nitrogen fluxes in rural catchments. *Ecological Modelling*. Elsevier : Volume 137 (1), pages 93-105

Beaujouan, V., Durand, P., Ruiz, L., Aurousseau, P., Cotteret, G., 2002 : A hydrological model dedicated to Topography-based simulation of nitrogen transfer and transformation. Rationale and application to the geomorphology-denitrification relationship. *Hydrological Processes* : vol.16 (2), p.493-507

BV-WPS, 2013 : http://geowww.agrocampus-ouest.fr/demo2013

Castronova, A.M,, Goodall, J.L., Elag, M.M., 2013 : Models as web services using the Open Geospatial Consortium (OGC) Web Processing Service (WPS) Standard. *Environmental Modelling & Software*, Volume 41, March 2013, Pages 72-83

CE, 1994 : Charte de l'environnement. http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000790249

Chafchafi, A., Bargeot, L., 2013 : Document de synthèse des travaux de l'axe 5 du RMT Sols et Territoires : Mettre à Disposition des Données Sols sur les Territoires. *Document de travail, Chambre régionale d'Agriculture Rhône-Alpes et AgroSup Dijon / EDUTER / CNERTA / DISI.*

CSEB, 2008 : Portail cartographique diffusant des informations sur la qualité des eaux (nitrate, pesticides) des bassins versants de Bretagne. *Conseil Scientifique de l'Environnement de Bretagne (CSEB), Aquascop (Bureau d'étude)*. http://geowww.agrocampus-ouest.fr/web/?page_id=171

DC, 2013 : DataCatalogs.org : Data Catalogs Dot Org aims to be the most comprehensive list of open data catalogs in the world. It is curated by a group of open data experts from around the world. A project initiated during the Open Knowledge Foundation Conference 2011. http://datacatalogs.org/

DGF, 2011 : data.gouv.fr : Plateforme française d'ouverture des données publiques. *Etalab.gouv.fr (Service du Premier Ministre Chargé de l'Ouverture des Données Publiques),* http://www.data.gouv.fr/A-propos

Directive INSPIRE, 2007: http://inspire.jrc.ec.europa.eu/

Feng, M., Liu, S., Euliss, N.H., Young, C., Mushet, D.M., 2011 : Prototyping an online wetland ecosystem services model using open model sharing standards. *Environmental Modelling & Software*, Volume 26, Issue 4, April 2011, Pages 458-468

geOrchestra, 2009 : http://www.georchestra.org/

GéoSAS, 2010 : http://geowww.agrocampus-ouest.fr/web/

Mérot, P., 2002 : Typologie fonctionnelle des zones humides de fond de vallée : les concepts de zones humides potentielles, effectives et efficaces. *Zones Humides Infos,* Volume 36, p.16

OGC, 1994 : Open Geospatial Consortium. http://www.opengeospatial.org/

OGC WPS standard, 2007 : http://www.opengeospatial.org/standards/wps

ONU, 2010 : Projet de résolution A/64/L.63/Rev.1 de l'Organisation des Nations Unies. http://doc.un.org/DocBox/docbox.nsf/GetFile?OpenAgent&DS=A/64/L.63/REV.1&Lang=F&Type=PDF

SdB, 2013 : Sols de Bretagne (Websol Bretagne). http://geowww.agrocampus-ouest.fr/portails/portail.php?portail=websol&mode=viewer&viewer=http://geowww.agrocampus-ouest.fr/sviewer&wmc=http://geowww.agrocampus-ouest.fr/wmc/bzh_FamilleDominante.wmc

Squividant, H., 1994 : MNTSurf : logiciel de traitement des modèles numériques de terrain. *Document E.N.S.A.R.*

Tortrat, F., Aurousseau, P., Squividant, H., Gascuel-Odoux, C., Cordier, M.O., 2004 : Modèle Numérique d'Altitude (MNA) et spatialisation des transferts de surface : utilisation de structures d'arbres reliant les exutoires de parcelles et leurs surfaces contributives. *Bulletin SFPT,* 172 p.128-136

Vidae, 2009 : VIsualisation de Données Agro-Environnementales. http://geowww.agrocampus-ouest.fr/web/?page_id=103

Vidae, 2013 : http://geowww.agrocampus-ouest.fr/portails/portail.php?portail=vidae

Vinson, J., 2003 : Calculs des flux annuels et des flux spécifiques annuels d'azote nitrique des principaux fleuves et rivières de Bretagne. *Bretagne Environnement.* http://www.bretagne-environnement.org/Media/Documentation/Bibliographies/Calculs-des-flux-annuels-et-des-flux-specifiques-annuels-d-azote-nitrique-des-principaux-fleuves-et-rivieres-de-Bretagne

SPATIODATA, une approche alternative pour la gestion des données spatialisées relatives aux bâtiments

Cyril Carré[1], Vincent Delfosse[1], Roland Billen[1], Aurélie Jeunejean[2], Christelle Boulanger[2], Anne Vandervreken[2], Pierre-Henri Lefèbvre[3], Benoît Michel[4], Pierre Leclercq[2]

1. Unité de Géomatique, Université de Liège, Allée du 6 Août, 17, 4000 Liège, Belgique
{cyril.carre,vincent.delfosse,rbillen}@ulg.ac.be

2. Lucid Group, Université de Liège
Chemin des Chevreuils, 1, 4000 Liège, Belgique
{aurelie.jeunejean,christelle.boulanger, a.vandevreken,pierre.leclercq}@ulg.ac.be

3. Belgian Building Research Institute, Avenue Pierre Holoffe, 21, 1342 Limelette, Belgique
phl@bbri.be

4. ICTeam, Université Catholique de Louvain, Euler Building, Avenue Georges Lemaître, 4, 1348 Louvain-la-Neuve, Belgique
benoit.michel@uclouvain.be

RESUME.
Les applications mobiles exploitant les informations relatives à l'intérieur du bâtiment sont en plein essor. SpatioData est un projet de recherche visant à développer une plateforme collaborative pour le partage efficace des informations liées aux bâtiments et adaptable aux diverses activités relatives au bâtiment. Cet article présente les besoins exprimés par les intervenants du bâtiment, ainsi qu'un descriptif des standards pouvant être utilisés et des solutions commerciales existantes. Le modèle de données SpatioData et l'architecture de la plateforme sont ensuite décrits.

ABSTRACT.
There are an increasing number of mobile applications supporting specific activities related to the indoor building environment. SpatioData is a research project aiming at the development of a collaborative platform for the effective sharing of building-related data and supporting a wide range of different activities. After presenting the initial requirements of the system, this paper provides a review of the standards for the building modeling and the commercial solutions currently available. Then, it describes the SpatioData building model, as well as the general system architecture.

MOTS-CLES : *Modèle intérieur, modèle d'information du bâtiment, BIM, modèle spatial, base de données en graphe, NoSQL, ontologie des espaces intérieurs, services Web, application pour l'intérieur du bâtiment.*

KEYWORDS: *Indoor model, building information model, BIM, spatial model, graph database, NoSQL, ontology for indoor space, Web Services, Indoor application.*

1. Introduction

Les travaux de rénovation dans un bâtiment existant sont souvent problématiques. Plusieurs informations relatives à ce bâtiment sont souvent lacunaires, perdues ou indisponibles au moment du travail. Par exemple, l'emplacement des câbles n'est pas connu, il n'y a plus trace de ce qu'il y a dans le mur à l'endroit où il faut forer, ou encore il n'est pas aisé de trouver quel disjoncteur relever pour couper l'électricité d'une certaine pièce. Pourtant toutes ces informations ont été disponibles au moment de la construction ou d'une rénovation précédente du bâtiment. Le projet SpatioData propose une solution pour encoder ces informations au moment de leur production avant qu'elles ne soient perdues. Il s'agit en fait de constituer l'historique du bâtiment.

SpatioData se présente sous la forme d'une plate-forme Web, accessible en temps réel à partir d'une station de travail, d'un smartphone ou d'une tablette. Il permet d'accéder à un modèle partagé du bâtiment et d'obtenir ou de fournir des données le concernant. C'est un projet interdisciplinaire qui regroupe des spécialistes de la construction, des architectes, des informaticiens, des géomètres et des ergonomes. Il est financé par la Région wallonne belge et a commencé en février 2011 et se terminera en janvier 2014.

Le scénario typique envisagé est celui d'un gestionnaire de bâtiments qui impose l'usage de SpatioData aux ouvriers intervenants sur son parc immobilier afin de constituer l'historique des travaux effectués. Il est également possible pour un artisan d'encoder ses interventions via l'application ou encore un particulier qui va prendre des photographies de la construction de sa maison et les archiver dans le système.

Face à cette diversité, développer une application unique est complexe. Nous avons choisi l'option de baser le système sur une architecture client-serveur.

Contrairement aux solutions existantes, la particularité du système est de permettre d'encoder des données incomplètes, contradictoires et ambiguës, voire fausses par moment. Les informations que peut d'ailleurs accueillir le système sont variées : plans, croquis, photographies, textes, fichiers informatiques, etc.

La particularité de l'interface utilisateur est qu'elle ne nécessite pas la saisie de longs formulaires fastidieux. Nous avons travaillé tout particulièrement l'ergonomie et la rapidité de la saisie de l'information. L'interface utilisateur sera adaptée aux métiers et aux besoins spécifiques des professionnels. En restant proche de l'utilisateur, la solution SpatioData ne nécessite pas l'apprentissage d'un logiciel complexe. L'intervenant encode facilement et rapidement ses informations et accède aux informations pratiques ou techniques dont il a besoin.

Bien que le système se veuille collaboratif, l'aspect sécurité des données a été une ligne directrice tout au long du développement du projet. Il sera ainsi possible de restreindre l'accès aux données et de ne pas les partager. Cette limitation est nécessaire pour respecter la vie privée des clients de l'entreprise. De plus, une compagnie ne veut pas forcément voir ses données internes diffusées dans la

plateforme. La mise en œuvre de ces aspects de sécurité liés à une information localisée est néanmoins un challenge au niveau de son implémentation

La première section expose un état de l'art des modèles de bâtiments existants. Il s'ensuit une revue des solutions existantes de gestion patrimoniale, concurrentes de SpatioData. La deuxième partie présente le système SpatioData dans sa globalité puis détaille plus spécifiquement le modèle de bâtiment au cœur du système. Enfin, nous concluons sur les perspectives de développement.

2. Besoins auxquels doit répondre le système

Les besoins du système ont été établis après une phase d'étude auprès des divers professionnels du secteur du bâtiment : architectes, entrepreneurs, chefs de chantier, maçons, électriciens, plombiers, gestionnaires de bâtiment, notaires, agents immobiliers, etc.

De cette étude, faite d'enquêtes et d'entretiens auprès d'une vingtaine d'entreprises, sont ressortis les critères suivants :

- un non-spécialiste doit pouvoir utiliser simplement le système sans formation préalable ;
- le système doit s'adapter aux besoins et méthodes de travail propres à chaque professionnel. En fonction de l'utilisateur, l'interface d'encodage et d'accès aux données doit être personnalisable afin de correspondre au mieux à ses pratiques métiers ;
- l'information doit être accessible immédiatement aussi bien du bureau que sur chantier ;
- l'information doit être facilement partagée entre les différents intervenants.
- le système doit pouvoir recevoir et enregistrer de l'information même si elle est incomplète ;
- la saisie de l'information doit être rapide, par exemple en proposant des formulaires simples, des listes d'objets avec saisie prédictive, etc. ;
- le système doit pouvoir accueillir des données très variées, comme des photographies, des mémos sonores, ainsi que les nombreux documents et rapports rencontrés tout au long de la vie du bâtiment ;
- le positionnement spatial est très important. Aussi il faut pouvoir placer facilement et rapidement un objet sur un plan ;
- le système doit être extensible afin de gérer et d'encoder des données non prévues initialement (extensibilité du modèle de données).

4

3. Etat de l'art : modèles et solutions existantes

3.1. Principaux modèles de bâtiments existants

Depuis une quinzaine d'années, les BIM (Building Information Model) ont pris une part importante lors de la phase de construction d'un bâtiment. Le BIM est un modèle de donnée commun servant à échanger et à partager les informations du bâtiment entre différents logiciels.

Dans le domaine de la construction, l'IFC (Industry Foundation Classes) est le modèle de bâtiment standard soutenu par la buildingSMART alliance. L'IFC a été initialement conçu pour organiser la phase de construction du bâtiment et partager les informations entre les participants. La norme a ensuite évolué afin de s'étendre aux phases d'exploitation, de maintenance et de démolition du bâtiment (Liebich, 2009). Cette norme définit l'entièreté du bâtiment et comporte plus de 600 classes d'objets.

En géographie, le CityGML (City Geography Markup Language), norme de l'OGC, s'impose comme le standard pour la modélisation et la représentation des objets urbains 3D. L'échelle de travail est la ville. Cependant l'intérieur du bâtiment peut être modélisé (Gröger, Kolbe et al., 2012).

De nombreux travaux présentent l'IFC et le CityGML ainsi que différents méta-modèles et techniques pour transférer les informations d'un modèle à l'autre (El-Mekawy, Östman et al., 2011, Gröger et Plümer, 2012, Nagel, Stadler et al., 2009).

L'IFC et le CityGML ne sont pas les seuls standards utilisés pour le bâtiment. INSPIRE est une directive du Parlement et du Conseil européen établissant une infrastructure d'information géographique dans la Communauté européenne, entrée en vigueur en mai 2007. Elle a pour but de faciliter les échanges de données dans le domaine de l'environnement. Les thématiques qui nous intéressent sont la modélisation des bâtiments et la gestion des adresses (INSPIRE, 2010, INSPIRE, 2013).

A ce jour, il n'existe ni norme, ni standard de facto pour le positionnement à l'intérieur d'un bâtiment. De nombreuses solutions propriétaires existent, par exemple Ekahau Real Time Location System (Ekahau, 2012), SenionLab - Indoor Positioning and Navigation (SenionLab, 2012), etc. Ces solutions proposent des systèmes de positionnement par *fingerprinting* WiFi permettant de positionner un téléphone ou une tablette dans un bâtiment. De même que pour les BIM, de nombreux travaux ont été produits, par exemple un modèle indoor générique (Lee, 2009) ou un modèle de navigation intérieure (Meijers, Zlatanova et al., 2005).

Avec le développement des smartphones, le domaine du positionnement intérieur est en plein essor. Nous pouvons citer l'initiative de Google qui propose depuis 2011 les plans de grands aéroports, de centres commerciaux, etc. (Google, 2011). L'In-Location Alliance, un consortium de vingt-deux constructeurs, s'est aussi formé afin de développer la géolocalisation à l'intérieur des bâtiments et d'offrir une interopérabilité grâce à des standards communs (Nokia, 2012). Depuis mars 2012,

un standard de modélisation indoor, porté par l'OGC, est en cours d'élaboration : il s'agit de l'IndoorGML (OGC, 2013).

3.2. Principaux concurrents

Le projet SpatioData est un projet de recherche mais aussi un projet opérationnel. Dans ce cadre particulier, l'étude de la concurrence est une étape importante dans la démarche scientifique.

Les principaux concurrents de SpatioData sont des logiciels de gestion patrimoniale, comme par exemple Abyla (Abyla, 2012), ArchiDATA (ArchiDATA, 2012) ou ISI for you (Isiom, 2012). Certains de ces logiciels exploitent plus particulièrement le BIM : ACTIVe3D (ACTIVe3D, 2012), EcoDomus PM & EcoDomus FM (EcoDomus, 2012). Il existe aussi des solutions telles que Vizelia (Vizelia, 2012), axée sur l'éco-performance des parcs immobiliers et le pilotage de la maintenance des bâtiments (GMAO). Le principal inconvénient de ces solutions est qu'elles nécessitent une gestion centralisée et un personnel qualifié, ce qui est incompatible avec les capacités des petites et moyennes structures (PME, collectivités publiques, etc.).

Il existe cependant des solutions plus proches des artisans, utilisables sur smartphone, comme Monbati (MonBati, 2012) et Mobics (Mobics, 2012) qui permettent respectivement de faire du suivi de chantier et de la maintenance des installations de chauffage.

Plusieurs constatations peuvent être faites sur ces solutions. Ces dernières nécessitent un encodage complet et détaillé des données bâtiment alors que dans la plupart des cas, l'information disponible est parcellaire, obsolète, voire erronée. Ces solutions gèrent la plupart du temps des informations textuelles et descriptives d'un problème mais n'acceptent pas d'autres types de données tels que photos, images, détails de la solution adoptée, géoréférencement de l'intervention, etc. Ces systèmes sont fermés et très rigides. Les données sont encodées et manipulées par des logiciels propriétaires. Il est donc impossible de développer ses propres programmes pour manipuler différemment les données enregistrées. Il n'est pas, non plus, possible d'ajouter des données non prévues par le système. De plus, les données demandées par la plupart des logiciels sont coûteuses à acquérir. En effet, peu de bâtiments possèdent des plans 3D au format IFC. La majorité des solutions de gestion patrimoniale n'utilise pas, non plus, les normes pour gérer le bâtiment, elles utilisent leur propre modèle interne, mais restent cependant compatibles en import et export avec les standards du marché (IFC, DXF…).

4. Approche SpatioData

4.1. Positionnement par rapport aux standards existants

Si le modèle de bâtiment de SpatioData s'inspire des différents standards évoqués ci-dessus, plusieurs caractéristiques du projet ont amené au développement d'un modèle dédié.

Tout d'abord, le modèle doit répondre à des activités et à des utilisateurs très variés, allant des considérations techniques d'un électricien aux aspects légaux d'un notaire, en passant par le sens commun d'un usager. Le modèle au cœur de SpatioData doit refléter la vision commune du bâtiment que ces différents utilisateurs partagent. C'est au travers de ce langage ou référentiel commun que ces différents acteurs pourront ranger et partager leurs informations. Les standards existants s'adressent généralement à des professionnels et supportent généralement des activités ciblées. Ainsi, la norme IFC, orientée construction, n'est pas la plus indiquée pour supporter l'activité des notaires.

Ensuite, pour permettre le développement des nombreuses applications clientes envisagées, le serveur SpatioData doit fournir une plateforme pouvant être exploitée par des informaticiens externes avec des niveaux d'expertises forts variés sur les normes liées au bâtiment. A cette fin, il apparaît plus indiqué de proposer un modèle aussi simple et intégré que possible, plutôt que d'offrir une accumulation de différentes normes souvent complexes. Google l'a bien compris et propose une API de programmation très simple pour accéder à son service de cartographie Google Maps ; en moins d'une heure, il est possible pour un programmeur d'utiliser les principales fonctions de l'API.

SpatioData étant une base de données collaborative, il est important que le système soit capable de supporter des données éparses et incomplètes sur le bâtiment. Deux utilisateurs pourraient même encoder des informations contradictoires dans le système. SpatioData n'a pas pour ambition de proposer un système strict et non ambigu, comme c'est souvent le cas des standards présentés.

Enfin, il y a de nombreuses façons de considérer le modèle à construire. Il s'agit d'un modèle BIM au sens le plus large du terme, puisque sa finalité est de capturer de nombreuses informations sur le bâtiment. Le système peut être vu également comme une base de données documentaire « classique », puisqu'il va principalement accueillir des photos, notes, rapports d'interventions ou tous autres documents utiles. Mais le système est aussi une base de données spatiale, puisque nous voulons exploiter les informations de localisation au maximum lorsqu'elles sont disponibles. Le choix d'un standard bien précis aurait orienté le modèle vers un de ces domaines au détriment des autres.

4.2 Modèle de bâtiment SpatioData

De ces observations, nous avons tiré quatre principes généraux afin de nous guider pour le choix du niveau d'abstraction adéquat de notre modèle :

- la simplicité : pour répondre aux besoins des utilisateurs variés mais aussi des développeurs externes envisagés, il est nécessaire d'avoir un modèle simple et facile à appréhender ;
- l'expressivité : le modèle doit être capable de représenter et de lier les innombrables éléments possibles du bâtiment;
- la généralité : le modèle servant de base à tous les acteurs et à toutes les activités liées au bâtiment, il doit proposer des concepts compréhensibles et exploitables par tous ;
- l'extensibilité : puisque des applications clientes viendront répondre à des besoins précis et variés, il est nécessaire qu'elles puissent étendre le modèle par leurs propres concepts et modèles « métiers ».

Pour répondre à ces besoins, le modèle sera constitué de 2 niveaux. Le premier est un modèle simple, formé de quelques classes à peine, d'un niveau d'abstraction intermédiaire et responsable de la logique de composition des différents concepts du modèle. Le second est constitué d'une ontologie, venant décrire toute la richesse des éléments dont il est question dans le bâtiment. Cette ontologie présente la première possibilité d'extensibilité du système : une application cliente traitant d'électricité, par exemple, peut compléter l'ontologie existante de son vocabulaire métier, sans que cela affecte la logique de composition proposée par le niveau 1. Les figures 1 et 2 illustrent les éléments clés du modèle SpatioData.

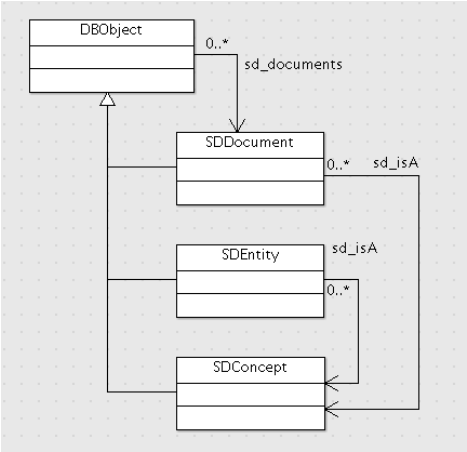


Figure 1. Organisation des concepts principaux

8

La figure 1 montre la hiérarchie des classes de haut niveau du modèle. Trois types d'objets viennent se combiner :

- *SDEntity* est la classe de base de toute entité spatiale. Ces entités sont détaillées ci-dessous ;
- *SDDocument* représente tous les documents supportés par le système, que ce soit des photos, des annotations, ou des documents plus complexes comme des formulaires ou des rapports d'interventions ;
- *SDConcept* est la classe décrivant les éléments des ontologies. Tout document ou toute entité est plus précisément typé par un ou plusieurs concepts présents dans les ontologies.

La figure 2 illustre les entités spatiales. Elles constituent une partie importante du modèle et leur niveau d'abstraction a été choisi avec soin. Outre la classe de base *SDEntity*, ce diagramme ne comprend que 3 classes :

- *SDSpace* représente les espaces du modèle, par exemple les bâtiments eux-mêmes, les couloirs ou les locaux. Les espaces peuvent maintenir des relations d'adjacence avec d'autres espaces, ils peuvent contenir d'autres entités spatiales et peuvent être liés à des frontières.
- *SDFrontier* représente les différentes frontières des bâtiments, que ce soit les murs, les planchers, les toits, etc. Il est également possible de définir des frontières virtuelles. Par exemple, un utilisateur peut décomposer son loft en des zones fonctionnelles plus précises, avec un espace salon séparé virtuellement d'un espace cuisine.
- *SDElement* représente les éléments plus « atomiques » du bâtiment, comme les prises électriques, le mobilier, les extincteurs, etc.

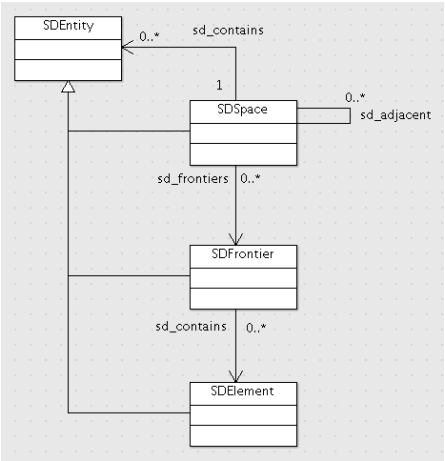


Figure 2. Modèle de données SpatioData

Cette logique de composition se retrouve dans beaucoup de standards existants, comme les IFC. On la retrouve aussi dans des travaux sur la conception architecturale (Eastman et Siabiris, 1995, Ekholm et Fridqvist, 1996). Ce modèle simple peut être rapidement compris par des développeurs n'ayant aucune expérience particulière dans le bâtiment. La logique de décomposition est découplée de la complexité liée à l'énumération quasi infinie des entités concrètes que l'on peut rencontrer.

Il est à noter que le schéma des entités n'est pas spécifique au monde du bâtiment. Il serait facile d'adapter ce système pour d'autres domaines en changeant les ontologies descriptives.

4.3. Base de données en graphe et services Web

Depuis quelques années, on assiste à l'émergence de bases de données dites NoSQL, aujourd'hui utilisées par de grands acteurs du Web comme Facebook ou Google. Pour le projet SpatioData, une base de données en graphe a été utilisée. Dans ce type de base de données, les données ne sont plus organisées en tables selon un schéma strict, mais plutôt en nœuds et arcs libres, qui acceptent des paires clé/valeur, sans schéma prédéfini. L'approche que nous avons adoptée ainsi que les bénéfices tirés de cette base de données en graphe vont être résumés ici.

UML ayant été choisi comme formalisme de modélisation du projet, nous avons développé notre propre mapping UML vers la base de données en graphe (Delfosse, Billen et al., 2012). Cette tâche fut assez simple, une structure en graphe épousant presque directement les diagrammes de classe UML. Le mapping UML-graphe est donc beaucoup plus commode que le mapping UML-relationnel. Aussi, du point de vue de la recherche, la base de données ne devant pas suivre un schéma strict, il a été bien plus facile de faire évoluer notre modèle, comparé à un système SQL traditionnel.

Mais l'avantage majeur est apparu lorsqu'il a été question de décider des services Web que le serveur de données SpatioData allait proposer. Puisque le serveur SpatioData est principalement axé sur la modélisation des données plutôt que sur une logique métier laissée à la responsabilité des applications clientes, le choix a été fait de fournir des services Web dits « orientés UML ». Chaque type de relation UML présente dans un diagramme de classes se traduit par des services Web permettant de la naviguer. Ainsi, les services développés dépendent de la norme UML et pas d'un modèle particulier de données. D'autres services Web sont venus compléter le système pour offrir des services plus complexes comme l'authentification et la gestion des droits d'accès ou pour cacher des logiques plus complexes comme celles des index spatiaux.

Tant la base de données que les services sont construits autour de la norme UML générale plutôt que d'un modèle UML spécifique. Il devient alors possible aux applications clientes d'étendre, avec leurs propres schémas UML, ceux déjà définis dans SpatioData. Par exemple, une application cliente responsable du recensement du matériel lié au gaz sur un campus peut décider de lier à tout objet *SDSpace* une

référence vers un autre objet dont elle aura elle-même défini les propriétés pour refléter les données à collecter. Référez-vous à (Delfosse, Billen et al., 2012) pour plus d'informations sur cette approche originale.

Cette stratégie apporte au système un niveau d'extensibilité supplémentaire, puisque la structure même des données peut être étendue. Bien sûr, cette approche n'est pas exempte de problème et il convient d'encadrer au mieux ces extensions pour éviter un développement anarchique des données du système. A cette fin, chaque application cliente doit préfixer les classes et les propriétés qui lui sont propres d'un tag bien spécifique. Aussi, les applications clientes sont authentifiées et leurs activités sont suivies pour éviter tout abus. Enfin, les développeurs tiers, codant des plug-ins pour SpatioData, sont encadrés et encouragés à soumettre leurs modèles au consortium SpatioData afin de valider l'approche entreprise et, le cas échéant, de partager ces modèles avec d'autres développeurs travaillant sur des domaines similaires.

Malgré les problèmes qui peuvent résulter de cette stratégie, cette dernière ouvre un champ inattendu de recherche. Le système SpatioData était initialement pensé pour permettre à des utilisateurs de partager toutes les informations disponibles sur le bâtiment. Or, il offre maintenant la possibilité de faire partager non seulement les données mais aussi les modèles de données liés aux très nombreuses activités du bâtiment. Ainsi, SpatioData peut être vu comme un système de construction et de partage collaboratif de modèles UML dédiés au bâtiment.

4.4 Le système global

SpatioData se base sur l'architecture classique client-serveur. Le serveur est responsable de la base de données et de fournir des services Web compatibles avec les standards du W3C (World Wide Web Consortium), à savoir SOAP (Simple Object Access Protocol) et WSDL (Web Services Description Language). D'autres standards sont utilisés par notre architecture comme le WKT pour l'échange des géométries. Ces technologies permettent à des entreprises indépendantes de développer leurs propres applications clientes SpatioData, ce qui offre une très grande interopérabilité au système.

Les applications clientes peuvent être très variées. Elles sont responsables de porter la logique métier et d'offrir des interfaces dédiées aux nombreuses activités envisagées. Dans le cadre du projet SpatioData nous avons développé un client principal, codé en JavaScript et HTML5. Il est ainsi exécutable dans n'importe quel navigateur Web compatible et fonctionne sur poste fixe, smartphone ou tablette.

L'interface et les fonctions du client SpatioData sont personnalisables par l'utilisation d'un système de plug-ins. Cette conception sous forme modulaire est employée dans de nombreux logiciels, AutoCAD, ArcGIS… Ainsi, une société tierce peut soit développer une application client dédiée, soit étendre le client SpatioData en créant un module supplémentaire propre à un métier ou à un besoin spécifique.

La figure 3 détaille les composants de la plateforme SpatioData évoqués dans cet article.

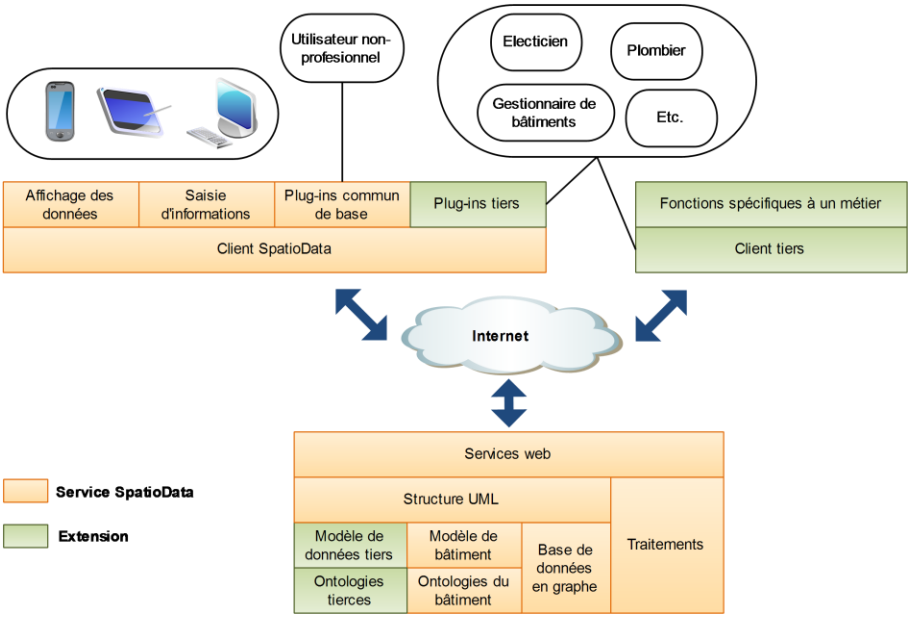


Figure 3. Plateforme SpatioData

5. Conclusion et perspectives

SpatioData propose une plateforme Web collaborative et adaptable pour un partage facile et efficace de toutes informations relatives au bâtiment. Il a l'ambition d'être un référentiel commun pour une multitude d'applications clientes, chacune dédiée à une activité spécifique. Par ailleurs, il est possible d'étendre la structure de la base de données et de définir de nouveaux domaines d'application par l'utilisation des ontologies.

Dans cet article nous avons détaillé principalement le modèle de bâtiment, cependant le système comporte aussi une gestion des adresses, un module plan, etc. Les prochaines étapes sont la validation de la robustesse du client et la résolution des différents problèmes techniques liés à la montée en charge du serveur et les aspects de sécurité liés aux données spatialisées.

Le projet entre dans sa phase de déploiement, le système va être lancé et être accessible par tous avant la fin de l'année 2013.

Remerciements

Le projet SpatioData est financé par la Région wallonne dans le cadre du programme WIST 3 N°1017094.

12

Bibliographie

Abyla (2012). http://www.abyla.fr.

ACTIVe3D (2012). http://www.active3d.net.

ArchiDATA (2012). http://www.archidata.com.

Delfosse, V., R. Billen et P. Leclercq (2012), UML as a schema candidate for graph databases, *NoSQL Matters 2012*, Koln, Germany.

Eastman, C. M. et A. Siabiris (1995), A generic building product model incorporating building type information, *Automation in construction* 3(4): 283-304.

EcoDomus (2012). *EcoDomus software & services: BIM & COBie Lifecycle Data Management*, http://www.ecodomus.com.

Ekahau (2012). *Ekahau Real Time Location System (RTLS)*, http://www.ekahau.com/products/real-time-location-system/overview.html.

Ekholm, A. et S. Fridqvist (1996), Modelling of user organisations, buildings and spaces for the design process, *CIB REPORT*: 165-178.

El-Mekawy, M., A. Östman et K. Shahzad (2011). Towards interoperating cityGML and IFC building models: a unified model based approach, *Advances in 3D Geo-Information Sciences*, Springer: 73-93.

Google (2011). *A new frontier for Google Maps: mapping the indoors*, http://googleblog.blogspot.be/2011/11/new-frontier-for-google-maps-mapping.html.

Gröger, G., T. H. Kolbe, C. Nagel et K.-H. Häfele (2012), *OGC City Geography Markup Language (CityGML) Encoding Standard*, Open Geospatial Consortium

Gröger, G. et L. Plümer (2012), CityGML–Interoperable semantic 3D city models, *ISPRS Journal of Photogrammetry and Remote Sensing* 71: 12-33.

INSPIRE, I. f. S. I. i. E. (2010), *D2.8.I.5 INSPIRE Data Specification on Addresses - Guidelines* INSPIRE Thematic Working Group Addresses.

INSPIRE, I. f. S. I. i. E. (2013), *D2.8.III.2 INSPIRE Data Specification on Buildings - Draft Technical Guidelines*, INSPIRE Thematic Working Group Buildings.

Isiom (2012). *ISI for you*, http://www.isiom.fr/solution/ify.html.

Lee, J. (2009), *Indoor Spatial Data Model: Space Space oriented Data Model oriented Data Model*, 8 th Workshop on Indoor Spatial Awareness.

Liebich, T. (2009), *IFC 2x Edition 3 Model Implementation Guide Version 2.0*, Dresden: Building SMART International Modeling Support Group.

Meijers, M., S. Zlatanova et N. Pfeifer (2005), 3D geoinformation indoors: structuring for evacuation, *Proceedings of Next generation 3D city models*.

Mobics (2012). *L'union fait la force, MOBICS la différence*, http://www.ubbu-ics.be/fr/mobics.html.

MonBati (2012). *Le projet MonBati*, http://www.monbati.com.

Nagel, C., A. Stadler et T. H. Kolbe (2009), Conceptual requirements for the automatic reconstruction of building information models from uninterpreted 3D models, *Academic Track of Geoweb 2009 Conference, Vancouver*.

Nokia (2012). *Accurate Mobile Indoor Positioning Industry Alliance, called In-Location, to promote deployment of location-based indoor services and solutions* http://press.nokia.com/2012/08/23/accurate-mobile-indoor-positioning-industry-alliance-called-in-location-to-promote-deployment-of-location-based-indoor-services-and-solutions/.

OGC (2013). *IndoorGML SWG* http://www.opengeospatial.org/projects/groups/indoorgmlswg.

SenionLab (2012). http://www.senionlab.com.

Vizelia (2012). *Vizelia - GMAO - Gestion Energie Bâtiment - Gestion Patrimoine Immobilier - Pilotez durablement votre patrimoine*, http://www.vizelia.com.

Une combinaison de graphes spatiaux pour la modélisation de la structure spatiale des territoires bâtis

Mathieu Domingo, Rémy Thibaud, Christophe Claramunt

Institut de Recherche de l'Ecole Navale (IRENav), BCRM Brest, CC 600, Ecole Navale, 29240 Brest, cedex 9

RESUME.
Cet article décrit un modèle de représentation de la structure des territoires bâtis basé sur la théorie des graphes. Afin de rendre compte des différents éléments de l'espace, il intègre trois types d'entités géographiques à savoir les parcelles, les routes et le bâti. Utilisant les propriétés topologiques décrites au niveau de chacun de ces types d'entités, mais également entre des entités de type différents, il produit un graphe connexe permettant d'analyser la structuration du bâti au travers de son interconnexion avec le parcellaire et les voies de circulation. Les opérateurs de graphe génériques n'étant pas adaptés à notre modèle, nous décrivons ensuite un opérateur de centralité personnalisé. Celui-ci permet d'attribuer aux entités du bâti des valeurs calculées et pondérées en prenant en compte non seulement l'adjacences de ces entités, mais aussi leurs relations distantes.

ABSTRACT.
This paper introduces a graph-based model for the representation of spatial structures. It integrates three complementary spatial entities: parcels, roads and buildings. Topological relations between entities of similar or different types generate a graph that favors the analysis of structural properties. A centrality operator calculating weighted values for buildings is introduced and applied to a sample that reveals several local and global patterns.

MOTS-CLES : Graphe, Syntaxe spatiale, Structure spatiale.

KEYWORDS : Graph , Space syntax, Spatial Structure

1. **Introduction**

La représentation et l'analyse de la structuration et de la dynamique des espaces géographiques restent encore des problématiques ouvertes dans de nombreux domaines liés à l'aménagement, la gestion et la planification des territoires. L'objectif consiste à étudier la relation entre la structure et le fonctionnement du territoire, afin d'en extraire des constats et des scénarios d'évolution. Dans le cas des environnements bâtis, la syntaxe spatiale a permis de mettre en place une approche structurelle autour de l'analyse des réseaux. Se basant principalement sur (Hillier et Hanson, 1984), livre fondateur de la syntaxe spatiale, les recherches actuelles se concentrent d'abord sur la morphologie du territoire pour en déduire l'empreinte de processus économiques et sociaux. Une critique souvent formulée à propos de la syntaxe spatiale est de limiter l'analyse des territoires à l'analyse des réseaux de rues et de routes, et de se focaliser sur l'analyse des connectivités de ces réseaux pour structurer l'espace (Ratti, 2004). Ces critiques suggèrent de prendre en compte d'autres paramètres tels que l'attractivité du bâti et du parcellaire, et des contextes moins connectés (îlots ou hameaux isolés) dont le développement ne dépend pas de leur niveau d'intégration global au territoire mais de celui du niveau local (Buhl *et al.*, 2006). Il faudrait dans ce sens étendre la modélisation des territoires en y intégrant le bâti et le parcellaire, et potentiellement tout autre information géographique pouvant influencer la façon dont se structure l'espace. L'intérêt est de pouvoir étendre l'analyse de la structure spatiale à des niveaux d'abstractions différents des simples voies de circulation. Cette orientation est plus particulièrement adaptée à l'observation de la structuration des territoires ruraux. En effet, contrairement au domaine urbain ou le lien entre l'organisation du bâti et celui des rues est très fort, en domaine rural la structuration du territoire s'articule autour d'îlots bâti (bourg, hameaux, villages, …) au sein desquels vient s'entrelacer le réseau de circulation assurant le lien entre ces îlots. Ainsi, la structuration s'articule tout autant autour de la morphologie du parcellaire que celui des réseaux. En conséquence, nous proposons ici d'étendre l'approche d'analyse de la structure des territoires à partir d'un modèle de représentation qui prenne en compte la morphologie des réseaux de circulation et celle du bâti et du parcellaire. Cette structuration propose une approche qui repose sur des concepts topologiques et de connexion entre entités représentées. Il s'agit d'une approximation qui pourra être précisée par des points de vue provenant des applications pour tenir compte par exemple de la typologie de ces connexions et par exemple de la présence d'obstacles (haies, clôtures, barrières). Concernant les associations du parcellaire et du bâti avec le réseau de transport, la typologie du réseau routier et son importance pourra également être qualifié.

Au regard des travaux précédents nous avons choisi de baser notre approche sur la théorie des graphes (Sriti *et al.*, 2005, Del Mondo *et al.*, 2013) et de la syntaxe spatiale (Hillier et Hanson, 1984, Jiang et Claramunt, 2002). Ces approches étant intimement liées, elles semblent bien adaptées à la représentation, la gestion et l'analyse de relations entre entités géographiques. Nous définissons trois types

d'entités : le bâti, le parcellaire et les voies de circulation. Les relations spatiales se focalisent autour de concepts topologiques et se limitent aux relations de connexion. Cette orientation, qui écarte les relations de type quantitative (distance euclidienne ou réticulaire), évite d'introduire dans le modèle des seuils de distances entraînant des approximations dans les notions de voisinage et de proximité. Elle permet également d'éviter que le graphe de relations soit fortement contraint par ces seuils et soit non stationnaire (Badariotti *et al.,* 2007). L'autre originalité de l'approche est l'utilisation d'une combinaison de plusieurs graphes pour représenter le territoire, dans l'objectif d'obtenir un graphe total connexe. En effet, dans le cas du bâti, l'utilisation exclusive de relations de connexion entraine la construction d'un graphe non connexe (tous les bâtiments et autres constructions ne sont pas joints topologiquement). La combinaison de ces trois graphes permet de pallier cette limitation en assurant un lien entre toutes les entités construites, par le biais des graphes de parcelles et de voies de circulation. Dans ce cadre nous présentons ici les éléments conceptuels et formels qui définissent les opérateurs et les associations spatiales, et permettent d'aboutir à la combinaison d'un graphe connexe pour modéliser la structure d'un territoire habité. Dans un second temps, l'article décrit des opérateurs spécifiques permettant d'extraire les propriétés du graphe. Enfin, après avoir décrit l'implémentation du modèle sur un cas réel, nous mettons en évidence son utilisation pour la mesure locale de degré.

2. Définition des éléments de l'espace E

Dans le modèle proposé, le territoire (exemple de la Figure 1) est représenté par un espace E défini dans $\mathbb{R}^2$ dans lequel sont localisées des régions. L'espace E considéré est défini dans un espace topologique modélisé par un ensemble de points (point set). Nous considérons trois types de régions correspondant respectivement au bâti (maisons), au parcellaire (parcelles) et aux voies de circulation (routes) modélisées ainsi :

– un ensemble fini M de maisons représentées par des régions m_i de E avec $1 \leq i \leq n$ $|M = \{m_1,\ldots, m_n\}$ où $n \in \mathbb{N}$

– un ensemble fini P de parcelles représentées par des régions p_j de E avec $1 \leq j \leq m$ $|P = \{p_1,\ldots, p_m\}$ où $m \in \mathbb{N}$

– un ensemble fini R de routes représentées par des régions r_k de E avec $1 \leq k \leq o$ $| R = \{r_1,\ldots, r_o\}$ où $o \in \mathbb{N}$

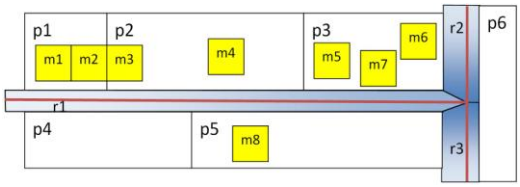


Figure 1 : Représentation schématique d'un territoire habité avec $M=\{m_1,m_2,\ldots,m_8\}$ $P=\{p_1,p_2,\ldots,p_6\}$ et $R=\{r_1,r_2,r_3\}$

Une première propriété particulière de cet espace E est que les maisons étant toujours situées sur une ou des parcelles, l'union des régions maisons de M sera incluse dans l'union des régions parcelles de P d'où $M \subset P$. Une seconde propriété est que l'ensemble des parcelles P et l'ensemble des routes R forment une partition de E ($P \cap R = \emptyset$ et $P \cup R = E$). Ces deux propriétés impliquent que l'ensemble des maisons et l'ensemble des routes sont spatialement disjoints ($\forall\ m_i \in M, \forall\ r_k \in R \mid m_i \cap r_k = \emptyset$) mais sont liés par l'intermédiaire des parcelles.

2.1 Operateurs et association spatiale

2.1.1 Propriétés topologiques

Toute région r est constituée de son intérieur r^0 et de sa frontière ∂r tels que définis par (Egenhofer et Franzosa, 1991). Pour ne pas introduire dans le modèle des seuils de distances entraînant des approximations dans les notions de voisinage et de proximité, les relations entre les régions de l'espace sont basées sur les relations topologiques du RCC8 (Cohn *et al.*, 1997) qui comprennent sept cas de connexion et un cas de déconnexion.

La mise en place du graphe spatial repose sur l'exploitation d'opérateurs permettant d'établir le type de relation entre les régions. La formalisation de ces opérateurs repose sur l'utilisation d'opérateurs génériques permettant de préciser si les régions sont adjacentes, s'intersectent ou sont incluses l'une dans l'autre. L'opérateur d'union est défini comme un opérateur ensembliste sur un espace E représenté par un ensemble de points, et les opérateurs d'inclusion et d'intersection comme des opérateurs Booléen.

Plus formellement, soit x un point de notre espace, A et B des ensembles de points de notre espace :

– **Union** : $A \cup B = \{x \mid x \in A \vee x \in B\}$

– **Inclusion** : $A \subset B$ est vrai ssi $\forall\ x \in A \Rightarrow x \in B$

– **Intersection** : $A^0 \cap B^0$ est vrai ssi $\exists\ x \in E \mid x \in A^0 \wedge x \in B^0$. Ainsi dans le cas où A et B sont seulement connectés par frontières l'intersection de leurs intérieurs est vide.

2.1.1.1 Propriétés topologiques entre maisons et parcelles

Les caractéristiques de l'espace E permettent de définir des propriétés entre les maisons de M et les parcelles de P en utilisant les relations définies dans le RCC8 (Figure 2)

– **Propriété 1** : toute maison est incluse dans une parcelle ou l'union de plusieurs parcelles :

$$\forall\ m_i \in M, \exists p_j, \ldots p_o \in P \mid m_i \subset (p_j \cup \ldots \cup p_o) \text{ avec } 1 \leq i \leq n \text{ et } 1 \leq j \leq o \text{ avec } n,o \in \mathbb{N}$$

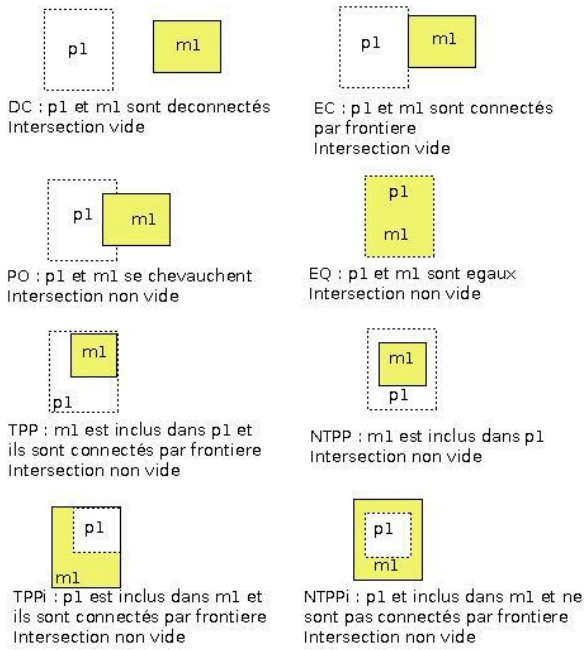


Figure 2 : relations du RCC8 entre maison(m_1) et parcelle(p_1)

– **Propriété 2** : quelles que soient deux maisons m_i et m_j de M alors, soit m_i est déconnectée de m_j , soit m_i est connectée par frontière à m_j

$\forall\ m_i, m_j \in M \Rightarrow m_i\ DC\ m_j \vee m_i\ EC\ m_j$ avec $1 \leq i \leq n$ et $1 \leq j \leq n$ et $i \neq j$ avec $n \in \mathbb{N}$

– **Propriété 3** : quelles que soient deux parcelles p_k et p_l de P alors, soit p_k est déconnectée de p_l , soit p_k est connectée par frontière à p_l

$\forall\ p_k, p_l \in P \Rightarrow p_k\ DC\ p_l \vee p_k\ EC\ p_l$ avec $1 \leq k \leq o$ et $1 \leq l \leq o$ et $k \neq l$ avec $o \in \mathbb{N}$

2.1.1.2 Propriétés topologiques entre parcelles et routes

– **Propriété 4** : quelles que soient deux routes r_i et r_j de R alors, soit r_i est déconnectée de r_j , soit r_i est connectée par frontière à r_j

$\forall\ r_i, r_j \in R \Rightarrow r_i\ DC\ r_j \vee r_i\ EC\ r_j$ avec $1 \leq i \leq n$ et $1 \leq j \leq n$ et $i \neq j$

– **Propriété 5** : L'ensemble des routes forme une région connectée : à partir de toute route, il est possible d'atteindre n'importe quelle autre route. Nous ne considérons pas le cas d'un réseau déconnecté, par exemple une ile séparé du continent.

$\forall\ r_i, r_j \in R, \exists\ (r_k \cup \ldots \cup r_m) \in R \mid (r_i \cup (r_k \cup \ldots \cup r_m))\ EC\ r_j$

– **Propriété 6 :** quelles que soient une parcelle p_i de P et une route r_j de R alors soit p_i est déconnectée de r_j soit p_i est connectée par frontière à r_j

$\forall\ p_i \in P, r_j \in R \Rightarrow p_i\ DC\ r_j \vee p_i\ EC\ r_j$ avec $1 \leq i \leq n$ et $1 \leq j \leq o$

2.1.2 Opérateurs d'adjacences et d'intersections

Afin de construire le graphe qui intègre à la fois le bâti, les parcelles et les voies de circulation, il nous faut identifier les opérateurs qui permettent de décrire les relations entre les parcelles, entre les routes, et entre les maisons. De plus, les parcelles étant utilisées pour relier les maisons et les routes, il faut également identifier les relations entre maisons et parcelles et entre parcelles et routes.

– **$VOISIN_M(m_i)$** donne les maisons qui sont adjacentes à une maison m_i.

- $VOISIN_M : M \rightarrow 2^M$
- $VOISIN_M (m_i) = \{m \in M \mid m \text{ EC } m_i\}$

– **$VOISIN_P(p_i)$** donne les parcelles qui sont adjacentes à une parcelle p_i.

- $VOISIN_P : P \rightarrow 2^P$
- $VOISIN_P (p_i)= \{p \in P \mid p \text{ EC } p_i\}$

– **$VOISIN_R(r_i)$** donne les routes qui sont adjacentes à une route r_i.

- $VOISIN_R : R \rightarrow 2^R$
- $VOISIN_R (r_i)= \{r_j \in R \mid r_j \text{ EC } r_i\}$

– **$INT_{M-P}(m_i)$** donne les parcelles qui s'intersectent une maison m_i

- $INT_{M-P} : M \rightarrow 2^P$
- $INT_{M-P} (m_i)= \{p \in P \mid m_i \cap p \neq \emptyset\}$

– **$INT_{P-M}(p_i)$** donne les maisons qui s'intersectent avec une parcelle p_i

- $INT_{P-M} : P \rightarrow 2^M$
- $INT_{P-M} (p_i) = \{m \in M \mid m \cap p_i \neq \emptyset\}$

– **$ADJ_{R-P}(r_i)$** donne les parcelles qui sont adjacentes à une route r_i.

- $ADJ_{R-P} : R \rightarrow 2^P$
- $ADJ_{R-P} (r_i)= \{p \in P \mid p \text{ EC } r_i\}$

– **$ADJ_{P-R}(p_i)$** donne les routes qui sont adjacentes à une parcelle p_i.

- $ADJ_{P-R} : P \rightarrow 2^R$
- $ADJ_{P-R} (p_i)= \{r \in R \mid r \text{ EC } p_i\}$

2.1.3 Association spatiales

Dans le cadre de la mise en place des graphes, les relations entre les nœuds du graphe sont basées sur l'utilisation de fonctions d'association qui permettent d'identifier si deux régions sont liées. À partir de la définition des opérateurs VOISIN, INT et ADJ nous pouvons définir les associations élémentaires suivantes :

– **$Asso2M(m_i,m_j)$** : Une maison m_i est associée à une maison m_j si et seulement si m_i et m_j sont adjacentes :

- $Asso2M : M \times M \rightarrow$ Booléen
- $Asso2M(m_i,m_j)$=vrai ssi $m_j \in VOISIN_M(m_i)$ ($\Leftrightarrow m_i \in VOISIN_M(m_j)$)

– **Asso2P(p_i,p_j)** : Une parcelle p_i est associée à une parcelle p_j si et seulement si p_i et p_j sont adjacentes :

- Asso2P : P x P → Booléen
- Asso2P(p_i,p_j)=vrai ssi $p_j \in VOISIN_P(p_i)$ ($\Leftrightarrow p_i \in VOISIN_P(p_j)$)

– **Asso2R(r_i,r_j)** : Une route r_i est associée à une route r_j si et seulement si r_i et r_j sont adjacentes

- Asso2R : R x R → Booléen
- Asso2R(r_i,r_j)=vrai ssi $r_j \in VOISIN_R(r_i)$ ($\Leftrightarrow r_i \in VOISIN_R(r_j)$)

– **AssoMP(m_i,p_j)** : Une maison m_i est associée à une parcelle p_j si et seulement si leur intersection est non vide.

- AssoMP : M x P → Booléen
- AssoMP(m_i,p_j) = vrai ssi $p_j \in INT_{M\text{-}P}(m_i)$ ($\Leftrightarrow m_i \in INT_{P\text{-}M}(p_j)$)

– **AssoPR(p_i,r_j)** : Une parcelle p_i est associée à une route r_j si et seulement si p_i et r_j sont adjacentes.

- AssoPR: P x R → Booléen
- AssoPR (p_i,r_j) = vrai ssi $p_i \in ADJ_{R\text{-}P}(r_j)$ ($\Leftrightarrow r_j \in ADJ_{P\text{-}R}(p_i)$)

3. Modèle multiple de graphe de l'espace E

Un graphe G est un couple (S, A) tel que S est un ensemble non vide de sommets et A un ensemble d'arêtes représentant une relation *r*. Une arête de G est une paire de sommets de G.

Pour représenter et manipuler efficacement nos données spatiales, à partir des associations spatiales que nous venons de définir, nous définissons un modèle de graphe composé de 3 graphes simples représentant respectivement l'ensemble des routes, l'ensemble des maisons et l'ensemble des parcelles (Figure 3).

3.1 Graphes simples

– **Graphe de maisons adjacentes** G_M= (S_M, A_M) tel que :

- S_M = M
- A_M = {(m_i,m_j) ∈ M | Asso2M(m_i,m_j) =vrai} représente les maisons adjacentes dans E

– **Graphe de parcelles adjacentes** G_P= (S_P, A_P) tel que :

- S_P = P
- A_P= {(p_i,p_j) ∈ P | Asso2P(p_i,p_j) =vrai} représente les parcelles adjacentes dans E

– **Graphe de routes adjacentes** G_R= (S_R, A_R) tel que :

- S_R=R
- A_R= {(r_i,r_j) ∈ R | Asso2R(r_i,r_j) =vrai} représente les routes adjacentes dans E

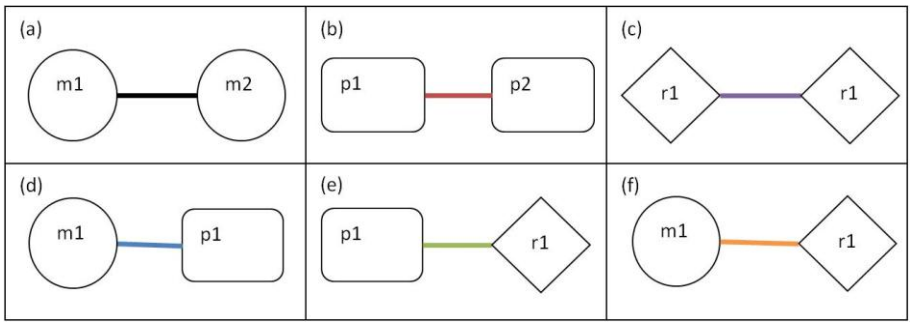


Figure 3 : Représentation des différents liens possibles dans le modèle de graphe entre : (a) 2 maisons adjacentes,(b)2 parcelles adjacentes,(c)2 routes adjacentes, (d) 1 maison et 1 parcelle qui s'intersectent,(e)1 parcelle et 1 route adjacentes, (f)1maison et 1 route sont liées à une même parcelle

3.2 Combinaisons de graphes simples

Il n'existe pas de lien topologique direct dans notre modèle entre les maisons et les routes. Afin de construire un graphe combinant les maisons, les parcelles et les routes, nous devons donc établir le lien entre les maisons et les routes via les parcelles. En conséquence nous définissons le graphe reliant les maisons et les parcelles ainsi que le graphe reliant les parcelles et les routes.

– **Graphe de maisons et de parcelles** $G_{MP}= (S_{MP}, A_{MP})$ tel que :

- $S_{MP}= M \cup P$
- $A_{MP}= A_M \cup A_P \cup \{(m_i,p_j) \in M \times P \mid AssoMP(m_i,p_j) = vrai\}$ représente les associations des maisons avec les parcelles qu'elles intersectent et les associations de maisons adjacentes et de parcelles adjacentes de E (Figure 3a b d)

– **Graphe de parcelles et de routes** $G_{PR}= (S_{PR}, A_{PR})$ tel que :

- $S_{PR}= P \cup R$
- $A_{PR}= A_P \cup A_R \cup \{(p_i,r_j) \in P \times R \mid AssoPR(p_i,r_j) = vrai\}$ représente les associations des parcelles avec les routes qui leurs sont adjacentes et des associations de parcelles adjacentes et de routes adjacentes de E (Figure 3b c e)

3.3 Graphe complet

Le graphe final est le résultat de la combinaison des graphes précédents et nous permet d'établir un graphe connexe incluant les différents types d'associations qui existent entre les maisons, les parcelles et les routes.

– **Graphe total connexe** $G_{MPR}= (S_{MPR}, A_{MPR})$ tel que :

- $S_{MPR}= M \cup P \cup R$

- A_{MPR}= A_M ∪ A_P ∪ A_R ∪ A_{MP} ∪ A_{PR} ∪ {(m_i,r_j) ∈ M x R | ∃p_k ∈ P, AssoMP(m_i,p_k) =vrai ∧ AssoPR(p_k,r_j) =vrai } représente les associations des maisons et des routes respectivement associées à une même parcelle en plus des associations de maisons adjacentes, parcelles adjacentes, routes adjacentes de E ainsi que les associations des maisons avec les parcelles qu'elles intersectent et les associations des parcelles avec les routes qui leurs sont adjacentes. (Figure 3 complète)

3.4 Traitement complet d'un exemple à travers le modèle de graphe proposé

Afin d'illustrer la mise en œuvre du graphe complet, nous utilisons la configuration simplifiée du territoire de la Figure 1. Celui-ci regroupe 8 maisons, 6 parcelles et 3 routes.

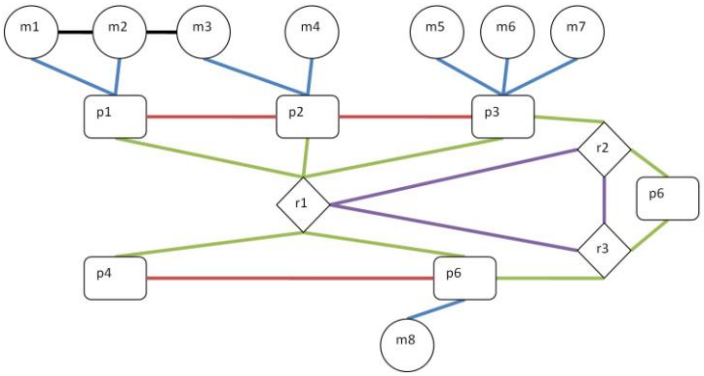


Figure 4 : Graphe total[1] correspondant au territoire illustré dans la Figure 1

Les différents sous-graphes, et le graphe total sont construits comme suit et représentés sur la Figure 4 :

Graphe de maisons adjacentes : G_M= (S_M, A_M) avec S_M = {m_1; m_2; m_3; m_4; m_5; m_6; m_7; m_8} et A_M = {(m_1, m_2); (m_2, m_3)}

Graphe de parcelles adjacentes : G_P= (S_P, A_P) avec S_P = {p_1; p_2; p_3; p_4; p_5; p_6} et A_P = {(p_1, p_2); (p_2, p_3); (p_4, p_5)}

Graphe de routes adjacentes : G_R= (S_R, A_R) avec S_R = {r_1; r_2; r_3} et A_R = {(r_1, r_2); (r_2, r_3); (r_1, r_3)}

Graphe de maisons et parcelles : G_{MP}= (S_{MP}, A_{MP}) avec S_{MP} = S_M ∪ S_P et A_{MP} = A_M ∪A_P ∪{(m_1,p_1); (m_2,p_1); (m_3,p_2); (m_4,p_2); (m_5,p_3); (m_6,p_3); (m_7,p_3); (m_8,p_5)}

Graphe de parcelles et de routes : G_{PR}= (S_{PR}, A_{PR}) avec S_{PR} = S_P ∪ S_R et A_{PR} = A_P ∪A_R ∪{(p_1,r_1); (p_2,r_1); (p_3,r_1); (p_4,r_1); (p_5,r_1); (p_3,r_2); (p_5,r_3); (p_6,r_2); (p_6,r_3)}

Graphe final : G_{MPR}= (S_{MPR}, A_{MPR}) avec S_{MPR} = S_M ∪S_P ∪S_R et A_{MPR} = A_M ∪A_P ∪ A_R ∪A_{MP} ∪ A_{PR} ∪ {(m_1,r_1); (m_2,r_1); (m_3,r_1); (m_4,r_1); (m_5,r_1); (m_6,r_1); (m_7,r_1); (m_8,r_1); (m_5,r_2); (m_6,r_2); (m_7,r_2); (m_8,r_3)}

[1] Par souci de visibilité les arcs entre les maisons et les routes ne sont pas représentés

4. Analyse du graphe

Afin d'illustrer l'impact de cette approche sur la caractérisation de la structure d'un territoire, nous proposons d'observer l'extraction d'une mesure locale de degré concernant les maisons (Freeman, 1979). Les graphes sont établis à partir des données d'une partie du territoire de la commune de Guisseny (Figure 5) : 18 maisons, 22 parcelles et 5 tronçons de route. La mesure de degré est un indicateur local, mesuré sur le graphe G=(S, A), qui est défini pour chaque sommet comme le nombre de connexions directes vers les autres sommets.

$C_{Deg}(S_i) = Deg(S_i) = |\{ S_j \in S \mid (S_i, S_j) \in A \}|$

Cette mesure est adaptée dans le cas où le graphe est connexe et regroupe des entités géographiques d'un seul type. Ce dernier point n'est pas satisfait dans le cas d'un graphe de maison. Afin d'étendre cette mesure à l'analyse de la structuration des maisons, il est nécessaire de modifier cet opérateur.

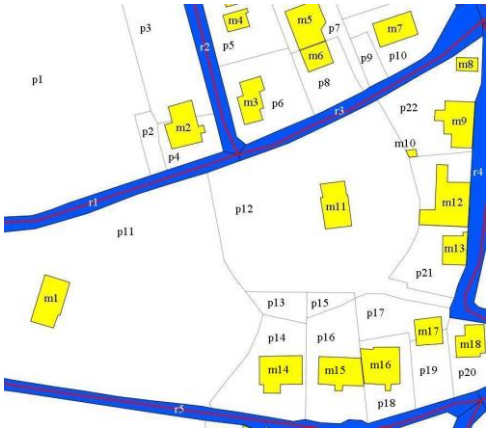


Figure 5 : un exemple d'espace E (Guisseny, Finistère Nord) avec
$M = \{m_1, m_2, \ldots, m_{17}, m_{18}\}$, $P = \{p_1, p_2, \ldots, p_{21}, p_{22}\}$, et $R = \{r_1, r_2, r_3, r_4, r_5\}$

A partir du graphe total que nous avons défini, toutes les maisons sont connectées et le graphe est connexe. Cependant les liens qui associent chaque maison sont différents selon que celles-ci sont adjacentes ou distantes. Dans ce dernier cas, le lien inclus les parcelles et les routes. Nous proposons de quantifier ces différences par une fonction $Poids(m_i, m_j)$ qui permet de mesurer l'importance du lien qu'il existe entre 2 maisons. Afin d'évaluer le potentiel de cette approche nous proposons de hiérarchiser les connexions tels que : les maisons qui se touchent ont les connexions les plus fortes et celles connectées uniquement par la route les connexions les plus faibles. Cette approche qualitative permet de renforcer les associations des maisons connectées, mais est réalisé uniquement à titre de démonstration du potentiel et des capacités d'adaptation du modèle.

Selon les cas la fonction Poids(m_i,m_j) retournera :

- Poids(m_i,m_j) = 5, si m_i et m_j sont adjacentes et sur la même parcelle : $(m_i,m_j) \in A_m \wedge \exists p_k \in P \mid (m_i,p_k) \in A_{MP} \wedge (m_j,p_k) \in A_{MP}$
- Poids(m_i,m_j) = 4 si m_i et m_j sont adjacentes mais sur des parcelles différentes : $(m_i,m_j) \in A_m \wedge \nexists p_k \in P \mid (m_i,p_k) \in A_{MP} \wedge (m_j,p_k) \in A_{MP}$
- Poids(m_i,m_j) = 3 si m_i et m_j sont sur la même parcelle mais ne sont pas adjacentes : $(m_i,m_j) \notin A_m \wedge \exists p_k \in P \mid (m_i,p_k) \in A_{MP} \wedge (m_j,p_k) \in A_{MP}$
- Poids(m_i,m_j) = 2 si m_i et m_j sont sur des parcelles adjacentes mais ne sont ni adjacentes ni sur la même parcelle : $(m_i,m_j) \notin A_m \wedge \nexists p_k \in P \mid (m_i,p_k) \in A_{MP} \wedge (m_j,p_k) \in A_{MP}$,$\wedge \exists p_k,\ p_l \in P \mid (m_i,p_k) \in A_{MP} \wedge (m_j,p_l) \in A_{MP} \wedge (p_k,p_l) \in A_P$
- Poids(m_i,m_j) = 1 si m_i et m_j sont reliés à une même rue mais ne sont ni adjacentes ni sur des parcelles adjacentes : $(m_i,m_j) \notin A_m \wedge \nexists p_k \in P \mid (m_i,p_k) \in A_{MP} \wedge (m_j,p_k) \in A_{MP}$,$\wedge \nexists p_k,\ p_l \in P \mid (m_i,p_k) \in A_{MP} \wedge (m_j,p_l) \in A_{MP} \wedge (p_k,p_l) \in A_P$, $\exists r_k \in R \mid (m_i,r_k) \in A_{MPR} \wedge (m_j,r_k) \in A_{MPR}$
- Poids(m_i,m_j) = 0 sinon.

Dans ce contexte, la mesure de centralité pour une maison correspond à la somme des poids obtenus vers toutes les autres maisons :

$$C_{MPR}(m_i) = \sum_{j \neq i} Poids(m_i, m_j)$$

Appliqué par exemple à la maison m_2 de la Figure 5, nous obtenons :

$C_{MPR}(m_2)$ = Poids(m_2,m_1)+ Poids(m_2,m_3)+ Poids(m_2,m_4)+ Poids(m_2,m_{11}) = 1+1+1+1 = 4

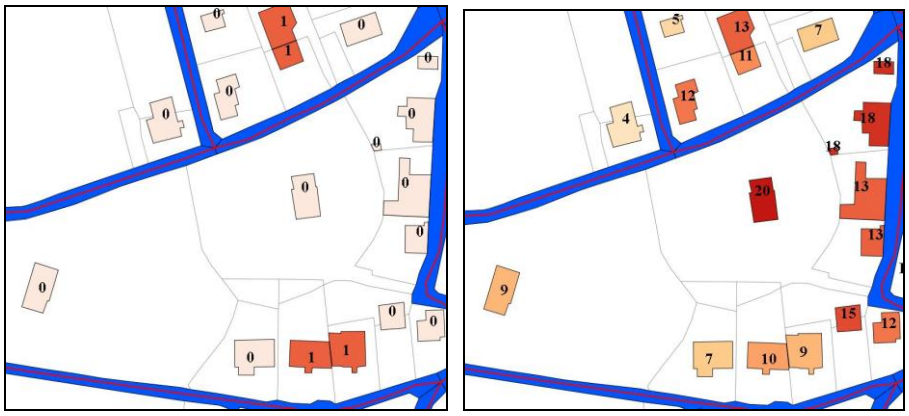


Figure 6 : Application de C_{Deg} à gauche et C_{MPR} à droite sur la Figure 5

Dans le cas du territoire de Guisseny (Figure 6), C_{Deg} met en évidence quatre maisons et montre un lien fort entre les maisons m_5, m_6 et m_{15}, m_{16} de la Figure 5.Ce

résultat donne une vision incomplète et trop localisé qui ne permet pas de refléter toute la structure à l'échelle de toute la zone. Au contraire, le modèle multi graphe aboutit à des mesures de C_{MPR} permettant d'élargir le point de vue et de mieux visualiser la structure du bâti en attribuant un poids à toutes les maisons. Il met notamment clairement en évidence la position centrale de m_{11} avec C_{MPR} (m_{11})= 20. L'intérêt de cette approche est de mettre en évidence les différents degrés d'associations entre les entités spatiales du graphe. La maison m1 possède par exemple un poids dans le graphe plus important que les maisons m2 et m14, alors que ces dernières apparaissent spatialement moins isolés. En effet la maison m1 appartient à une parcelle connectée à plusieurs routes, donc potentiellement en relation avec un nombre de maisons plus important. Cette nouvelle vision de la structuration de l'espace souligne l'importance des maisons isolées rarement souligné dans les approches conventionnelles.

5. Conclusion

Dans l'objectif d'étudier la relation entre la structure et le fonctionnement du territoire et afin d'en extraire des constats et des scénarios d'évolution, cet article propose un modèle de graphe résultant de la combinaison d'un graphe de maisons, d'un graphe de parcelles et d'un graphe de routes. Cette approche permet de générer un graphe connexe permettant d'analyser la structuration du bâti au travers de son interconnexion avec le parcellaire et les voies de circulation. Nous proposons un opérateur d'analyse de centralité C_{MPR} adapté à ce contexte. Celui-ci permet d'attribuer aux entités du bâti des poids de centralités calculées et pondérées en prenant en compte non seulement l'adjacence de ces entités, mais aussi leurs relations distantes.

Présenter A la suite de ce travail conceptuel préliminaire, une étude plus approfondie de cet opérateur devrait permettre de mettre en avant sa capacité à raisonner à plusieurs niveaux d'échelle dans l'espace. En effet il est possible de l'appliquer uniquement à un sous ensemble de maisons, par exemple l'ensemble des maisons liées à une rue, les maisons d'un ensemble de parcelles (voir même seulement à G_M). De même, il ouvre des perspectives sur l'identification d'éléments structuraux tels que l'extraction de cluster ou la mesure d'identificateurs de la compacité des territoires.

Bibliographie

Badariotti D., Banos A. et Moreno D. (2007). *Conception d'un automate cellulaire non stationnaire à base de graphe pour modéliser la structure spatiale urbaine: le modèle Remus*, Cybergeo : revue européenne de géographie / European journal of geography, Article 403, edition en ligne, http://www.cybergeo.eu/index10993.html

Buhl J., Gautrais J., Reeves N., Sole R. V., Valverde S., Kuntz P., and Theraulaz G. (2006). Topological patterns in street networks of self-organized urban settlements, *The European Physical Journal* B, 49, 513 – 522.

Cohn A.G. ,Bennett B., Gooday J., Gotts N.M. (1997). Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 1:275–316.

Del Mondo G., Rodriguez M. A., Claramunt C., Bravo L., Thibaud R. (2013). Modelling Consistency of Spatio-temporal Graphs, *Data & Knowledge Engineering, Elsevier*, 84(1), 59-80.

Egenhofer M., Franzosa R. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Science*, 5(2):161–174.

Freeman L. C. (1979). Centrality in social networks conceptual clarification. *Social networks*, 1(3), 215-239.

Hillier B., Hanson J. (1984), *The Social Logic of Space*, Cambridge University Press: Cambridge.

Jiang B., Claramunt C. (2002). Integration of space syntax within GIS: new perspectives for urban morphology, *Transaction in GIS*, 6(3), 277-293.

Ratti C. (2004), Space syntax: some inconsistencies, *Environment and Planning B: Planning and Design*, 31(4), 501–511.

Sriti M., Thibaud R., Claramunt C. (2005). A fuzzy identity-based temporal GIS for the analysis of geomorphometry changes, *Journal of Data Semantics, Springer Verlag, Special issue on spatial data semantics*, 3(1), 81-99.

Une méthode de construction de données spatio-temporelles appliquée aux rues anciennes de Paris.

Bertrand Dumenieu[1], Julien Perret[1], Anne Ruas[3]

1. Laboratoire COGIT, IGN
73avenue de Paris, 94165 Saint-Mandé France
bertrand.dumenieu@ign.fr, julien.perret@ign.fr

3. Institut français des sciences et technologies des transports, de l'aménagement et des réseaux (IFSTTAR).
Ministère de l'Ecologie, du Développement Durable, des Transports et du Logement
IFSTTAR, 58, boulevard Lefebvre, 75732 Paris Cedex 15 - France
anne.ruas@ifsttar.fr

RESUME. Cet article présente une méthode semi-automatique permettant de construire une structure de graphe spatio-temporelle à partir de données vecteur extraites de cartes historiques sur la ville de Paris. La méthode proposée ici est basée sur une méthode métaheuristique par recuit simulé permettant de construire des graphes spatio-temporels entre plusieurs sources de données. La méthode est appliquée ici aux rues de la ville. Si le processus est générique, une spécialisation en est proposée ici par l'ajout d'un niveau hiérarchique permettant de construire des groupes de rues liées dans le temps et l'espace.

ABSTRACT. This paper proposes a semi-automatic algorithm that allows building spatio-temporal graphs from historical vector data extracted from early cartographic sources.

The algorithm is applied to the street networks of the city of Paris in the late XVIIIth and XIXth centuries. Although generic, the algorithm is here extended to efficiently deal with the streets by adding a hierarchical level that build spatio-temporal clusters.

MOTS-CLES : appariement, SIG historique, graphes spatio-temporel

KEYWORDS: feature-matching, historical GIS, spatio-temporal graphs

1. Introduction

De plus en plus de chercheurs en sciences humaines et sociales s'appuient sur les SIG afin d'appuyer leur propos, visualiser ou partager leurs données. L'Histoire, notamment sociale, demande d'interroger les espaces dans lesquels prennent place les activités humaines. Dès lors les outils de la géomatique prennent une place importante au sein de l'étude historique en cela qu'ils permettent de situer les objets d'étude. Appropriés par l'historien, ils deviennent donc un support à l'étude historique.

Nous nous centrons dans cette étude essentiellement sur les XVIIIe et XIXe siècles Paris ayant subi ses principales transformations sur cette période. Bien que les sources soient pour ces deux siècles nombreux et de qualité, leurs particularités empêchent leur intégration au sein d'un SIG classique. En effet, les dynamiques à l'œuvre dans la ville ancienne sont au cœur de la plupart des études sur la ville .Paris ne doit plus être considérée dans sa dimension spatiale seulement mais également temporelle. Il s'agit dès lors de stocker et représenter au sein d'un SIG non plus un instantané de l'espace de Paris, mais une structure spatio-temporelle.

Nous proposons dans cet article une méthodologie d'intégration de données historiques fondée sur une méthode semi-automatique d'appariement spatio-temporel en nous concentrant sur les rues de Paris. La partie 2 s'attache à présenter sources cartographiques et la méthodologie utilisée pour en extraire des objets spatiaux et temporels. La partie 3 expose la structure de données spatio-temporelle, puis le processus semi-automatique est exposé en partie 4. Enfin, la méthode est appliquée aux rues de Paris dans la partie 5.

2. Des sources cartographiques aux observations historiques

2.1. Géoréferencement des sources

De très nombreuses cartes et plans décrivent l'espace ancien de Paris (Pinon, 2004) aux XVIIIe et XIXe siècles. La plupart de ces sources cartographiques sont d'une qualité suffisante pour être géoréferencées et intégrées au sein d'un SIG. Nous nous intéressons ici plus précisément à la période allant de la Révolution à la fin du XIXe siècle, riche en transformations et en plans.

Nous avons sélectionné quelques sources de référence réparties sur toute le siècle et contenant les rues nommées de Paris. Ces sources ont été également choisies car elles représentent la ville de Paris dans sa globalité à l'échelle la plus grande possible.

Ainsi, nous avons sélectionné les trois sources suivantes, dont des extraits sont présentés en figure 1:

– l'atlas de Verniquet, levé entre 1785 et 1795 et gravé entre 1793 et 1799;

– le cadastre Vasserot, levé entre 1808 et 1836;

– le plan de Poubelle, levé entre 1885 et 1888;

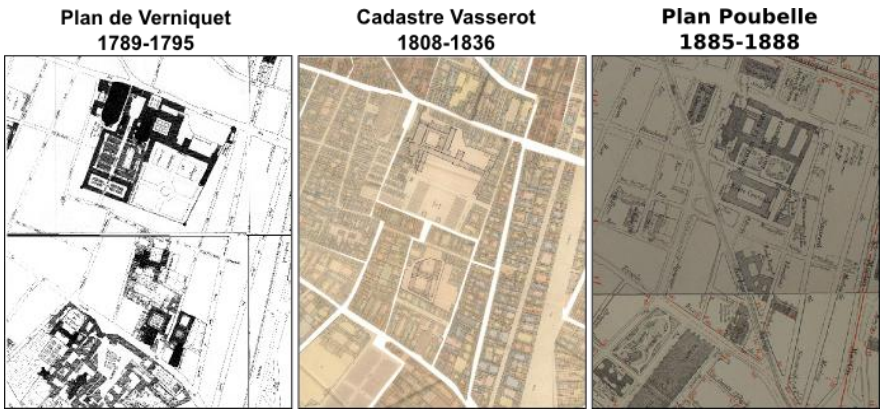


Figure 1. Les sources cartographiques utilisées

Les plans de Verniquet et Poubelle ont ensuite été géoréferencés de façon à minimiser les déformations au sein des plans, et d'assurer une cohérence maximale entre les planches composants chacun des plans (72 pour Verniquet, 52 pour Poubelle). Le plan de Vasserot quant à lui est directement issu des données du projet ALPAGE (Noizet et Grosso, 2011).

2.2. Extraction des observations historiques

Les sources cartographiques sont ensuite vectorisées manuellement de façon à en extraire des objets géographiques vecteurs intéressants pour l'étude. Chaque objet peut être enrichi par des données attributaires. Les objets vectorisés, tout comme les sources à partir desquelles ils sont saisis ne sont pas seulement spatiaux, mais également temporels. Il est alors nécessaire non seulement de considérer cette temporalité, mais également de la stocker, sans quoi aucune logique d'évolution ne peut être retrouvée.

Nous considérons que la temporalité d'une source cartographique correspond à sa période de levé, un intervalle de temps durant lequel on considère que l'espace réel est représenté. Bien sur, cette période est rarement connue avec précision, les plans principaux faisant l'objet de plusieurs levés successifs dont on ne connait pas nécessairement exactement la répartition spatiale ou la période d'exécution. Ne pas prendre en considération cette incertitude pourrait amener à un graphe faux. Ce risque est d'autant plus présent lorsque plusieurs plans se chevauchent temporellement, ou sont inclus les uns dans les autres.

Afin d'exprimer cette incertitude, nous généralisons les intervalles de temps des sources par l'utilisation d'intervalles flous. N'ayant généralement pas

d'informations temporelles autres que des dates, l'utilisation d'intervalles flous trapézoïdaux permet d'exprimer de façon satisfaisante les temporalités des sources. Enfin, nous considérons que nous n'avons au moment de l'extraction pas d'autre informations que celle fournie par la source. Ainsi, la temporalité des objets extraits d'une source correspond à la temporalité de la source.

Nous permettons cependant de définir une temporalité plus précise mais toujours bornée par celle de leur source. Ce choix est motivé par le fait que l'on considère les objets vectorisés non pas comme une représentation d'une entité du monde réel qui serait directement retranscrite dans une carte, mais comme une simple observation d'une entité faite pendant le tracé de la source.

Chaque objet, défini sur les dimensions spatiale, temporelle et sémantique (Rodier et Saligny, 2010), que nous nommons « observation historique» est défini par le triplet $(\{\alpha 1, \dots, \alpha i\}, \omega, t)$ avec :

– $\{\alpha 1, \dots, \alpha i\}$: les données attributaires de l'observation ;

– ω : la géométrie de l'observation de type point, ligne, surface ;

– t : la temporalité floue de l'observation ;

3. Reconstruction de données spatio-temporelles

3.1. Une structure de graphe spatio-temporel

L'extraction effectuée depuis l'ensemble des sources cartographiques nous fournit un ensemble d'observations historiques E (voir figure 3).

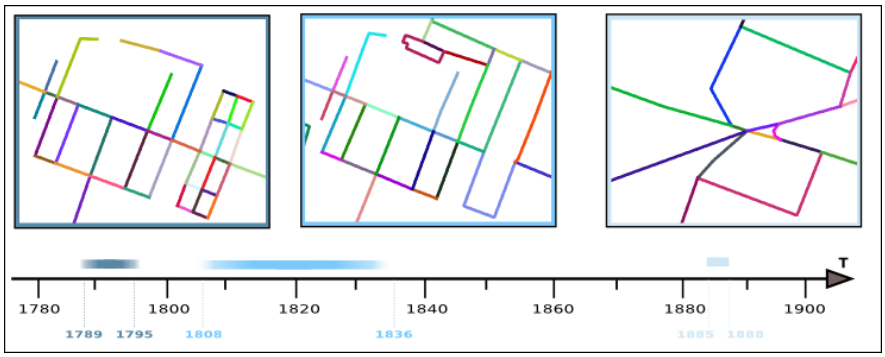


Figure 3. Observations historiques : l'ensemble des tronçons de rue autour de Saint-Martin des Champs

Ces observations ne permettent pas pour autant de retranscrire les dynamiques de l'espace ancien : il ne s'agit toujours que de traces éparses d'entités évoluant dans le temps. Cependant, si l'on considère un ensemble d'observations d'une même entité,

l'œil retrouve naturellement la suite logique des évolutions. Ainsi, il crée les relations de filiation qui apparaissent entre les observations historiques. La figure 4 montre un chemin possible entre des observations de tronçons de rue. Bien sur, le tracé de ces chemins dépend de l'expertise de l'observateur : il s'agit d'un processus subjectif.

Il devient alors particulièrement intéressant de reproduire ce processus informatiquement. En effet, si les observations historiques ne sont pas des objets spatio-temporels, ces chemins temporels permettent de représenter formellement les logiques d'évolution de l'espace.

On peut exprimer les chemins temporels entre les observations historiques sous la forme d'un graphe $G(V,E)$, dont chaque arête $e = (o1, o2) \in E$ est un lien dans le temps entre deux observations $o1, o2 \in V$. Les liens et le graphe sont nommés respectivement *liens de filiation* et *graphe de filiation*. Nous considérons ici un temps linéaire, faisant de G un graphe orienté.

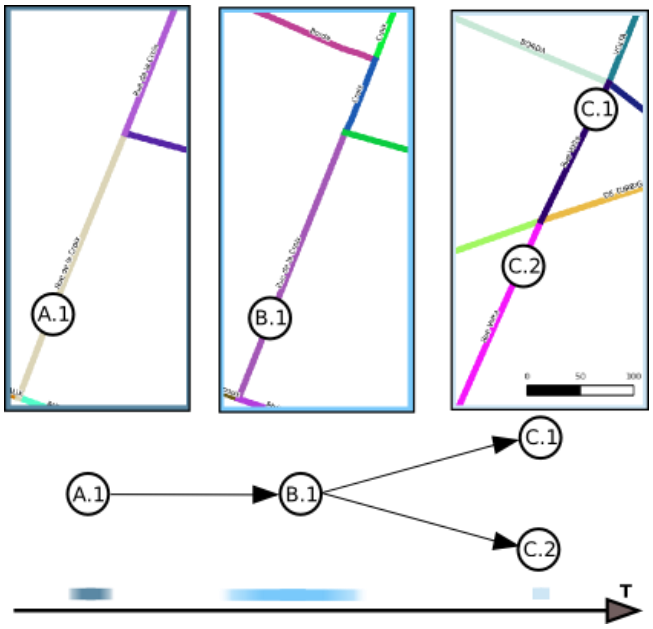


Figure 4. Un exemple de graphe de filiations

3.2. Ajout de la sémantique des transformations

Il est possible de créer sur G des liens de filiation plus complexes qu'une filiation simple. Par exemple, la figure 4 illustre le cas d'un tronçon se séparant en deux objets plus tardivement. Ces configurations, que l'on retrouve dans la majorité

des méthodes d'appariement de données (Tong, 2010) prennent ici une signification particulière puisqu'elles décrivent non plus des correspondances simples mais les structures de transformations de l'espace. Il devient alors important d'exprimer formellement ces configurations, utiles tant pour le processus d'appariement que pour ajouter du sens aux liens tracés.

L'enrichissement du graphe avec la sémantique des transformations de l'espace rejoint les très nombreux travaux portant sur la représentation et la modélisation des changements de l'espace. La question de la représentation des dynamiques de l'espace au sein d'un SIG pose question depuis longtemps (Langran et Chrisman, 1988 ; Couclelis, 1992 ; Peuquet 2001), aboutissant à la proposition de nombreux modèles de bases de données spatio-temporelles (Pelekis, 2005, Silva, 2010).

Les travaux portant sur la représentation du changement (Hornsby et Egenhofer, 2000 ; Stell et Worboys, 2008) ont participé à l'intérêt plus récent porté à la théorie des graphes. Axé vers la représentation visuelle des transformations, les travaux de Renolen (2000) puis Stefani (2010) visent ainsi à former des graphes illustrant l'évolution d'objets spatiaux. Plusieurs propositions visant à modéliser des phénomènes spatio-temporels à l'aide de graphes ont également été proposés. Ainsi, Spéry (2004) propose un modèle de base de données temporel pour étudier l'évolution de parcelles cadastrales. Plus récemment, Del Mondo (2010) propose un modèle de graphe spatio-temporel permettant d'exprimer les relations de spatiales, spatio-temporelles et temporelles d'équivalence et de filiation entre des objets issus d'ensembles d'objets à des temps différents. Le modèle est étendu (Stell et al., 2011) par l'utilisation de bigraphes permettant d'exprimer les relations spatio-temporelles par des règles de réécriture.

Nous nous limitons ici à l'expression des transformations de l'espace les plus courantes, illustrés dans la figure 5 ci-dessous. De cette façon, nous conservons une structure relativement simple pouvant être utilisée dans un processus d'appariement, tout en représentant les transformations de base.

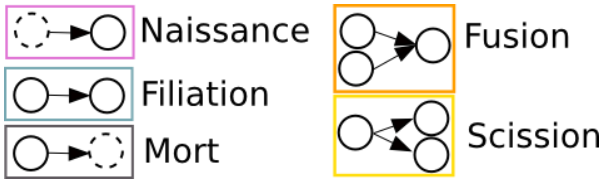


Figure 5. Les 5 schémas sélectionnés

Les schémas de mort et de naissance d'une observation historique correspondent respectivement au début et à la fin d'un chemin temporel. Les schémas de fusion et de scission permettent d'exprimer les remembrements de l'espace. Le schéma de filiation correspond quant à lui à un lien de filiation simple.

Ainsi exprimés sur notre graphe, ces schémas de filiation ne correspondent toutefois pas nécessairement à des transformations. En effet, les schémas de scission, fusion et filiation peuvent également exprimer une relation d'équivalence entre les observations.

Il est possible d'exprimer ces schémas sur le graphe G de façon formelle. Nous définissons $H = \{h1, h2, \dots hn\}$ une partition du graphe G telle que chaque partition h corresponde à un schéma, nommé structure de filiation. La figure 6 illustre ce partitionnement. Nous pouvons désormais tracer un graphe de filiation G muni d'une partition H décrivant les évolutions d'un ensemble d'observations historiques.

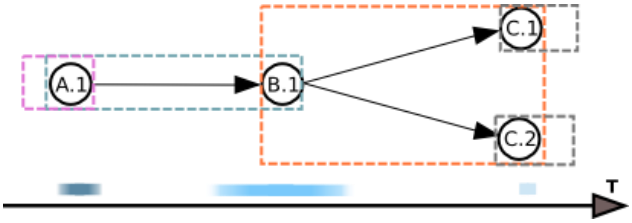


Figure 6. Partition du graphe de filiations

3.2. Gestion des lacunes des sources historiques

En l'état, le graphe G ne permet pas de considérer les imperfections des sources historiques. Cependant, il est possible de prendre en compte des données lacunaires. Renolen identifie un schéma de réincarnation correspondant à un objet disparaissant, puis réapparaissant plus tardivement. Nous pouvons utiliser ce schéma pour modéliser les lacunes des données.

Par exemple dans la figure 7, l'observation C_1 semble être une réincarnation de l'observation A_1. En effet, l'entité dont A_1 et C_1 sont les traces est décrite dans les sources A et C, mais la source B ne possède aucune observation de cette entité. Cette absence peut être due à deux phénomènes : il s'agit soit d'une véritable réincarnation au sens de Renolen, soit d'une lacune de la source B. Le lien de filiation $e = (A1, C1)$ pourra donc désigner soit une lacune, soit une réincarnation.

8

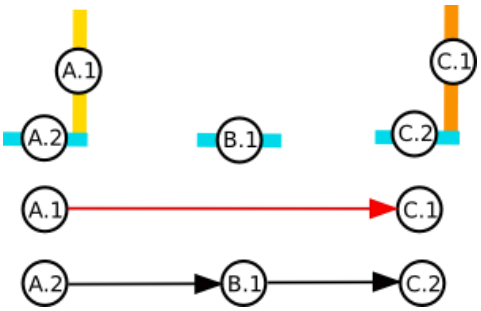


Figure 7. Un exemple de donnée lacunaire

Toutefois, la détermination de la sémantique exacte de ce lien ne pouvant être réalisée à partir des sources seules, elle sera confiée à postériori à un expert historien.

3.3. Vraisemblance d'un graphe

Le problème revient à construire le graphe le plus vraisemblable G*, se rapprochant le plus possible de la réalité historique. La vraisemblance d'un graphe de filiation est une notion complexe, dépendant fortement du point de vue de celui qui construit ce graphe. Pourtant, il est nécessaire de pouvoir mesurer cette vraisemblance afin de pouvoir nous assurer de construire G*. Cette vraisemblance n'est pas une mesure unique, mais l'agrégation de différentes estimations portant sur l'évolution des formes géométriques, la cohérence temporelle et la logique d'évolution des différents attributs.

Afin de quantifier ces jugements, nous considérons un ensemble de critères de vraisemblance spécialisés sur une des trois dimensions des observations historiques : l'espace, la sémantique et le temps. Ces critères correspondent à des distances, métriques ou similarités et permettent de prendre en considérations les incertitudes liées aux données (imprécisions géométriques, lacunes, incertitudes temporelles) (Dumenieu, 2013). De manière à utiliser la sémantique ajoutée par les structures de filiation, les critères sont exprimés sur les partitions de G.

4. Une méthode de reconstruction semi-automatique

4.1. Un problème d'optimisation multicritère

Construire le graphe G* manuellement ne pose pas de difficultés lorsque l'on dispose de peu d'observations, et que les imperfections des données restent faibles. Mais, dès lors que l'espace considéré et le nombre de sources historique grandissent, l'approche manuelle n'est plus envisageable. Les incertitudes des données,

notamment temporelles, et les conflits entre sources rendent la construction plus ardue. Enfin, l'expertise d'historiens différents peut aboutir à des graphes différents, obligeant à répéter la construction. Dès lors, il est intéressant d'automatiser la création du graphe G*, sa vraisemblance étant calculable par différents critères. La prise en compte de l'expertise d'un historien va alors se traduire par le choix des critères de vraisemblance.

La construction de G* peut être ramenée à un problème d'appariement de données géographiques. Cependant, à la différence des approches classiques d'appariement (Mustière et Devogele, 2008 ; Safra *et al.*, 2013), la nature multisource et temporellement incertaines des données considérées oblige à effectuer un appariement entre plus de deux sources de données. G* est construit à partir d'un graphe vide G, les nœuds étant les observations historiques considérées. Il est alors possible de le construire itérativement par la répétition de deux actions :

– ajouter une arête dans G ;

– supprimer une arête dans G ;

Appliquer l'une de ces opérations crée un nouveau graphe dont on peut évaluer la vraisemblance.

Le processus de construction apparaît comme un problème d'optimisation combinatoire où l'objectif est de maximiser la vraisemblance d'un graphe de filiation pour un ensemble d'observations historiques, l'optimum étant G*. La notion de vraisemblance nous oblige à considérer une approche multicritères. Il ne nous est cependant pas possible d'envisager l'utilisation d'approches adaptées à l'appariement de données géographiques comme celle proposée par Olteanu (Olteanu *et al.*, 2010) en raison de la taille de l'espace de recherche considéré.

Pour résoudre un tel problème, une première méthode serait d'effectuer l'énumération complète des solutions afin de choisir celle ayant la plus grande vraisemblance. Bien sur, le coût de cette énumération est prohibitif, d'autant plus que la plupart des graphes possibles sont peu pertinents. Une solution est alors d'explorer l'espace de recherche tout en convergent vers des parties intéressantes de cet espace, de façon à trouver la solution optimale sans énumérer l'ensemble des solutions. De nombreuses stratégies, nommées métaheuristiques, s'adressent à ce problème. Ces méthodes permettent de parcourir l'espace de recherche d'un problème tout en tendant vers l'optimum global, ou, dans le cas de méthode multicritères, vers un ensemble de solutions dont aucune n'est meilleure que les autres. Cet ensemble est nommé « front de pareto » du problème. De nombreuses méthodes métaheuristiques dédiées à l'optimisation ont été proposées dans la littérature (Fouskakis et Draper , 2002), majoritairement suivant un modèle évolutionnaire (Jaimes *et al.*, 2007) ou un modèle de recuit simulé (Suman et Kumar, 2004). Nous nous intéressons plus particulièrement ici aux modèles de recuit simulé, s'inspirant du principe physique du recuit où un matériau est successivement réchauffé et refroidi afin d'obtenir la meilleure structuration atomique possible.

Plusieurs comparaisons entre les deux approches (Manikas *et al.*, 2000) ont montré que le recuit simulé a la particularité d'obtenir rapidement des solutions satisfaisantes, les algorithmes génétiques, plus lents, tendent vers de meilleures solutions. Le recuit simulé est également une méthode d'optimisation locale, qui nous parait particulièrement adaptée dans notre cas, où une partie de l'espace de recherche peut être rapidement exclue. En effet, de nombreux liens de filiation potentiels entre deux observations peuvent être immédiatement disqualifiés (par exemple, deux rues trop éloignées). Parmi les différents modèles de recuits simulés proposés dans la littérature, nous avons choisi l'approche de Suppapitnarm (Suppapitnarm *et al.*, 2000), présentant l'avantage de ne pas agréger les différents critères au sein du processus en définissant une température par objectif.

Cette méthode est fondée sur les relations de dominances entre les solutions, de façon à approximer le front de pareto du problème. Une solution a domine une solution b si a est au moins aussi bonne que b pour tous les critères et meilleure pour au moins un.

4.2. Algorithme de recuit simulé

Rappelons que nous cherchons ici à obtenir le graphe de filiation le plus vraisemblable, cette vraisemblance dépendant de différents critères.

Considérons un graphe G à optimiser à l'aide de m critères de vraisemblance. On a alors :

$C_G = (c_1, c_2, \ldots, c_m)$ Le vecteur de valeurs des critères exprimés sur G.

$C_G = (c_1, c_2, \ldots, c_m)$ Le vecteur de températures associées aux m critères.

A chaque étape du recuit simulé, un nouveau graphe est généré en ajoutant ou supprimant un arc du graphe courant. Le vecteur de valeurs du nouveau graphe est calculé et comparé avec les solutions du front de pareto. Si cette nouvelle solution n'est dominée par aucune solution du front de pareto alors elle y est ajoutée.La solution est retenue comme état courant du recuit simulé comme ceci :

– Si la solution est dans le front de pareto, elle devient l'état courant.

– Sinon, la solution est acceptée selon la probabilité suivante, définie par Suppapitnarm :

$$p = min(1, \prod_{i=1}^{|C|} e^{-\Delta s_i / T_i}) \quad (1)$$

De façon à parcourir le mieux possible l'espace de recherche, l'algorithme effectue toutes les N_1 itérations un saut, en sélectionnant aléatoirement une solution du front de pareto et en l'affectant comme état courant.

Suivant le modèle de Supappnitarm, les températures sont d'abord initialisées à une grande valeur. Après une étape de burn-in effectuée pendant N_2 itérations du recuit, chaque température est redéfinie par l'écart type des valeurs du critère auquel elle correspond (Salamon, 2002). Toutes les N_3 itérations une règle de décroissance géométrique des températures est appliquée:

$\forall t \in T$, i l'itération courante, $t_{i+1} = \alpha * t_i$ (2)

Nous avons choisi empiriquement de poser $\alpha = 0.975$, cette valeur déjà utilisée par Suman nous permettant d'obtenir les meilleurs résultats. Le recuit simulé est enfin stoppé au bout de N_3 itérations ou s'il n'y a plus d'améliorations des solutions.

Les méthodes multicritères ne fournissent pas une solution unique. Autrement dit, la méthode de recuit simulé ne donne généralement pas un unique graphe G*, mais l'ensemble des solutions du front de pareto. Il est donc nécessaire de choisir par la suite une solution parmi l'ensemble proposé. Ici, la solution choisie dépend fortement de l'avis de l'historien. Pour cette raison, nous avons pour l'instant privilégié un choix manuel effectué par un expert. Une réflexion sur les méthodes de décision multicritères reste toutefois à mener.

4. Application aux rues de Paris

4.1 Ajout d'un niveau hiérarchique

Nous avons utilisé jusqu'ici les tronçons de rue comme observations historiques. Cependant, cette modélisation ne permet pas de considérer les cas de transformation n-m (ou réallocations), le schéma n'étant pas représenté dans notre graphe. Considérer non plus uniquement les tronçons suivant un modèle « spaghetti » (Rigaux *et al.* 2002) mais aussi les rues permet d'ajouter d'enrichir notre modèle.

Travailler avec les rues pose nécessairement la question de l'identité des objets considérés (Horsnby et Egenhofer, 2000). Les rues étant des objets principalement spatiaux, nous avons choisi de les former uniquement d'après leurs géométries. Pour ce faire, nous nous basons sur les travaux de Jiang (2011) pour construire les composantes du réseau gardant une certaine linéarité. L'auteur propose pour construire ces composantes ou « natural roads » 3 méthodes, le calcul de la linéarité d'un rue variant de l'une à l'autre. Les composantes sont construites itérativement, un nouveau tronçon étant ajouté selon une des stratégies suivantes :

– « self fit » : un tronçon connexe est ajouté arbitrairement s'il respecte une contrainte d'angle ;

– « self best fit » : le tronçon connexe d'angle de déviation le plus faible est ajouté ;

– « every best fit » : le meilleur tronçon connexe est ajouté s'il n'appartient à aucune autre meilleure composante ;

Nous nous sommes basés pour les rues de Paris sur l'approche « every best fit », plus proche vision classique d'une rue.

Nous construisons ensuite l'ensemble des composantes pour chaque réseau de rues considéré comme illustré par la figure 8 où chaque couleur correspond à une composante. Nous considérons dès lors chaque composante comme un cluster de tronçons pouvant former une transformation n-m. De cette façon, nous ajoutons un niveau hiérarchique aux observations historiques.

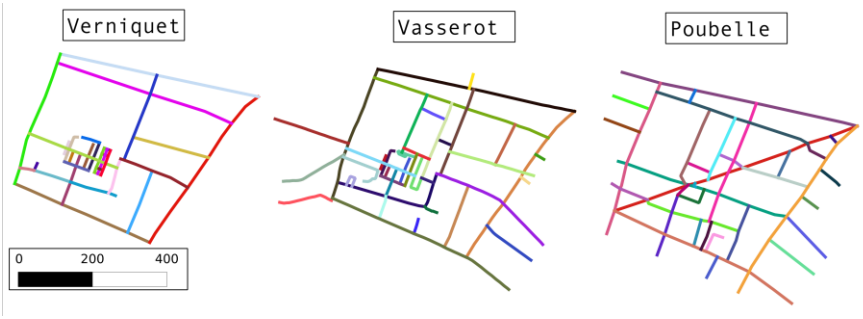


Figure 8. Natural roads construites sur l'espace de Saint-Martin des Champs

Afin d'intégrer ce niveau hiérarchique au sein du processus d'appariement, il est nécessaire d'identifier quels seront les composantes de chaque réseau dont les tronçons pourront être appariés. Si l'on considère des réseaux peu décalés et peu déformés, on peut considérer que des composantes proches géométriquement ont de fortes chances de contenir des tronçons qui seront mis en relation.

Nous proposons ici d'utiliser une classification hiérarchique ascendante (Sibson, 1973) afin de former des clusters spatio-temporels regroupant des composantes potentiellement en relation dans le temps.

Puisque nous cherchons à regrouper les composantes proches géométriquement, nous avons utilisé la distance de Fréchet partielle correspondant à une distance entre courbes comme indice d'agrégation de la classification. Trois dissimilarités interclasse ont alors été évaluées : le saut maximum, minimum et le lien moyen, le choix dépendant du type de classification voulue. Dans notre cas, nous utilisons principalement le critère de saut minimum, nous permettant de former des clusters même dans le cas où les sources se trouvent graduellement décalées.

La classification hiérarchique ascendante construit un dendrogramme dans lequel il est nécessaire d'effectuer une coupe pour obtenir un nombre de classes voulues. Dans notre cas, nous ne connaissons pas par avance le nombre de groupes de composantes à construire. Afin de palier ce problème, nous effectuons une coupe selon une valeur de dissimilarité interclasse minimum. La figure 9 illustre des coupes effectuées pour plusieurs valeurs de β, pour le critère de saut minimum.

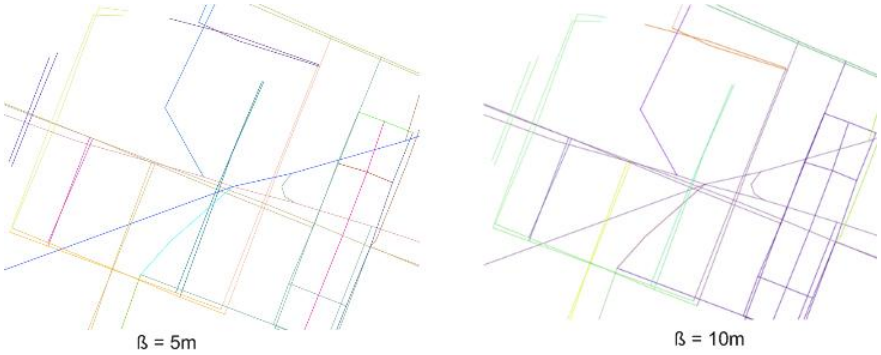


Figure 9. Groupes de tronçons, toutes les sources sont superposées.

A partir des groupes de composantes ainsi créés, nous formons les groupes d'observations historiques correspondants. Le processus d'appariement est alors effectué sur chaque groupe d'observations. Ce processus présente plusieurs avantages. Tout d'abord, il s'agit de coupes locales dans l'espace de recherche, ayant pour effet d'accélérer la convergence du recuit simulé. De plus, chaque groupe étant indépendant, nous pouvons envisager une parallélisation du processus. Enfin, il est possible de relancer un appariement non satisfaisant de façon très locale sans reconsidérer le réseau entier.

4.2 Résultats d'appariement

Afin d'appliquer la méthode de construction du graphe d'évolutions au réseau des rues, un critère a été choisi pour chaque dimension :

– **Spatiale** : distance de Fréchet partielle (Devogele, 2005) ;

– **Sémantique**: double métaphone sur les noms des rues (Phillips, 2000) ;

– **Temporelle** : distance floue proposée par Cheng (1998) ;

La méthode est appliquée au quartier de Saint-Martin des Champs, aujourd'hui le CNAM, à partir des trois sources considérées jusqu'ici. L'annexe 1 illustre ainsi la meilleure solution, choisie par l'expert parmi les solutions du front de pareto.

14

Chaque couleur et numéro correspond à un ensemble de tronçons de rue ayant des transformations communes. Les paramètres suivants ont été utilisés :

$$\beta = 5,\ N_1 = 10000,\ N_2 = 0.15 * N_1,\ N_3 = 0.1 * N_1$$

Cet espace est bien connu des historiens, ce qui permet de comparer nos résultats avec une vérité terrain donnée par les documents historiques existants. Ainsi, les mesures de précision, rappel et fscore sont exposées dans le tableau 1 pour différentes valeurs des paramètres.

Tableau 1. Précision, rappel et fscore selon plusieurs valeurs de paramètres

(β, N_1, N_3)	Précision	Rappel	Fscore
(5, 5000, 0.1* N_1)	0.75	0.73	0.74
(5, 10000, 0.1* N_1)	0.76	0.86	0.80
(5, 5000, 0.5* N_1)	0.78	0.66	0.71
(5, 10000, 0.5* N_1)	0.75	0.70	0.72
(10, 5000, 0.1* N_1)	0.78	0.82	0.80
(10, 10000, 0.1* N_1)	**0.81**	**0.93**	**0.86**
(10, 5000, 0.5* N_1)	0.75	0.69	0.72
(10, 10000, 0.5* N_1)	0.76	0.84	0.80

D'après le tableau 1, le meilleur résultat (93% d'appariement sur les 3 sources) est obtenu avec les paramètres β =10, N_1 = 10000, N_3 = 0.1* N_1. Plus la valeur de β est faible, plus il y a de risque de sur-découper l'espace de recherche et d'élaguer des liens de filiations vraisemblables. En effet, le processus étant séquentiel, les liens élagués par la classification par rues ne pourront jamais être testés lors de l'exploration de l'espace de recherche. Etant donné la sensibilité de ce critère, il est préférable de ne fixer une valeur trop basse, bien que cela puisse ensuite allonger le temps de convergence du recuit simulé. Il nous paraît ici approprié pour de futur travaux de corréler la valeur de avec le décalage géométrique des sources historiques.

Les valeurs de N_1 et N_3 tendent à montrer qu'il est possible de plonger rapidement vers la zone contenant les meilleures solutions par un refroidissement rapide (N_3 étant faible). Cependant, la convergence au sein de cette zone est plus longue, nécessitant d'effectuer un grand nombre d'itérations (N_1). N_1 et N_3 sont des paramètres propres au recuit simulé et, de par la nature de l'approche, ne peuvent être fixés qu'empiriquement. Afin de fixer ces paramètres pour le problème des rues de Paris, nous envisageons de mettre en place une méthode d'apprentissage

supervisée permettant à des experts historiens de fixer ces paramètres en évaluant des résultats donnés par le processus.

Les écarts à la vérité terrain s'expliquent principalement par le fait que les critères choisis ne permettent pas de décrire parfaitement le processus d'évolution des rues. En effet, la seule distance géométrique et la proximité des noms (le critère de temps ayant surtout un effet sur la création de sauts temporels) ne sont pas totalement satisfaisants. Ainsi, la distance de Fréchet partielle est sensible aux cas où deux tronçons se recouvrent partiellement. Le double métaphone quant à lui permet de bien évaluer la similarité entre deux noms de rue, même si l'orthographe a changé, mais évalue mal les changements de noms. Afin d'enrichir ce dernier, un critère utilisant les dictionnaires de noms de rues a été implémenté et doit être évalué.

5. Conclusion

Dans cet article nous avons présenté une méthode de construction semi-automatique de graphes spatiotemporels à partir de données vecteur temporalisés extraites de sources cartographiques anciennes. Afin d'appliquer ce processus aux rues de Paris anciennes, nous avons proposé d'ajouter un niveau hiérarchique basé sur les « natural roads » proposées par Jiang (Jiang et al., 2011). Si l'article expose ici une application sur les rues de Paris, la méthode proposée est générique. En effet, seuls les critères de vraisemblance dépendent du type de données considérées. Ainsi, les travaux futurs porteront sur l'application de cette méthode à d'autres types de données historiques spatiales : parcelles, adresses, etc. La méthode spécifique aux réseaux de rue doit faire cependant encore l'objet d'analyses comparatives avec d'autres méthodes d'appariement de réseaux, notamment celles traitant de données historiques (par exemple celle de Costes (Costes *et al.*, 2012)).

Si la structure de graphe historique et la méthode d'appariement spatio-temporel présentée dans cet article visent à proposer aux historiens méthode d'intégration de données au sein d'un SIG historique, les graphes ainsi créés peuvent eux-mêmes être sujet d'études. En effet, le graphe issu du processus d'appariement porte de nombreuses informations portant sur les transformations de l'espace, constituant autant de données spatio-temporelles. Parmi les possibilités offertes par de telles données, l'analyse des rythmes, des densités de transformation et de leur répartition permet d'envisager l'utilisation de ces graphes pour extraire automatiquement des évènements (Horsnby et Cole, 2007). De futurs travaux s'intéresseront également à la possibilité d'intégrer les graphes ainsi générés dans des modèles de données spatio-temporels existants.

Bibliographie

Cheng C. (1998), *A* new approach for ranking fuzzy numbers by distance method, *Fuzzy Sets and Systems*, vol. 95, n°3, mai 1998, pp 307-317.

Costes B., Grosso E., Plumejeaud C. (2012). Géoréferencement et appariement de données issues des cartes de Cassini : Intégration dans un référentiel topographique actuel, *Colloque International de Géomatique et d'Analyse Spatiale SAGEO'12*, novembre 2012, Liège, Belgique.

Couclelis, H. (1992). People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS. *In Theories and methods of spatio-temporal reasoning in geographic space*, 65-77. Berlin: Springer Verlag

Del Mondo G., Stell J., Claramunt C., Thibaud R. (2010): A Graph Model for Spatio-temporal Evolution. *J. UCS 16*(11): 1452-1477

Devogele T. (2002), *New Merging Process for Data Integration. Based on the Discrete* Fréchet Distance, *Advances in Spatial Data*, Springer, pp 167-181.

Dumenieu B. (2013), Automatic reconstruction of spatio-temporal data from historical maps, *Worshop on integrating 4D, GIS and cultural heritage*, Leuven, Belgique.

Fouskakis D., Draper D. (2002), Stochastic optimization : a review, *International Statistical review*, vol. 70, N° 3, p315-349.

Hornsby K., Egenhofer M. J. (2000), Identity-Based Change: A Foundation for Spatio-Temporal Knowledge Representation, *International Journal of Geographical Information Science*, vol. 14, 2000.

Hornsby K., Cole S. (2007): Modeling Moving Geospatial Objects from an Event-based Perspective. *T. GIS 11*(4): 555-573

Jaimes A. L., Coello C. A. (2002), Multi-Objective Evolutionary Algorithms : a review of the state of the art and some of their applications in chemical engineering, *World Scientific*, p 61-90.

Jiang B., Zhao S., Yin J. (2008), Self-organized natural roads for predicting traffic flow: a sensitivity study, *Journal of Statistical Mechanics: Theory and Experiment*, July, 2008.

Langran G., Chrisman N. R. (1988). A Framework For Temporal Geographic Information. *Cartographica The International Journal for Geographic Information and Geovisualization* 25(3):1-14.

Manikas T. W., Cain J. T. (1996), Genetic Algorithms vs. Simulated Annealing: A Comparison of Approaches for Solving the Circuit Partitioning Problem, *Computer Science and Engineering Research.*

Mustière S., Devogele T., 2008, Matching networks with different levels of detail, *GeoInformatica*, Vol.12 n°4, pp 435-453.

Noizet H., Grosso E., "The ALPAGE project: *Paris and its suburban area at the intersection of history and geography (9th-19th century)"*, in digital proceedings of the 25th International Cartographic Conference (ICC'11), 3-8 July 2011, Paris, France.

Olteanu-Raimond A.-M., Mustière S., Ruas A., 2009, Fusion des connaissances pour apparier des données géographiques, *Revue Internationale de Géomatique* vol.19 n°3, septembre 2009, pp.321-349

Peuquet D. J. (2001). Making Space for Time: Issues in Space-Time Data Representation. *Geoinformatica* 5:11-32.

Philips L. (2000), The Double Metaphone Search Algorithm, *C/C++ Users Journal*, juin 2000.

Pinon P., Le Boudec B., Carré D. (2004), Les plans de Paris : Histoire d'une capitale, *Le Passage.*

Renolen A. (2000). History History graphs: Conceptual modelling of spatiotemporal data, *Transaction in GIS*, vol. 4, n°1, p 23-42.

Rigaux P., Scholl M., Voisard A. (2002). *Spatial databases : with application to GIS*, Morgan Kaufmann.

Rodier X., Saligny L. (2010), *Modélisation des objets historiques selon la fonction, l'espace et le temps pour l'étude des dynamiques urbaines dans la longue durée*, Cybergéo, http://cybergeo.revues.org/23175

Safra E., Kanza Y., Sagiv Y., Doytsher Y., Ad hoc matching of vectorial road networks**,** *International Journal of Geographical Information Science*, vol. 27, n° 1, 1 January 2013 , pp. 114-153

Salamon P., Sibani P., Frost R. (2002), Facts, Conjectures, and Improvements for Simulated Annealing (SIAM Monographs on Mathematical Modeling and Computation), *Society for Industrial and Applied Mathematic*.

Sibson R. (1973), SLINK : an optimally efficient algorithm for the single-link cluster method, *The Computer Journal*, vol. 16, N° 1.

Spéry L., Claramunt C., Libourel T. (2001): A Spatio-Temporal Model for the Manipulation of Lineage Metadata.*GeoInformatica* 5(1): 51-70

Suman B., Kumar P. (2002), A survey of simulated annealing as a tool for single and multiobjective optimization, *Journal of the Operational Research Society*, vol. 57, p1143-1160.

Suppapitnarm A., Seffen K. A., Parks, G. T., Clarkson P. J. (2000), A simulated annealing algorithm for multiobjective optimization. *Engineering Optimization, 33*. pp. 59-85.

Stefani C., De Luca L., Véron P., Florenzano M. (2010), Modeling buildings historical evolutions, *Proceedings of Focus K3D conference on Semantic 3D Media and Content*, Sofia Antipolis, France.

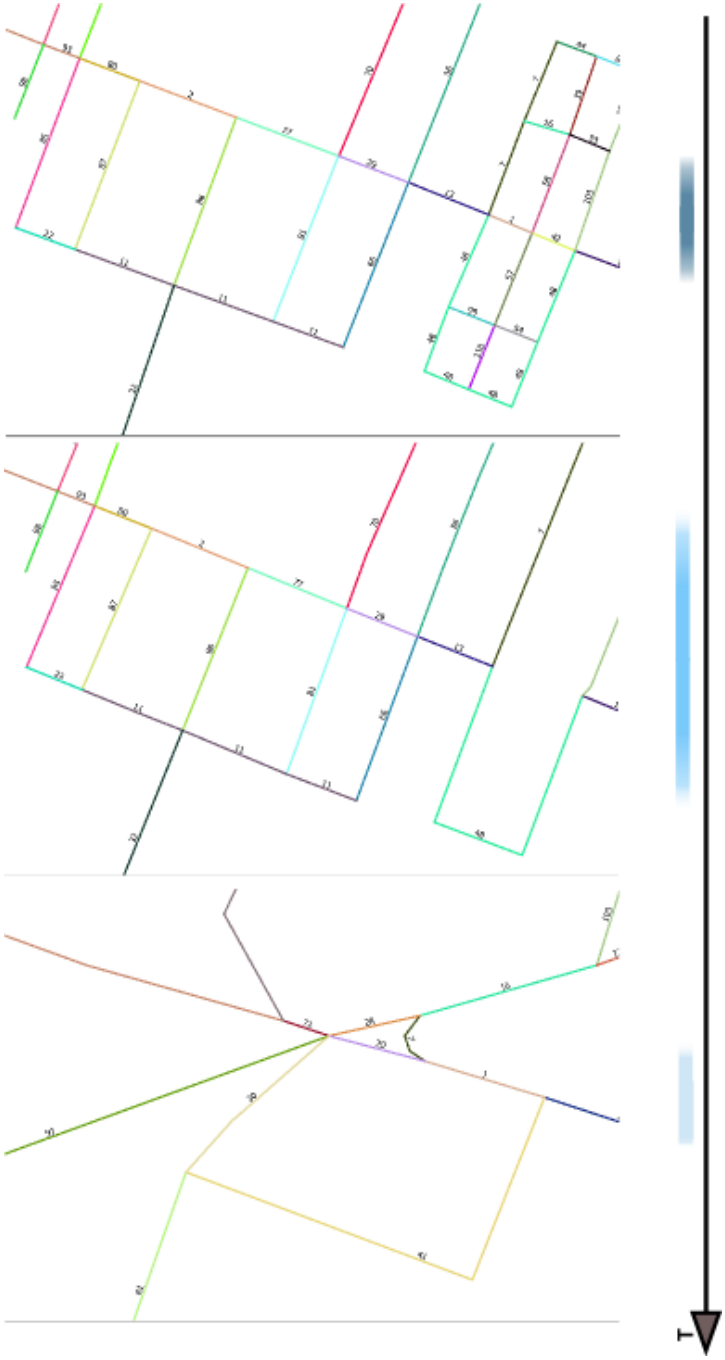
Stell J., Worboys M. (2008). A Theory of Change for Attributed Spatial Entities. In Proceedings of the 5th international conference on Geographic Information Science, *GIScience '08*, pp. 308-319, Berlin, Heidelberg. Springer-Verlag.

Stell J., Del Mondo G., Thibaud R., Claramunt C.: Spatio-temporal Evolution as Bigraph Dynamics. *COSIT 2011*: 148-167

Tong X et al. (2009). A probability-based multi-measure feature matching method in map conflation. *International Journal of Remote Sensing 30*(20):5453-5472.

18

Annexe 1. Composantes connexes du graphe de filiations



SESSION 6 :

Modélisation et fouille de données

Managing flash flood hazards susceptibility coupling GIS and Cellular Automata

Experimentations in Upper-Normandy (France)

Johnny Douvinet[1], Florian Mallet[1], Aurélie Escudier[2], Aurélien Christol[1], Daniel Delahaye[3], Patrice Langlois[4]

1. UMR CNRS 7300 ESPACE, Université d'Avignon et des Pays de Vaucluse
74 rue Louis Pasteur – 84029 Avignon Cedex 1 – France
johnny.douvinet@univ-avignon.fr, aurelien.christol@wanadoo.fr, florian.mallet@etd-univ-avignon.fr
2. Service Central Hydrométéorologique d'Appui à la Prévision des Inondations,
42 avenue Gaspar Coriolis – 31057 Toulouse
aurélie.escudier@developpement-durable.gouv.fr
3. UMR CNRS 6554 LETG-Géophen, Université de Caen Basse-Normandie
1 Esplanade de la Paix – 14032 Caen Cedex 2 – France
daniel.delahaye@unicaen.fr
4. UMR CNRS IDEES 6566, Université de Rouen
Bâtiment IRED, 7 Rue Thomas Becket - 76821 Mont Saint Aignan
Patrice.langlois@univ-rouen.fr

ABSTRACT. This paper proposes a new way for measuring the flash flood susceptibility of small basins (< 20km^2) in Upper-Normandy (France) combining GIS-data and Cellular Automata modelling. Shortly following high rainfall (from 50 to 100 mm in less than 6 hours), these floods are characterized by a violent onset, a rapid rising time and a surge rushing down the main valley just a few minutes after rains have peaked. But forecasting phenomena in space and time still remains delicate because the quantitative errors of radar disturb rainfall-runoff simulations. To anticipate possible hazards without depending on meteorological predictions, we implement GIS-data in the CA model RUICELLS to define possible peak-flow discharges, specific discharges and time-of-flowing out for 16 predefined rainfall intensities. Simulations obtained on 151 basins confirm that the flash floods susceptibility depends on the spatial and complex interactions between land use, rains and morphology. Methodological investigations also question the potential effectiveness of a future flash flood alert system in this department.
RESUME. En implémentant des données SIG à échelle fine dans une simulation hydrologique par automate cellulaire, cette étude propose de mesurer la susceptibilité aux crues rapides de petits bassins-versants (< 20km^2). Apparaissant suite à des pluies de forte intensité (de 50 à 100 mm en moins de 6 heures), ces crues s'accompagnent d'une « vague de crue » qui dévale

rapidement les talwegs des drains principaux avant d'affecter les exutoires et de provoquer des dommages lorsque ces derniers sont urbanisés. Le mode de fonctionnement de ces crues est connu mais leur anticipation dans le temps et dans l'espace demeure néanmoins délicate, d'autant plus que les erreurs dans les estimations des radars météorologiques perturbent les prévisions hydrologiques associées. Afin de s'affranchir des « alertes pluviométriques », nous avons fait appel à une démarche de modélisation par anticipation pour estimer les débits de pointe, les débits spécifiques et les temps d'écoulement probables selon 16 intensités de pluie différentes. Les simulations obtenues sur 151 bassins montrent que les interactions spatiales entre l'utilisation du sol, l'organisation morphologique du bassin et l'intensité des pluies sont primordiales car elles conditionnent les relations entre zones de production du ruissellement et zones d'infiltration. Ces investigations méthodologiques permettent aussi de questionner la faisabilité d'un système d'alerte aux crues rapides à l'échelle de la Seine-Maritime.

KEYWORDS: *flash flood, dry valleys, alert system, Upper-Normandy, France.*

MOTS-CLES: *crues rapides, vallons secs, système d'alerte, Seine-Maritime, France*

1. Introduction

Over the last two decades, flash floods posed serious risks in sedimentary areas (Boardman, 2004; Evrard *et al*., 2007), especially in small and dry basins located in the department of Upper-Normandy, in the northern part of France (Delahaye *et al.,* 2001). Shortly generated after high rainfall intensities, such floods are characterized by a sudden onset, a rapid rising time and a surge rushing down the main valley just a few minutes after rains have peaked (Douvinet *et al.,* 2013). These floods present features quite those similar to others occurring within western France or in Flanders (Auzet *et al*., 2005; Evrard *et al,* 2007), but are different from Mediterranean floods (Reid, 2004; Schmitz and Cullmann, 2008; Ortega and Heydt, 2009; Morin *et al.,* 2009). Even if hazards threaten people and cause increasing losses of infrastructures, predicting the time of occurrence and the precise location of a given phenomenon remains difficult in this department for several reasons: 1) Measurements and field-based experimentations are rarely collected in dry basins. 2) The phenomena are still insufficiently documented and difficult to monitor in real time because they produce numerous destructive effects to measuring devices. 3) The rarity of events and long recurrence intervals make statistical analysis and calibration of models difficult (Ferraris *et al.,* 2002). 4) Recurrent changes in velocity, roughness and water height introduce many uncertainties in estimation of peaks of discharge or of flowing-out (Douvinet and Delahaye, 2010).

With the double aim of increasing available information and facing the problem of anticipating spatial occurrence of such floods, we propose to experiment a new way of integrating GIS data in a Cellular Automata, namely RUICELLS (Delahaye *et al*, 2001; Douvinet *et al*., 2013). Our objectives are to simulate the hydrological responses of many basins using various environmental data (topography, land use and various rainfall intensities) and at managing their susceptibility by identifying the probable rainfall thresholds (critical rains) required to generate flash floods from local (basin) to departmental (Upper-Normandy) scales. Interestingly, the forecasters should use the critical thresholds for anticipating probable damage without uniquely

depending on the rainfall predictive models that try to forecast precise location and time-duration of future intense rains. In this sense, our method pursues the same objectives expected in USA with the *Flash Flood Guidances* (Estupina *et al.*, 2005).

Efforts in meteorological observations recently provide some relevant details on timing and location of convection in storm environments. Meteo-France experiments its own "alert to intense rain at the municipal scale" (namely APIC) and develops its own reanalysis method combining rain measurements with radar data (Lobligeois *et al.,* 2012). However, these efforts provide information on rainfall in real time but cannot predict rainfall in advance. Predictions with ARAMIS or PANTHERE radar networks remain difficult for forecasters. So we decided to experiment a new way by anticipation. We choose to use RUICELLS as we do not want a complicated and irreproducible approach, and our aims are also different from soil erosion models, in which many parameters are required (Nearing *et al*, 2005). One important parameter in these latter is the Manning roughness coefficient that comprises a flow resistant factor and depends on land cover in flooded areas (De Roo, 1999). But this factor is is usually not available in advance. We also want to control the "end-to-end" process that controls the transformation of input to output data (it is not the case for "black-box models" such as the ARPEGE or AROME Meteo-France models). This enables to envisage different configurations to measure the sensitivity to the model to initial conditions we cannot experiment in a realistic way. Our model like other Cellular Automata (Coulthard and Van De Wiel, 2006; Ménard and Marceau, 2006; Fonstad, 2006; Van de Wiel *et al.*, 2007; Douvinet *et al.*, 2013) is simple to implement over many basins. In this study, we hope at least to detect levels at which to issue alerts in a critical situation (intense rains during spring or summer after previous rainy days).

2. Study sites

In this study, we focus on 151 basins (figure 1) classified in three categories: 67 small basins (< 5km^2), 54 entities of medium size (ranged from 5 to 9 km^2) and 30 "bigger" basins (from 9 to 20km^2). The earlier study of 49 basins affected by flash floods over the period 1983-2005 (Douvinet, 2006) allows us to identify a number of properties which make basins susceptible to flash flooding and extrapolation of such features over the entire department of Upper-Normandy identified 151 basins with similar characteristics. Firstly, all these basins present small size (< 20km^2) and the highest slopes for this region (ranged from 2 to 15%). Generally, these small valleys are connected to major humid valleys in a short distance (< 2km) and this creates an order gap in the Strahler network, explaining the short time-duration of hydrological responses. Recent works (Douvinet et Delahaye, 2010; Douvinet *et al.,* 2009, 2013) highlight that the morphology plays a double controlling influence on flash floods dynamics: at regional scale, the morphology defines a number of basins ready to react in case of high intensities (so we need to work at large scale); at local scales (during one event and in a given basin), the dynamics of water flows and kinematics of the surges are strongly controlled by micro-topography (relief, depression, slopes angle and forms) and organization of networks whatever the basin size. Secondly, although selected basins contain differences in land use and degree of urbanization,

they have nonetheless common morphostructural features as they all belong to the Parisian Basin. Landscapes consist of successive subhorizontal to slightly undulating plates (Mathieu et al, 1997) incised by "dry valleys" inherited from the Quaternary periglacial periods (Lahousse et al., 2003). Thirdly, dominant soils (luvisols) are highly vulnerable to erosion due to the small amounts of organic matter (< 2%) and clay (< 15%) but high rates of silt (> 70%). This sensitivity has been demonstrated in many studies (Auzet et al., 1995; Mathieu et al., 1997; Le Bissonnais et al., 2002; Souchère et al., 2005). 64% of the 151 basins present high susceptibility to erosion during summer and 54% during spring. The surface degradation under raindrop effect induces strong reduction of infiltration water capacity, and the progressive disappearance of soil roughness concentrates runoff water. That is why the soil erosion has been yet studied in this area, where agricultural practices also increase since the last decade the runoff production (Souchère et al., 2005).

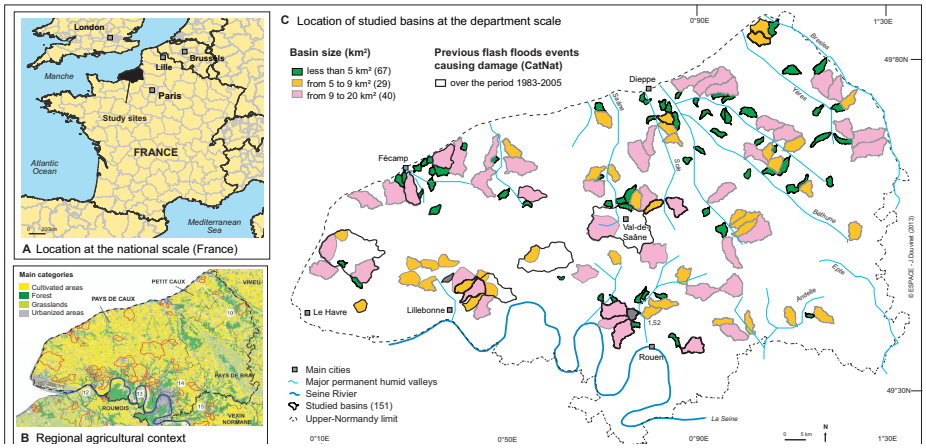


Figure 1. Location of the 151 studied basins in Upper-Normandy (northern France).

3. Material and data

3.1. Land Cover Maps (LCM) based on GIS-data cross-combined

Two types of GIS-data were used to produce the Land Cover Map (LCM). The Corine Land Cover delineates real-world land cover objects (lakes, cultivated fields, meadows, forests, industrial areas or natural areas for example). CLC (with 46 hierarchical levels of classification) has been produced by the Agency of Development at the European scale and derived from satellite images and digital cartography. Therefore, this data is not sufficient as it doesn't accurately delineate areas of less than 50 hectares and the precision of CLC is not enough to detect the runoff production sources. Then we improved CLC (2006) using the Geographical Parcel-Based File (2010). GPBF was initially created to help farmers apply for European Common Agriculture (ECA) funding. Gratefully provided by the DREAL-

Normandie, the GIS-data with 114 categories precisely locates each agricultural use (wheat, corn, flax, potatoes, etc) for each season (spring or winter periods). The cross-combined CLC and GPBF data reveals the complexity of land use patterns at the departmental scale, and the precision at local scales, which is also relevant. We present an overview of such LCM on a small basin (figure 2). The Mesnil-Val basin (5.33km^2) has an elevation ranging from 210 m to 26 m over a distance of 1.1 km. Its geology comprises of Cretaceous calcareous rocks throughout most of the basin. Grasslands exist on the middle parts of the basin, where slopes are higher, and interact with sugar beet, maize, springer peat or wheat in downstream and upstream parts. On May 10, 2000, following an event in which 75 mm of rain fell in two hours, the main inundated areas were located at the final outlet. The precision of this cross-combined LCM is readily interesting. We do not need to use more precise earth observation multi-spectral images derived for example from IKONOS-2 (Van der Sande *et al.*, 2003) because the spatial resolution is sufficient to detect or distinguish the main land cover. Errors due to geometric intersections between the GIS-data were nonetheless erased after field-experiments. On this basin, the quality of the final LCM was validated after a detailed field survey in June 2011. After this validation, we decided to launch the simulations on all the 151 studied basins. The digital data (RGP-2010) may not reflect recent (February 2013) changes in land use, but we depend on the data availability. This gap between the creation of digital data and the reality, currently pointed out in modeling approach (Kirkby *et al.*, 2005; Nearing *et al.,* 2005), remains unsolvable without new surveys.

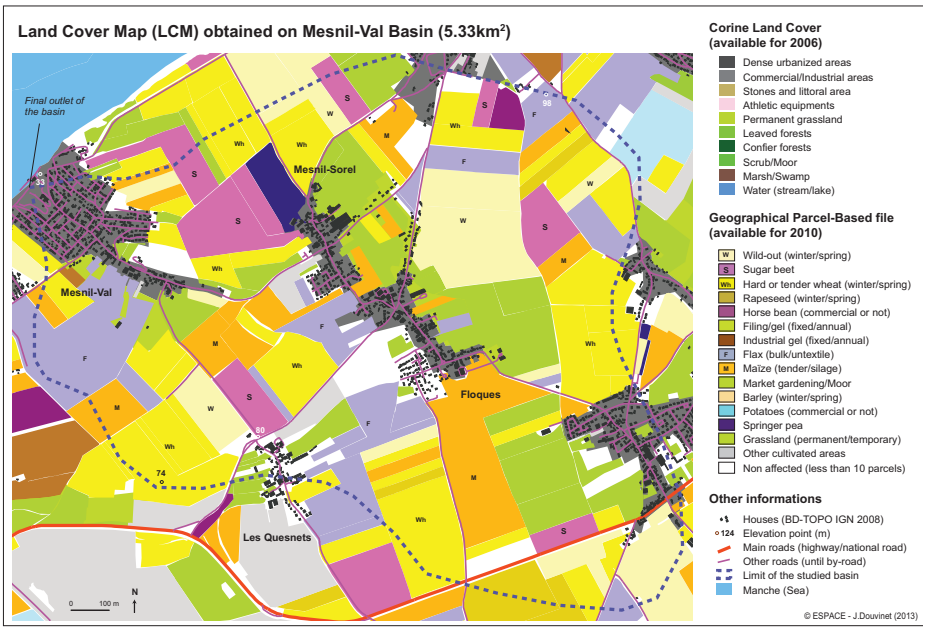


Figure 2. Land Cover Mapping combining Corine Land Cover and RGP SIG-data.

3.2. Choice and values of the infiltration water excess capacities

Numerous studies (Cerdan *et al,* 2002; Joannon, 2004; Souchere, 2005) have been conducted in Upper-Normandy to define regional infiltration and water excess capacities according to the land use. In this study we consider as critical the situation where soils are completely saturated at the beginning of the simulations. Then just a few minutes after rains fallen, surface flows quickly occur. Values implemented in RUICELLS surely underestimate the flash flood susceptibility but we assume this led to identify the critical gentle rains and alerts the forecasters as soon as possible. Implemented coefficients have been simplified for the main different land use type (table 1). Several data have been changed according to field-experiments we led after older flash flood events (the June 16, 1997 flood for example, Delahaye *et al.,* 2001). We pay particular attention to the most sensitive cultivated areas in terms of runoff production: values appear weak for maize (3mm in 1h) or potatoes (4mm in 1h) whereas forests and permanent grasslands have the highest infiltration capacities (50mm in 1h). One part of the initial rains have also been cut off (5 mm for each 1h) to take into account of the influence of soil and its porosity on the infiltration water excess capacities (Cerdan *et al*., 2002). Coefficients are unfortunately deterministic and fixed during the simulation process, unlike reality. We take into account effects of the rainfall intensities on land use types in changing both the cumulative amount of rains (50 mm) and the time-duration (1 or 6 hours).

Land use type (LCM)	Saturated infiltration (ARHEN, 2005)	Coefficients implemented in *RUICELLS*
Winter wheat	5 - 10	5
Silage corn	5 - 25	3
Winter barley	5 – 10	5
Spring barley	10 – 20	12
Rapeseed	5 – 10	5
Pea-grass	5 – 10	10
Flax	5 – 10	5
Production freeze	50	50
Industrial freeze	5 - 10	5
Other freezeland	/	50
Fodder	50	50

Land use type (LCM)	Saturated infiltration (ARHEN, 2005)	Coefficients implemented in *RUICELLS*
Permanent grass	50	50
Temporary grass	50	50
Orchard	50	50
Sugar beet	2 – 5	3
Potatoes	2 – 5	4
Industrial areas	/	10
Sunflower	/	15
Oil-producing	/	15
Seed areas	/	10
Unknwon area	/	100
Urban areas	/	5

Table 1. Infiltration excess capacities linked to the Land Cover Map.

3.3. Choice and values of the rainfall intensities

Meteo France estimated that rainfall intensities ranged from 42.1 to 44.5 mm in one day with a return period of 10 years for summer / spring periods (Meteo France, 2009). Therefore, these rainfall statistics remain delicate and strongly depend on the quality of monitoring stations and on the spatial distribution of the official stations, in comparison with the spatial or temporal variability of the thunderstorms. Use of

climatic statistics will then integrate strong uncertainties in our modelling approach. Consequently, we prefer defining 16 theoretical rains (ranging from 20 to 60 mm in 1h to 6h), in order to test the effects of these intensities on all the basins and to count the number of basins with high discharges for each one. The location of all sensitive areas may help forecasters in charge of flash flood prediction.

3.4. Implementation in the CA RUICELLS and simulation fallouts

LCM, infiltrations and rains are now implemented in RUICELLS. We previously presented this CA model (Douvinet *et al*, 2008; 2013) but several parameters need to be detailed for new readers. Experimentations in CA water flow modelling provide several suggestions and recommendations. The main difficulty in these models is to establish a link between topographic and hydraulic variables (Crave and Davy, 2001). Moreover, square lattices induce problems for simulating the runoff routing (Palacio-Vélez *et al*., 1998; Mita *et al*., 2001) as surface flows do not follow the real drainage. Different studies also highlighted the critical influence of the DEM cell size on the accuracy of extracted networks (O'Callaghan and Mark, 1984; Vogt *et al*., 2003). So we choose a lattice based on triangular, regular and interconnected cells based on a Digital Elevation Model (DEM), and we define rules to simulate the constant interactions between basin form, slope and the drainage network.

The structure of RUICELLS can be summarised as follows. The first step creates a topological mesh in triangular finite elements. In this, the direction of the steepest slope gives the downstream direction of flow and this information for each cell is draped over the Digital Elevation Model (DEM). The lowest diagonal was chosen to obtain more realistic surface flows. Each cell contains a pointer to its lowest downstream neighbour. The second step assigns one hydrological rule for each cell. The triangular facets represent elementary cells on hill-slopes, linear portions the thalwegs, and nodes the local closed depressions. In combining these rules, we aim to simulate interactions between these various surface-water flows (Murray and Paola, 1994). Each cell is linked to upstream/downstream cell(s) by a flow graph to form a cellular unit. The connectivity is directed by morphological links structured by the mesh as well as the neighbourhood topology of cells. Contrary to other CA, pathways are not only guided by the neighbourhood or vicinity conditions. Although two kinematics cascades were proposed by O.L. Palacios-Velez *et al*. (1998), we assume that linear and spatial runoffs are dependent. One synchronous advection operator avoids the problem of order in calculations. It means that surface runoff or flows used during the simulation flow occur in each cell at the same moment. So RUICELLS is based on a generalised CA model, in which cells have different facets and in which flow pathways represent real effects of morphological structure and not only its topology. Finally, the structure of RUICELLS is different from original CA.

Once the DEM (of 50 meter) is converted to a triangular lattice, the LCM is generated from the GIS server. The land use covering up to 50% of one cell (250m^2) is automatically defined in the mesh. After this process the user merges the infiltration water excess capacities linked to the LCM and downloads the predefined rainfall intensity he wants to test. We conserve the property defined as *locality* in

classical CA: the transition rules operate on cells directly based on local neighbourhood. In our case, we do not use the Moore (4) or Von Neumann (8) neighbourhood because surface flow follows the downstream direction as defined previously. Cellular Routing Scheme depends both on surface flows from each cell and on the updating of values of all sub-states. Flow is routed downstream until the final boundary is reached. At the end, the CA RUICELLS proposes a map indicating the cumulative runoff amounts and a graph showing evolution of discharges through time, mentioning both the time of flow-out and the peak-flow discharges (figure 3).

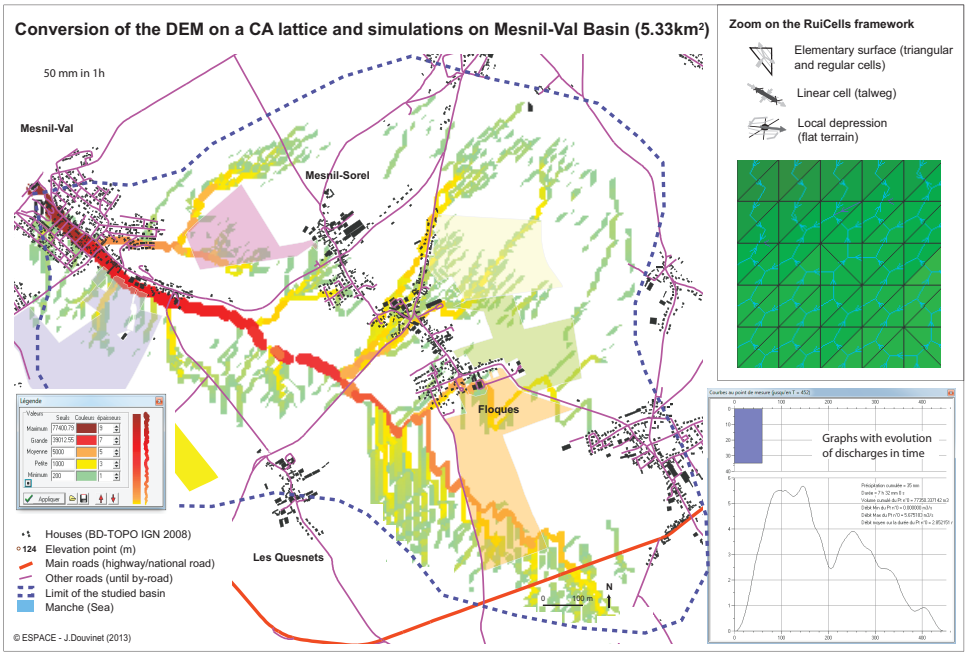


Figure 3. Triangular mesh and zoom on the framework used in RUICELLS.

4. Results

4.1. High susceptibility after 50 mm fell in 1h

Results obtained for peak-flows (Q) have been analysed for the different rainfall intensities, then reported at the department scale to render comparisons easy. Only 10 basins responding 30 mm in 1h have peaks-flows > 4m^3/s, but the number of susceptible basins increases for higher intensities, especially responding 40 mm in 1h (68 basins with Q > 4m^3/s) and 50 mm in 1h (106 basins with peak discharge Q > 4m^3/s, of which 22 basins with Q > 10 m^3/s; figure 4). The susceptibility therefore decreases for rainfalls more spread over time (33 basins with Q > 4m^3/s in case of a rain of 50 mm in 2h, of which 7 basins with Q > 10m^3/s). In 10 basins, the peak-

flow discharge regularly increases during the simulation and it can be explained by the fact that sugar beet, corn, maize or flax dominate land cover (on the Mesnil-Val basin for example) and grasslands occupy a big fraction of the basin area. These basins are well identified after 30 mm in 1h and react even if rains are more displayed in time. In contrast, 15 basins weakly react after 30 or 40 mm in 1h, but peaks strongly increase after 50 mm in 1h. Grasslands represent a small portion of basin size and reduce runoff productions coming from upstream for gentle rainfall intensities but are inefficient to slow cumulative amounts for very high rains. Other basins present intermediary behaviours between these extremes. The maps of peak-flows presented at the department scale permit us to follow the spatial distribution of basin response to rainfall. In several parts (south of Dieppe, north of Rouen, east of Le Havre), basins with similar response appear concentred in space, especially along the Manche Sea and Seine River. Flash floods tumble down the dry valleys across the fossil networks established during inherited periglacial periods (Boogart et al., 2003; Larue, 2005). This information is important for the risk forecasters because an extended thunderstorm (covering 10km^2 or more) can generate several flash floods at the same moment, while the response in isolated basins (in the eastern part of the department) may be easier to manage and prevent. Mapping the susceptibility finally highlights the variability and complexity of functioning thresholds, but also the non-linear relations between basin sizes, rainfall intensities and peak-flows.

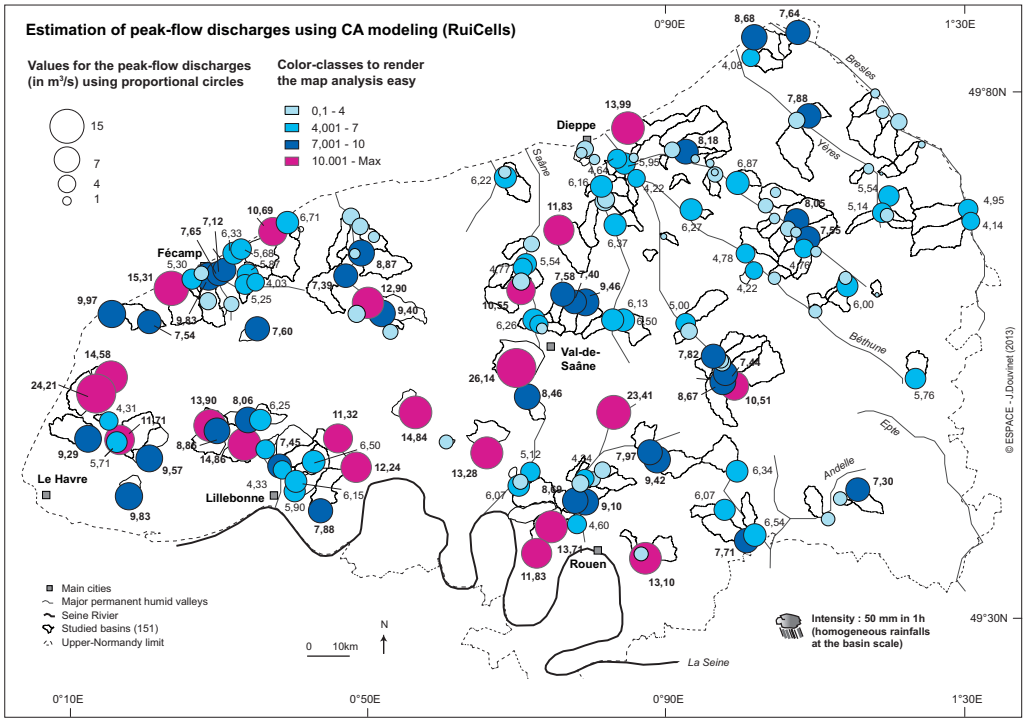


Figure 4. Peak flow discharges obtained after a rain of 50 mm fell in 1h.

4.2. High potential for the smallest basins

Previous studies carried out on Mediterranean flash floods (Borga *et al.,* 2007; Gaume *et al.,* 2009) highlight that flash floods are dangerous or strongly destructive when the specific peak-flow discharges (Qs), calculated by dividing discharges by the basin size, exceed at least 0.7m^3/s/km^2. We also defined thresholds in the north of France: one survey (Douvinet and Delahaye, 2010), conducted just a few days after the June 28, 2005 flood recorded in the basin of Hautot-sur-Mer (1.4 km^2), identified peak-flows which ranged from 1.2 m^3/s to 5.4 m^3/s, with specific stream powers varying between 215 and 511 W/m^2 in different cross-sections. The surveys provide a good assessment of the minimum energy required to initiate the incisions. Runoff contributions from upstream areas explain the appearance of a turbid wave through the main dry valley as well as important damage observed at the outlet of different sub-catchments. Consequently, we decided to calculate the probable Qs for all the 151 studied basins. We only map results after 50 mm fell in 1h (figure 5).

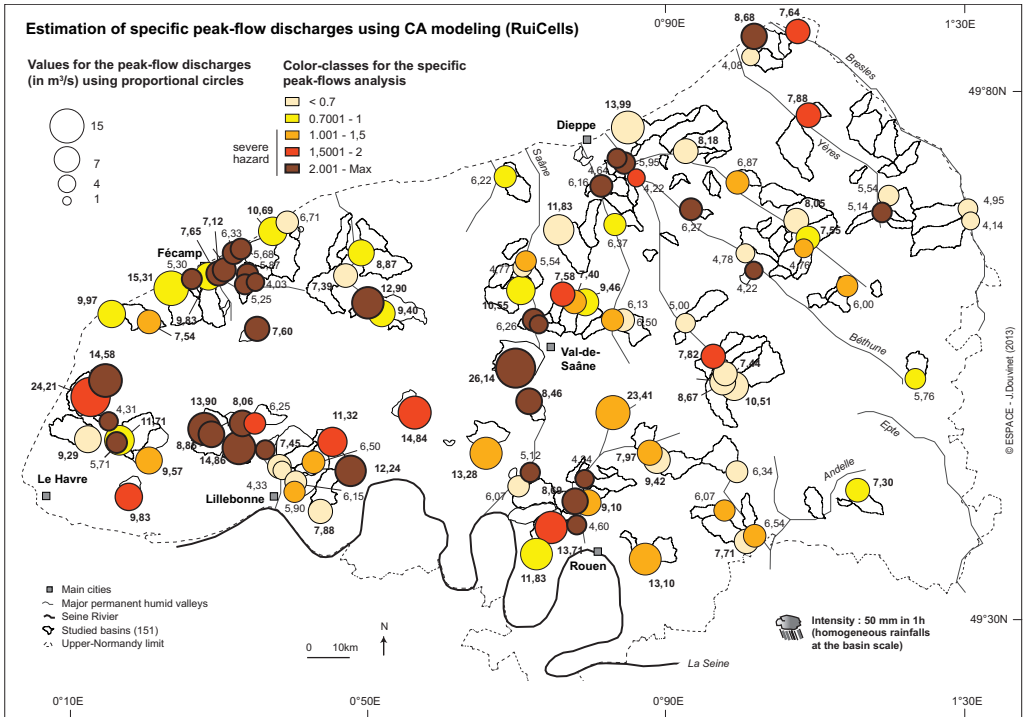


Figure 5. Specific peak flow discharges obtained after a rain of 50 mm fell in 1h.

This estimation of Qs supports another view of the flash flood susceptibility. After 50 mm fell in 1h, Qs exceed 1 m^3/s/km^2 in 64 basins, of which 34 basins with Qs > 2 m^3/s/km^2 (table 4). Most of the latter (56 %) have small size (the Hanouard basin of 2.75km^2 has Qs = 4.3 m^3/s/km^2). The increase of width in floodplains and

the decrease of slope gradients negatively influence the other 52 Qs values. But carefully, if the basin scale negatively affects Qs, shouldn't bigger basins they have smaller destructive capacity. These maps also suggest vigilance for risk forecasters that: 1) gentle peak-flow discharges (Q) do not automatically induce weak specific peak-flows (Qs); 2) the identification of five basins (figures 6 and 9) grouping high Qs and significant urbanization. In the past, events produced damage over the period 1983-2005 and in the future these areas can be affected again by severe damage in case of high rainfalls, unless dams or large projects have been recently built.

4.3. Does stakeholders have time to alert?

Another important question for the risk managers concerned the time they have to alert local authorities or the population for evacuation and protection in hazardous areas. The French Ministry of Environment and the General Delegation on Majors Risks (DGPR, 2011) focuses on this since the dramatic flash floods occurred in the western coast of France (49 deaths on February 28, 2010) and in southern part (25 fatalities June 15, 2010). The way proposed in this paper is to provide estimates that allow the forecasters to better know timing of peak-flows in all the 151 studied basins. The estimation of time-response depends here on the implementation of rains at the beginning of the simulation. So we prefer define it as "times-of-flow-out".

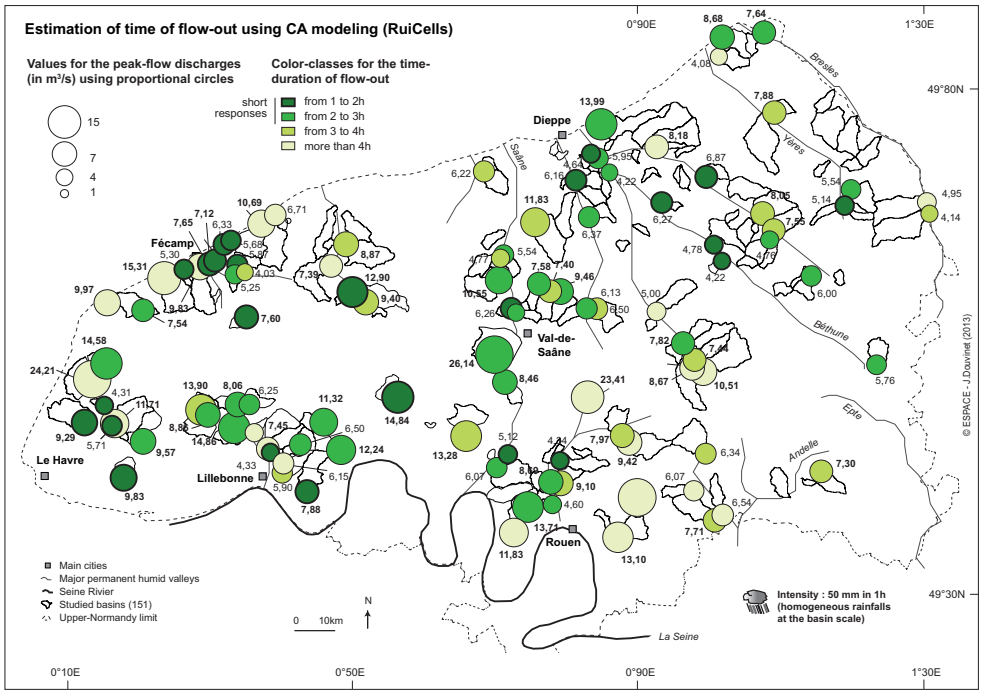


Figure 6. Time-of-flow-out obtained after a rain of 50 mm fell in 1h.

The estimation of Tf provides additional information. After 40 mm fell in 1h, 11 basins respond in less than two hours. But only one of these has an important peak discharge (Le Hanouard Q = 8,33 m^3/s). Most of the other basins with notable peak discharge response have Tf > 2h or Tf > 3h. In these cases forecasters have time to inform the local actors of a possible flash flood risk. A scaling effect appears: more the basin size increases more the time-of-flow-out extends – with the exception of the Fontenay and Notre-Dame-de-Gravenchon basins (in the western part of the department). The number of basins on which the time for alert and information is limited changes with higher rainfall intensity. With 50 mm in 1h (figure 6) 42 basins have Tf < 3h and 22 of these have Tf < 2h. Although the latter has small discharges, their specific peak-flows can be important. The 20 basins with Tf between 2h and 3h also need attention from forecasters. Indeed, the basins can generate flash floods at the same moment as they are concentrated in space. Flood phenomena can propagate in the major humid valley connected in the downstream and cities or urbanized areas located in the floodplains can be indirectly affected by such events. We observe this risk in Bethune valley (in several but disconnected basins), in the upstream part of the Saâne River or locally along the Seine River.

5. Conclusion and discussion in brief

The anticipation of flash floods in in Upper-Normandy is hampered by a lack of hydrological, meteorological and geomorphological knowledge. The rareness and violence of such events make the measurement of hydrological responses and behavior after intense rains difficult. In this study, we propose to measure the hazard susceptibility from local (basin) to global (department) scales by combining GIS-data within the CA model RuiCells. The numerical simulations obtained on the 151 studied basins highlight that the spatial interactions between land use, morphology and rainfall intensities appear of prime importance to explain this sensitivity compared to the basin size or to the average land use percent.

The number of basins ready to react strongly varies according to the rainfall parameters. When 50 mm fell in 1h, the discharges become important on 112 basins (73%) but forecasters need to pay particular attention to 14 basins combining high specific peak-flow discharges (Q > 1 $m^3/s/km^2$) and short time-of-flow-out (less than 2 hours). Flow should propagate downstream into larger valleys, especially near Fecamp, Lillebonne or Dieppe. Interestingly, the simulations highlight areas where the most dangerous flash floods previously occurred over the period 1983-2005 were registered. When the rains decrease (40 mm in 1h) or are extended (50 mm in 2h), the number of sensitive basins greatly decreases. Even if a rain of 50 mm in 1h may never occur on the entire department, forecasters can now better anticipate the susceptible areas from departmental to local scales. If they have sufficient time to inform stakeholders, the latter might inform and alert population. The validation of simulations is therefore required before these forecasts can be disseminated and for this, field-experiments and surveys are expected during the next few years.

Two main questions should be addressed in the next studies. On one hand, these floods present a high sediment concentration that remains difficult to define. Indeed, managers and official services clean the flooded urbanized areas and erase deposit signs before they can be surveyed and studied. If sediment sources are well known (soil erosion, destabilization of slopes and mass movements, incision in networks, overthrusting of debris, vegetal and artificial elements adding to solid fluxes), quantifying precisely the sediment budget is an insolvable problem in these small and ungauged areas (Douvinet et al., 2013). On the other hand, we will estimate the specific stream powers in a few cross sections within the 151 studied basins. These values could provide a good assessment of the minimum energy required to induce incision. Runoff contributions from upstream can explain the appearance of a turbid wave observed at the outlet of the sub-catchments. Then, we could better understand why some high-incised gullies occur in less than 10 meters in several basins.

Bibliography

Anquetin S., Ducrocq V., Braud I., Creutin J.D. 2009. Hydrometeorological modelling for flash flood areas. The case of the 2002 Gard event in France. *Flood Management* 2, 101-110.

Auzet A.-V., Boiffin J., Ludwig D., 1995. Concentrated flow erosion in cultivated catchments: influence of soil surface. *Earth Surface Processes and Landforms* 20, 759-767.

Coulthard T.J., Van de Wiel M.J., 2006. A cellular model of river meandering. *Earth Surface Processes and Landforms* 31, 123-132.

Crave A., Davy P., 2001. A stochastic "precipiton" model for simulating sedimentation dynamics. *Computers & Geo-sciences* 27, 815-827.

Cerdan O., Le Bissonnais Y., Couturier A., Bourennane H., Souchère V., 2002. Rill erosion on cultivated hillslopes during two extreme rainfall events in Normandy, France. *Soil & Tillage Research*, 67, 99-108.

Delahaye D., Guermond Y., Langlois P., 2001. Spatial interaction in the runoff process. *Proceedings of the 12th ECTG2001, European Colloquium on Theoretical and Quantitative Geography,* Saint-Valéry-en-Caux, France, 2001, http://www.cyber- geo.eu/index3795.html.

De Roo, A.P.J., 1999. LISFLOOD: a rainfall-runoff model for large river basins to assess the influence of land use changes on flood risk. In: Balabanis, P. et al. (Eds.), Ribamod: River Basin Modelling, Management and Flood Mitigation. Concerted Action, European Commission, EUR 18287 EN, pp. 349–357.

Douvinet J., 2008. *Les bassins versants sensibles aux « crues rapides » dans le Bassin parisien. Analyse de la structure et de la dynamique de systèmes spatiaux complexes.* Thèse de docto- rat en géographie, université de Caen Basse-Normandie, 381 p.

Douvinet J., Delahaye D., 2010. Caractéristiques des « crues rapides » du nord de la France (Bassin parisien) et risques asso- ciés. Gé*omorphologie : relief, environnement, processus* 1, 73-90

Douvinet J., Delahaye D., Langlois P., 2013. Measuring surface flow concentrations using a cellular automaton metric: a new way of detecting the potential impacts of flash floods in sedimentary context. Gé*omorphologie, relief, environnement, processus* (GRPE), 1, 27-46.

Estupina-Borell V., Chorda J., Dartus D., 2005. Prévision des crues éclair. *Comptes Rendus Geoscience*s, 337, 1109-1119.

Evrard O., Persoons E., Vandele K., Van Wesemael B., 2007. Effectiveness of erosion mitigation measures to prevent muddy floods: A case study in Belgian, *Agriculture, Ecosystems and Environment*, 118, 149-158.

Ferraris L., Rudari R., Siccardi F., 2002. The uncertainty in the prediction of flash floods in the Northern Mediterranean Environment. *Journal of Hydrometerology* 3, 714-727.
Fonstad M. A., 2006. Cellular automata as analysis and synthesis engines at the geomorphology-ecology interface. *Geomorphology* 77, 217-234.
Gaume E., Bain V., Bernardara P., Newinger O., Barbuc M., Bateman A., Blaskovicova L., Bloschl G., Borga M. Dumitrescu A., Daliakopoulos I., Garcia J., Irismescu A., Kohnova S., Koutroulis A. Marchi L., Matreat S., Medina V., Preciso E., Sempre-Torres D., Strancalie G., Szolgay J., Tsnais I., Velasco D., Viglione A., 2009. A compilation of data on European flash floods. Journal of Hydrology 367, 70-78.
Joannon A., 2004. Coordination spatiale des systèmes de culture pour la maîtrise de processus écologiques. Cas du ruissellement érosif dans les bassins versants agricoles du Pays de Caux, Haute- Normandie. Thèse de Doctorat à l'INA-PG, INRA SAD, 393 p. + annexes.
Kirkby M.J., Bracken L.J., Shannon J., 2005. The influence of rainfall distribution and morphological factors on runoff delivery from dryland catchments in Spain. *Catena*, 136-156.
Lahousse P., Pierre G., Salvador P.-G., 2003. Contribution à la connaissance des vallons élémentaires du nord de la France: l'exemple de la creuse des fossés (Authieule, plateau picard). *Quaternaire* 14, 189-196.
Larue J.-P., 2005. The status of ravine-like incisions in the dry valleys of the Pays de Thelle (Paris basin, France). *Geomorphology* 68, 242-256.
Mantilla R., Gupta V.K., Mesa O.J., 2006. Role of coupled dynamics and real network structures on Hortonian scaling of peak flows. *Journal of Hydrology* 75, 1-13.
Mathieu R., King C., Le Bissonnais Y., 1997. Contribution of multi-temporal SPOT data to the mapping of a soil erosion index. The case of loamy plateaux of northern France. *Soil Technology* 10, 99-110.
Ménard A., Marceau D.J., 2006. Simulating the impact of for- est management scenarios in an agricultural landscape of southern Quebec, Canada, using a cellular automata. *Landscape and Urban Planning* 16, 99-110.
Mita D., Catsaros W., Gouranis N., 2001. Runoff cascades, channel network and computation hierarchy determination on a structured semi-irregular triangular grid. *Journal of Hydrology* 244, 105-118.
Morin E., Jacoby Y., Navon S., Bet-Halachmi E., 2009. Towards flash flood prediction in the dry Dead Sea region uti- lizing radar rainfall information. *Advances in Water Resources* 32, 1066-1076.
Murray A.B., Paola C., 1994. A cellular model of braided rivers. *Nature* 371, 54-57.
Nearing M. A. Jetten V., Baffaut C., Cerdan O., Couturier A., Hernandez M., Le Bissonnais Y., Nichols M. H., Nunes N. J., Renschler C. S., Souchere V., Van Oost K., 2005. Modeling response of soil erosion and runoff to changes in precipitation and land cover, *Catena*, 61, pp. 131-154.
O'Callaghan J.F., Mark D.M., 1984. The extraction of drainage networks from digital elevation map. Computer vision, *Graphics and Image Processing* 28, 323-344.
Ortega J.E., Heydt G.G., 2009. Geomorphological and sedi- mentological analysis of flash floods deposits. The case of the 1997 Rivillas flood (Spain). *Geomorphology* 112, 1-14.
Palacios-Vélez O.L., Gandoy-Bernasconi W., Cuevas-Renaud B., 1998. Analysis of surface runoff and the computation order of unit elements in distributed hydrological models. *Journal of hydrology* 211, 266-274.
Reid I. (2004) – *Flash flood*. In Goudie A. (Ed.) *Encyclopedia of Geomorphology*. Routledge, London, 1156 p.
Schmitz G.H., Cullmann J., 2008. PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments. *Journal of hydrology*, 360, 1-14.
Van der Sande C.J., De Jong S.M., De Roo A.R., 2003. A segmentation and classification approach of IKONOS-2 imagery for land cover mapping to assist flood risk and flood damage assessment, *Journal of Applied Earth Observation and Geoinformation*, 4, 217–229.

Vogt J.V., Colombo R., Bertolo F., 2003. Deriving drainage network and catchments boundaries as a new methodology com- bining digital elevation data and environmental characteristics. Geomorphology 53, 281-298.

OPITER : Fouille de données d'opinion pour les territoires

Eric Kergosien[1, 2], Pierre Maurel[2], Mathieu Roche[1], Maguelonne Teisseire[1, 2]

1. LIRMM, CNRS, Université Montpellier 2
161, rue Ada, 34392 Montpellier, Cedex 5, FRANCE
prenom.nom@lirmm.fr

2. IRSTEA - TETIS
500, rue J. F. Breton, 34093 Montpellier, Cedex 5, France
prenom.nom@teledetection.fr

RESUME.

De nombreux travaux ont été réalisés sur la fouille de données d'opinion dans des contextes spécifiques tels que les critiques de films, les évaluations de produits commerciaux, les discours électoraux … Dans cet article, nous nous posons la question de l'adéquation de ces méthodes aux documents associés à l'aménagement des territoires. Après avoir souligné les limites des propositions actuelles et les verrous soulevés par les données textuelles associées, nous proposons la méthode OPITER (Détection d'OPInion dans les textes liés à l'aménagement du TERritoire). Nous posons ainsi les premières bases d'un projet ambitieux et les pistes prometteuses associées.

ABSTRACT.

A great deal of research on opinion mining and sentiment analysis from textual datas has been done in several specific contexts such as movie reviews, commercial product evaluations, campaign speeches, etc. In this article, we raise the issue of how appropriate these methods are for documents related to lands use planning. After highlighting the limitations of existing proposals and discussing issues related to textual data, we present a method called OPITER (Opinion mining in texts related to land-use planning). So, we are laying the first foundations of an ambitious project and associated promising direction.

MOTS-CLES : Fouille de données d'opinion, territoire, corpus, lexique

KEYWORDS: Opinion mining, territories, corpus, lexicon

1. Introduction

Au-delà de sa stricte définition d'entité administrative et politique, le territoire, selon Guy Di Méo (1998), témoigne d'une "appropriation à la fois économique, idéologique et politique de l'espace par des groupes qui se donnent une représentation particulière d'eux-mêmes, de leur histoire, de leur singularité". Dans ce contexte éminemment subjectif, la caractérisation et la compréhension des perceptions d'un même territoire par les différents acteurs est difficile, mais néanmoins particulièrement intéressante dans une perspective d'aménagement du territoire et de politique publique territoriale. La prise en compte des sentiments et opinions de la population constitue un double enjeu politique et démocratique, à la fois pour apprécier la progression de l'identification au nouveau territoire dans un objectif d'assoir une légitimité politique, mais aussi pour que cette identité collective émergente transforme une population passive en sujets politiques, individuels ou collectifs, acteurs de leur territoire et engagés dans un projet partagé de développement territorial. Dans le premier cas, la mesure des sentiments et opinions relève surtout d'une logique de marketing territorial à visée interne (promotion du nouveau territoire) et passe le plus souvent par des enquêtes d'opinion, de satisfaction et des sondages. Dans le deuxième cas, la prise en compte des sentiments et opinions peut répondre à de multiples finalités de la part des élus, de l'administration territoriale, mais aussi de la population comme : accéder aux représentations que les individus et les collectifs se font d'un territoire ou de certaines de ses composantes thématiques ; savoir ce qu'ils pensent d'un projet politique global (une vision à moyen ou long terme du territoire) ou de projets particuliers d'aménagement dans le plan physique ; anticiper les résistances ou au contraire les soutiens à des projets ; se rapprocher de ceux qui partagent les mêmes opinions pour forger des coalitions ; analyser les convergences ou les écarts entre les perceptions du territoire par des acteurs et enfin l'image de ce territoire véhiculée dans l'espace numérique et la réalité mesurable sur le terrain.

La fouille de données d'opinion a donné lieu à de nombreux travaux qui trouvent leurs applications dans des domaines variés tels que l'analyse automatique de critiques et de commentaires de film, d'articles politiques, de tweets … Les méthodes proposées reposent sur des techniques statistiques ou de traitement automatique du langage. Dans la plupart des cas, elles utilisent un lexique ou dictionnaire de mots d'opinion avec, ou non, une polarité qui y est associée. Les corpus sont dédiés à l'expression d'une opinion qui peut être neutre s'il n'est pas possible d'en identifier la portée positive ou négative. Dans le contexte de l'aménagement des territoires, certes les informations publiées sur le web (blog, fora…) ou dans les médias de façon plus générale (journaux, comptes rendu d'organisation de l'état …) expriment un sentiment ou un ressenti mais les subtilités associées montrent les limites des approches de fouille d'opinion existantes. Il est donc nécessaire de trouver une approche alternative, permettant de dépasser les verrous identifiés afin de proposer une démarche spécifique, orientée domaine, qui permette d'extraire de façon automatique ou semi-automatique des sentiments là où il est difficile de les exprimer, de les identifier et de les restituer.

Dans cet article, nous tentons de répondre à ces questions dans le contexte de l'aménagement des territoires en particulier et à partir des limites des méthodes existantes dans la littérature, nous posons les premières bases d'une approche générique afin de définir un vocabulaire d'opinion spécialisé.

L'organisation de cet article est la suivante. Un panorama des méthodes de fouille de données d'opinion est présenté dans la Section 2 avec un focus particulier sur les données textuelles spécifiques associées aux territoires. Dans la Section 3, nous décrivons la méthode OPITER (détection d'OPInion dans les textes liés à l'aménagement du TERritoire). La Section 4 reporte les expérimentations réalisées sur le corpus étudié pour offrir un premier bilan de ce contexte spécifique. Puis, en Section 5, nous dressons les conclusions et traçons les nombreuses perspectives associées.

2. Etat de l'art

Le travail proposé s'inscrit dans un projet plus large, le projet SENTERRITOIRE dont l'objectif est la réalisation d'un environnement décisionnel fondé sur une analyse automatique des textes liés à l'aménagement du territoire. Le projet s'est concentré dans un premier temps sur l'extraction automatique des Entités Spatiales ES (Tahrat *et al.*, 2013) en proposant une méthode hybride utilisant des patrons linguistiques dans un processus d'apprentissage. Pour cela, nous nous appuyons sur le modèle Pivot (Lesbegueries, 2007) qui permet d'interpréter la plupart des ES exprimées en langage naturel dans les textes. Dans ce modèle, l'ES est constituée d'au moins une Entité Nommée (EN) et d'un nombre variable d'indicateurs spatiaux précisant sa localisation. Ces entités spatiales nécessitent ensuite d'être enrichies par des informations caractérisant plus largement le concept de « territoire » afin d'analyser automatiquement l'utilisation de ces concepts dans les textes, et les perceptions qu'ils véhiculent.

Concernant la **fouille de données d'opinion**, les travaux menés par la communauté scientifique sont assez récents. Dans le traitement du langage naturel (TAL), l'analyse de la subjectivité et des opinions exprimées par des personnes dans les textes (journaux, documents techniques, blogs, commentaires des internautes, lettres aux éditeurs, etc.) est connue comme l'analyse des sentiments ou des opinions. Une étude exhaustive de ces travaux est présentée dans (Pang et Lee, 2008) ainsi que dans des numéros spéciaux de revues (Roche et Poncelet, 2009 ; El-Bèze et al, 2010). Parmi ces travaux, la détection de la subjectivité vise à séparer les déclarations des avis sur des faits (Kim et Hovy, 2004). La reconnaissance de polarité tente de classer les textes selon la positivité ou la négativité des opinions qui y sont exprimées. Deux approches principales peuvent être identifiées : l'une fondée sur le recensement des termes positifs et négatifs dans chaque texte (Turney, 2002), et l'autre sur l'apprentissage automatique à partir de textes annotés (Esuli et Sebastiani, 2006). Les approches hybrides semblent proposer les meilleurs résultats (Kennedy et Inkpen, 2006).

Dans toutes ces approches plusieurs descripteurs doivent être utilisés, tels que des mots, des n-grammes de mots (Pak et Paroubek., 2010), les mots modifiés (Joshi et al., 2011), et ainsi de suite. Ces différents descripteurs peuvent être exploités par des méthodes d'apprentissage automatique en s'appuyant sur des corpus annotés. Beaucoup de corpus sont disponibles dans le cadre de challenge d'analyse de texte tels que TREC (Text Retrieval Conference), mais seuls quelques uns sont annotés en prenant en compte les opinions et la polarité. En outre, plusieurs méthodes de classification peuvent être regroupées en systèmes de vote proposés par (Plantié *et al.*, 2008) ou en appliquant des méthodes de renforcement et / ou « sacs de mots » (Fan *et al.*, 2011). Les méthodes de renforcement intègrent une phase d'apprentissage à partir du "vécu", de l'expérience passée évaluée mathématiquement. Les méthodes « sacs de mots » adoptent une phase d'apprentissage à partir des mots simples du corpus pris séparément les uns des autres afin de discriminer les textes pertinents des textes non pertinents. Dans ce type d'approche, les mots vides (mots fonctionnels) ne sont en général pas pris en compte dans le processus. D'autres approches s'appuient sur des méthodes incrémentales pour l'analyse des sentiments (Wiebe et Riloff, 2011). Outre la classification de textes d'opinion, des travaux se sont concentrés sur la construction automatique de vocabulaires spécialisés liés à l'opinion (Harb *et al.*, 2008, Duthil *et al.*, 2011). Les approches incrémentales proposées reposent sur des méthodes de fouille du web afin d'apprendre un vocabulaire d'opinion lié à un thème ou sous-thème donné.

Concernant la problématique applicative associée à **l'analyse des ressentis** des acteurs à propos de **l'aménagement des territoires**, le panorama des travaux existants soulignent l'implication de différentes communautés. En effet, depuis une trentaine d'années, le concept de territoire a été largement utilisé et discuté, sous des acceptions diverses, tant, dès l'origine, par les éthologues et écologues, que par les géographes (Sack, 1986), les sociologues (Barel, 1981), les économistes (Pecqueur, 2007) les politistes (Allies, 1980), les agronomes (Deffontaines *et al.*, 2001), ainsi que les philosophes (Deleuze & Guattari, 1980)… La géographie en particulier a été particulièrement prolixe, partageant, pour aller vite, sa production entre une analyse du territoire « sujet politique » et une analyse du territoire « sujet social » (Vanier, 2009) : d'un côté, la géographie sociale a analysé la dimension identitaire du territoire, les rapports d'appartenance et d'ancrage (Buléon & Di Méo, 2005) ; de l'autre, une géographie plus politique s'est efforcée d'éclairer la dimension de représentation du territoire, à travers une analyse des dispositifs d'action publique (Debarbieux et Vanier, 2002). Pour faire face à ses acceptions multiples, il convient souvent dès lors d'en préciser le sens par l'ajout d'un qualificatif : territoires biophysiques (bassin versant, grand paysage …), territoires politico-administratifs (commune, Etat, Europe …), grands territoires (Grand Paris, Arc Atlantique, …), territoires appropriés (Bretagne, Béarn, Pays Basque, Larzac, …), territoires mobiles des nomades traditionnels (chameaux …) ou modernes (TGV, avion …), territoires numériques générés par les réseaux de télécommunication, territoires virtuels sur le Web qui peuvent parfois prolonger et étendre dans le cyberespace des représentations des territoires physiques. Cependant, ces différents territoires ne

constituent pas des ensembles disjoints, bien au contraire, tellement ils sont désormais étroitement imbriqués.

3. Approche OPITER : Détection d'Opinion dans les textes liés à l'aménagement du TERritoire

Notre objectif est d'identifier, de la façon la plus automatique possible, l'opinion véhiculée dans les textes liés à l'aménagement du territoire. Les méthodes classiques de fouille d'opinion trouvent leur limite dans le domaine choisi car le vocabulaire utilisé s'éloigne de celui des corpus habituellement traités (Voir section 4, les scores de classification de référence). Nous avons testé une approche sac de mots en appliquant des algorithmes classiques de classification par apprentissage supervisé. Notons ici que l'apprentissage supervisé est une technique d'apprentissage automatique où l'on cherche à produire automatiquement des règles à partir d'une base de données d'apprentissage contenant des exemples (en général des cas déjà traités et validés). Compte tenu notamment de la faible proportion du corpus d'entrainement et de la diversité des thématiques abordées, les résultats se sont révélés peu satisfaisants (entre 50 à 55%).

C'est pourquoi nous proposons une démarche générique, détaillée dans les sous-sections suivantes, qui intègre et valorise de façon harmonieuse les vocabulaires adoptés afin de calculer un score d'opinion pour les documents traités. OPITER est ensuite évalué de façon expérimentale sur un corpus spécifique et ses résultats sont discutés en section 4.

3.1. Principe global

OPITER repose sur l'identification du vocabulaire propre à une opinion afin d'être en mesure d'attribuer une polarité aux documents étudiés. Cette polarité est positive ou négative. Notons qu'une approche non supervisée fondée sur la construction de vocabulaires d'opinion ne nécessite pas d'avoir à disposition un corpus conséquent et annoté comme données d'apprentissage contrairement aux méthodes classiques de fouille d'opinion (Torres-Moreno et al., 2007). Néanmoins, dans un objectif d'évaluation de la méthode proposée, une version du corpus a été étiquetée (positif/négatif).

Pour identifier une opinion dans un texte, nous définissons trois types de vocabulaire :

1. *Vocabulaire pivot d'opinion généraliste* (VPOG) : liste de mots polarisés généralistes ;
2. *Vocabulaire d'opinion contextualisé* (VPOC) : liste de mots qui seront polarisés au regard du contexte dans lesquels ils se trouvent ;
3. *Vocabulaire pivot d'opinion spécialisé* (VPOS) : liste de mots polarisés liés au domaine de l'aménagement du territoire.

Dans les sections suivantes, nous décrivons tout d'abord comment construire les différents vocabulaires et détaillons la méthode d'attribution d'un score d'opinion aux documents. Ce score d'opinion est fondé sur une méthode mixte, à la fois statistique et linguistique.

3.2. Construction des dictionnaires d'Opinion

3.2.1 Vocabulaire pivot d'opinion généraliste

Le vocabulaire dit général dispose de nombreux mots ayant une polarité clairement définie. Par exemple, les mots "aimer", "beau", "intéressant" sont clairement associés à une opinion positive tandis que les mots "mauvais", "laid" et "insuffisant" ont une polarité négative. Les lexiques sont construits manuellement, semi-automatiquement ou de manière contributive. Ainsi, leur richesse et leur "fiabilité" sont très dépendantes de leur mode d'acquisition. Leur utilisation conjointe ou séparée peut être délicate. Il s'agit alors de trouver la meilleure combinaison afin d'obtenir une liste de descripteurs linguistiques d'opinion pertinente et adaptée à la problématique étudiée. Les différentes combinaisons proposées seront évaluées sur le corpus en section 4.

Il existe peu de lexiques d'opinion spécifiques au français et nous avons choisi les trois lexiques suivants :

Lexique 1 : le lexique *GeneralInquirer français* (Vincze & Bestgen, 2011), version traduite du Ge- neralInquirer[1], contient des informations syntaxiques, sémantiques et pragmatiques sur une liste de mots polarisés. La polarité permet d'expliciter pour chaque mot si il est positif ou négatif. Cette liste est disponible en version française après traduction, lemmatisation[2] et vérification par deux juges (Vincze & Bestgen, 2011). Au final, le lexique contient 1246 mots positifs et 1527 mots négatifs.

Lexique 2 : le lexique *LIWC français* (Piolat *et al*., 2011) est la traduction du dictionnaire anglais du Linguistic Inquiry and Word Count[3] (LIWC). Les mots ne sont pas stockés sous la forme de lemme et de nombreuses flexions dues à la conjugaison des verbes augmentent la taille du lexique. On trouve notamment dans les catégories des mots proposés, ceux ayant trait aux émotions positives et négatives.

Lexique 3 : le lexique *JeuxDeMots* (Lafourcade, 2007) est un lexique français étendu à toutes les parties du discours (nom, verbe, adjectif, adverbe) mais également à un grand nombre d'entités nommées (personnes, lieux, marques, événements). Il est défini sur la base d'un jeu en ligne nommé LikeIt[4], permettant de

---

[1] http://www.wjh.harvard.edu/~inquirer/
[2] analyse lexicale du contenu d'un texte regroupant les mots d'une même famille (pluriel, verbe à l'infinitif, etc.)
[3] http://www.liwc.net/
[4] http://www.jeuxdemots.org/likeit.php

capturer les informations de polarité (opinion positive, négative ou neutre), et plus de 250000 termes sont ainsi potentiellement polarisables. A l'heure actuelle, 27529 termes on été polarisés avec plus de 218739 opinions exprimées.

Afin de consolider ces données d'opinion, nous proposons dans un premier temps de fusionner les différents lexiques en un vocabulaire d'opinion généraliste VPOG. Un score de fiabilité pourra alors être attribué à chaque type de fusion. Par exemple, si un mot est présent dans les trois lexiques positifs (ou les trois lexiques négatifs), une fiabilité élevée pourra être accordée à ce mot. Ainsi, nous proposons d'accorder un score de fiabilité S_i selon les configurations suivantes :

- GeneralInquirer inter ∩ LIWC ∩ JeuxDeMots = S_1 ;
- GeneralInquirer ∩ LIWC = S_2 ;
- GeneralInquirer ∩ JeuxDeMots = S_3 ;
- LIWC ∩ JeuxDeMots = S_4 ;
- GeneralInquirer = S_5 : Mots du GeneralInquirer distincts des autres lexiques ;
- LIWC = S_6 : Mots du LIWC distincts des autres lexiques ;
- JeuxDeMots = S_7 : Mots du JeuxDeMots distincts des autres lexiques.

De manière globale, nous pouvons déterminer trois types de scores de fiabilité : un score élevé (S_1) pour un mot présent dans les trois lexiques, un score moyen (S_2, S_3, S_4) lorsqu'un mot est présent dans deux lexiques et un score faible (S_5, S_6, S_7) lorsqu'un mot est présent dans un seul des lexiques. Ces scores seront rigoureusement évalués en Section 4.

Notons qu'un pré-traitement initial a consisté à supprimer les mots présents à la fois dans les lexiques positifs et négatifs.

3.2.2 Vocabulaire d'opinion contextualisé

Les mots issus du vocabulaire VPOG vont constituer des "opinions pivots". C'est-à-dire que les mots situés dans le même contexte, a priori neutres, seront alors polarisés. Concrètement, si un mot de type adjectif, nom, verbe ou adverbe se situe "proche" d'une opinion pivot (i.e. selon une fenêtre donnée), il sera à son tour polarisé. Le score de polarité qui lui sera attribué dépendra de deux facteurs : le score du mot pivot provenant du vocabulaire d'opinion généraliste et la distance avec ce dernier (cf. formule (1)).

$$\text{ScoreMotVoisin(Mot)} = \frac{\sum \frac{\text{ScoreMot(MotPivot)}}{d}}{\sum \text{MotsPivots}} \qquad (1)$$

Avec

- *d : le nombre de mots entre le mot traité et le mot pivot*

Pour chaque mot non pivot candidat à intégrer le vocabulaire d'opinion *MotC* (ayant la catégorie grammaticale verbe, nom, adjectif ou adverbe), on identifie l'ensemble des mots pivots voisins (i.e. dans une fenêtre de mots voisins) et pour chaque mot pivot sélectionné, un score de polarité *SPVoisin* est calculé en divisant son score de polarité par la distance avec *MotC*. Le score *ScoreMotVoisin* de *MotC* correspond alors à la moyenne de ces scores *SPVoisin*.

Le vocabulaire VPOC obtenu à cette étape est ainsi constitué des mots pivots du VPOG et des mots situés dans le même contexte que ces mots pivots.

3.2.3 Vocabulaire pivot d'opinion spécialisé

Rappelons que nous souhaitons mettre en exergue un vocabulaire d'opinion spécialisé à notre domaine. Pour déterminer un tel vocabulaire, nous avons mis en place un module permettant l'identification de descripteurs linguistiques représentatifs sur la base d'un corpus expertisé et du vocabulaire contextualisé VPOG défini en Section 3.2.2.

Pour chaque descripteur d'opinion présent dans le vocabulaire VPOG, nous comptabilisons le nombre de documents positifs *nbPos* et négatifs *nbNeg* dans lesquels il apparaît. Un premier critère est défini (cf. formule (2)) pour ne garder que les descripteurs qui sont présents de façon significative dans le corpus.

$$nbPos + nbNeg < \log(nbDocument) \tag{2}$$

Avec

- *nbDocument le nombre de documents constituant le corpus.*

Nous assignons aux descripteurs sélectionnés un score de pondération *scoreP* (cf. formule (3)) basé sur leur facteur de discrimination et sur la proportion des documents positifs et négatifs du corpus. Cela permet de prendre en compte un corpus déséquilibré en nombre de documents positifs et négatifs. Nous considérons alors qu'un descripteur est représentatif si son apparition dans un type de document est plus significative que dans un autre. Ainsi, sur la base des scores de représentativité calculés, nous supprimons les descripteurs dits ambigus, c'est-à-dire ceux répartis quasi également dans les documents positifs et négatifs (cf. formule (3)).

$$scoreP(M) = \frac{\max(nbPos,1)}{\max(nbNeg,1)} \times \frac{nbTotalDocNeg}{nbTotalDocPos} \tag{3}$$

Avec

- *M le mot analysé*
- *nbTotalDocNeg, NbTotalDocPos les nombres de documents négatifs et positifs du corpus*

A partir du score de pondération défini, un score de représentativité *scoreR* est attribué à chaque descripteur *M* en fonction des classes (positive et négative) (cf. formule (4)). Une polarité positive est allouée à un descripteur si sa valeur de

pondération est supérieure ou égale à un seuil T_r et négative si sa valeur de pondération est inférieure à $1/T_r$.

$$Si\ scoreP(M) \geq T_r\ \ Alors\ scoreR_{pos}\ = scoreP\ (M) \qquad (4)$$

$$Si\ scoreP(M) < \frac{1}{T_r}\ \ Alors\ scoreR_{neg}\ = \frac{1}{scoreP(M)}$$

Le seuil utilisé dans les expérimentations est T_r = 65/35, indiquant que l'ensemble des descripteurs, qui n'ont pas une distribution d'au moins 65% dans l'une des deux classes de documents, sont supprimés. La validation manuelle de ces descripteurs a permis de constituer un vocabulaire spécialisé VPOS. A titre d'exemple, les mots tels que « consensus », « négociation », « tradition » ont ainsi été associés au vocabulaire d'éléments positifs spécialisé du domaine de l'aménagement du territoire.

3.3. Attribution des scores d'opinion pour chaque document

Une fois les différents vocabulaires définis, il s'agit d'attribuer un score de polarité à chaque document. Au préalable, deux types de pré-traitements sont réalisés qui prennent en compte à la fois des informations statistiques (section 3.3.1) et linguistiques (section 3.3.2). Ensuite, le score d'opinion peut être calculé.

3.3.1 Pré-traitement statistique de l'approche OPITER

Les mots issus des lexiques ne sont pas nécessairement adaptés pour une analyse textuelle fine des documents. En effet, la plupart des mots très fréquents ne sont pas pertinents car trop généraux (par exemple, "faire", "dire", "avoir"). A contrario, les mots très spécifiques (par exemple, les mots des lexiques qui n'apparaissent qu'une ou deux fois dans le corpus) n' apportent pas une information utile dans l'analyse des documents.

Ainsi, ces types de mots trop généraux et trop spécifiques ont été éliminés des lexiques que nous utilisons par la suite. Concrètement, lorsque des mots sont répartis dans de nombreux textes des corpus étudiés, la confiance que l'on peut accorder à leur polarité en est affectée. Afin de mesurer la discrimination de ces mots, nous avons calculé leur *idf* (inverse document frequency) (cf. formule (5)). Ceci nous a permis de supprimer des vocabulaires définis les mots ayant une valeur de discrimination faible (*idf* inférieur à un seuil S_{idf}). Le seuil appliqué dans nos expérimentations est S_{idf}= 0.5.

$$idf_i = \log \frac{|D|}{|\{d_j : t_i \in d_j\}|} \qquad (5)$$

Avec

- $|D|$: nombre total de documents dans le corpus
- $|\{ d_j : t_i \in d_j \}|$: nombre de documents où le terme t_i apparaît

Outre les critères statistiques, trois types de pré-traitements linguistiques de l'approche OPITER sont présentés dans la section suivante.

3.3.2 Pré-traitement linguistique de l'approche OPITER

Le premier pré-traitement linguistique consiste à ne conserver dans le lexique d'opinion que les mots ayant une catégorie grammaticale donnée : adjectif, verbe, nom et adverbe. De tels mots sont reconnus pour être porteurs de sentiment et/ou d'opinion (Sokolova et al., 2008, Taboada et al., 2006). Cette sélection a été effectuée à partir des données étiquetées automatiquement avec TreeTagger (Schmid, 1994). Dans ce premier pré-traitement, nous intégrons une étape intermédiaire permettant d'identifier la pondération adéquate selon la catégorie grammaticale des mots constituant le vocabulaire obtenu.

Par ailleurs, des traitements plus fins ont été appliqués sur la base de configurations linguistiques précises. Ainsi, la négation a été prise en compte en modifiant la polarité des mots suivis d'une négation. Par exemple, dans la phrase "Ce projet n'est pas intéressant", le mot "intéressant", initialement identifié comme positif dans nos lexiques, apportera une étiquette négative compte tenu de la négation qui précède.

3.3.3 Attribution d'un score global

Au regard de l'ensemble des paramètres et prétraitements présentés, un score global de polarité est alors attribué à chaque objet textuel (mot, phrase et document). Dans un premier temps, le score des mots pivots est calculé en fonction de leur présence dans un lexique *l* d'opinion négatif ou positif (formule (6)). Cette polarité sera pondérée selon un score de fiabilité $S_{i,\,l}$ discuté en section 3.1.1.

$$scoreMot(MortPivot) = S_{i,l} \times polarite\ (MotPivot, l) \qquad (6)$$

Avec

- *polarité (MotPivot, l)* $\in \{-1, 1\}$

Le score des phrases est alors calculé en fonction des scores attribués à chaque mot (cf. formule (7)). Une pondération est attribuée aux mots selon leur catégorie grammaticale (paramètre *Cat*).

$$scorePhrase(P) = \frac{\sum Cat \times scoreMot(M)}{\sum Cat \times M} \qquad (7)$$

Enfin, le score global du document (formule (8)) est obtenu en calculant la moyenne des scores des documents.

$$scoreDocument(D) = \frac{\sum scorePhrase(Phrase)}{\sum Phrase} \qquad (8)$$

4. Expérimentations

4.1. Constitution du corpus

Nous avons sélectionné un ensemble de documents sur l'aménagement du territoire du Bassin de Thau qui a été réparti en deux classes d'opinion : positive et négative. L'opinion exprimée dans ces textes est liée à la constitution d'une nouvelle agglomération propre à ce territoire. Nous avons uniquement choisi des articles de journaux et nous avons écarté (1) les textes ambigus présentant des opinions diverses, (2) les textes ne dégageant aucune information de polarité (typiquement ceux contenant trop d'informations factuelles). L'indexation (attribution à un document d'une opinion positive ou négative) a été réalisée et validée par les experts géomaticiens de l'aménagement du territoire travaillant sur le projet SENTERRITOIRE. Nous avons ainsi constitué le corpus CSENTERRITOIRE de 99 documents classés selon leur polarité (positive ou négative).

4.2. Description des expérimentations

Nous présentons le protocole expérimental s'appuyant sur des lexiques d'opinion généralistes pour faire émerger un vocabulaire d'opinion du territoire implicitement décrit dans un corpus documentaire. Ce protocole, décrit en figure 1, se décompose en 4 étapes, exécutées de façon incrémentale. Les étapes 2 et 3 forment un processus automatisé sans expertise du corpus. La quatrième étape, quant à elle, extrait un vocabulaire spécialisé du domaine en s'appuyant sur l'expertise des géographes travaillant sur le projet SENTERRITOIRE.

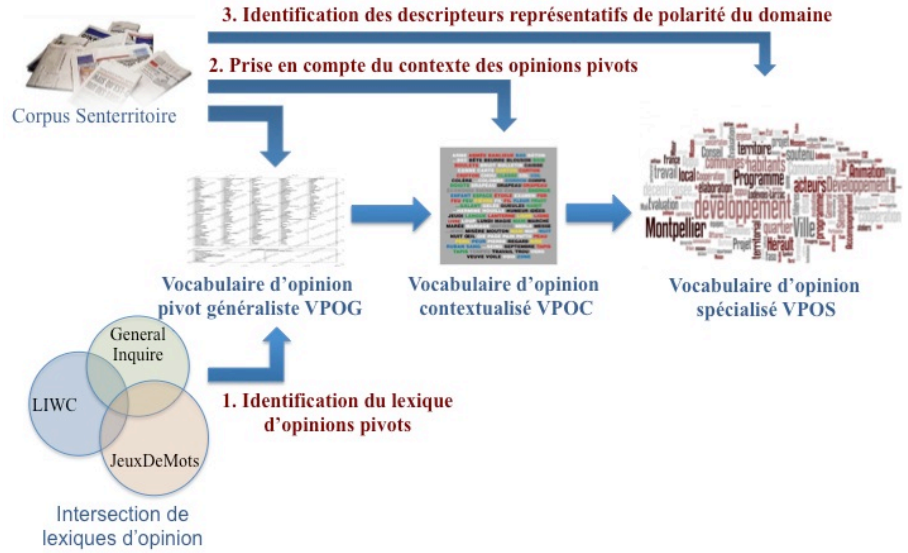


Figure 1. Méthodologie OPITER

Nous présentons maintenant les résultats expérimentaux obtenus en appliquant la démarche OPITER. A chaque étape, les résultats sont évalués en comparaison à la classification manuelle réalisée et validée par les experts du domaine.

4.2.1. Première évaluation de référence

La première étape permet d'obtenir une évaluation de référence sur la base de l'ensemble des mots pré-sélectionnés selon leur catégorie grammaticale lors des pré-traitements linguistiques. Le score de classification des documents polarisés est obtenu en utilisant les trois lexiques choisis (*GeneralInquirer*, *LIWC* et *JeuxDeMots*) (cf. tableau 1). Les différents lexiques sont testés de façon indépendante et nous ne pondérons pas le score des mots en fonction de leur catégorie grammaticale.

Tableau 1. Calcul du score global à partir des opinions pivot des lexiques

Lexique	Score de classification correcte
GeneralInquirer	**57,5%**
LIWC	54,5%
JeuxDeMots	51,5%

Nous remarquons que le lexique *GeneralInquirer*, plus complet que le *LIWC*, apporte de meilleurs résultats de marquage des opinions (57.57%). Le lexique *JeuxDeMots*, bien que complet, est moins efficace dans cette première étape et cela s'explique par le fait qu'il n'est pas défini comme un dictionnaire d'opinion. Il contient un vocabulaire relatant différents domaines qui ont été polarisés par de nombreux et différents contributeurs. Ces résultats soulignent que ces lexiques ne sont pas définis à l'origine pour décrire un domaine cible tel que celui de l'aménagement du territoire.

Nous analysons dans les sous-sections suivantes les expérimentations menées sur le corpus CSENTERRITOIRE afin de construire un vocabulaire spécialisé pour le marquage des opinions liées à notre domaine d'application, i.e. l'aménagement du territoire.

4.2.2 Construction du vocabulaire d'opinion pivot généraliste

Nous proposons tout d'abord de fusionner les trois lexiques à disposition afin d'identifier un vocabulaire généraliste nous permettant d'améliorer le marquage des opinions dans un corpus de type CSENTERRITOIRE. Dans cette seconde étape, nous comparons les sept lexiques distincts présentés en section 3.1 qui sont ensuite utilisés pour marquer les opinions dans les documents via le calcul du score de polarité. Afin d'identifier la configuration optimale, nous faisons varier de 0 à 3 l'ensemble des scores de fiabilité S_1, S_2, S_3, S_4, S_5, S_6 et S_7 et un extrait des résultats est présenté en Tableau 2 (une plus grande amplitude sur la pondération des scores, de 0 à 10, n'a pas été pertinente expérimentalement).

Tableau 2. Résultats pour la construction du vocabulaire d'opinion généraliste

Tests	GI-LIWC-JDM (S_1)	GI-LIWC (S_2)	GI-JDM (S_3)	LIWC-JDM (S_4)	GI (S_5)	LIWC (S_6)	JDM (S_7)	Score correct
1	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**52,5%**
2	1	1	1	1	0	0	0	59,6%
3	3	0	2	0	1	0	1	54,5%
4	3	0	0	2	0	1	1	54,5%
5	3	2	0	0	1	1	0	63,6%
6	**3**	**2**	**0**	**0**	**1**	**0**	**0**	**64,6%**

Nous remarquons tout d'abord que l'intégration de l'union de l'ensemble des lexiques, test 1, diminue le score de classification correcte à 52,5%. En conservant le lexique intersection des trois lexiques, nous supprimons successivement dans les tests 3, 4 et 5 les lexiques *LIWC*, *GeneralInquirer*, et *JeuxDeMots*. Nous obtenons alors un résultat de 63,6% en utilisant un lexique d'opinion généraliste ne faisant pas intervenir le lexique très étendu *JeuxDeMots*. Nous remarquons, test 6, que les scores sont meilleurs lorsqu'il s'agit (1) de ne garder que le lexique intersection des trois lexiques, le lexique intersection de *GeneralInquirer* et *LIWC* ainsi que le lexique complément *GeneralInquirer*, et (2) de pondérer les mots présents en fonction du niveau d'intersection des lexiques définis. En nous appuyant sur ce vocabulaire d'opinion généraliste VPOG ainsi construit, l'approche OPITER nous permet d'obtenir un score de classification de **64,6%.**

4.2.3. Construction du vocabulaire d'opinion pivot contextualisé

Dans une troisième étape, nous proposons d'enrichir le vocabulaire pivot d'opinion généraliste en prenant en compte le **contexte** des opinions pivots. Pour ce faire, nous attribuons un score de polarité aux mots identifiés comme voisins des mots pivots comme discuté section 3.2. Concrètement, un score de polarité est attribué à chaque mot de type nom, verbe, adjectif et adverbe, ayant un ou plusieurs voisins de pivot(s). Ce score correspond à la moyenne des scores des mots pivots présents dans une fenêtre de distance donnée. Dans les expérimentations, les tests sont réalisés avec des fenêtres de distance de 1 à 4 mots (cf. Tableau 3).

Tableau 3. Résultats pour la construction du vocabulaire d'opinion contextualisé

Fenêtre	GI-LIWC-JDM	GI-LIWC	GI-JDM	LIWC-JDM	GI	LIWC	JDM	Score correct
0	**3**	**2**	**0**	**0**	**1**	**0**	**0**	**64,6%**
1	3	2	0	0	1	0	0	64,6%

4	**3**	**2**	**0**	**0**	**1**	**0**	**0**	**65,6%**

Au regard des résultats obtenus et plus particulièrement lorsque nous étendons la prise en compte du contexte à une fenêtre de 4 mots précédents et suivants un mot cible, le vocabulaire pivot généraliste VPOG, enrichi des mots proches des mots pivots, nous permet d'améliorer l'identification de la polarité avec un score de **65,6%** de documents bien classés.

4.2.4. Construction du vocabulaire d'opinion pivot spécialisé

La quatrième et dernière étape définit un vocabulaire pivot d'opinion spécialisé à partir du vocabulaire contextualisé VPOC, et de descripteurs représentatifs identifiés par *validation croisée*. Cette méthode repose sur l'idée d'un découpage (aléatoire) des données en deux sous-échantillons. Supposons dans un premier temps ce découpage fixe, nous utilisons la première partie des données (appelée échantillon d'entraînement, de taille Ne) pour construire une estimation de la « vraie » courbe avec chacun des modèles en compétition. Ensuite, avec les données restantes (formant l'échantillon de validation, de taille Nv), on évalue la qualité de chacune des estimations en les comparant aux valeurs observées. Les évaluations ne sont donc pas réalisées sur les données ayant permis de construire les estimations.

Rappelons que nous entendons par descripteurs représentatifs un ensemble de mots permettant de distinguer des documents positifs de documents négatifs. Dans un premier temps, l'ensemble des mots du corpus de type adjectif, nom et verbe sont candidats à devenir descripteurs. Nous appliquons ensuite les deux filtres présentés section 3.2.3 (cf. formules (2) et (4)) afin de ne garder que les descripteurs représentatifs. Ce seuil nous permet d'éliminer à la fois les descripteurs trop peu présents dans le corpus et ceux qui sont présents dans autant de documents négatifs que de documents positifs. Enfin, nous considérons un descripteur comme représentatif si sa distribution dans le corpus polarisé est d'au moins 65% en faveur d'une des deux classes (positive ou négative) (cf. formule (4)). Ce seuil permet de sélectionner les descripteurs les plus discriminants. Ainsi, chaque descripteur sélectionné a un score de représentativité permettant de distinguer les deux polarités (cf. Formule (3)). Les tests réalisés sont présentés en Tableau 4.

En sélectionnant l'ensemble des 527 descripteurs obtenus par validation croisée, et en les pondérant à 1 en fonction de leur valeur de représentativité (cf Tableau 4, test a), nous obtenons un vocabulaire d'opinion nous permettant d'améliorer l'identification de la polarité des documents de façon très significative (81,8%). Rappelons ici qu'attribuer une valeur 1 signifie que le descripteur est considéré comme un descripteur provenant du vocabulaire contextualisé d'opinion. En pondérant les descripteurs avec leur score de représentativité *scoreR*, nous améliorons significativement les scores d'identification de la polarité (cf. Tableau 4, test b).

Tableau 4. Résultats pour la construction du vocabulaire d'opinion spécialisé

Tests	Score de représentativité	Score
a	1	81,8%
b	*scoreR (Mot)*	91,9%

L'ensemble de nos expérimentations nous permet d'identifier un vocabulaire spécialisé d'opinions pivots représentatif du corpus analysé. Des tests réalisés sur un deuxième ensemble de documents décrivant l'aménagement du Bassin de Thau offrent des résultats sensiblement similaires.

4.2.5 Application de pré-traitements linguistiques

Sur la base du vocabulaire d'opinion spécialisé VPOS, nous avons appliqué les règles linguistiques consistant à (i) traiter la négation, puis à pondérer les mots d'opinions du lexique généraliste selon (ii) leur polarité, et (iii) leur catégorie grammaticale. Une analyse du corpus met en exergue que 18% des phrases contiennent des tournures négatives, et qu'il y a 2,5 fois plus de mots positifs que de mots négatifs dans le corpus. Concernant la négation, un premier prétraitement permet d'inverser la polarité des mots pivots présents entre les tournures de négation (ne…pas, ne…plus, etc.), et dans une fenêtre de cinq mots à la suite de la négation. Les expérimentations menées avec prise en compte de la négation n'améliorent pas les premiers résultats obtenus. En effet, le vocabulaire spécialisé généré nous permet d'obtenir un score de classification des documents de notre corpus de 87,9% contre 91,2% sans la négation. Nous travaillons actuellement à affiner les règles de marquage des mots pivots présents dans les tournures afin d'améliorer ces premiers résultats. Ceci nécessite une analyse syntaxique plus fine des documents.

Un dernier pré-traitement consiste à pondérer les mots pivots en fonction de leur catégorie grammaticale *Cat* (cf. formule (6)) comme dans les travaux de (Wang et al., 2013). Ce dernier pré-traitement linguistique n'a pas permis d'améliorer les scores de classification même si les tests réalisés ont montré que, dans notre corpus, les mots de types nom et verbe se révèlent prédominants pour l'identification des opinions.

5. Conclusion et perspective

L'application de notre approche OPITER sur le corpus défini donne des premiers résultats très encourageants. En effet, nous remarquons qu'en intégrant un vocabulaire d'opinion généraliste défini sur la base de l'intersection de trois lexiques d'opinion de la communauté, nous améliorons de façon significative les scores d'identification de la polarité des documents (57,5% en s'appuyant uniquement sur le *GeneralInquirer* qui semble être le plus adapté contre 64,6% en utilisant le vocabulaire d'opinion pivot généraliste défini). Notons que l'ajout de descripteurs

représentatifs, identifiés par validation croisée sur notre corpus de tests, nous permet d'améliorer significativement les résultats pour arriver à un score de 91,9% de documents bien classés. Nous identifions par ce biais un vocabulaire complet spécialisé adapté pour le marquage des opinions.

Les perspectives à ce travail sont nombreuses. Tout d'abord, nous souhaitons mettre en évidence le caractère générique de l'approche proposée. Pour cela, nous devons conduire le même protocole expérimental sur d'autres jeux de données relatifs à des territoires autres que celui du Bassin de Thau. Nous souhaitons également étendre la démarche à d'autres ressources textuelles telles que les blogs, les sites web qui sont d'intérêt et retracent le ressenti en fonction des catégories d'acteurs engagés dans le processus de planification territoriale. Au niveau du projet SENTERRITOIRE, l'étape suivante consiste à travailler sur l'identification et l'extraction de relations entre les entités spatiales extraites (Tahrat et al., 2013) et le vocabulaire spécialisé d'opinion pivot généré via l'approche OPITER. Enfin, les évolutions temporelles des différentes opinions de chacun des acteurs sont une information précieuse qu'il faut être en mesure de restituer.

Remerciements

Ces travaux sont en partie financés par le labex NUMEV et la Maison des Sciences de l'Homme de Montpellier (MSH-M). Nous en profitons également pour remercier Cédric Lopez[5] et Bernard Laval[6] pour leur contribution dans les travaux présentés.

[5] Chercheur à l'entreprise VISEO, http://www.viseo.net/cedric-lopez
[6] Etudiant en master 2 spécialité IPS, Université Montpellier 2

Bibliographie

Alliès P. (1980). *L'Invention du territoire*, Grenoble : Presses universitaires de Grenoble, 184 pages.

Barel, Y. (1981). Modernité, code, territoire*, Annales de la recherche urbaine*, 10-11, pp.3-21.

Buléon, P. et Di Méo, G. (2005). *L'espace social*. Armand Colin.

Debarbieux B. et Vanier M. (2002) *Ces territorialités qui se dessinent*, Editions de l'Aube, Datar, 267 pages.

Deffontaines J.-P., Marcelpoil E. et Moquay P. (2001). *Le développement territorial : une diversité d'interprétations*, Dans Maurel, P. Lardon, S., Piveteau, V. (Eds), Représentations spatiales et développement territorial, Paris, Hermès Sciences Publications.

Deleuze G. et Guattari F. (1980). *Mille plateaux*, Editions de Minuit. coll. Capitalisme et schizophrénie. 645 pages

Di Méo G. (1998) *Géographie sociale et territoires*, Belin ,Paris.

Duthil B., Trousset F., Roche M., Dray G., Plantié M., Montmain J. and Poncelet P. (2011). Towards an Automatic Characterization of Criteria, *Proceedings of DEXA'11 (International Conference on Database and Expert Systems Applications)*, Vol 1, Springer-Verlag, LNCS, p. 457-465, Toulouse, France.

El-Bèze M, Jackiewicz A., Hunston S. (2010). *Traitement Automatique des Langues*, Volume 51 Numéro 3.

Esuli A. et Sebastiani F. (2006). *SentiWordNet:* A Publicly Available Lexical Resource for Opinion Mining*. In 5th Conference on Language Resources and Evaluation*, pp. 417-422.

Fan W., Sun S. and Song G. (2011). Sentiment Classification for Chinese Netnews Comments Based on Multiple Classifiers Integration, *In the Int. Joint Conf. on Comp. Sciences and Optimization*, pp. 829-834,.

Harb A., Plantié M., Roche M., Dray G., Trousset F. and Poncelet P. (2008). Détection d'opinion. Comment déterminer les adjectifs d'opinion d'un domaine donné, *Revue Documents Numériques*, No 1-2, Vol 11, pp. 37-61.

Joshi A., Balamurali P., Bhattacharyya P., and Mohanty R. (2011). C-feel-it: a sentiment analyzer for microblogs, *In Proceedings of HLT*, pp. 127-132.

Kennedy A. and Inkpen D. (2006). Sentiment Classification of Movie Reviews Using Contextual Valence Shifters, *Computational Intelligence,* 22(2), pp.110-125.

Kim S-M and Hovy E. (2004). Determining the sentiment of opinions, *In Proceedings of COLING*, pp. 1367-1373,

Lafourcade, M. (2007). Making people play for Lexical Acquisition. *In Proc. SNLP 2007, 7th Symposium on Natural Language Processing*. Pattaya, Thailande, 8 p.

Pak A., Paroubek P. (2010). Microblogging for Micro Sentiment Analysis and Opinion Mining, *TAL* 51(3), 75-100

Pang B. and Lee L. (2008). Opinion mining and sentiment analysis, *Found. and Trends in IR*., 2(1-2), pp. 1-135.

Pecqueur B. (2007). L'économie territoriale : une autre analyse de la globalisation, *Alternatives économiques, l'économie politique* (1) 33, pp. 41-52.

Piolat A., Booth, R. J., Chung C. K., Davids M. and Pennebaker J. W. (2011). La version française du dictionnaire pour le LIWC : modalités de construction et exemples d'utilisation. *Psychologie Française*, 56(3) :145-159.

Plantié M., Roche M., Dray G., and Poncelet P. (2008). Is a Voting Approach Accurate for Opinion Mining?, *In proceedings of DaWaK 08 (Int. Conf. on Datawarehousing and Knwoledge Discovery)*, p.413-422.

Roche M. and Poncelet P. (2009). Fouille de Données d'Opinions, *Numéro spécial RNTI (Revue des Nouvelles Technologies de l'Information)*, 202 pages, volume E-17.

Schmid, H. (1994). *Probabilistic Part-of-Speech Tagging Using Decision Trees*.

Sokolova, M. and Lapalme, G. (2008). Verbs speak loud : Verb categories in learning polarity and strength of opinions. *In Conference of the Canadian Society for Computational Studies of Intelligence*, pages 320-331.

Taboada, M., Anthony, C., and Voll, K. (2006). Creating semantic orientation dictionaries. *In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*.

Tahrat S., Kergosien E., Bringay S., Roche M., Teisseire M. (2013). Text2Geo : des données textuelles aux informations géospatiales, *Actes des 13ièmes journées d'Extraction et Gestion des Connaissances, EGC 2013*, Toulouse, pp. 407-412.

Tahrat S., Kergosien E., Bringay S., Roche M., Teisseire M. (2013). Text2Geo: from textual data to geospatial information. *In the 3rd International Conference on Web Intelligence, Mining and Semantics (WIMS)*.

Tesson F. (2004). Quand le politique parle du territoire, c'est pour le fabriquer : utilité scientifique et vertus pédagogiques de la géographie sociale pour comprendre la relation du politique à l'espace, *Espaces et sociétés d'aujourd'hui*, Rennes.

Torres-Moreno, J.-M., M. El-Bèze, F. Béchet, et N. Camelin (2007). Comment faire pour que l'opinion forgée à la sortie des urnes soit la bonne ? application au défi DEFT 2007. *In Actes de l'atelier de clôture du 3ème défi fouille de textes*, Grenoble, France.

Turney P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, *In Proceedings of ACL*, pp. 417-424.

Vanier M. (1995). La petite fabrique de territoires en Rhône-Alpes : acteurs, mythes et pratiques, *Revue de géographie de Lyon*. Vol. 70 n°2, 1995. Les nouvelles mailles du pouvoir local. pp. 93-103.

Vanier, M. (2009). *Territoires, territorialité, territorialisation - Controverses et perspectives*, PUR.

Vincze, N. et Bestgen, Y. (2011) *Identification de mots germes pour la construction d'un lexique de valence au moyen d'une procédure supervisée*.

Wiebe J. and Riloff E. (2011). Finding Mutual Benefit between Subjectivity Analysis and Information Extraction, *IEEE Transactions on Affective Computing*, 2 (4): 175-191.

Usage de modèles dynamiques et spatialement explicites pour la spatialisation de scénarios prospectifs normatifs.

Application à la dynamique des plantations forestières industrielles dans le sud chilien.

Nicolas Maestripieri[1], Thomas Houet[1], Martin Paegelow[1]

1. Laboratoire GEODE, UMR 5602 CNRS, Université de Toulouse 2 Le Mirail
5 allées A. Machado, 31058 Toulouse cedex 9
nicolas.maestripieri@univ-tlse2.fr

RESUME. *Caractéristiques de la géoprospective, la simulation et la spatialisation des scénarios prospectifs permettent aux sociétés d'anticiper et de réduire leurs incertitudes face aux changements futurs d'occupation et d'utilisation du sol. Malgré une diversité d'approches et de techniques, la construction des scénarios prospectifs suit deux cheminements temporels (exploratoire et normatif). Cet article s'intéresse spécifiquement à la spatialisation des scénarios normatifs et à la pertinence d'une approche exploratoire (dynamique) face à l'application systématique d'une démarche rétrospective (synchronique).Ces scénarios sont appliqués dans un contexte de gestion forestière dans le sud chilien. Ils mettent en évidence deux images contrastées du futur, avec, d'un côté, la préservation de la forêt native et, de l'autre, l'invasion de plantations industrielles exotiques.*

ABSTRACT. *Simulation and spatialization of prospective scenarios are characteristics of geoprospective. They allow societies to anticipate and to reduce their uncertainties regarding future land use and land cover changes. Despite a diversity of approaches and techniques, the construction of scenarios follows two temporal courses (exploratory and normative).This paper specifically focuses on the spatialization of normative scenarios and the relevance of exploratory (dynamic) approach faced with the systematic use of retrospective (synchronic) process. These scenarios are applied in the framework of forest management in southern Chile. They highlight two contrasting images of the future, with, on one hand, the preservation of native forest and, on the other hand, invasion of exotic timber plantations.*

MOTS CLES : *Géoprospective, scénarios normatifs, modèle spatialement explicite, approche exploratoire, sud chilien*

KEYWORDS : *Géoprospective, normative scenario, spatially explicit models, exploratory approach, southern Chile*

1. Introduction

Une prise de décision, quelle soit d'ordre privé ou public, se confronte au niveau de connaissance acquise et à la confiance que nous lui accordons (Sigel *et al.*, 2010). Cet état détermine le degré d'incertitude qu'un individu ou une société dispose et mobilise pour, *a posteriori*, faire face aux problématiques environnementales et socio-économiques (aléas climatiques, changements d'occupation et d'utilisation du sol, crise économique, rendements agricoles, etc.). L'incertitude amène la société à chercher des points de repères et à se projeter dans le futur pour mieux maîtriser son présent et optimiser ses choix/stratégies pour l'avenir. De fait, l'appréhension de futurs possibles (Phdungsilp, 2011), plausibles (Wilkinson, 2009 ; Amer *et al.*, 2013) ou désirables (Godet, 2010) permet de mieux faire face à des événements incertains (extrêmes ou non) dans la mesure où chaque futur, une fois caractérisé, pourra s'accompagner de mesures d'adaptation ciblées. Il est question de plusieurs futurs -appelés *scenario development* (Bishop, 2007), *strategic foresight* ou prospective (Godet, 2010)-, et non d'un futur probable -équivalent aux *foresight studies*- où le degré d'incertitude est faible. L'élaboration de scénarios s'est généralisée à travers le monde afin de s'adapter à des phénomènes complexes et incertains. Il existe trois grandes écoles de construction de scénarios. Les deux premières sont anglo-saxones avec l'approche de la logique intuitive proposée par Kahn dans les années 1950, et la *probabilistic modified trends (PMT) school* développée par Helmer et Gordon en 1966. La prospective, en tant que spécificité française (Roubelat, 1994), remonte aux années 1950 avec la création par Gaston Berger du Centre International de prospective et du World Futures Studies Federation par Bertrand de Jouvenel dans les années 1960. Cette approche permet de mettre en lumière des enjeux environnementaux et socio-économiques à de multiples échelles spatiales. Elle considère l'avenir comme un « espace » d'action, exploratoire ou normatif, qui se bâtit par et pour un individu (ou des groupes d'individus). Or, si ces scénarios peuvent s'appliquer à plusieurs échelles spatiales, ils ne permettent pas de modéliser et de localiser ces processus. De fait, la géoprospective (Gourmelon *et al.*, 2012 ; Voiron-Canicio, 2012 ; Houet et Gourmelon, 2013) donne la capacité aux sociétés de s'organiser et de s'adapter face à un environnement complexe et dynamique.

L'objectif de cet article est de démontrer la pertinence et la performance d'un modèle dynamique et spatialement explicite, suivant un cheminement temporel exploratoire (contrairement à une approche synchronique), pour spatialiser des scénarios prospectifs normatifs. L'utilisation de chaînes markoviennes couplées à une évaluation multicritères (EMC) permettra de mener à terme notre raisonnement. A la suite de cette introduction, la section 2 se penche sur la typologie des scénarios et leur méthode de construction. L'aire d'étude, le contexte de la recherche et la méthode de spatialisation utilisées sont exposées dans une troisième section. Les résultats sont présentés dans la section 4, puis sont comparés et discutés dans une dernière section.

2. De la construction des scénarios à leur spatialisation

2.1. Typologie des scénarios : exhaustivité et confusion sémantique

La littérature scientifique traitant des scénarios rend complexe l'élaboration d'une typologie claire et univoque. De nombreux auteurs expriment cette confusion sémantique : Bishop *et al.* (2007) et van Notten et Rotmans (2003) remarquent que la catégorisation des scénarios est large et qu'une variété de scénarios est souvent regroupée au sein d'une même catégorie. Certains parlent même de « *methodological confusion* » (Wilkinson, 2009) voire de « *methodological chaos* » (Stewart, 2008). L'élaboration de typologies synthétiques des différentes écoles (ou approches), et surtout des méthodes et techniques, expliquent en partie ce « chaos » méthodologique. Ces synthèses proposent généralement un état de l'art exhaustif pour *in fine* présenter une typologie issue du croisement des différents auteurs. De fait, plus le spectre sémantique s'élargit et s'approfondit plus les confusions se multiplient. Ces confusions doivent être résolues afin de créer une cohésion au sein de la communauté scientifique et d'amener, auprès du grand public, les valeurs issues du développement et de l'analyse des scénarios (Bishop *et al.*, 2007).

Fabrice Hatem (1993) distingue quatre approches en *opposition* pour construire un scénario :

- *Forecasting vs backcasting.* Börjeson *et al.* (2006)[1] fait respectivement référence aux scénarios exploratoires (avec deux sous-catégories – *external* et *strategic*) et aux scénarios normatifs (*transforming* et *backcasting*). La typologie de van Notten *et al.* (2003) distingue les scénarios descriptifs des scénarios normatifs, et affine son analyse selon le cheminement temporel : exploratory ou *forecasting* scenarios vs anticipatory ou *backcasting* scenarios.

- variantes vs scénarios synthétiques,

- scénarios « sans surprise » vs contrastés [ou tendanciels vs contrastés (Masini et Médina Vasquez, 2000)]

- approche littéraire vs formalisée [ou qualitative vs quantitative (European Environment Agency, 2000)].

Si la typologie des scénarios prospectifs ne fait pas consensus, leur méthode de construction, bien que variable selon les auteurs, s'appuie :

> « *sur une base méthodologique commune (ou « méta-méthode ») qui a été synthétisée sous l'impulsion de Godet (1992) et dénommée « la méthode des scénarios* » » (Houet *et al.*, 2008).

[1] Les scénarios ne sont pas des prédictions (Wilkinson, 2009; Amer *et al.*, 2013) mais Börjeson *et al.* (2006) inclut dans sa typologie des scénarios un volet *predictive* qui s'oppose à la prospective d'une part, car il ne s'intéresse pas à l'appréhension des multiples futurs et à leurs incertitudes, et d'autre par parce que le raisonnement porte sur le court terme. D'après Medina Vásquez et Ortegón (2006) l'objectif de la prédiction au sens strict du terme est d'aboutir à une déclaration non probabiliste avec un degré de confiance absolue sur le futur.

2.2. La Prospective et la méthode des scénarios

Pour Godet (1986) la méthode des scénarios fait consensus pour appréhender les futurs possibles (*les futuribles*). Cependant, ces derniers ne sont pas le seul intérêt de la prospective car ils mettent de côté les alternatives stratégiques ou les futurs souhaitables (*futurables*) (Godet, 2010). La méthode générale de construction de scénarios de Godet se décline en deux grandes phases : la construction de la « base » et la construction des scénarios (Hatem, 1993).

Construire la « base » consiste dans un premier temps à définir le système, c'est-à-dire à identifier les principales variables (économiques, politiques, sociales, techniques) qui le composent. Il y a les variables externes (caractérisant l'environnement du système) et les variables internes (caractérisant le système étudié *stricto sensu*). Le prospectiviste doit ensuite analyser les relations causales entre chacune des variables et déterminer celles qui sont susceptibles d'évoluer dans le futur avec certitude et incertitude. Enfin, la dynamique du système doit être étudiée, en menant (i) une analyse rétrospective afin de comprendre les dynamiques passées et d'identifier les acteurs et les variables qui ont joué un rôle ; et (ii) une analyse du « jeu des acteurs » pour mettre en exergue leurs stratégies, leurs interrelations et la génération possible de conflits ou d'alliances. Une fois que l'image de l'état actuel du système est établie, la construction des scénarios est mise en œuvre.

Cette phase passe tout d'abord par la considération et l'élaboration de dimensions d'incertitudes (des risques de ruptures, des signaux faibles de changements, etc.). La dimension d'incertitude est définie par Hatem (1993) comme :

> *« une incertitude considérée comme décisive pour l'avenir du système, et dont les réalisations possibles seront utilisées comme hypothèses de base pour la construction des scénarios ».*

Chaque élément sera constitué en suite logique qui permettra de lier le présent au futur (*horizon temporel*).

2.3. Cheminement temporel : backcasting vs forecasting

En faisant abstraction du large spectre sémantique susmentionné, la géoprospective (et la recherche scientifique en général) construit des projets et cherche à atteindre des objectifs spécifiques qui renvoient aux grands principes de l'inférence scientifique : la déduction, l'induction et l'abduction. Ces inférences correspondent au cheminement temporel des prédictions et des scénarios.

La déduction, qui appartient au monde de la rationalité et de l'exactitude (Blecic, 2008), est caractéristique de la prédiction. Si la conclusion d'une inférence déductive découle des prémisses alors une simulation prédictive sera la résultante d'observations passées. L'optimisation de ces résultats passe par une calibration rigoureuse du modèle, en cherchant à reproduire une situation connue (passé ou présent) (Batty et Torrens, 2005).

Un scénario n'étant pas une prédiction, notre intérêt porte sur l'inférence abductive et inductive (figure 1). La première, qui définit les scénarios prospectifs exploratoires (*forecasting*), est proche de la déduction mais se différencie de par sa

nature incertaine. Des hypothèses plausibles sont construites à partir d'observations et sont susceptibles d'être vérifiées *a posteriori* (Catellin, 2004). Le scénario part d'une situation présente, connue, et progresse vers le futur en cherchant à répondre à la question : que pourrait-il arriver ?

L'induction ou homo-constructivisme (Wilkinson, 2009) correspond au scénario normatif (*backcasting*) qui décrit un futur préférable, désirable ou non, et remonte de manière rétrospective jusqu'au présent. L'attitude est proactive (Godet, 2010) et cherche à atteindre des objectifs prédéfinis par le biais de planification stratégique. Ainsi, la question qui se pose alors au prospectiviste est : comment une cible spécifique peut être atteinte ?

Inférence	Abductif		Inductif	
Méthode de scénario	Exploratoire (*forecasting*)		Normatif (*backcasting*)	
Cheminement temporel	★ FORECASTING		BACKCASTING ★	
	Présent	Futur	Présent	Futur
Méthode de spatialisation	Exploratoire		Rétrospectif	
			Hypothèse ← Image du futur	
	Hypothèse → Image du futur		SIG	
			Exploratoire	
	Chaîne markovienne et évaluation multicritères		Hypothèse → Image du futur	
			Chaîne markovienne et évaluation multicritères	

Figure 1. Inférence, cheminement temporel et spatialisation des scénarios prospectifs

2.4. Spatialisation des scénarios: la place des modèles

Une fois le scénario définit, l'objectif du modélisateur est de retranscrire spatialement l'ensemble des hypothèses, en simplifiant les composantes du système étudié, sans pour autant perdre toute complexité. Plusieurs approches de modélisation existent et leur diversité dépend de la grande variété des questions de recherche. Verburg *et al.* (2006) présentent plusieurs types de modèles :

- dynamiques et statiques (mise à jour des variables modélisées au cours du temps vs pas de modification des données introduites au cours de la simulation) ;

- descriptifs et normatifs (simulation exploratoire basée sur les dynamiques passées vs optimisation des configurations spatiales en fonction d'objectifs) ;

- déductifs et inductifs (identification théorique des relations entre processus et facteurs vs corrélation statistique entre les changements d'occupation du sol et les variables explicatives).

Les scénarios exploratoires sont généralement dynamiques et peuvent se baser sur une approche quantitative et qualitative. L'intérêt est de pouvoir simuler des changements d'occupation du sol tout en paramétrant les probabilités de transitions entre plusieurs états, ce que permettent les modèles stochastiques et ceux à base d'intelligence artificielle. Cependant, l'approche quantitative est parfois critiquée de par son incapacité à se détacher des données historiques, et parce qu'elle laisse supposer que les mêmes tendances prévaudront dans le futur (Amer *et al.*, 2013). Or,

il est possible de mettre en œuvre des scénarios exploratoires contrastés, afin de créer une rupture avec les tendances passées.

Les scénarios normatifs sont dans l'ensemble narratifs, qualitatifs et la représentation cartographique des hypothèses se fait de façon synchronique. Par exemple, dans le cadre du Schéma d'Aménagement et de Gestion des Eaux (SAGE) en Bretagne, Houet *el al.* (2008) ont utilisé un Système d'Information Géographique afin de spatialiser un scénario normatif. D'après les auteurs les SIG, en plus de faciliter la transmission d'information avec les parties prenantes et les gestionnaires, permettent de localiser et représenter les changements d'occupations du sol et les structures paysagères à l'aide de requêtes multicritères, attributaires et/ou spatiales.

La section suivante montre que les scénarios normatifs (*backcasting*) peuvent être spatialisés en suivant la même approche que celle utilisée pour les scénarios exploratoires.

3. Matériel et méthode

3.1. Site d'étude et contexte

La forêt native, intégrée à l'Ecorégion Valdivienne (35°S–48°S), fait partie de ces écosystèmes ayant la plus haute priorité de conservation dans le monde. L'application du Decreto-Ley 701 (D.L. 701) en 1974, sous la dictature d'Augusto Pinochet, généra l'expansion de monocultures de pins et d'eucalyptus de la 8ème Région du Bio-Bio au sud de la 10ème de Los Lagos (figure 2) mettant en exergue un gradient nord-sud. L'intensification des pratiques sylvicoles, combinée à la substitution des écosystèmes forestiers natifs, soulève de nombreuses inquiétudes d'ordre environnemental et socio-économique. Ainsi, aux problèmes de compaction des sols, de perte de fertilité, d'acidification des sols et de l'eau, de perte de biodiversité et d'augmentation des flux de sédiments suite aux coupes rases, s'ajoutent des inquiétudes en matière de développement local, de chômage, de déscolarisation et de pauvreté.

Ces pratiques ont cours, depuis le début des années 1990, dans la commune de San Juan de la Costa. Elle se localise au nord-ouest de la 10ème Région de Los Lagos (figure 2), entre les parallèles 73°47'30"–73°18'50" (longitude ouest) et 40°14'14"–40°44'54" (latitude sud). Bien que le taux de plantations soit nettement inférieur à celui des communes de la 8ème et 9ème Région, les équilibres écologiques y sont menacés. En effet, plus de 70% de la superficie communale est couverte par de la forêt native. De plus, la question du foncier prend une importance considérable. Cette commune assiste à un rachat progressif par de grands propriétaires terriens et des entreprises forestières engendrant une concentration et une capitalisation des terres. Les plantations exotiques sont essentiellement exploitées par quatre grandes entreprises : la Forestal Tornagaleones S.A. (Groupe Nueva), la Forestal Anchile Ltda. (Daio Paper Corporation), la Forestal Valdivia S.A. et la Forestal Los Lagos S.A. (Groupe Angelini). Elles se partagent plus de 14 000 hectares alors que les 880 petits propriétaires vivent sur 7 430 hectares de terres morcelées.

La question foncière, l'absence de projet communal, de développement territorial et l'incitation au boisement via le D.L. 701 dégradent les conditions de vie des petits propriétaires et des communautés autochtones Mapuche. La commune est

qualifiée comme l'une des plus pauvres du Chili et affiche l'Indice de Développement Humain le plus faible en 2003 (0,510) (PNUD, 2004).

L'objectif est de sensibiliser et d'aider à la prise de conscience le grand public, les décideurs et les acteurs du secteur forestier, en leur donnant la possibilité de se projeter dans le futur, et ainsi de mieux comprendre l'impact des prises de décisions actuelles sur l'évolution du couvert forestier.

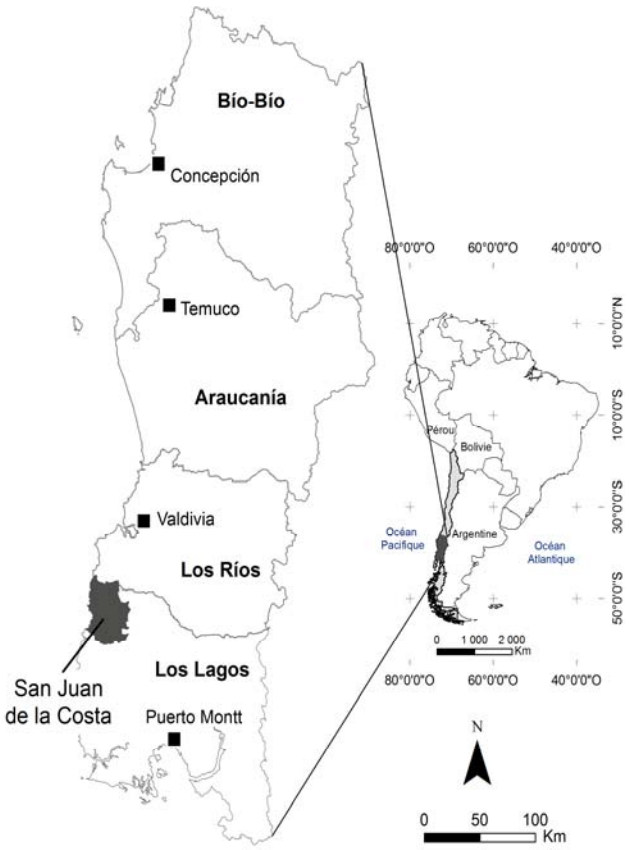


Figure 2. La commune de San Juan de la Costa (10ème Région de Los Lagos - Chili)

3.2. Approche méthodologique

3.2.1. Les trajectoires d'évolution et les facteurs de changements

La compréhension du système forestier dans le sud chilien passe au préalable par le suivi des plantations forestières industrielles et de la forêt native à plusieurs échelles spatio-temporelles. Bien que les scénarios soient construits à l'échelle communale, il est nécessaire de situer ces dynamiques locales dans un contexte régional et national, impliquant de fait des images satellites à moyennes (MODIS) (Paegelow *et al.*, 2012) et haute résolution spatiale (LANDSAT). En effet, les grandes tendances d'évolution observables à l'échelle régionale ne permettent pas d'appréhender la multiplicité et la complexité des variables qui influencent la

dynamique de ces monocultures (propriété foncière, topographie, qualité et aptitude du sol, habitat, occupation du sol, réseau routier, etc.) à l'échelle locale.

Les principaux résultats obtenus révèlent une concentration remarquable des plantations industrielles à l'ouest de la route Panaméricaine et à proximité de la côte Pacifique par la présence de ports d'embarcations. L'expansion des monocultures suit également un gradient nord-sud principalement sur la précordillère côtière, avec un rythme de changement soutenu mis en valeur par l'analyse des rotations sylvicoles à l'échelle sous-régionale.

A l'échelle de la commune de San Juan de la Costa, la dynamique des plantations industrielles résulte de processus multiformes opérant à plusieurs échelles. Ces processus sont alimentés par l'interrelation de facteurs de nature économique, sociale, environnementale, technique et politique. Une analyse systémique couplée à des entretiens participatifs a permis d'identifier et de hiérarchiser ces facteurs. Entre autres, la promulgation du D.L. 701 a engendré l'explosion du secteur forestier et s'est accompagné, lors de son actualisation en 1998, de la mise en place de mécanismes de financements destinés aux petits propriétaires et d'assurances forestières. Sur un plan économique, la faible rentabilité de la forêt native en tant que ressource commercialisable, les coûts de boisement, le coût du foncier jouent aussi un rôle important, dans la mesure où ils influent (directement ou indirectement) sur la vente et/ou la location de petites et moyennes propriétés par des entreprises forestières.

3.2.2. Les scénarios

Deux scénarios normatifs et contrastés (état futur du système) sont présentés et mettent en exergue des situations extrêmes, faites de ruptures, avec un degré de probabilité faible. Chacune des hypothèses est élaborée sur plusieurs critères : (i) la connaissance du terrain et du secteur forestier chilien, (ii) des entretiens participatifs avec les acteurs du secteur forestier, (iii) des analyses d'experts, (iv) des discussions informelles ponctuées de rumeurs et (v) d'imagination tempérée. Si les scénarios sont construits en partie avec les acteurs, l'étape de spatialisation est réalisée exclusivement par le scientifique.

Le scénario normatif « éco-centré » est un scénario qualifié de préservation. Le système forestier tel qu'il existait en 2008 est révolu. Une récession économique à l'échelle internationale et nationale engendre une chute du secteur forestier marquée par une explosion des prix des produits forestiers. Les très grands propriétaires terriens (entreprises forestières) vendent leurs terres qui seront redistribuées aux communautés autochtones Mapuches. Les zones boisées par des espèces natives s'accroissent sur l'ensemble de la commune et sont, pour la grande majorité, soumises à une gestion soutenable. Seuls des fragments de plantations exotiques occupent la précordillère côtière. Dans ce scénario, l'horizon temporel est fixé à 2035 afin de prendre en compte la croissance des espèces végétales arborées natives (30 ans pour le .*Nothofagus dombeyi*).

Le scénario normatif intensif. Le secteur forestier à l'échelle internationale est en plein boom et les politiques forestières (D.L. 701) incitent les propriétaires (du petit propriétaire à la multinationale) à intensifier leur production. La 10ème Région de Los Lagos voit la mise en service de la Ruta Costera et la construction d'une nouvelle usine de cellulose. La Ruta Costera est un projet qui permettrait, selon le Ministère des Travaux Publics (MOP), de générer d'importants avantages sociaux,

tels que l'intégration des zones isolées, des temps de déplacement réduit et le développement de nouvelles zones d'attraction touristique. La route est un agent potentiel de déforestation sur le piémont occidental de la Cordillère de la Côte (Maestripieri et Paegelow, 2013). De plus, elle joue un rôle majeur dans les dynamiques d'occupation du sol et notamment sur l'implantation des monocultures. En d'autres termes, son absence est aussi une des principales raisons pour laquelle (i) le couvert végétal natif reste relativement peu exploité et dégradé et (ii) les plantations industrielles ne progressent pas sur la Cordillère de la Côte.

Dans ce scénario, la surface attribuée aux plantations explose littéralement alors que la forêt native est dégradée systématiquement et n'évolue pas. L'horizon temporel est fixé à 2026, ce qui correspond à deux cycles de rotations d'eucalyptus à partir de la dernière date connue (2008).

3.2.3. Spatialisation des scénarios

Si les scénarios normatifs sont d'inférence inductive, la façon de restituer le scénario répond à une inférence abductive. L'ensemble des scénarios sont modélisés de façon exploratoire (du présent vers le futur) avec CA-Markov, suite à une phase de calibration du modèle (Maestripieri et Paegelow, 2013).

CA-Markov est un modèle prédictif (ici utilisé à des fins prospectives) combinant chaînes de Markov (quantité), évaluation multicritère (EMC) (localisation) et filtrage, une procédure qualifiée par les éditeurs du logiciel d'automate cellulaire (AC) (Eastman, 2006). Les quantités de changements futurs sont déterminées par les chaînes markoviennes par la production de matrices de transition et un ensemble d'images de probabilité conditionnelle issues d'observations passées. La spatialisation des hypothèses s'ajuste à la disponibilité des données, et s'appuie sur l'évaluation multicritère pour pondérer les facteurs. La variation du poids des facteurs, de même que la détermination surfacique, va dépendre de l'interprétation des hypothèses. L'objectif de l'EMC est de construire des cartes d'aptitude ou de probabilité en intégrant un ensemble de critères mesurables et cartographiables. Ces cartes pouvant ainsi servir à l'élaboration de choix stratégiques d'intervention. Le modélisateur contrôle le processus en identifiant et en caractérisant les facteurs explicatifs grâce à une approche experte.

La pondération des facteurs s'est appuyée sur la technique de comparaison par paires (*pairwise comparaison*) dans le contexte de processus décisionnel appelé Analytical Hierarchy Process (AHP) (Saaty, 1977). Chaque type d'occupation du sol (plantations industrielles, forêt native et autre) correspond à un objectif et chaque facteur se voit attribuer un poids allant de 0 (aptitude nulle) à 255 (aptitude forte). Ces derniers traduisent l'importance d'un facteur par rapport à tous les autres et contrôlent comment les facteurs se compensent entre eux (Eastman, 2006). Le poids relatif de chaque facteur est calculé par la méthode des vecteurs propres (*eigenvalue method*), « *indiquant le niveau de priorité pour chaque élément dans la hiérarchie* » (Wolfslehner *et al.*, 2005). Enfin, une pondération selon les poids d'ordre est appliquée. Cette technique permet un double choix – celui de la compensation entre facteurs et la prise de risque – et donne, par conséquent, le contrôle sur le degré d'optimisme (attitude face au risque)[2]. La pondération s'applique différemment à

[2] L'attitude face au risque s'échelonne de 1 (prise de risque nulle et pas de compensation) à 5 (prise de risque maximale et pas de compensation), en passant par une prise de risque

chaque pixel selon le classement des aptitudes, de la plus faible (1er rang) à la plus forte (dernier rang) (Paegelow et Camacho Olmedo, 2005).

4. Résultats

4.1. L'approche exploratoire

4.1.1. Le scénario de préservation

Ce scénario (figure 3) exclut la création de nouvelles plantations exotiques ainsi que le reboisement des superficies existantes après la coupe. Les entreprises forestières sont inaptes à l'apparition de plantations (aptitude nulle) et les autres propriétaires ont une aptitude très faible (75 pour les grands propriétaires et 20 pour les moyens et petits). La stratégie « pessimiste » est choisie pour l'objectif PLANTATION. Une superficie de 5 400 ha est attribuée aux plantations en 2035. La forêt native n'évolue pas à l'intérieur des (très) grandes propriétés (aptitude de 25), mais peut apparaître chez les moyens et petits propriétaires (aptitude de 255). L'objectif FORET NATIVE opte pour une stratégie « optimiste » (avec une prise de risque élevée et peu de compensation) en donnant un poids d'ordre plus important aux pixels ayant une valeur d'aptitude élevée, telles que les propriétés et la distance aux forêts native existantes. Enfin, l'objectif AUTRE[3] intègre les propriétés et donne une aptitude élevée aux entreprises forestières afin de favoriser l'apparition de zones non-boisées. La stratégie n°4 est appliquée pour donner plus de poids aux facteurs propriétés.

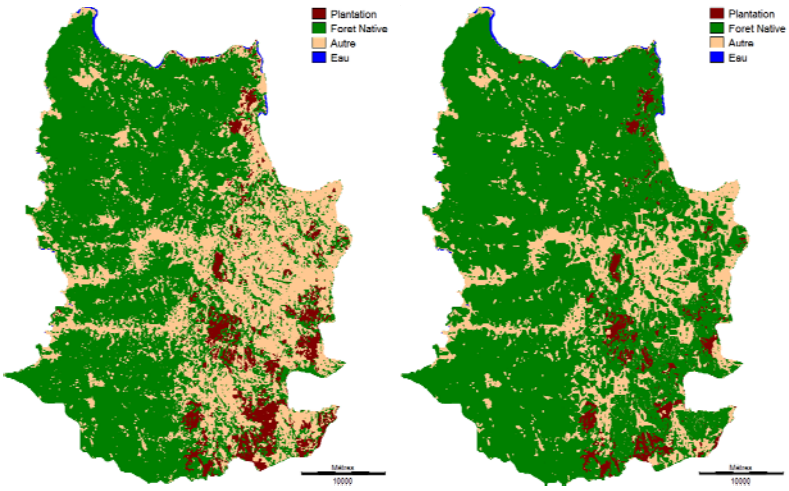


Figure 3 : Occupation du sol en 2008 (à gauche) et scénario normatif de préservation en 2035 (à droite)

moyenne et une compensation totale entre les facteurs (entre le ET et le OU booléen) (Paegelow et Camacho Olmedo, 2005; Makropoulos et Butler, 2006).

[3] Afin de simplifier le processus de simulation prospective et de mettre en évidence la dynamique des plantations industrielles, les entités « forêts de seconde génération, prairies, buissons et sols nus » sont fusionnées sous l'entité « Autre ».

4.1.2. Le scénario de production intensive

Dans ce scénario (figure 4) l'objectif PLANTATION donne une aptitude maximale à tous les types de propriétés et opte pour une stratégie optimiste, c'est-à-dire en donnant un poids d'ordre aux pixels ayant une valeur d'aptitude élevée. La détermination surfacique se fonde sur le taux de variation annuelle moyenne des plantations entre 1999 et 2008 (9,98 %/an), variation qui sert à projeter leur superficie 18 ans plus tard (2026). Les plantations exotiques atteignent 54 973 ha en 2026. La même opération est effectuée en intégrant le facteur « Ruta Costera » dans le modèle et en lui affectant un poids très élevé. A titre d'exemple, les propriétés ont un vecteur propre de 0,27 alors que celui de la Ruta Costera est de 0,19. L'objectif FORET NATIVE attribue une aptitude plus faible à la catégorie d'occupation du sol « autre », réduisant la probabilité d'y voir apparaître de la forêt native.

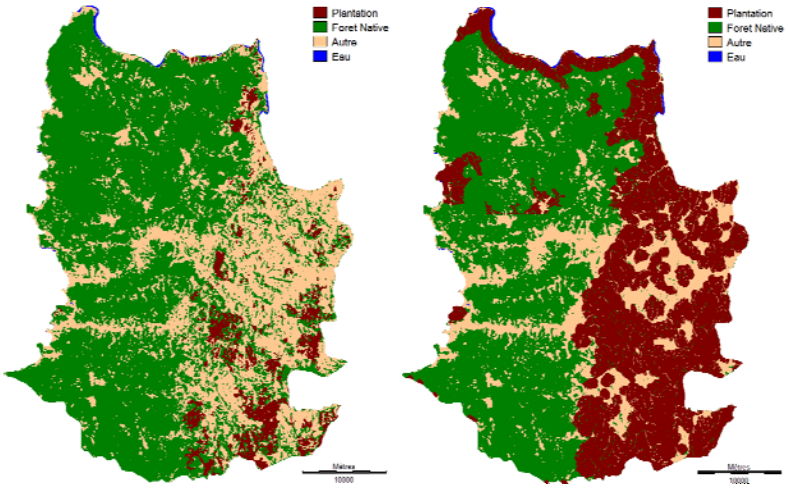


Figure 4. Occupation du sol en 2008 (à gauche) et scénario normatif intensif en 2026 (à droite)

4.2. Analyse des scénarios

Le scénario de préservation montre que la forêt native se développe sur la précordillère côtière (à l'est) et remplace principalement les zones catégorisées comme « autre ». Une partie des plantations exotiques est remplacée par des espèces natives. Les communautés autochtones, après avoir récupérées leurs terres, ont systématiquement boisé leur parcelle en espèces natives et y ont appliqué une politique de gestion soutenable afin de valoriser la ressource.

Au contraire, le scénario intensif traduit parfaitement l'invasion de la précordillère côtière par les plantations industrielles, laissant quelques îlots de terres non boisées. La construction de la *Ruta Costera* à l'ouest de la commune renforce le réseau routier existant en 2008 et facilite l'accessibilité des entreprises dans les terres, générant par conséquent l'apparition de grandes plantations au sein même de

la Cordillère. Ce scénario met en évidence les répercussions (environnementales et sociales) qui pourraient être engendrées.

La place occupée par le facteur foncier est déterminante dans les deux scénarios. Les données numériques disponibles font état d'une concentration des propriétés sur la précordillère ce qui explique la part occupée par la forêt native et les plantations dans cette zone. A l'inverse, la Cordillère de la côte est très faiblement affectée par cette variable mais reste tributaire du développement de nouveaux axes de communications.

5. Discussion et conclusion

Les images obtenues correspondent, dans une certaine mesure, aux hypothèses émises dans chacun des scénarios, néanmoins plusieurs limites ont été identifiées.

La mise en place d'une approche exploratoire afin de spatialiser un scénario normatif ne remet pas en cause sa pertinence. Cependant il peut être légitime de considérer ce scénario comme un scénario exploratoire en situation extrême qui ne se détache pas des tendances observées. Selon Börjeson *et al.* (2006) un ajustement marginal du développement actuel n'est pas suffisant, et il est nécessaire de rompre la tendance afin d'atteindre l'objectif fixé.

Dans ce sens, la limite de l'approche exploratoire se traduit par la dépendance du modélisateur face à la capacité du modèle à faire varier les paramètres de simulation (prise en compte des cycles de rotation et de déboisement par exemple). Nous avons vu que l'évaluation multicritère permet de pondérer les facteurs afin d'intégrer les hypothèses émises mais surtout pour tendre au plus près d'une image souhaitée. Bien que l'image du couvert forestier en 2035 soit riche d'enseignement elle ne rompt pas totalement avec la logique d'occupation du sol observée en 2008. Si les scénarios normatifs ne posent pas de réelles contraintes de conception, leur spatialisation (exploratoire) reste tributaire des données disponibles et de la performance du modèle employé.

Par conséquent, cette approche peut se concevoir comme une combinaison de l'approche normative (synchronique) dans la construction du scénario (image du futur) et de l'approche exploratoire dans le processus de spatialisation des hypothèses (du présent vers le futur).

Le choix du modèle dans une démarche prospective est une étape importante et doit prendre en compte plusieurs critères. Ainsi, après avoir élaboré l'image d'un futur souhaité et présenté les différentes hypothèses, le modélisateur doit s'assurer d'avoir à disposition une base de données complète qui permette de retranscrire spatialement chacune des hypothèses énoncées. Au cours du processus de simulation, le modèle doit être développé de sorte à ce qu'il rompe avec l'occupation du sol observée à la dernière date connue, afin de se démarquer des scénarios exploratoires contrastés. Cela implique de pouvoir intervenir en « temps réel » dans le modèle pour faire varier les paramètres. De plus, il est important de pouvoir intégrer de nouvelles variables ou des variables actualisées (parcelles cadastrales par exemple) à des instants clés de la simulation et qui auront été prédéterminés.

Remerciements

Les auteurs sont reconnaissants au projet ECOS-Sud Chili n° C07H03 intitulé "La forêt de la cordillère côtière continentale dans le sud du Chili : Dynamiques contemporaines et modélisations prospectives".

Bibliographie

Amer M., Daim T.U. et Jetter A. (2013). A review of scenario planning. *Futures*, vol. 46, p. 23-40.

Batty M. et Torrens P.M. (2005). Modelling and prediction in a complex world. *Futures*, vol. 37, p. 745–766.

Bishop P., Hines A. et Collins T. (2007). The current state of scenario development: an overview of techniques. *Foresight*, vol. 9, n° 1, p. 5-25.

Blecic I. et Cecchini A. (2008). Design beyond complexity: Possible futures—Prediction or design? (and techniques and tools to make it possible). *Futures*, vol. 40, p. 537–551.

Börjeson L., Höjer M., Dreborg K-H, Ekvall T. et Finnveden G. (2006). Scenario types and techniques: Towards a user's guide. *Futures*, vol. 38, n° 7, p. 723-739.

Catellin S. (2004). L'abduction : Une pratique de la découverte scientifique et littéraire. Critique de la raison numérique. HERMÈS n°39, CNRS Editions, Paris, p. 179-184.

Eastman J.R. (2006). *IDRISI Andes. Guide to GIS and Image Processing*, Clark University, Worcester.

European Environment Agency (2000). Scenarios as tools for international environmental assessments. Environmental issue report, n° 24, 31 p.

Godet M. (1986). Introduction à la Prospective. Seven key ideas and one scenario method. *Futures*, p. 134-157.

Godet M. (1992). *De l'anticipation à l'action*, Dunod, Paris.

Godet M. (2010). Future memories. *Technological Forecasting & Social Change*, vol. 77, p. 1457-1463.

Gourmelon F., Houet T., Voiron-Canicio C. et Joliveau T. (2012). La géoprospective. Points de vue conceptuels et applicatifs de l'apport du spatial à la prospective. *L'Espace Géographique*, n°2, p. 97-98.

Hatem F. (1993). *La Prospective. Pratiques et Méthodes*, Economica, Gestion, Paris.

Houet T., Hubert-Moy L. et Tissot C. (2008). Modélisation prospective spatialisée à l'échelle locale : approche méthodologique. *Revue Internationale de Géomatique*, vol. 18, n°3, p. 345-373.

Houet T. et Gourmelon F. (2013) (accepté). La géoprospective – Apport de la dimension spatiale aux approches prospectives. *Cybergéo. European Journal of Geography*.

Maestripieri N. et Paegelow M. (2013) (accepté). Validation spatiale de deux modèles de simulation : l'exemple des plantations industrielles au Chili. *Cybergéo : European Journal of Geography*.

Makropoulos C.K. et Butler D. (2006). Spatial ordered weighted averaging: incorporating spatially variable attitude towards risk in spatial multi-criteria decision-making. *Environmental Modelling & Software*, vol. 21, p. 69-84.

Masini E.B. et Médina Vasquez J. (2000). Scenarios as Seen from a Human and Social Perspective. *Technological Forecasting and Social Change*, vol. 65, p. 49–66.

Medina Vásquez J. et Ortegón E. (2006). *Manual de prospectiva y decisión estratégica: bases teóricas e instrumentos para América Latina y el Caribe*, CEPAL, Santiago.

Paegelow M. et Camacho Olmedo M.T. (2005). Possibilities and limits of prospective GIS land cover modelling—a compared case study: Garrotxes (France) and Alta Alpujarra Granadina (Spain). *International Journal of Geographical Information Science*, vol. 19, n° 6.

Paegelow M., Maestripieri N., Sáez Villalobos N. et Toro Balbotín D. (2012). Détection des plantations forestières par imagerie satellite MODIS dans le Sud du Chili, http://mappemonde.mgm.fr/num34/articles/art12204.html

Phdungsilp A. (2011). Futures studies' backcasting method used for strategic sustainable city planning. *Futures*, vol. 43, p. 707-714.

Roubelat F. (1994). *La prospective aux Etats-Unis. De la rigueur au divertissement*.

Saaty T. (1977). A scaling method for priorities in hierarchical structures. *Journal of mathematical psychology*, vol. 15, n°. 3, p. 234-281.

Sigel K., Klauer B. et Pahl-Wostl C. (2010). Conceptualising uncertainty in environmental decision-making: The example of the EU water framework directive. *Ecological Economics*, vol. 69, n°3, p. 502–510.

Stewart C.C. (2008). Integral scenarios: Reframing theory, building from practice. *Futures*, vol. 40, p. 160–172.

van Notten P.W.F. et Rotmans J. (2003). An updated scenario typology. *Futures*, vol. 35, p. 423-443.

Verburg P.H., Kok K., Pontius Jr. R.G. et Veldkamp A. (2006). *Modeling Land-Use and Land-Cover Change. Land-Use and Land-Cover Change. Local Processes and Global Impacts*. New York/Berlin, Springer, p. 117-135.

Voiron-Canicio C. (2012). L'anticipation du changement en prospective et des changements spatiaux en géoprospective. *L'Espace géographique*, n°2, p. 99-110.

Wilkinson A. (2009). Scenarios Practices: In Search of Theory. *Journal of Futures Studies*, vol. 13, n°3, p. 107-114.

Wolfslehner B., Vacik H. et Lexer M.J. (2005). Application of the analytic network process in multi-criteria analysis of sustainable forest management. *Forest Ecology and Management*, vol. 207, p. 157-170

Etude de faisabilité d'une cartographie multi niveaux des végétations du Nord-Ouest de la France

Vanessa Sellin

Conservatoire Botanique National de Brest
52 allée du Bot, F-29200 Brest, France
v.sellin@cbnbrest.com

RESUME.
Cette contribution présente les résultats acquis dans le cadre d'une étude associant deux laboratoires universitaires (LETG et ECOBIO) et le Conservatoire Botanique National de Brest (CBNB), sur le thème de l'évaluation des potentialités des images aériennes de la BDORTHO® IRC de l'IGN et des images satellitaires SPOT5 et Worldview-2 pour la cartographie des grands types de végétations de l'Ouest de la France. Quatre sites ont été retenus pour tester les procédures de classification et vérifier leur reproductibilité. Une typologie emboîtée à trois niveaux hiérarchiques est mise en place. L'approche orientée-objet non supervisée est privilégiée pour les traitements. Les meilleurs résultats sont obtenus à partir des images Worldview-2, puis la BDORTHO® IRC et enfin les images SPOT5.

ABSTRACT. This paper presents results obtained in the framework of a study involving two university laboratories (LETG and ECOBIO) and the National botanical Conservatory of Brest (CBNB). The topic of this study is to evaluate potentialities of BDORTHO® IRC aerial imageries (IGN) and SPOT5 and Worldview-2 satellite imageries, for mapping major vegetation types in western France. Four sites were chosen to test classification procedures and verify their reproducibility. Nested classification with three hierarchical levels is implemented. An unsupervised object-oriented approach is preferred for treatments. Best results are obtained from Worldview-2 imageries, then BDORTHO ® IRC and finally SPOT5 imageries.

MOTS-CLES : classification orientée-objet, eCognition, végétation, BDORTHO® IRC, SPOT5, Worldview-2

KEYWORDS: Object-oriented classification, eCognition, végétation, BDORTHO® IRC, SPOT5, Worldview-2

1. Introduction

Face aux demandes des politiques publiques, un nombre croissant de cartes de la végétation est réalisé par classification d'images aériennes et satellitaires (Augello et Capogrossi, 2007 ; Lucas *et al.*, 2011) qui permet une économie et une reproductibilité par rapport à la photo-interprétation.

En France, le Ministère en charge de l'environnement a mis en place un programme de cartographie nationale des habitats (CarHab) et plusieurs initiatives en région allant dans le même sens ont été engagées (SIRS, 2009 ; Brunet et Renaud, 2011).

Les progrès réalisés en télédétection concernant la qualité des images et les performances des procédures de classification ont justifié la multiplication des études de cartographie de la végétation par traitement automatique d'images. Leurs résultats attestent l'intérêt de l'approche orientée-objet pour l'identification, sur des images à haute résolution spatiale, des grands types de végétation (Allenbach *et al.*, 2010 ; Lucas *et al.*, 2011).

C'est dans ce contexte que cette étude de faisabilité a été mise en place avec l'objectif d'étudier la possibilité de produire, par des techniques de traitements d'images aériennes et satellitaires, des cartes de grands types de végétation au 1/25 000 (Sellin *et al.*, 2013).

Cette étude a fait l'objet d'un partenariat scientifique entre le Conservatoire Botanique National de Brest (CBNB) et deux unités de recherche : l'UMR6554 CNRS LETG et l'UMR6553 ECOBIO.

2. Données et méthodes

Trois types images (BDORTHO® IRC, SPOT5, Worldview-2) ont été sélectionnés et traités, couvrant quatre sites de l'Ouest de la France.

La majeure partie des traitements mis en œuvre utilisent l'approche orientée-objet non supervisée sous eCognition. Seules les procédures d'analyse de l'image Worldview-2 utilisent en complément la méthode pixel supervisée sous ENVI. Plusieurs couches d'information géographique vectorielles (bâti, routes, surfaces en eau, zone intertidale, sables dunaires, zones à dominante humide, parcellaire, forêts, haies, fourrés nains, plantations) et matricielles (images de texture) ont été utilisées dans les procédures de traitement (Eisfelder *et al.*, 2009 ; Allenbach et *al.*, 2010).

Sur les quatre sites sélectionnés, trois ont permis d'élaborer et de tester les procédures de cartographie (complexe du Menez-Hom dans le Finistère, marais de la Sangsurière et de l'Adriennerie dans la Manche, dunes de la Sauzaie et marais du Jaunay en Vendée). Le quatrième site, le plus vaste et le plus riche en types de végétations (havre de Saint-Germain–sur-Ay et landes de Lessay dans la Manche), a servi à la validation de la reproductibilité des procédures de traitement.

La typologie emboîtée des végétations réalisée par le CBNB (Delassus et Magnanon, 2013) sert de référence. Elle permet d'articuler la typologie phytosociologique utilisée sur le terrain, avec une typologie physionomique (structurale). Les classes de végétation pouvant être identifiées par traitement automatique sont déclinées en trois niveaux : « Occupation du sol » (3 classes),

« Grands types de végétations » (10 classes) et « Types de formations végétales » (23 classes).

Pour la validation des classifications, des levés au GPS ont été réalisés sur le terrain en 2011 et 2012. Ils ont été complétés par les relevés phytosociologiques les plus récents réalisés par le CBNB sur ces sites et de quelques points réalisés par photo-interprétation à partir des ortho-images.

3. Résultats

Cette étude (Sellin *et al.*, 2013) a permis d'aboutir pour chaque image à :

– Une procédure de classification permettant d'identifier un certain nombre de classes de végétation, à des niveaux plus ou moins fins de la typologie utilisée,

– une carte par niveau (occupation du sol, grands types de végétations, types de formations végétales), accompagnée de la matrice de confusion correspondante,

– un catalogue présentant les traitements par végétation, accompagné de l'ensemble des procédures de classification issu du logiciel eCognition.

Les résultats ont conduit à divers constats :

– Une méthode automatique de cartographie des végétations a été produite pour chaque image ; elle est reproductible sur l'ensemble de l'Ouest de la France en l'état mais seulement pour certains grands types de végétations.

– Globalement, et quel que soit le niveau de validation, c'est avec l'image Worldview-2 que les résultats sont les plus concluants, puis la BDORTHO® IRC et enfin les images SPOT5. Sur le site des dunes de la Sauzaie et des marais du Jaunay, les coefficients kappa au niveau « Types de formations végétales » sont les suivants : Worldview-2 : 0,73 ; BDORTHO® IRC : 0,71 ; SPOT5 : 0,60.

– Les végétations les mieux identifiées par cette méthode sont les herbiers marins, les prairies hygrophiles, les pelouses sur sable (non fixé et fixé), les végétations des prés salés et les milieux aquatiques avec ou sans prairies et pelouses amphibies. La distinction entre les forêts et les fourrés est difficile. De même que la distinction de leurs types écologiques. La classe des fourrés nains est bien identifiée, mais les confusions sont nombreuses aux niveaux typologiques inférieurs. Au sein de la classe des végétations herbacées, les prairies mésophiles et hygrophiles sont fréquemment sous ou sur-détectées, les végétations des roselières et cariçaies ainsi que les végétations des franges et lisières ne sont, quant à elles, pas détectées.

– La méthode orientée-objet a montré un réel intérêt pour la caractérisation des grands types de végétation et pour la délimitation précise de plusieurs types de formations végétales. Des cartes préalables aux inventaires de terrain pourraient être envisagées afin de faciliter les travaux de cartographie de terrain opérés par les phytosociologues.

4. Conclusion

Suite à ces résultats, des propositions pour améliorer les classifications du niveau

« Types de formations végétales » ont été avancées : l'acquisition de données vectorielles lorsqu'elles existent, exploitables au 1/25 000 (Kamagata, 2008) ; une adaptation de la typologie ; l'utilisation de la photo-interprétation en post-classification (Eisfelder *et al.*, 2009) ; l'acquisition d'images multi-temporelles de plusieurs capteurs (Lucas *et al.*, 2011) ; l'acquisition de données altimétriques de type LiDAR (Genç *et al.*, 2004).

Remerciements:

Cette étude a été réalisée dans le cadre d'une convention CNRS LETG-CBNB, en partenariat avec plusieurs membres du CBNB, des UMR LETG et ECOBIO. Elle a bénéficié de financements européens (FEDER) et régionaux (Agence de l'eau Seine-Normandie, Régions et DREAL de Basse-Normandie, Bretagne et Pays de la Loire) ainsi que d'un soutien ISIS pour l'acquisition d'images SPOT.

Bibliographie

– Allenbach K., Burnier J., Lehmann A., Martin P., Ray N., Wyler N. (2010). *Cartographie des milieux naturels dans le périmètre Vuache-Sion-Laire*. Rapport technique. DGAT-Projet d'Agglo. 98 p., 4 annexes.

– Augello R., Capogrossi R. (2007). Il sistema informativo geografico : " Carta della Natura ". p. 1–6. In: *Atti 11a Conferenza Nazionale ASITA*. Torino.

– Brunet J.-B., Renaud J. (2010). *Méthodologie de cartographie des milieux naturels et semi-naturels. Apport de la télédétection satellitale - Milieux ouverts d'altitude*. Cemagref-UR Ecosystèmes montagnards-Centre de Grenoble. 74 p., 2 annexes.

– Delassus L., Magnanon S. (2013). Une classification emboîtée de la végétation de Basse-Normandie, Bretagne et des Pays de la Loire. Brest. *Journal de Botanique*. A paraitre.

– Eisfelder C., Kraus T., Bock M., Werner M., Buchroithner M. F., Strunz G. (2009). Towards automated forest-type mapping - a service within GSE Forest Monitoring based on SPOT5 and IKONOS data. *International Journal of Remote Sensing*, vol. 30, n°19, p. 5015-5038.

– Genç L., Dewitt B., Smith S. (2004). Determination of Wetland Vegetation Height with LiDAR. *Turkish Journal of agriculture and forestry*, vol. 28, n°1, p. 63–71.

– Kamagata N., Hara K., Mori M., Akamatsu Y., Li Y., Hoshino Y. (2008). Object-based classification of IKONOS data for vegetation mapping in Central Japan. In: Blaschke Th., Lang S., Hay G.J. (2008). Object-Based Image Analysis. *Springer*. p.459-475.

– Lucas R., Medcalf K., Brown A., Bunting P., Breyer J., Clewley D., Keyworth S., Blackmore P. (2011). Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, p. 81-102.

– Sellin V. (2013)- *Cartographie des grands types de végétation par télédétection : étude de faisabilité (Bretagne, Basse-Normandie et Pays de la Loire).* Brest : Conservatoire botanique national de Brest, Brest : UMR6554 LETG CNRS. A paraitre.

– SIRS (2009). *Evaluation des conditions de la biodiversité dans le Kent et le Nord Pas de Calais : Photo-interprétation, Base de données et cartographie régionale et transfrontalière des habitats naturels.* Cahier des clauses techniques particulières. SIRS. 30 p., 4 annexes.

Modélisation spatiale de la température dans un site viticole des Coteaux du Layon

Cyril Bonnefoy[1], Malika Madelin[2], Hervé Quénol[1]

1. LETG-Rennes-Costel UMR CNRS 6554, Université Rennes 2
Place du Recteur Henri Le Moal, 35043 Rennes cedex, France
cyril.bonnefoy@uhb.fr herve.quenol@uhb.fr

2. PRODIG UMR CNRS 8586, Université Paris Diderot
2 place Jussieu 75251 Paris Cedex05, France
malika.madelin@univ-paris-diderot.fr

RESUME. *Un site viticole des Coteaux du Layon a été équipé d'un réseau de capteurs thermiques afin d'étudier la variabilité spatiale de la température au sein du vignoble. Une partie de l'étude visait à mettre en évidence les facteurs locaux jouant dans cette variabilité et d'établir le rôle et le poids de chacun. Pour cela, une modélisation multi-critères à partir de régressions linéaires multiples a été mise en place. Cette modélisation a été réalisée pour les températures minimales et maximales journalières de la saison 2009. Les résultats montrent que plus de 70% de la variabilité de la température minimale est expliquée pour près de 2 nuits sur 3 et en premier lieu par l'altitude. Ce résultat confirme la présence d'inversions thermiques fréquentes sur le coteau viticole. La qualité des modèles pour les températures maximales est largement inférieure démontrant une complexité des phénomènes impliqués dans cette variabilité. L'intégration de nouvelles variables et de la non-linéarité dans les modèles devraient permettre d'en améliorer la qualité.*

ABSTRACT. *A fine networks of thermal data loggers has been set up in a viticultural site of the « Coteaux du Layon » to assess spatial variability of temperature within the vineyard. One part of the study aimed to point out wich local factors prevail in the temperature variability and to establish role and weight of each factors. That is why a multicriteria modeling based on multiple linear regressions has been set up. This modeling has been realized for the daily minimum and maximum temperatures during the 2009 season. Results show more than 70% of the variability of minimum temperatures is explained for 2/3 of the nights and first by the elevation. This result confims frequent thermal inversions in this vineyard. The quality of models for maximum temperatures is widely inferior, proving the complexity of phenomenons involved in this variability. New variables should be added in models and non-linear regressions should be tested in order to improve models quality.*

MOTS-CLES : *Coteaux de Layon, réseau de capteurs, facteurs locaux, modélisation multicritères, températures, inversions thermiques*

KEYWORDS: *Coteaux du Layon, data loggers network, local factors, multicrietria modeling, temperatures, thermal inversions*

2

1. Introduction

La variabilité spatiale de la température dans un site viticole des Coteaux du Layon, situés au sud-ouest d'Angers, a été étudiée. Ce site expérimental est réputé pour les vins liquoreux qui y sont produits avec notamment l'appellation Grand Cru Quarts-de-Chaume. Un réseau fin de 21 capteurs thermiques a été installé au printemps 2009 en prenant en compte les différentes informations topographiques et environnementales dont nous disposions afin de représenter au mieux le site de 600 ha d'un point de vue climatique. L'objectif de cette étude est notamment de déterminer les facteurs locaux qui permettent d'expliquer au mieux la variabilité spatiale de la température, grâce à la construction de modèles multicritères.

2. Données d'entrée utilisées pour la modélisation

3.1. Les données de températures issues des capteurs thermiques

Les relevés des données de températures d'avril 2009 à octobre 2011 étaient disponibles pour la construction de notre modèle. Cependant, un travail de sélection des données a été réalisé, afin d'optimiser le poids statistique de nos calculs. La saison végétative 2009 a été retenue, puisqu'étant celle avec le moins de pertes de données et donc le maximum de séries de températures complètes sur l'ensemble de la période avril-octobre. Expliquer des écarts de températures journaliers trop faibles entre les capteurs n'ayant pas réellement de sens, une seconde sélection de journées avec une variabilité spatiale marquée a été réalisée. Le seuil arbitraire de 1,5°C entre le 2ème et le 8ème décile a été fixé aussi bien pour les températures minimales que pour les maximales. Après cette nouvelle sélection, le nombre de journées modélisables est passé de 197 à 47 pour les températures minimales et de 197 à 25 pour les maximales.

3.2. Les données topographiques issues du modèle numérique de terrain

Les différents paramètres descripteurs, issus du Modèle Numérique de Terrain (MNT) de moyenne vallée de la Loire, ont été extraits pour chaque capteur et leur sélection a été faite suite à des tests de régressions simples avec la température. Au final, l'altitude, la pente, l'orientation nord-sud (U) et l'orientation est-ouest (V) ont été retenues.

3. Méthodologie de construction des modèles et d'interprétation des résultats

Différentes méthodes de régression multiple sont proposées par les différents logiciels de statistiques disponibles. Pour cette étude, nous avons travaillé avec le logiciel de statistique R et nous avons choisi la méthode de régression multiple pas à pas avec sélection des variables grâce au critère AIC (Akaike Information Criterion). Dans notre cas, nous avons 4 prédicteurs potentiels soit 24 ou encore 16 modèles possibles (dont 1 qui ne sort aucun descripteur) pour chaque mois ou dates

modélisés. Le critère AIC permet ainsi de faire le tri entre ces variables et de sélectionner le meilleur modèle. L'algorithme permettant l'obtention des différents modèles journaliers a été créé sous le logiciel R et la procédure a été automatisée. Une attention toute particulière a été portée aux coefficients de détermination des modèles, indiquant leur qualité, et aux coefficients de régression standardisés, permettant de mesurer le poids et le rôle de chaque variable dans l'équation.

4. Résultats

4.1. Modélisation des températures minimales journalières

Sur les 47 journées présélectionnées, seulement 3 n'arrivent pas à être reproduites par un modèle car aucune relation significative n'existe entre les descripteurs et la température minimale. Pour ces 3 journées, les stations météorologiques indiquent des précipitations. L'analyse des bulletins climatiques journaliers de Météo-France montre que ces 3 journées sont toutes marquées par l'arrivée d'un front froid par le nord-ouest du pays qui a engendré une certaine couverture nuageuse sur les Pays-de-la-Loire et des précipitations à la station d'Angers. Ce temps plus perturbé peut expliquer que les facteurs locaux ne prévalaient pas sur ces journées. Pour les autres dates, les différents modèles arrivent à expliquer plus de 70% de la variabilité de la température minimale pour près de 2 nuits sur trois (soit 31 jours). La vérification des conditions synoptiques pour l'ensemble des journées, montre des conditions anticycloniques bien présentes sur les journées les mieux reproduites par les modèles (R^2>0,5), alors que les journées les moins bien reproduites correspondent à des situations globalement anticycloniques sur plusieurs jours mais avec des faiblesses des hautes pressions engendrant le passage de fronts plus ou moins actifs par le nord-ouest du pays. Dans le détail, l'altitude intervient pratiquement pour chaque journée et le coefficient de régression est positif, indiquant des cas d'inversions thermiques systématiques. La valeur des coefficients de régression standardisés, associés à l'altitude, sont pratiquement dans tous les cas très nettement supérieurs aux coefficients de régression des autres variables. L'altitude est donc la première variable explicative de la variabilité de la température minimale. La pente n'intervient que dans 6 cas avec un coefficient de régression négatif. L'orientation nord-sud (U) intervient pour 7 journées avec des coefficients négatifs sauf le 6 septembre. Enfin, l'orientation est-ouest (V) ressort pour la plupart des journées modélisées avec des coefficients de régression systématiquement négatifs, indiquant une baisse de la température pour les capteurs avec une orientation est.

4.2. Modélisation des températures maximales journalières

Les valeurs des coefficients de détermination des modèles journaliers pour les températures maximales sont très inférieures à celles des températures minimales. La complexité des champs thermiques semble dépendre de très nombreux facteurs non pris en compte par nos modèles. Ces modèles journaliers font donc ressortir l'altitude pour 4 cas seulement et la pente ressort pour l'ensemble des journées sauf 3 et ses régresseurs sont tous positifs, traduisant des températures plus élevées sur

pentes fortes. L'orientation nord-sud est retenue par 7 modèles et les coefficients de régression sont tous négatifs, d'où des températures plus fraîches sur les versants nord. Enfin, l'orientation est-ouest ressort pour 10 modèles avec des régresseurs positifs qui expliquent des températures plus élevées sur les versants exposés à l'est.

5. Conclusions et perspectives

Cette étude a permis de quantifier les relations existantes entre les différents facteurs locaux liés à la topographie et la variabilité de la température dans site expérimental des Coteaux du Layon. L'ensemble des modèles multicritères, réalisés ici, explique une part plus ou moins importante de la variabilité de la température, et suivant les dates ces modèles sont perfectibles. La spatialisation de la température minimale est beaucoup moins complexe que celle de la température maximale. En effet, la part des facteurs locaux lors des nuits calmes, sans vent et avec ciel dégagé, est importante. Des inversions thermiques se mettent souvent en place et des contrastes thermiques marqués peuvent alors être observés entre la vallée et les plateaux. Cependant, dans certains cas, les plateaux à une altitude supérieure à 70 m se retrouvent au-dessus de la couche d'inversion et la relation entre altitude et la température minimale n'est donc pas totalement linéaire sur l'ensemble de l'étendue altitudinale du site. Des modélisations non linéaires devraient être alors expérimentés afin d'améliorer la modélisation pour certaines journées. La spatialisation de la température maximale est beaucoup plus complexe. En journée, la turbulence de l'air se met en place, en lien avec l'énergie solaire et les inversions thermiques disparaissent. La relation avec l'altitude s'inverse donc mais pour le site des Coteaux du Layon, le dénivelé étant très faible, cette relation ne ressort pas comme évidente. Des facteurs comme l'inclinaison de la pente et son orientation vont expliquer en partie, la variabilité spatiale de la température maximale. Dans les modèles journaliers développés dans cette étude, d'autres éléments locaux devraient certainement permettre de mieux modéliser les champs thermiques. Ces facteurs environnementaux seraient à relier avec les positions d'abri de certaines parcelles et la ventilation plus importante de certains coteaux. Des indices tels que l'ouverture du paysage, l'exposition aux vents pourraient être calculés et intégrés aux modèles. Enfin, le nombre de capteurs utilisé dans le cadre de cette expérimentation pourrait être augmenté afin d'améliorer la couverture spatiale et donc la représentativité du site viticole.

Fouille d'images géoréférencées avec RapidMiner

Thomas Guyet[1,3], Hervé Nicolas[2,3], Boussad Ghedamsi[1,3], Élise Athané[2,3]

1. AGROCAMPUS-OUEST/IRISA, Laboratoire d'Informatique
65 rue de Saint-Brieuc, F-35042 Rennes, France
thomas.guyet@agrocampus-ouest.fr

2. AGROCAMPUS-OUEST/INRA, UMR1069 Sol Agro et hydrosystème Spatialisation
65 rue de Saint-Brieuc, F-35042 Rennes, France
herve.nicolas@agrocampus-ouest.fr

3. Université Européenne de Bretagne

RÉSUMÉ.
Cet article présente l'intégration de données spatiales dans le logiciel de fouille de données RapidMiner. RapidMiner est un outil pour concevoir interactivement des chaînes de traitements, orientées vers la fouille de données. Dans le cadre de ce travail, nous avons implémenté une extension de RapidMiner pour traiter des données spatiales. On bénéficie ainsi d'une large gamme de méthodes de fouille de données déjà à disposition pour répondre à une question d'analyse sur des données spatiales. L'article illustre le fonctionnement de l'extension par une comparaison de méthodes de classification sur un problème de détection de plante invasive (la Jussie) à partir d'une image aérienne hyperspectrale.

ABSTRACT.

This article presents the integration of spatial data in the data mining software RapidMiner. RapidMiner is a data mining tool to design data processing chains. In this work, we implemented an extension of RapidMiner to process spatial data. It benefits from the wide range of data mining methods already available and it enables spatial data users to answer to their analysis questions. The paper illustrates the extension by a comparison of classification methods on a problem of detecting invasive plant (Jussie) from an hyperspectral image.

MOTS-CLÉS : Fouille de données, images de télédétection, chaîne de traitement, évaluation

KEYWORDS: Data mining, remote sensing images, processing chain, evaluation

1. Introduction

L'augmentation continuelle de la quantité de données spatialisées, la démocratisation de leur accès et de leurs usages fait croître l'intérêt de ces données pour de nombreuses applications allant de la prédiction en agronomie au géo-marketing. Des outils d'analyse de données sont nécessaire aider à l'exploitation de cette masse de données. Les outils d'analyse de données intègrent à la fois des méthodes de traitement et prétraitements des données, des méthodes d'extraction d'information et des outils de visualisation. Les méthodes de fouille de données désignent quant à elle, l'ensemble des méthodes algorithmiques et statistiques qui résolvent des tâches élémentaires d'extraction d'information, p. ex. classification, clustering, regression (voir (Miller et Han, 2009) pour une revue des questions actuelles en fouille de données spatiale). Par la suite, on s'intéresse principalement à des tâches d'analyse d'images géoréférencées.

Les outils d'analyse de données prenant en compte l'information spatiale peuvent être organisés en deux grandes classes : d'une part, les outils dédiés à l'analyse et à la visualisation de données spatiales (tels que les Systèmes d'Information Géographiques, SIG, ou les outils d'analyse d'images géoréférencées) et, d'autre part, les outils de fouille de données.

La première catégorie offre l'avantage d'être bien adaptées aux données géographiques, nécessitant des traitements efficaces en gestion de la mémoire et du temps de calcul, p. ex. ENVI ou IDRISI. Ces outils intègrent des méthodes dédiés aux tâches courantes d'analyse d'images tels que des méthodes d'apprentissage supervisé. Deux problèmes se posent aux utilisateurs, d'une part, le choix du paramétrage de ces algorithmes. La forte dépendance des paramètres aux données rend cette question du paramétrage difficile. D'autre part, les outils intégrés ne permettent pas aux utilisateurs d'explorer d'autres méthodes d'apprentissage existantes. L'Orfeo Toolbox (Inglada et Christophe, 2009) offre des possibilités étendues pour le traitement de l'information spatiales. Mais cette boîte à outils n'intègre que très peu de méthodes d'analyse de données.

La seconde catégorie offre une large gamme d'algorithmes d'analyse de données ainsi que des fonctionnalités pour les comparer et pour choisir le paramétrage des algorithmes qui convienne le mieux aux données. Cependant, ces outils sont souvent difficiles à utiliser pour les non-informaticiens, des logiciels comme R, Weka (Hall *et al.*, 2009) ou Sci-Learn (Pedregosa *et al.*, 2011) nécessitent de la programmation pour mener des analyses de données. De plus, ils n'intègrent que rarement des traitements explicites sur la dimension spatiale des données géographiques.

Il apparaît nécessaire de proposer des outils facilitant l'expérimentation de méthodes de fouille de données potentiellement complexes sur des données géographiques pour des utilisateurs cherchant à analyser ces données. On ne cherche

pas à reconstruire un outil capable de traiter directement les images, mais plutôt d'aider au prototypage rapide de la méthode d'analyse. Un tel outil doit donner la possibilité d'expérimenter différentes méthodes d'analyse (choix des algorithmes et des paramètres) et de comparer leurs performances sur des jeux de test. La mise en production du traitement des données devant être menée dans un second temps par l'implémentation d'un programme dédié ou depuis les plate-forme dédiées.

L'approche proposée dans cet article est l'intégration de fonctionnalités liées aux données spatiales dans un outil de fouille de données disposant de la capacité de mise en œuvre de chaînes complètes d'analyse de données. On cherche ainsi à palier les difficultés rencontrées par l'une ou l'autre des classes d'outils existants : (1) on facilite l'accès à une large gamme d'algorithmes de fouille de données sur des données spatiales, (2) on rend possible l'analyse automatique des performances de la méthode d'analyse pour faciliter les choix des paramètres.

Notre choix s'est porté sur l'utilisation de l'outil de fouille de données nommé RapidMiner. D'autres outils tels que Orange (Curk *et al.*, 2005) ou KNIME (Berthold *et al.*, 2007) offrent des fonctionnalités similaires, le choix de RapidMiner a été fait, d'une part, pour sa visibilité internationale et, d'autre part, par l'importance du support au développement d'extensions du logiciel.

À notre connaissance, aucune extension pour l'utilisation de données geospatiales n'a été proposée pour aucune de ces trois plate-forme. (Burget *et al.*) ont proposés l'extension IMMI de RapidMiner pour intégrer des fonctionnalités de fouille d'images visuelles, mais ces fonctionnalités ne permettent pas d'intégrer les aspects spatiaux des images géoréférencées.

2. Vers une extension de RapidMiner pour des données spatiales

2.1 Présentation du logiciel RapidMiner

RapidMiner (Mierswa *et al.*, 2006), anciennement appelé YALE (Yet Another Learning Environment) est un environnent d'apprentissage, de fouille de données, d'analyse prédictive et d'analyse de données. Actuellement, il est très largement utilisé[1] dans le domaine de la recherche, de l'éducation et en milieu d'entreprise. C'est un logiciel puissant pour mettre en place rapidement une chaîne complète de traitement de données, de la saisie des données à la visualisation des résultats d'analyse.

Une chaîne de traitements de RapidMiner est constituée d'un ensemble de « blocs » reliés entre eux pour représenter la succession des traitements. Chaque bloc correspond à un traitement : les blocs de chargement de données depuis un

[1]RapidMiner a été classé de 2010 à 2012 comme l'outil interactif de fouille de données le plus utilisé au niveau mondial (voir enquête de KDNuggets, http://www.kdnuggets.com)

fichier, les blocs de transformation des données, les blocs implémentant des algorithmes de fouille de données. En sortie de la chaîne de traitements, les données peuvent être visualisées ou exportées. La Figure 1 illustre l'interface du logiciel dont la zone centrale contient une chaîne de traitements simple.

Pour étendre les possibilités de traitement, des méta-blocs définissent des modalités de traitements spécifiques à certaines analyses de données. Par exemple, le méta-bloc *XValidation* facilite la réalisation de validations croisées tandis que le bloc *Optimize Parameters* explore certains paramètres d'une chaîne de traitements pour identifier la meilleure combinaison de paramètres relativement à une évaluation.

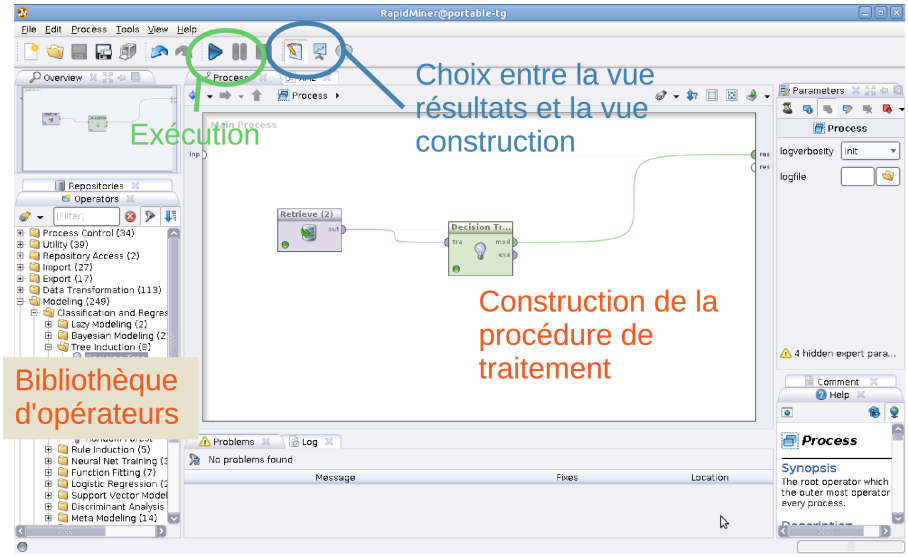


Figure 1. Illustration de l'interface de RapidMiner. Dans la zone centrale, la chaîne de traitement contient deux blocs. Les panneaux gauche et droite sont respectivement dédiés à l'ajout de bloc et au paramétrage du bloc sélectionné.

Pour que les blocs fonctionnent correctement, il est nécessaire de leur fournir en entrée des données correctement structurées. Les blocs implémentant des méthodes de fouille de données requièrent des données organisées de manière tabulaire, c'est-à-dire de données constituées d'enregistrements (une ligne dans un tableau) décrits par différents attributs (les colonnes du tableau).

RapidMiner est un logiciel conçu de manière modulaire, il offre la possibilité de programmer ses propres blocs. Une extension RapidMiner est un ensemble de blocs qui peuvent être utilisés dans l'interface de RapidMiner. Parmi les extensions plus plus populaires, on peut citer l'intégration de Weka ou de R.

2.2 Objectifs d'une extension pour les données spatiales

Les objectifs d'une extension RapidMiner pour les données spatiales est de pouvoir mettre en place les différentes méthodes d'extraction d'information à partir de données spatiales telles que listées par (Miller et Han, 2009) en s'appuyant les méthodes de fouille de données (non-spatiales) déjà existantes.

RapidMiner étant un logiciel de fouille de données ne supportant par le format des données géographiques, le défi principal est de rendre ces données accessibles depuis ce logiciel en implémentant une extension pour les données spatiales. Cette extension inclut des blocs de lecture et d'écriture de données. Pour garantir la communication entre nos blocs et les blocs préexistants, il a été nécessaire de respecter le principe d'utilisation des données tabulaires de RapidMiner.

Dans le cadre de cette première version de l'extension, nous nous sommes focalisé sur des tâches de classification d'images de télédétection (images satellite ou aériennes), de prédiction d'observations spatialisées dans le temps à partir de séquences d'images géoréférencées et d'évaluation des performances des deux tâches précédentes.

3. Extension GeoDM - Geospatial Data Mining

Dans cette section, on présente l'extension « Geospatial Data Mining » en introduisant le principe de transformation des données géoréférencées aux données tabulaires, puis on présente les différents modules de l'extension pour en décrire les fonctionnalités principales.

3.1. Principe de transformation des données géoréférencées aux données tabulaires

Le principe de transformation des données géoréférencées est l'utilisation d'un échantillonnage spatial, c'est-à-dire d'un ensemble de points définis par leurs coordonnées dans l'espace (coordonnées et système de coordonnées associé).

Chaque point d'un échantillon correspond à un enregistrement dans le format tabulaire. Les couches de données utilisées servent à définir les attributs qui caractérisent chacun des points de l'échantillon. La Figure 2 illustre la construction d'un jeu de données à partir d'un échantillon. Pour chaque point de l'échantillon, la valeur des attributs de chaque couche est récupérée à la position de ce point. Lorsque le point se trouve en dehors d'une couche, il est possible de donner une valeur par défaut à l'attribut ou de le définir comme « inconnu ».

L'intérêt majeur de cette méthode d'échantillonnage est d'utiliser des couches avec des représentations hétérogènes : couches vecteur et couches raster, couches

raster avec des résolutions spatiales différentes, couches avec des géoréférencements différents. Cette approche n'est néanmoins pas totalement générique. Elle est limitée par l'utilisation de couches de données au format vectorielle qui, d'une part, ne contiennent que des polygones et que, d'autre part, ces polygones ne se superposent pas. Dans le cas de superposition de polygones, le nombre d'attributs à prendre en compte pour un point qui se situe dans la zone de recouvrement ne serait pas le même que pour les autres points. Ces limitations restent néanmoins raisonnables par rapport aux usages prévus de l'extension GeoDM.

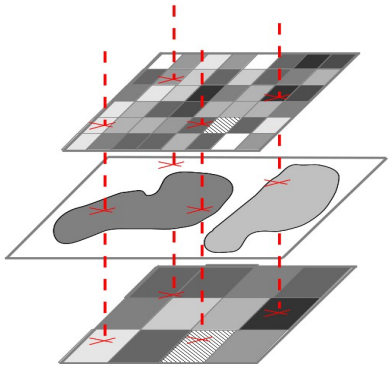


Figure 2. Illustration de la méthode d'échantillonnage. L'échantillon comprend ici 4 points visualisés en rouges (tirets-verticaux pointillés). Pour chaque point, on récupère les attributs pour les trois couches (deux raster et une vecteur).

Plusieurs types d'échantillonnages sont prévus pour répondre aux besoins d'analyses spatiales des données :

– l'échantillonnage selon un *transect* sert à l'analyse de l'influence d'un phénomène selon une dimension de l'espace : influence d'un objet sur une variable bio-physique, étude de la transition d'un état à un autre (ex. de l'urbain au rural) (Buckland, 2001).

– l'échantillonnage aléatoire assure un sous-échantillonnage d'une zone rectangulaire de manière à répéter des apprentissages avec différentes données,

– l'échantillonnage régulier correspondant, par exemple, aux centres des pixels d'une image raster géoréférencée,

– l'échantillonnage spécifique que peut définir un utilisateur par une couche de points.

3.2 Présentation générale

L'extension est organisée en trois modules :

– le module « I/O » contient des blocs destinés à ouvrir ou enregistrer des données géoréférencées. Actuellement, les données peuvent être chargées sous la forme d'images vectorielles (format Shapefile) et d'images raster (format GeoTiff). Ce module s'appuie sur les fonctionnalités de la GeoTools[2] et prend pleinement en charge les transformations géométriques entre différentes sources de données. L'enregistrement et la visualisation de données géoréférencées est possible uniquement en format raster (GeoTiff).

– le module « Sampling » contient les outils pour la génération et la manipulation d'échantillons (voir section précédente).

– le module « Data Transformation » contient des outils de transformation des données au format spécifique à l'extension. Il introduit des outils pour la gestion des attributs de coordonnées. En particulier, les attributs de coordonnées doivent être facilement masqués avant d'effectuer une tâche de fouille de données.

4. Exemples d'utilisation : comparaison de méthodes de classification supervisée sur des images hyperspectrales

Nous utilisons ici notre extension GeoDM dans le but d'identifier une bonne méthode de classification supervisée pour la classification d'images hyperspectrales[3]. De telles méthodes sont proposées dans des outils tels que ENVI, mais il peut être difficile de savoir laquelle est la mieux adaptée et quels paramètres des algorithmes d'apprentissage fonctionnent le mieux. On présente deux expérimentations permettant de répondre ces questions : une expérimentation comparant plusieurs algorithmes de classification supervisée, puis la comparaison de résultats obtenus avec différents paramétrages des SVM.

Notre objectif n'est pas de discuter les méthodes de classification elles-mêmes mais plutôt de mettre en évidence que l'utilisation de GeoDM facilite le choix des paramètres et la validation d'une telle méthode.

4.1. Problématique applicative et données

La jussie est une plante aquatique invasive que les gestionnaires de rivière surveillent particulièrement pour éviter son pullulement. Ce travail s'inscrit dans une étude sur la faisabilité d'une approche de la détection de la jussie par télédétection. Le site d'étude se trouve au confluent de l'Oust et de l'Aff, à Glénac (Morbihan).

Pour cette expérimentation, on dispose d'une image aérienne hyperspectrale comprenant 160 bandes spectrales ainsi qu'un fichier *shapefile* contenant une vérité terrain obtenue par des observations ponctuelles partielles. Quatre classes sont

[2]GeoTools : www.geotools.org

[3]Une image hyperspectrale est une image contenant un grand nombre de bandes. Chaque bande correspondant à la réponse du sol à une certaine longueur d'onde.

distinguées : *eau*, *jussie*, *nénuphar* et *autre végétation*. Le but de la classification de l'image est d'identifier en tout point de l'image la présence ou non de jussie.

Pour limiter les temps de calcul, seules les 15 premières bandes de l'image hyperspectrale sont utilisées. On utilise également une image contenant les 15 premières bandes de la transformation MNF, Minimum Noise Fraction (Boardman et Kruse, 1994). De manière similaire à une analyse en composantes principales, la transformation MNF est une transformation linéaire des données qui consiste à séparer le signal du bruit. Les premières bandes de l'image transformée contiennent ainsi les données les plus informatives de l'image originale.

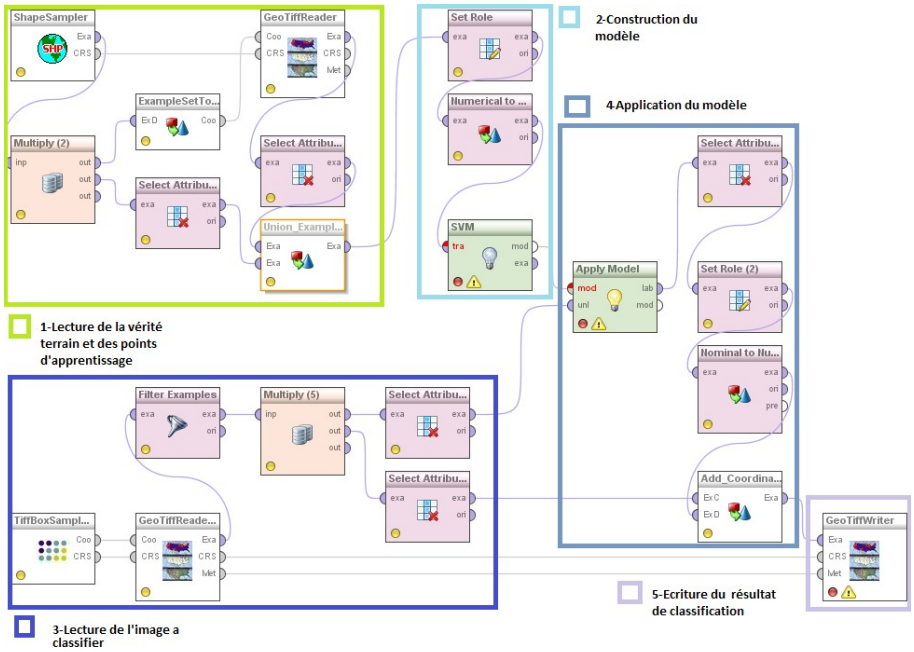


Figure 3. Illustration de la chaîne de traitements pour la classification supervisée d'une image géoréférencée. Les blocs blancs sont des blocs spécifiques à l'extension GeoDM. Les autres blocs sont des blocs prédéfinis dans RapidMiner.

4.2. Utilisation de l'extension GeoDM pour la classification supervisée d'une image hyperspectrale

On décrit le processus de la Figure 3 pour la classification supervisée de l'image hyperspectrale. Ce processus illustre une large gamme des fonctionnalités de l'extension GeoDM. La chaîne de traitements comporte cinq parties distinctes :

– les parties 1 et 2 effectuent l'apprentissage du classifieur. La partie 1 charge le jeu d'apprentissage depuis un fichier *shapefile* contenant la vérité terrain, et d'un

fichier GeoTiff contenant les informations spectrales. Le fichier de vérité terrain sert à définir les points d'échantillonnage. La partie 2 construit le classifieur. L'illustration montre l'usage d'un classifieur SVM, mais il existe de nombreux autres algorithmes déjà implémentés dans RapidMiner.

– la partie 3 et 4 classent une image à partir du classifieur construit précédemment. La partie 3 charge une image raster à classifier (sans vérité terrain ici), puis la partie 4 classe les pixels de l'image.

– finalement, la partie 5 enregistre l'image construite par classification dans un nouveau fichier GeoTiff.

Plus que le calcul d'une image résultant d'une classification, l'usage de RapidMiner facilite l'automatisation des comparaisons entre plusieurs modèles et entre plusieurs paramètres. Pour cela, on utilise des fonctionnalités d'automatisation disponibles dans RapidMiner telles que le bloc de validation-croisée ou le bloc de test des combinaisons de paramètres.

Pour illustrer les types de résultats recherchés nous avons construit deux chaînes de traitements dont les résultats sont présentés dans les sections suivantes. Pour des raisons de place, on ne présente pas ici les détails des chaînes de traitements mises en place. Plus d'information sur les chaînes de traitements qui peuvent être construites sont disponibles sur le site de l'extension[4].

4.3. Une chaîne de traitements pour la comparaison de différents algorithmes

La première expérimentation nous a conduit à comparer différents algorithmes de classification (*SVM*, *DecisionTree*, *NeuralNetworks* à trois couches, *k-NN*). Les évaluations sont effectuées à l'aide d'une validation croisées (10 validations) selon le critère kappa (κ). L'indice kappa est une valeur entre 0 et 1. Plus κ est proche de 1, meilleures sont les performances de classification.

Tableau 1. Indice kappa (κ) pour les classifications pour différents algorithmes.

	Image originale	**Image MNF**
SVM	0.151 +/- 0.013	0.778 +/- 0.014
Neural Network	0.902 +/- 0.012	0.991 +/- 0.003
k-NN	0.854 +/- 0.008	0.996 +/- 0.001
Decision Tree	0.753 +/- 0.025	0.968 +/- 0.006

[4] http://geoimagermp.gforge.inria.fr/

4.4. Une chaîne de traitements pour le paramétrage d'un classifieur SVM

On se place ici dans la situation où l'utilisateur souhaite utiliser un classifieur SVM pour construire ses classes. Le problème qui se pose est d'identifier les paramètres optimaux pour un classifieur *SVM*, c'est-à-dire le jeu de paramètres qui permettra d'obtenir le meilleur kappa. L'utilisateur à classiquement deux paramètres à donner γ et ε spécifiques aux *SVM* avec noyau Gaussien. Dans le cas de l'étude précédente, les paramètres utilisés étaient les paramètres par défaut.

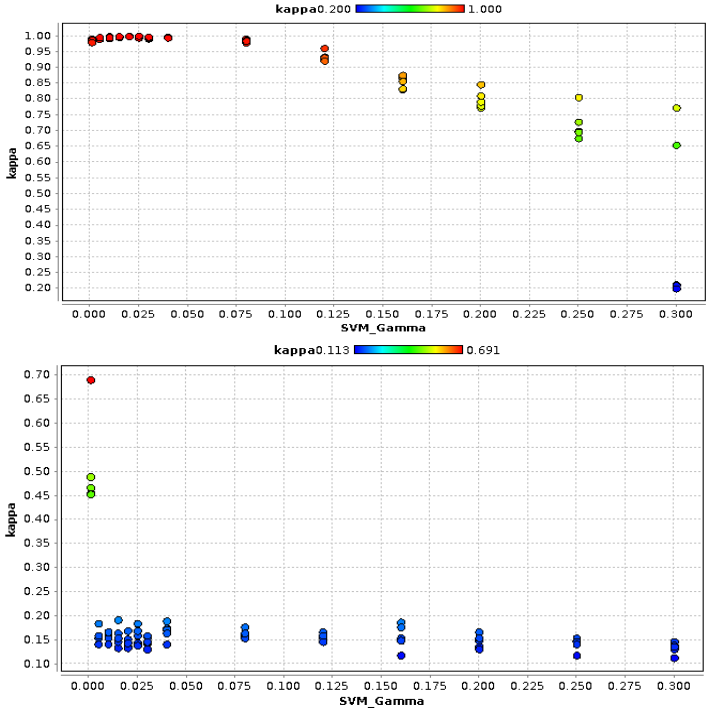


Figure 4. Résultats pour l'image originale (à gauche) et l'image MNF (à droite). Valeurs de κ en fonction de γ.

La méthode consiste à mettre en place une chaîne de traitement qui fait varier les paramètres γ et ε. Pour chaque combinaison de paramètres, une validation croisée évalue les performances du classifieur. Après lancement de cette chaîne de traitement, RapidMiner fournit la Figure 4 pour analyser les résultats pour l'image originale et pour l'image MNF.

On constate d'une part que les résultats obtenus avec l'image originale sont beaucoup moins bons qu'avec l'image MNF et, d'autre part, que les meilleurs

résultats sont obtenus pour γ=0,02 et ε=0,5 (κ=1). Ces valeurs pourront être réutilisées pour paramétrer un classifieur *SVM* dans ENVI par exemple.

Pour une image, l'ensemble des traitements a pu être réalisés en moins d'une minute. On obtient ainsi rapidement une indication importante sur le paramétrage qui guidera l'expert dans sa classification d'image.

L'objectif de ce papier n'étant pas de présenter une amélioration d'une méthode d'analyse d'image, on ne discute pas plus les résultats obtenus par les chaînes de traitements présentées. L'intérêt de cette expérimentation est de montrer que notre extension permet de mettre en place assez simplement une chaîne d'analyse de données à partir de données spatiales répondant à un besoin auquel les outils classiques ne répondent pas.

5. Perspectives et conclusion

Dans cet article, nous nous sommes intéressé à l'utilisation d'un outil de fouille de données, le logiciel RapidMiner, pour construire des chaînes de traitements pour l'analyse spatiale de données. Nous avons proposé une extension du logiciel qui intégre des données géoréférencées dans une chaîne de traitements usuelle. Cette approche a été illustrée pour l'identification d'une méthode optimale de classification d'une image hyperspectrale. Les chaînes de traitements mises en place permettent d'évaluer facilement et rapidement différents algorithmes de classification supervisée et d'identifier les paramètres optimaux d'un classifieur de type *SVM*.

Les applications visées de cette extension sont à la fois le prototypage rapide d'une méthode d'analyse d'images de télédétection, la mise en place de nouvelles méthodes d'analyse de données spatiales et, dans un but pédagogique, la facilitation de l'expérimentation des méthodes de fouille de données déjà existantes sur des données spatiales afin de sensibiliser à ses apports potentiels.

Cette extension a été utilisée dans un cadre pédagogique pour faire découvrir des méthodes de fouilles de données à des étudiants en formation de Master en Geoinformation. Cette usage a montré que l'outil RapidMiner nécessite un temps d'apprentissage. Malgré le caractère intuitif de l'utilisation des boîtes, les chaînes de traitements comportent de nombreux éléments « techniques » pour la transformation des données qui nécessitent une bonne compréhension du fonctionnement du logiciel. Une fois ce temps d'apprentissage passé, GeoDM a permis d'expérimenter différentes méthodes de fouille de données sur des problématiques réelles d'analyse d'images de télédétection. En pratique, les chaînes de traitements mises en place permettent de répondre à la question récurrente de choix du classifieur.

Seule un sous-ensemble limité des méthodes de fouille de données spatiales (Miller et Han, 2009) ont été actuellement intégrées à l'extension GeoDM. La version actuelle de l'extension est orientée vers cette tâche de classification d'images

de télédétection. La modularité de RapidMiner et la généricité de l'approche de spatialisation des données assurent l'intégration future d'un grand nombre de méthodes sous la forme de nouveau blocs ou méta-blocs sans avoir à réimplémenter les outils de base de l'apprentissage et de la fouille de données.

Remerciements :

Les auteurs souhaitent remercier Benjamin Bottner de l'Institut d'Aménagement de la Vilaine pour avoir fournis les images ainsi que Jacques Haury pour son aide à la compréhension des problématiques liées à la Jussie.

Bibliographie

Berthold M., Cebron N., Dill F., Gabriel T., Kötter T., Meinl T., Ohl P., Sieb C. Thiel K., Wiswedel B. (2007) KNIME: The Konstanz Information Miner,Studies in Classification, Data Analysis, and Knowledge Organization, Springer

Boardman J. W., Kruse F. A., (1994) Automated spectral analysis: a geological example using AVIRIS data, north Grapevine Mountains, Nevada, Proceedings of the Tenth Thematic Conference on Geologic Remote Sensing, pp. 407-418.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. (2001) Introduction to Distance Sampling: Estimating Abundance of Biological Populations. Oxford University Press. 432pp

Burget R., Karasek J., Smékal Z., Uher, V., Dostal, O. (2010) Rapidminer image processing extension: A platform for collaborative research. In proceedings of the 33rd International Conference on Telecommunication and Signal Processing, p. 114-118.

Curk T., Demšar J., Xu Q., Leban G., Petrovič U., Bratko I., Shaulsky G., Zupan B. (2005) Microarray data mining with visual programming. Bioinformatics, 1;21(3):396-398.

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

Inglada, J., Christophe E. (2009) The Orfeo Toolbox remote sensing image processing software, Geoscience and Remote Sensing Symposium, Volume 4, pp.733-736.

Mierswa I., Wurst M., Klinkenberg R., Scholz M., Euler T. (2006). YALE: Rapid Prototyping for Complex Data Mining Tasks, Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Miller H., Han J. (2009) Geographic Data Mining and Knowledge Discovery, Second Edition (Chapman & Hall/CRC).

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. (2011) Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12:2825-2830.

Visualisation de données relationnelles

Intégration des graphes spatiaux dans QGis

Thomas Guyet

Laboratoire d'Informatique, AGROCAMPUS-OUEST/IRISA UMR 6074
65 rue de Saint-Brieuc, F-35042 Rennes, France
thomas.guyet@agrocampus-ouest.fr

RÉSUMÉ.
Cet article présente l'implémentation d'une extension QGis pour la visualisation et la construction interactive de graphes spatiaux. Les graphes spatiaux sont des représentations de l'information spatiale sous la forme d'objets spatiaux reliés entre eux par des relations (spatiales ou non). Cette représentation est adaptée à la modélisation et à l'analyse d'information spatiale par des traitements informatiques (fouille de données, recherche de plus court chemins, etc). L'utilisation des graphes spatiaux pâtit de l'absence d'outils facilitant la construction et la visualisation intégrée. Dans cet article, on présente un extension QGis introduisant un nouveau type de couche, les GraphLayer. Ces nouvelles couches peuvent être intégrées dans les projets SIG. Elles offrent des fonctionnalités riches de visualisation et d'édition interactive.

ABSTRACT. This article presents the implementation of a QGis plugin for the visualization and the interactive construction of spacial graphs. Spatial graphs are accurate representations of spatial information through spatial objects linked by relationships (spatial or not). This representation is suited to the modeling and analysis of spatial information by computer processing (data mining, search for shortest paths, etc.). The use of spatial graph suffers from the lack of tools to facilitate the construction and integrated visualization. In this paper, we present a QGis plugin introducing a new type of layer: GraphLayer. These new layers can be integrated into any GIS projects. They offer rich functionality for visualization and interactive editing.

MOTS-CLÉS : Graphes spatiaux, information relationnelle, analyse visuelle , couche de graphe

KEYWORDS: Spatial graphs, relational information, visual analysis, graph layer

1. Introduction

La représentation de données est un élément important pour leur analyse (Shneiderman, 2001). Là où une mauvaise représentation noie l'information dans la masse des données, une représentation adéquate facilite l'identification d'éléments intéressants et rend possible leur interprétation. Ce problème de la représentation des données se pose aussi bien pour leur analyse automatique (par des algorithmes d'extraction de connaissance) que pour une analyse par un expert qui visualise les données (visual data mining). Dans le premier cas, il s'agit d'une question de représentation informatique de données, dans le second cas, il s'agit d'une question de représentation visuelle. L'analyse visuelle des données est facilité par la capacité des outils de visualisation à exprimer le maximum de contenu informationnel des données.

Les données spatiales se prêtent facilement à la visualisation et les systèmes d'information géographiques (SIG) offrent de grandes capacités de visualisation de l'information spatiale (Dodge *et al.*, 2008). Les questions relatives à la représentation des données font ainsi partie intégrante des problématiques de la géomatique (Miller, 2003). Le problème le plus largement discuté pour la représentation visuelle d'une carte est celui de la symbologie, c'est-à-dire le choix du mode de représentation visuelle des entités de la carte. Les évolution technologiques ont conduit récemment à poser de nouvelles questions de représentation d'information dans une carte : la représentation de l'information dynamique (évoluant dans le temps) (Harrower et Fabrikant, 2008) ou encore la représentation de l'information en 3D (Zlatanova *et al.*, 2004).

La représentation de l'information relationnelle dans les données spatialisées reste, quant à elle, assez peu étudiée dans le domaine des SIG. On entend ici par information relationnelle, des informations qui mettent en relation des entités (spatialisées ou non). Ces relations peuvent être de différentes natures : relations topologiques, fonctionnelles ou logiques (Guyet, 2010). La représentation pose à la fois la question du codage informatique d'une information relationnelle entre entités spatialisées et sa visualisation (représentation visuelle) dans un SIG. Actuellement, les seules possibilités de visualisation de l'information relationnelle consistent à détourner l'usage normal des outils de visualisation de SIG (représentation de données vectorielles). L'analyse visuelle de ce type de données est limité par l'absence d'outil dédié à leur visualisation. Il nous apparaît alors essentiel d'intégrer cette nouvelle possibilité dans les SIG actuels.

On s'intéresse dans cet article à visualiser l'information relationnelle dans un SIG en s'appuyant sur des représentations existantes de l'information relationnelle. On utilise une représentation de l'information relationnelle sous la forme d'un graphe et

on présente une extension de QGis[1] pour la construction et la visualisation de graphes spatiaux. L'objectif de ce travail est d'introduire un nouveau type de couche de données, les couches de graphes spatiaux, spécifiquement dédié à la visualisation d'informations relationnelles dans un projet SIG. Ce type de couche ajoute des capacités de visualisation, d'analyse automatique et d'analyse visuelle de l'information relationnelle.

2. État de l'art de la représentation de l'information relationnelle dans les SIG

Dans le domaine de la géographie et de l'environnement les entités spatiales ne peuvent être vues comme indépendantes, il est primordiale d'appréhender l'espace étudié dans son ensemble et donc de prendre en compte les relations qui lient les entités qui composent l'espace.

Les modèles proposés par (Pech, 2005) et (Guyet, 2010) sont basés sur une représentation de l'information spatiale à l'aide de trois sortes de relations binaires : les relations topologiques, fonctionnelles ou logiques. Ces deux modèles sont orientés vers des tâches de fouille de données. Les relations fonctionnelles incluent des relations spécifiques à un domaine d'application. Il peut s'agir de relations liées à la géographie comme des relations de transfert hydrologiques entre parcelles agricoles, ou bien de relations plus abstraites comme des relations d'appartenance à une exploitation agricole (une entité pouvant représenter une exploitation agricole). (Degenne *et al.*, 2010) introduit Ocelet, un outil pour la définition et la représentation d'informations relationnelles (dynamique) en environnement. Le formalisme de représentation d'Ocelet est plus riche que les approches précédentes. Les relations ne se limitent pas aux relations binaires. De part cette expressivité, Ocelet est orienté par la représentation des connaissance plutôt que par l'analyse des données représentée.

Les relations topologiques ont particulièrement été étudiées. D'un point de vue algébrique, les formalisme de représentation des relations spatiales qualitatives tels que le RCC8 (Region Connection Calculus) propose des outils de représentation et de raisonnement sur les relations topologiques (Renz, 2002). Le domaine des ontologies s'est également particulièrement intéressé à l'organisation des relations topologiques. (Katz et Grau, 2005) ont ainsi traduit la RCC8 dans le formalisme OWL en proposant OWL-DL. Le système ONTOAST (Miron *et al.*, 2007) a étendu le système AROM-ST pour représenter les relations topologiques mais également des informations d'orientation et de distance. L'Open Geospatial Consortium (OGC) a définit GeoSPARQL, une ontologie des relations topologiques des relations.

[1] QGis est un Système d'Information Geographique *opensource* largement utilisé, voir http://www.qgis.org/

Dans notre approche, on utilise les graphes comme support de la représentation de l'information relationnelle. Les formalismes qui utilisent des graphes sont contraints par la représentation exclusive de relations binaires. Toutes les relations du RCC8 sont des relations binaires et peuvent donc être représentées par un graphe. Les relations d'ordre supérieurs ne peuvent être représentées, mais sont peu utilisées. L'utilisation de graphes spatiaux est principalement orientée vers la représentation de réseaux (transports, électriques, etc.), de relations d'interactions (pour modéliser des informations en économie) ou de flux (flux migratoires, flux hydrologiques, etc.).

Peu de travaux s'intéressent à la visualisation de graphes spatiaux. (Pech, 2005) visualise ses graphes au moyen de l'outil GraphViz (Ellson *et al.*, 2003) spécialisé dans la représentation de graphes, mais ne permettant pas de représenter l'information spatiale tel que les SIG. Ocelet offre la possibilité de visualiser ses informations au sein de Google Earth à l'aide du format KML non-spécifique aux graphes. Des outils de visualisation tels que MBone, Constellation ou FlowMap sont utilisés pour représenter des graphes de flux (Munzner, 2000). À notre connaissance, il n'existe pas d'outils intégré à des SIG pour l'affichage générique de graphes. ArcGis comprend une extension pour l'analyse de réseaux de transports avec des fonctionnalités de recherche de plus court chemin. L'extension Roadgraph de QGis est également disponible pour le même type de fonctionnalité. Ce type de représentation ne nous semble pas approprié pour toutes les informations relationnelles. Notre objectif a donc été de proposer l'intégration de la visualisation de l'information relationnelle par des graphes spatiaux dans QGis.

3. Extension QGis pour la visualisation de graphes

3.1. Présentation générale

L'extension QGis pour la visualisation introduit un nouveau type de couche spécifique aux graphes. Chaque couche correspond à un fichier décrivant un graphe et à une modalité d'affichage (*cf.* section 3.2). Une couche ne contient qu'un seul graphe, mais ce graphe n'est pas nécessairement entièrement connecté.

Une couche de graphe peut être sauvegardée et chargée à l'intérieur d'un projet. On fournit également des outils pour l'édition de graphes (*cf.* section 3.3). Ces fonctions d'éditions permettent de créer un graphe directement dans QGis et de le modifier.

Le modèle d'un graphe est proche de celui proposé par Pech (Pech, 2005) et est illustré dans la Figure 1, les éléments principaux caractérisant un graphe sont les suivants :

– un graphe est défini par un ensemble de nœuds, d'arcs et il définit les attributs qui vont décrire les nœuds et les arcs, ainsi que le système de coordonnées dans

lequel sont exprimés les coordonnées du graphe. Le graphe dispose également d'une liste d'attributs/valeurs qui donnent des informations à l'échelle du graphe ;

– un nœud (node) est décrit par une liste de valeurs qui correspondent aux attributs qui le caractérisent. Un nœud est localisé dans l'espace, il précise les coordonnées auxquelles il doit être positionné ;

– un arc (edge) est décrit par ses deux nœuds d'extrémités (source et destination). Un arc est donc implicitement dirigé, mais il peut être considéré comme non-dirigé. Comme les nœuds, un arc porte également des valeurs pour les attributs qui le décrivent ;

– un attribut qui caractérise un nœud ou un arc peut être d'un des trois types suivants : des nombres entiers, des nombres réels et des chaînes de caractères. L'information relationnelle modélisée peut donc être qualitative ou quantitative.

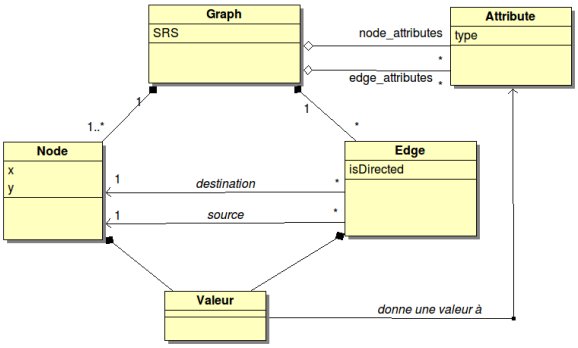


Figure 1. Schéma UML illustrant le modèle d'un graphe

3.2. Fonctionnalités de visualisation

Un graphe est visualisé spatialement par l'affichage des nœuds et des arcs. La visualisation des nœuds et des arcs peut être configurée pour répondre aux besoins de la représentation de l'information portée sur le graphe. Les caractéristiques géométriques suivantes peuvent être adaptés :

– caractéristiques des nœuds : taille d'affichage, largeur de la bordure, couleur de bord et couleur de fond

– caractéristiques des arcs : largeur de trait, couleur de trait, style de ligne.

Les fonctionnalités de visualisation permettent de définir chacune de ces propriétés pour l'ensemble des nœuds/arcs ou pour chaque élément individuel en fonction d'une valeur portée par celui-ci. La Figure 2 illustre l'édition de la propriété de couleur de fond d'un nœud. La couleur dépend ici de la valeur de l'attribut *cl*.

Cette répartition de couleur a été obtenue automatiquement par classification des valeurs.

Le mécanisme de gestion de l'affichage permet de configurer plusieurs propriétés de visualisation en même temps. On peut, par exemple, faire modifier la taille d'un nœud en fonction d'un premier attribut et sa couleur en fonction d'un autre.

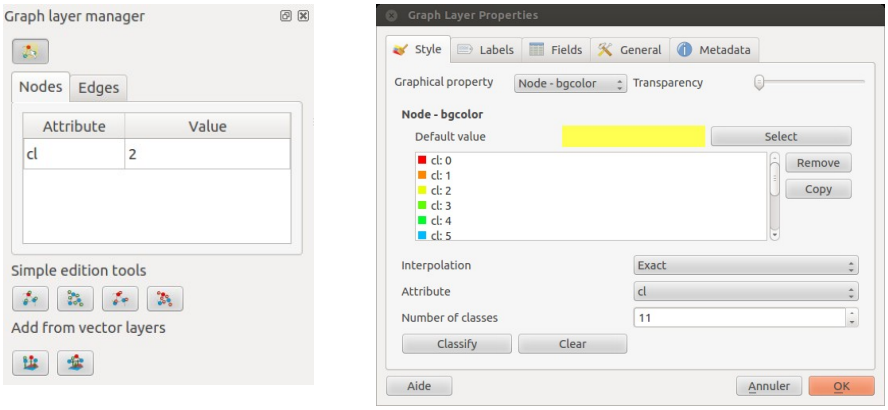


Figure 2. Sur la gauche, boîte à outils pour la consultation et l'édition d'une couche de graphe. Sur la droite, fenêtre de configuration de l'affichage d'un graphe. La fenêtre illustre la configuration de la couleur de fond d'un nœud à partir de l'attribut cl.

3.3. Fonctionnalités d'édition

L'édition des couches de graphes comporte deux types d'opérations :

– les opérations de modification manuelle des éléments du graphe : ajout/suppression d'un nœud, ajout/suppression d'un arc entre nœuds, modification des valeurs des attributs.

– les opérations de construction d'un graphe à partir de données provenant de couches vectorielles. Ces opérations permettent, d'une part, de construire des nœuds à partir d'une couche vectorielle en utilisant les centroïdes des géométries qui la composent et en récupérant les attributs. D'autre part, une opération similaire permet de récupérer les attributs d'une couche vecteur lors de la création d'arcs entre deux nœuds.

4. Exemples d'utilisation

On illustre maintenant la construction et la visualisation des graphes obtenus dans QGis sur deux exemples de cartes.

La carte de la Figure 3 représente un graphe de proximité entre les parcelles cadastrales dans un environnement urbain. Un arc indique une distance entre deux centroïdes de parcelles cadastrales inférieures à 100m. En milieu urbain, le nombre de parcelles répondant à ce critère de proximité peut être très important. Pour rendre le graphe plus lisible, nous avons construit uniquement les 5 arcs les plus proches. Cette couche de graphe a été obtenue intégralement par l'utilisation des fonctionnalités automatiques de notre extension QGis.

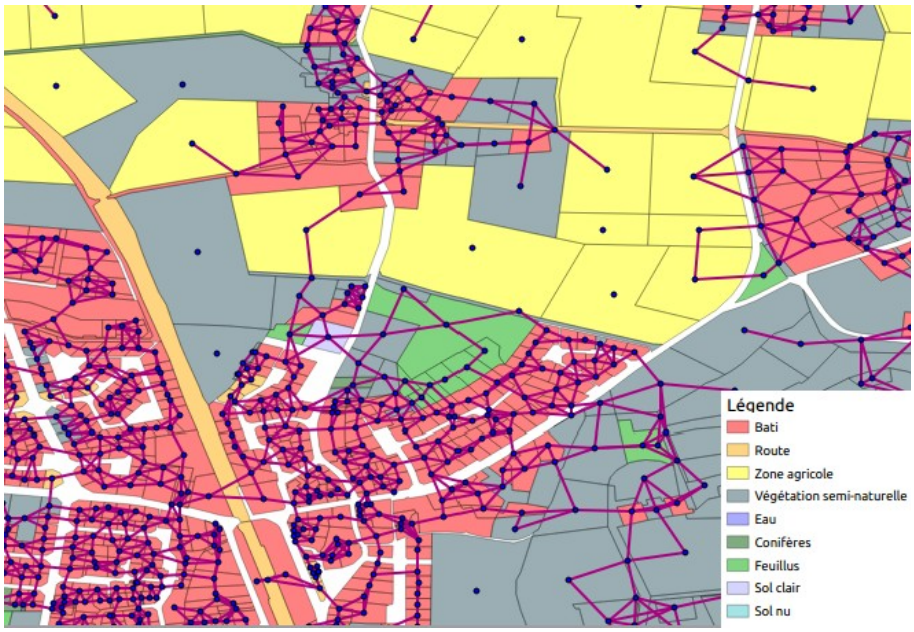


Figure 3. Illustration d'un graphe de proximité entre parcelles cadastrales. Chaque nœud représente une parcelle. Elle est connectée à au plus 5 parcelles distantes de moins de 100m.

Dans la Figure 4, nous illustrons les propriétés d'affichage de notre extension en représentant un graphe de connectivité entre les parcelles agricoles : un arc est indiqué lorsque deux parcelles sont connexes. Le graphe ne concerne que les parcelles agricoles dont l'information est portée par les nœuds. Les arcs représentent l'interface entre les parcelles connectées : type de haie ou absence de particularité (en bleu).

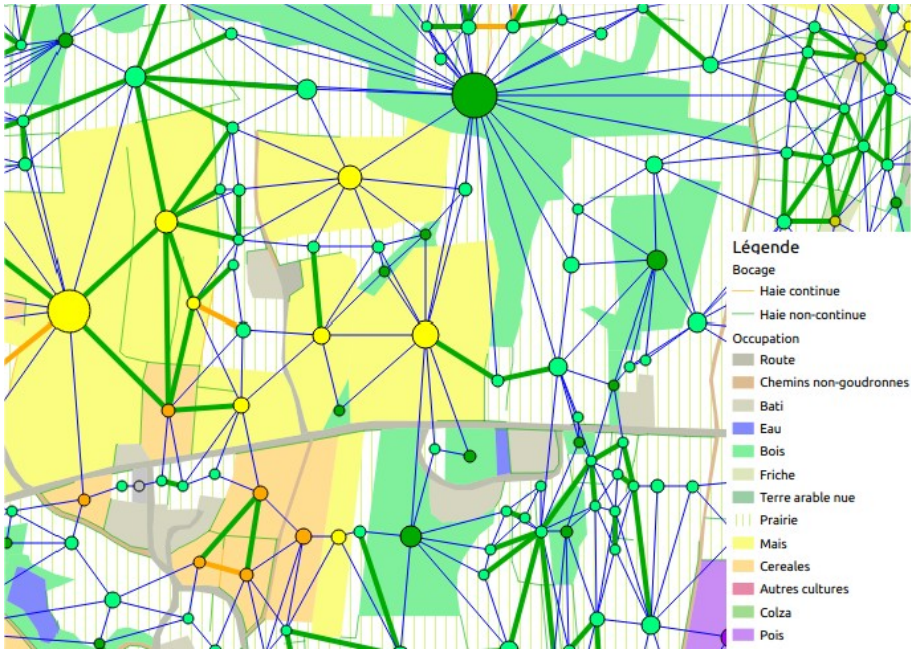


Figure 4. Graphe de connexité entre parcelles agricoles. La taille des nœuds est proportionnelle aux tailles de parcelle. La couleur des nœuds correspond à l'occupation du sol. La couleur des arcs indique le type d'interface.

5. Perspectives et conclusion

Dans ce papier, nous avons présenté un outil de visualisation de graphes spatiaux intégré dans QGis sous la forme d'un nouveau type de couche. L'intérêt de ce nouveau type de couche est de représenter une information relationnelle entre les entités spatiales.

La version du logiciel présentée dans cet article, implémentée en C++ sur la version de développement de QGis (v1.9.0), est une version préliminaire qui pourra être mise à disposition sur demande afin d'en étendre le modèle de données et les outils de génération par rapport aux usages qui peuvent être attendus d'un tel outil. À terme, il offrira également à chacun la possibilité de proposer ses propres traitements de graphes.

Ce travail s'inscrit dans un projet de réalisation d'un outil d'analyse automatique de graphes spatiaux. Dans un premier temps, des outils d'analyse des propriétés du graphe (statistiques sur les nœuds et les arcs du graphes : distribution des connectivités des longueurs d'arcs, etc.) permettront aux utilisateurs d'extraire des informations sur les caractéristiques générales du graphe. Dans un second temps, des

outils de fouille de données s'intéresseront à identifier des structures intéressantes dans le graphe : structures informationnelles, p. ex. SUBDUE (Holder *et al.*, 2002), ou structures fréquentes (Kuramochi et Karypis, 2002) ou bien de classer les graphes selon leurs structures (Horvàth *et al.*, 2004). Finalement, des outils d'interrogation riches permettront aux utilisateurs d'explorer de manière interactive les données. L'intégration de ces fonctionnalités dans QGis doit permettre de les rendre plus accessibles aux gestionnaires des données.

Remerciements

L'auteur tient à remercier l'unité SAD-Paysage de l'INRA et le laboratoire COSTEL du CNRS pour le parcellaire agricole et les données de bocage. Les données de zone urbaine ont été acquises par télédétection par le laboratoire Agrocampus-Ouest/INRA SAS dans le cadre du projet PayTal.

Bibliographie

Degenne P., Ait Lahcen A., Curé O., Forax R., Parigot D., Lo Seen D., (2010) Modelling the environment using graphs with behaviour: do you speak Ocelet?, International Congress on Environmental Modelling and Software, pp. 8.

Dodge M., McDerby M., Turner M. (2008) The Power of Geovisualization. In: Dodge, M., McDerby, M., and Turner, M., (eds.), Geographic Visualization: Concepts, Tool, and Applications, Wiley, pp. 1-10.

Ellson J., Gansner E., Koutsofios E., North E., Woodhull G. (2003) Graphviz and dynagraph – static and dynamic graph drawing tools, Graph Drawing Software, Springer-Verlag, pp. 127-148.

Guyet T (2010) Fouille de données spatiales pour la caractérisation spatiale de paysages en lien avec des fonctionnalités agro-écologiques, Spatial Analysis and GEOmatics, pp. 3.

Harrower M., Fabrikant S. I. (2008) The Role of Map Animation in Geographic Visualization. In: Dodge, M., Turner M., and McDerby, M., (eds.), Geographic Visualization: Concepts, Tools and Applications, Wiley, pp. 49-65.

Holder, L., Cook, D., Gonzalez, J., Jonyer, I. (2002). Structural pattern recognition in graphs. journal of Pattern Recognition and String Matching, vol. 13, pp. 93.

Horváth, T., Gärtner, T., & Wrobel, S. (2004). Cyclic pattern kernels for predictive graph mining. In proceedings of the conference on Knowledge discovery and data mining, pp. 158-167.

Katz Y., Grau B. (2005), Representing Qualitative Spatial Information in OWL-DL, Proceedings of OWL: Experiences and Directions.

Kuramochi, M., Karypis, G. (2002). Discovering frequent geometric subgraphs. In proceedings of International Conference on Data Mining, pp. 258-265.

Miller H., Wentz E. (2003), Representation and Spatial Analysis in Geographic Information Systems, Annals of the Association of American Geographers, 93(3):574–594

Miron A. D., Gensel J., Villanova-Oliver M., Martin H. (2007) Towards the Geo-spatial Querying of the Semantic Web with ONTOAST, Actes de W2GIS, pp. 121-136.

Munzner T. (2000) Interative visualization of large graphes and networks. Thèse de doctorat. Standford University.

Pech Palacio, M. (2005) Spatial Data Modeling and Mining using a Graph-based Representation, These de Doctorat, INSA Lyon.

Renz, J. (2002) Qualitative Spatial Reasoning with Topological Information, LNCS, Vol. 2293, Springer-Verlag.

Shneiderman B. (2001) Inventing Discovery Tools: Combining Information Visualization with Data Mining, Information Visualization(2001), In Discovery Science. Vol. 2226, pp. 17-28.

Zlatanova S., Rahman A. , Shi W. (2004) Topological models and frameworks for 3D spatial objects, Journal of Computers & Geosciences, Vol. 30, No. 4, pp. 419-428

SESSION 7 :

Traitements d'images

Change Detection in Floodable Areas of the Danube Delta using Radar Images

Simona Niculescu[1], Cédric Lardeux[2], Grégoire Mercier[3], Laurence David[1]

1.Laboratoire LETG-Brest, Géomer - UMR 6554 CNRS
Rue Dumont d'Urville, Technopole Brest Iroise, F-29470 Plouzané, France
simona.niculescu@univ-brest.fr

2. Office National des Forêts
2, avenue de Saint-Mandé, 75570 Paris Cedex 12
clardeux@gmail.com

3. Institut Télécom, Télécom Bretagne, Lab-STICC/CID - UMR 6285 CNRS
Technopole Brest-Iroise; CS 83818
F-29238 Brest cedex – France
gregoire.mercier@telecom-bretagne.eu

RESUME

Dans le delta du Danube, le risque d'inondation est un risque majeur. Depuis les années 2000, la zone humide côtière du delta a été frappée par des inondations en 2002, 2005, 2006 et 2010. En partant d'une série de données hydrologiques et d'observations satellite de 2009 et de l'été 2010, on utilise dans cet article l'information et les observations sur l'occupation du sol, notamment dans le domaine de l'inondable, pour étudier la problématique de la prévision du risque. L'un des objectifs principaux de la méthodologie développée consiste en la mise en œuvre de différents types de données pertinentes pour ce type d'analyse en proposant une chaîne méthodologique spécifique pour la détection du changement en termes de risque d'inondation. Les données du satellite japonais ALOS sont utilisées pour présenter une démarche méthodologique de traitement de données radar multidates fondée sur une analyse de l'entropie temporelle permettant de détecter des changements dans les zones inondables du delta du Danube.

ABSTRACT

In the wetlands of the Danube delta floodplain, flooding is a major natural risk. Since 2000, the coastal wetland has been seriously impacted by floods in 2002, 2005, 2006 and 2010. Using hydrological and satellite observations acquired in 2009 and during the summer of 2010, this paper tackles the issue of forecasting risk based on land cover information and observations. A major objective of this methodological work consists in exploring several types of data from the Japanese ALOS satellite. These data are used to illustrate a

multitemporal radar data processing methodology based on temporal entropy analysis enabling change detection in the floodable areas of the Danube delta.

MOTS-CLES : *delta du Danube, risque d'inondation, détection du changement, entropie temporelle, satellite ALOS, prévision, zone inondable.*

KEYWORDS: *Danube delta, flood risk, change detection, temporal entropy, ALOS satellite, forecasting, floodable areas*

1.Introduction

Assumptions related to global warming and its effects engender significant preoccupations with regards to the rise in heavy rainfall occurrences despite numerous doubts that still exist concerning rainfall response to global warming. In this respect, climate change may increase flooding and other risks throughout the whole water cycle in the years to come. Extreme weather conditions such as heavy rains causing flash floods are expected to become more frequent throughout Europe. Climate change is also expected to lead to a rise in the sea level (expected to be from 18 to 59 cm in 2100, IPCC, 2007) and shoreline erosion and, when accompanied by strong storm waves, will threaten low altitude coastal towns, which will be more likely flooded. According to GIEC, this phenomenon will probably affect up to 1.6 million more Europeans until 2070. The rise in sea level will also be damaging for the coastal wetlands.

The geographical landscape dealt with in this paper refers to deltas, i.e. recently formed ever-changing environments that are part of the shoreline and wetlands and that are influenced by absent or weak tides. These are generally characterized by salinity variations –freshwater from the river on the one hand and marine saltwater on the other hand – as well as by considerable biomass and productivity, due mainly to the abundance of nutrients present in the water and sediments. Despite their indisputable importance, these environments remain among the most endangered ecosystems in the world due to land reclamation, drying out, pollution or overexploitation of resources. The Danube delta, a coastal wetland of the Black Sea, cannot escape these dangers and, to preserve its resources, it has been declared a Biosphere Reserve (in 1993). The Danube delta is a representative flood hazard site since it has been affected by climate-related floods and hence by systematic rises of the water level.

Radar satellites are extremely useful for characterizing and mapping flood risks due to their ability to collect data and images through the cloud cover that usually accompanies floods. The purpose of this paper is to test the Japanese ALOS satellite, specifically the PALSAR microwave radar (*Phased Array type L-band Synthetic Aperture Radar*) with 15 m resolution permitting the discrimination of floodable areas. Change detection by temporal entropy based on radar images is the methodology proposed in this paper.

2. Study area

The Danube crosses Europe from Germany to the Black Sea where it forms the second largest delta in Europe after the one at the mouth of the Volga. The Danube delta area covers 4 455 km^2 and 79% (i.e. 3 510 km^2 without the Razim-Sinoe lagoon complex) of the delta area is located in southeastern Romania (figure 1). The Romanian portion of the delta is divided between the inland water area and the maritime area (figure 1). The inland section is located westward from a line connecting the Letea dunes (Periprava Village) with the Dunavat peninsula. The maritime section begins at the alignment of the Letea Caraorman and Crsanicol marine levees and runs all the way to the Black Sea. Our study site is located around Caraorman: the Caraorman fishing village, where fishermen berth their boats, the village and forest bearing the same name (figure 1). The village was set up on the Caraorman levee, 12 m above sea level.

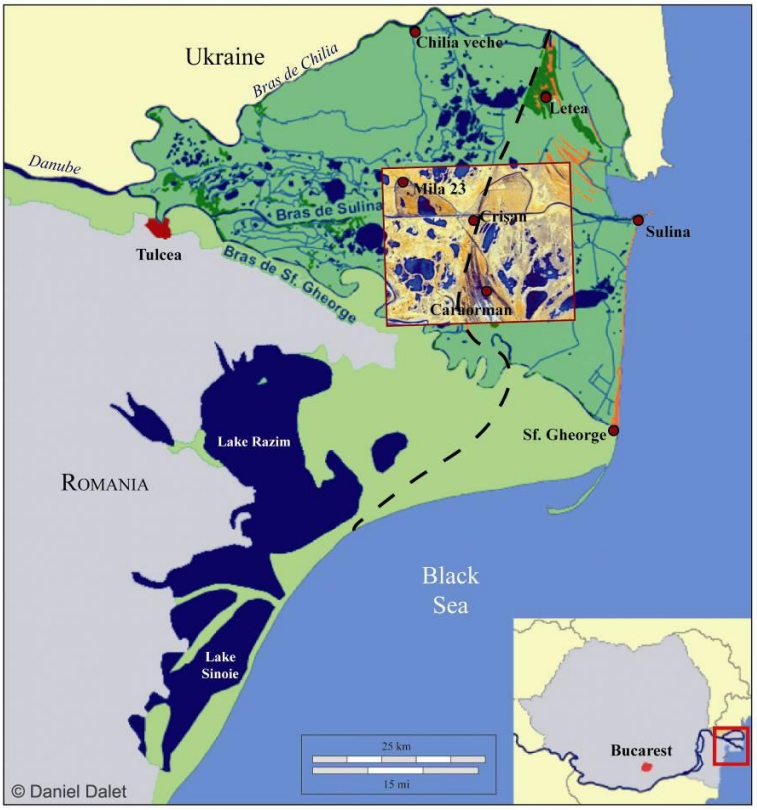


Figure 1. The Danube delta and geographical location of the training sample

3

At the outset, the hydrographic network of the delta is formed by three main arms: the Chilia arm (120 km long), which forms the natural border between Romania (south) and the Ukraine (north), the Sulina arm (64 km long) in the middle and the Sfantu-Gheorghe arm to the south, currently 70 km long. The delta plain is characterized by minor altitude differences: 20.5% of the delta is below the 0 level of the Black Sea, 54.6 % is between 0 and 1 m, 18.2% is between 1 and 2 m and only 6.7% is above 2 m of altitude. The low-lying areas between the Chilia-Sulina and Sulina-Sfantu-Gheorghe arms harbor many delta lakes taking up 8.9% of the whole delta area (Gastescu P. and Stiuca R., 2008). These lakes play an important role during rises in the Danube level, as the excess water accumulates here and then flows into the main arms when the water level is low. Another characteristic of this deltaic environment is the existence of ecosystems consisting of reed beds, mainly *Phragmites australis*. These reed beds make up a considerable part of the floodable delta and they are also thought to be the largest uninterrupted reed beds in the world.

Despite the moderate rainfalls in this area, the Danube delta is flooded every year. The historical analysis of the maximum discharge of the Danube (figure 2) in Tulcea reveals that considerable river discharges were recorded in 1895 (13 700mc/s), 1942 (13 387mc/s), 1970 (14 520mc/s), 1974, 2003, 2006 (15 900 mc/s) and 2010 (16 600mc/s). The annual multi-analysis (1931 – 2011) of the Danube water levels at the entrance of the delta (in Tulcea) (figure 2) also shows that most of the maximum river discharge episodes occur from March to July. The last two major flood rises, which occurred in 2006 and 2010 (figure 2), were preceded by heavy snowfalls during the winter and by relatively early snow melting on the mountains in the catchment area of the Danube. In July, the Danube level decreases significantly down to low levels and discharge rates at the end of summer and beginning of autumn (from August to February) (figure 2). According to these data, the floods of 2006 and 2010 were followed by periods of considerable low water (figure 2): in Tulcea, the Danube level was 1 m in August 2007 and 0.35 m in October 2011.

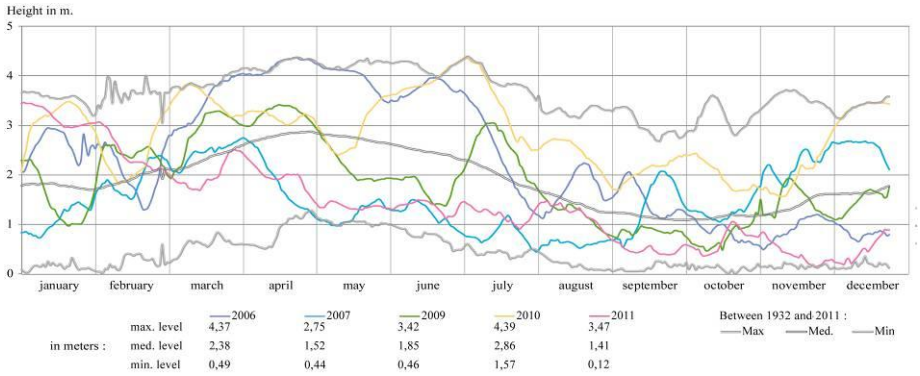


Figure 2. Danube daily levels at Tulcea in 2006, 2007, 2009, 2010 and 2011. Mean daily levels from 1931-2011.

4

The floods of 2010 affected more than 30 counties and 400 localities in the northern and eastern portions of Romania, the Danube delta included. More than 16 500 people were evacuated and more than 45 000 hectares of agricultural land and forests were submerged.

3. Data set

The satellite data included in this paper were acquired by the PALSAR instrument on the ALOS satellite. The Phased Array type L-band Synthetic Aperture Radar (PALSAR) is an active microwave sensor using the L-band frequency (24 cm) to achieve cloud-free and day-and-night land observations. PALSAR is a fully polarimetric instrument, operating in fine-beam mode with single polarization (HH or VV), dual polarization (HH+HV or VV+VH), or full polarimetry (HH+HV+VH+VV). This paper focuses on dual polarization data (dual HH and HV mode) (table 1).

The dual polarimetric mode allows image acquisition in two different linear polarization modes. The alternation between the two polarization configurations occurs on each individual impulse, which accounts for the phase difference between the two acquisitions.

Table 1. ALOS/PALSAR satellite data used

RADAR sensor instrument	Before / during / after the floods	Radar data	Technical characteristics	Danube water level at Tulcea
Japanese ALOS satellite PALSAR sensor	Before	15 June 2009	Dual mode: HH, HV L-Band: 24 cm Resolution: 15 m x15 m	2.52 m
		20 May 2010		3.70 m
	During	18 June 2010		4.37 m
	After	3 August 2010		3.06 m

Other types of data were used: the daily Danube levels in Tulcea in 2006, 2007, 2009 and 2010, the multi-annual mean of the Danube levels in Tulcea between 1931 and 2011, as well as the phenological reed calendar in the Danube delta (figure 5).

4. Methods

In order to detect the various types of surface changes resulting from the floods of 2010 in the Caraorman deltaic section and to characterize the floodable class by means of its backscattering mechanisms, we propose the following methodological analysis and image processing sequence: polarimetric indices extraction and implementation, calculation of the temporal entropy of the polarimetric indices, temporal entropy classification using support vector machine (SVM) and mono temporal floodable class classification using all the polarimetric indices (table 2) used.

Dataset		Danube water levels at Tulcea		Processing steps	Image processing	Objectives
Radar Dataset ALOS / PALSAR	06/15/2009	Normal condition (June 2009)	2.52 m	**Orthorectification** **Data filtering** **Polarimetric indices** **Temporal entropy of the polarimetric indices**	Extraction of the polarimetric indices (intensity) Calculation of the temporal entropy of the polarimetric indices Detection of change entropy of the polarimetric indices	Analysis of polarimetric change between the 2 dates Change indicators (entropy) applied to polarimetric indicators Change characterization (stable classes, changing classes)
	06/18/2010	Floods (June 2010)	4.37 m			
	Image resulting from temporal entropy			**Entropy SVM classification**	Classification into two classes: no change and radical change	Risk class detection (floodable areas): hydrophilic vegetation, halophilic vegetation psamophilic vegetation Stable class mask
	06/15/2009	Normal condition (June)	2.52 m	**Mono-temporal SVM classification**	Diachrony of the 4 dates	Floodable area detection based on the thematic mask extracted relying on the flooded areas of 18 June 2010 (Short vegetation (grassland), hydrophilic vegetation, psamophilic vegetation, mixed vegetation, mixed phragmites, forest, doinant phragmites, cut phragmites)
	05/20/2010	Flood onset	3.70 m			Flooded area detection
	06/18/2010	During	4.37 m			Flooded area detection and thematic mask of these areas
	08/03/2010	After	3.06 m			Floodable area detection and characterization

Table 2: Methodological sequence applied to ALOS/PALSAR images

6

Two major types of polarimetric indices can be extracted from fully polarized data or from dual polarization data. Coherent indices conserve the phase information while incoherent indices break up the phase information common to the resolution cells. The parameters of interest in these areas for radar modeling are basically the parameters describing the geometry and electromagnetic features of the ground vegetation and soil elements. Thus due to the incoherent nature of the study classes we will focus only on incoherent polarimetric indices.

Polarimetry relies on the vector-like nature of the wave (Papathanassiou and Cloude, 2001). It is designed to characterize targets by recording their responses in different radar wave emission and reception polarizations. In order to allow the separation of backscattering mechanisms in the same resolution cell, different target decomposition models were proposed (Touzi et al. 2002; Lee and Pottier, 2009). Their principle relies on polarimetric data projection on canonical bases associated with each type of backscattering mechanism resulting from the interactions between the radar waves and the excited targets. These methods differ depending on the coherent or incoherent nature of the targets. As far as coherent targets are concerned, we should mention Pauli decomposition (Cloude and Pottier, 1997), SDH decomposition (Krogager, 1999) and Cameron decomposition (Cameron et al., 1996). On the other hand, Cloude-Pottier (Entropie/Alpha), Van Zyl (Van Zyl, 1989, Cloude and Pottier, 1997), Freeman (Freeman and Durden, 1998) and Yamaguchi (Yajima et al, 2008) decompositions describe the non-coherent features of the so-called distributed targets (Cloude and Pottier, 1997; Touzi, 1997).

As for double polarization (fine beam dual mode), only the S_{hh} and S_{hv} terms are recorded coherently, thus reducing the polarimetric information acquired.

Starting from the K_{DP} target vector, the covariance matrix is expressed as:

$$K_L = \begin{pmatrix} Shh \\ Shv \end{pmatrix} \tag{1}$$

$$C = K_L \,.\, K_L{}^* = \begin{pmatrix} Shh \\ Shv \end{pmatrix}^{[Shh^* \,.\, Shv^*]} = \begin{bmatrix} Shh\ Shh^* & Shv\ Shh^* \\ Shh\ Shv^* & Shv\ Shv^* \end{bmatrix} \tag{2}$$

$$Shh \,.\, Shh^* = Ahh \,.\, e^{j.\varphi hh} \times Ahh \,.\, e^{-j\varphi hh} = Ahh^2 \,.\, e^{j(\varphi hh - \varphi hh)} = Ahh^2 \,.\, e^0 = Ahh^2 \tag{3}$$

$$Shh = Ahh \,.\, e^{j\varphi hh}$$

$$Shv = Ahv \,.\, e^{j\varphi hv} \tag{4}$$

Shh and Shv are the matrix terms corresponding to HH and HV polarizations. Ahh and Ahv are the amplitudes of the two polarizations, whereas φhh and φhv are their phase.

This configuration allows to extract the coherence degree of HH-HV polarizations as well as the $H_{DP}/A_{DP}/\alpha_{DP}$ parameters (S. Cloude, 2007).

7

By pointing out that polarimetric ***entropy H*** characterizes the level of depolarization of the backscattered wave, entropy is calculated pixel by pixel using a linear combination of the image cross-variance matrix values.

$$\boldsymbol{H_{DP}} = -\sum_{i=1}^{2} \boldsymbol{P_i log\ (P_i)} \quad (5)$$

Where $\boldsymbol{P_i} = \frac{\lambda i}{\sum_{i=1}^{2} \lambda i}$ and λi are the specific vectors of the coherency matrix.

In a second stage, the temporal analysis of some polarimetric indices (HH intensity, hv intensity and polarimetric entropy) using temporal entropy is carried out using this equation.

$$H_{Temp} = -\sum_{i=1}^{n} \mathbf{P_i} \boldsymbol{log}\ (\mathbf{P_i})$$

Where $\boldsymbol{P_i} = \frac{Bi}{\sum_{i=1}^{n} Bi}$ and bi correspond to the b indicator (HH, HV or polarimetric entropy in our case) observed on date n.

Therefore, temporal entropy has the same definition as polarimetric entropy, the only difference being that temporal entropy measures the changes undergone by the same index throughout time.

Thus, a low temporal entropy (close to 0) means that the indicator used is stable in time. On the other hand, a high entropy value, close to 1, shows that the indicator has undergone a significant change through time.
Moreover, we suggest a comparison between the temporal HH, HV polarization entropies and their polarimetric entropy in order to be able to characterize their changes. It is well known that HH polarization allows the characterization of the double rebound importance where HV polarization is particularly characteristic of the volume reflections. Thus, the temporal analysis of these polarizations together with the analysis of their polarimetric entropy will enable us to characterize the Danube water level evolution. Since the Danube delta is mainly covered by reeds, the water level variation and the reed density variation are characterized by a significant double rebound modulation and volume backscattering mechanism.
Thus, the temporal analysis of HH and HV entropy reveals values that can characterize the state of the changes affecting the studied areas. In some areas, before the floods of 18 June 2010, HH intensity was -7.7dB whereas HV intensity was -15.5. During the floods, HH intensity was -12.5dB versus -26.7dB for HV polarization. This shows that intensity decreases during flooding, which means that water response increases contrary to volume and double bounce scattering mechanisms. Therefore in the delta we can observe many different temporal signatures explained in figure 4 from the Results and discussion chapter

A change detection process was also conducted. Several authors have published change detection method assessments, for instance Singh (1989), Lunetta and Elvidge (1998), Mas (2000), Jensen (2004) and Lu et al. (2004), Inglada and Mercier (2007), Quin, Pinel-Puysségur, and Nicolas (2012).

Lu et al. (2004) classify the change detection methods into six categories: algebraic, transformation, classification, advanced, GIS approach, visual analysis and other techniques.

Therefore, we used the supervised SVM classification algorithm in addition to temporal entropy analysis in order to detect changes and to isolate and characterize the type of area subjected to flooding.

A supervised classification algorithm was then applied. The algorithm used was SVM (*Support Vector Machine*) based on the core theory, which allows the settlement of problems linearly inseparable. It also allows the taking into consideration of numerous heterogeneous parameters. The purpose of the SVM algorithm is to calculate the best hyperplane separating two classes or, similarly, to maximize the margin between the study classes, more details are given in Burges C-J, 1998. Thus, SVM are particularly relevant because SVM do not take into account any distribution assumption. In addition, using kernel RBF[1] of SVM allow classifying non linear classification problem.

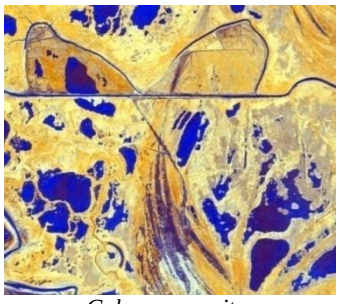
The presence of noise in the data can be taken into account by defining a distance allowing a certain data dispersion, thus relaxing the decision-making constraint (Lee J. S., Grunes M. R, de Grandi G., 1999; Lee et al., 2009).

The SVM principle was developed for a two-class problem, yet it may be extended to a higher number of classes according to the "one against one" (OAO) algorithm. This algorithm consists in the development of $\frac{Q(Q-1)}{2}$ hyperplanes separating each pair of classes. The final label is the one that was most commonly retained.
Whereas Q is the number of classes to consider, the UCT algorithm consists in the development of Q hyperplanes separating each class from the others Q-1.
As for the core used by the SVM method, please note that after several tests, better results are obtained with the same parameterization as Fukuda and Hirosawa, 2001, with a RBF kernel, with $\sigma = 0.5$ and a cost parameter equal to 1000.

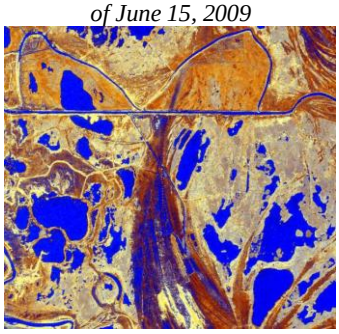
5. Results and discussion

Using two color composites (figures 3), combining HH and HV intensity with entropy, we propose the calculation of the entropy between two images, the first acquired on 15 June 2009 and the second on 18 June 2010 (figure 3), in order to detect the changes that occurred between the two dates. The 2009 date was chosen for improved mapping and comparison between the floodable class and the flooded class of 2010.

[1] The Radial Basis Function (RBF) kernel.



Color composites of June 15, 2009



Color composites of June 18, 2010

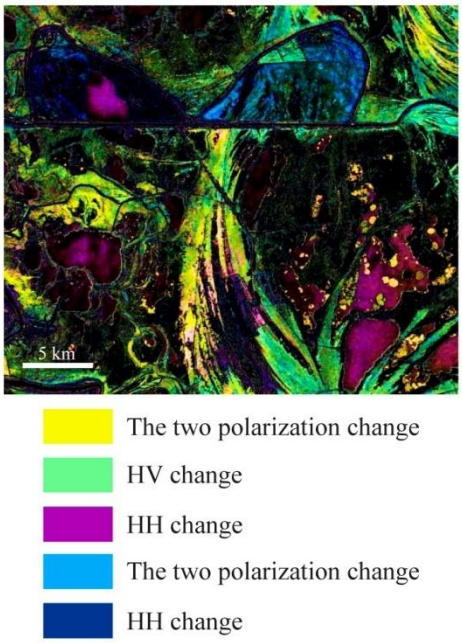


Figure 3. Temporal entropy. Change detection between June 15, 2009 and June 18, 2010.

On the resulting image (figure 3), the lighter the colors, the more numerous the changes recorded. Therefore, yellow expresses a large number of changes, indicating that the two polarizations change, green indicates a HV polarization change, whereas magenta reveals a stable HV polarization and a considerable HH change. Cyan indicates a change in the two polarizations and blue expresses HH change as compared to HV. Bearing in mind that polarimetric information is closely related to the backscattering mechanism, the changes actually reflect the backscattering mechanism changes depending on the changes affecting the studied areas (figure 4).

In order to better characterize the backscattering mechanism and hence to detect changes occurring in these areas depending on the Danube level changes and on the ground vegetation changes (*Phragmites spp.* in particular) before and during the floods, we propose an analysis of the major changes between the two polarizations in the various color composites (figure 4). The different colors used in this diagram correspond to the colors shown in figure 3 (temporal entropy).

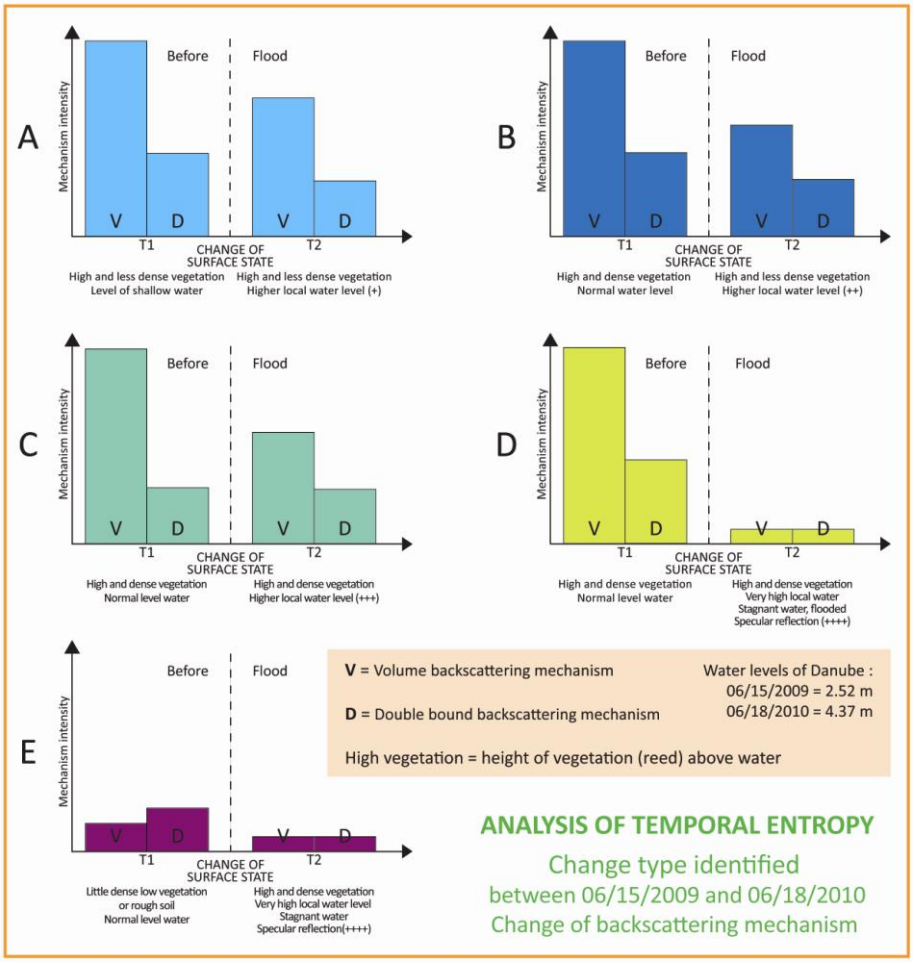


Figure 4: Analysis of temporal entropy

Relying on the field knowledge acquired in connection with the monitored areas and with the characteristics of volume backscattering and double data rebound in double polarization mechanisms, we distinguished between two types of areas in our analysis: the local water level and the vegetation. As far as the changes in the two polarizations (HH and HV) is concerned, which appears here in cyan blue (diagram A), the volume backscattering and double rebound mechanisms are more significant for the date prior to the floods. The areas associated with these backscattering mechanisms are the high and less dense vegetation and the low water level (2.5m), and the high and less dense vegetation and the higher local water level (+), respectively. The changes in the HH polarization as compared to the HV polarization, in blue (diagram B), is somewhat similar to the previous one, except for double rebound backscattering prior to the events, which is lower than the one recorded when the two polarizations changed (before and during the process). HV

polarization change, in green (diagram C), is characterized by considerable volume backscattering prior to the events, which decreases during the process (less than 50% compared to the one prior to the events). As for double rebound backscattering, it is about 30% lower than the volume backscattering prior to the events and more than 20% higher than the double rebound backscattering during the events, when the local water level is also higher (++). The two mechanisms occurring during the floods characterize a higher local water level (+++). Changes in both the volume backscattering and double rebound polarizations are equal during the floods and distinctly lower than the previous mechanisms and they characterize a very high local water level (the still water of the flooded areas) (diagram D in yellow). The changes in the HH polarization compared to the HV polarization, which remains stable, in magenta (diagram E), show very low volume backscattering prior to the events, when the vegetation is low and less dense, the soil is rough and the water level is normal. During the events, these changes are similar to the changes in the two polarizations in the same environmental conditions (high and dense vegetation and very high water level (++++)).

Reeds are an essential component of the deltaic landscape and *Phragmites australis* is the dominant species in the entire Danube delta. It is also accompanied by other hydrophilic species[2]. The combination of these plants varies and there are different plant communities. Hanganu *et al.* (2002) defined four main communities[3] that differ depending mainly on the soil types, pH, flooding conditions and salinity. The ecological dynamics of the reed beds can be characterized by several types of plant cover, each enjoying its own dynamics in the ecological sequence process. This process is mainly driven by the water levels, as it may be accelerated or stopped by events such as floods (2006, 2010) or drying outs (2007, 2011). The reeds studied by ALOS on June 15 and June 18 were in their development, regeneration and blooming stages (figure 5). The water level in Tulcea was 2.52 m on 15 June 2009 and 4.37 m on 18 June 2010.

[2] Typha angustifolia, Schoenoplectus lacustris, Sparganium sp. and Thelypteris palustris.

[3] Phragmitetum communis, Scipo-Phragmitetum, Bolboschoeno-Phragmitetum and Astero tripolii-Phragmitetum.

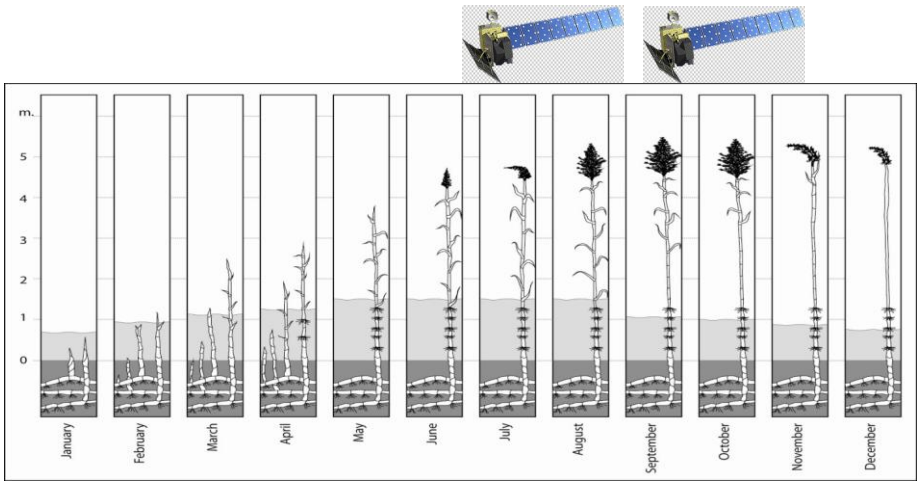


Figure 5. Correspondence between our ALOS data and the reed calendar in the Danube Delta

In a second phase, Support Vector Machine classifications are first applied to temporal entropy in order to extract 3 different classes characterizing change detection (gray and blue classes on next figure). Secondly, in order to determine thematic classes that are affected by flood, we apply an SVM classification for each PALSAR data base only on trained classes of 2009. Thus, the SVM model defined for this date could be applied to the other PALSAR data because PALSAR calibrations are well know to be good and also because we used exactly the same processing chain for each date. The stable classes (forest, delta lakes, cut phragmites, dominating phragmites) are well detected, whereas the flooding classes change the most (short vegetation (grasslands), mixed vegetation, hidrophilic vegetation, psamophilic vegetation, halophilic vegetation). The flooded class is also well illustrated by this image (figure 6).

13

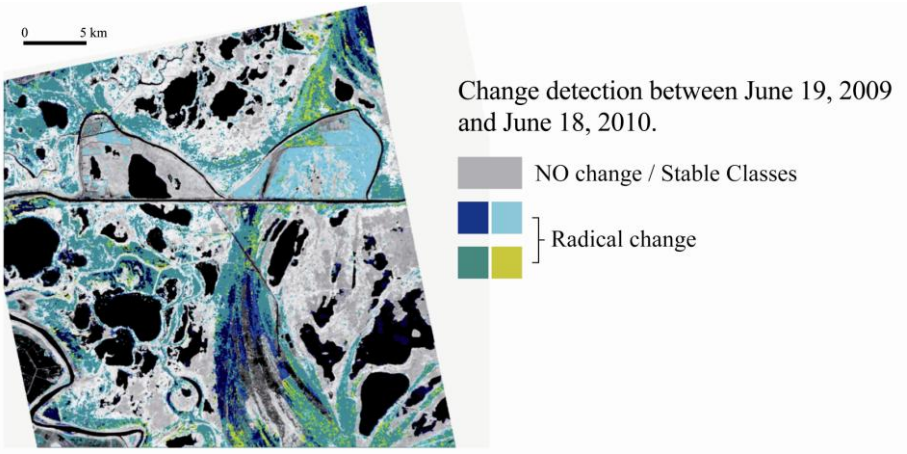


Figure 6. Change detection between June 15, 2009 and June 18, 2010

The more dynamic classes, such as the flooded and floodable classes, are approached through the diachronic analysis of the three dates considered important for these floods: 20 May 2010, 18 June and after the events of 3 August 2010 (figure 8). The date of 15 June 2009 is used to define the floodable classes in relation to the thematic mask of the flooded areas extracted from the image of June 18, when the floods reached their spatial extent peak. June 18 saw the widest flooded areas, especially in Caraorman Village, floods that were mainly due to groundwater oversaturation (rainfall floods). Therefore, as stated before, reference maps for each date were drafted, which suggested the same land cover classes (8 classes) (table 4). Differences between dates were recorded due to the fact that the local water level changed considerably between the four dates. The same thematic classes were used (table 4) but the water level altered the polarimetric response, which accounts for the differences between the floodable classes on certain dates (for instance, June 15, 2009 includes several classes in the floodable areas). Therefore, thematic masks of the floodable areas were extracted for each date and inserted into an image representing the mean of the intensity values on the dates taken into consideration (figure 7).

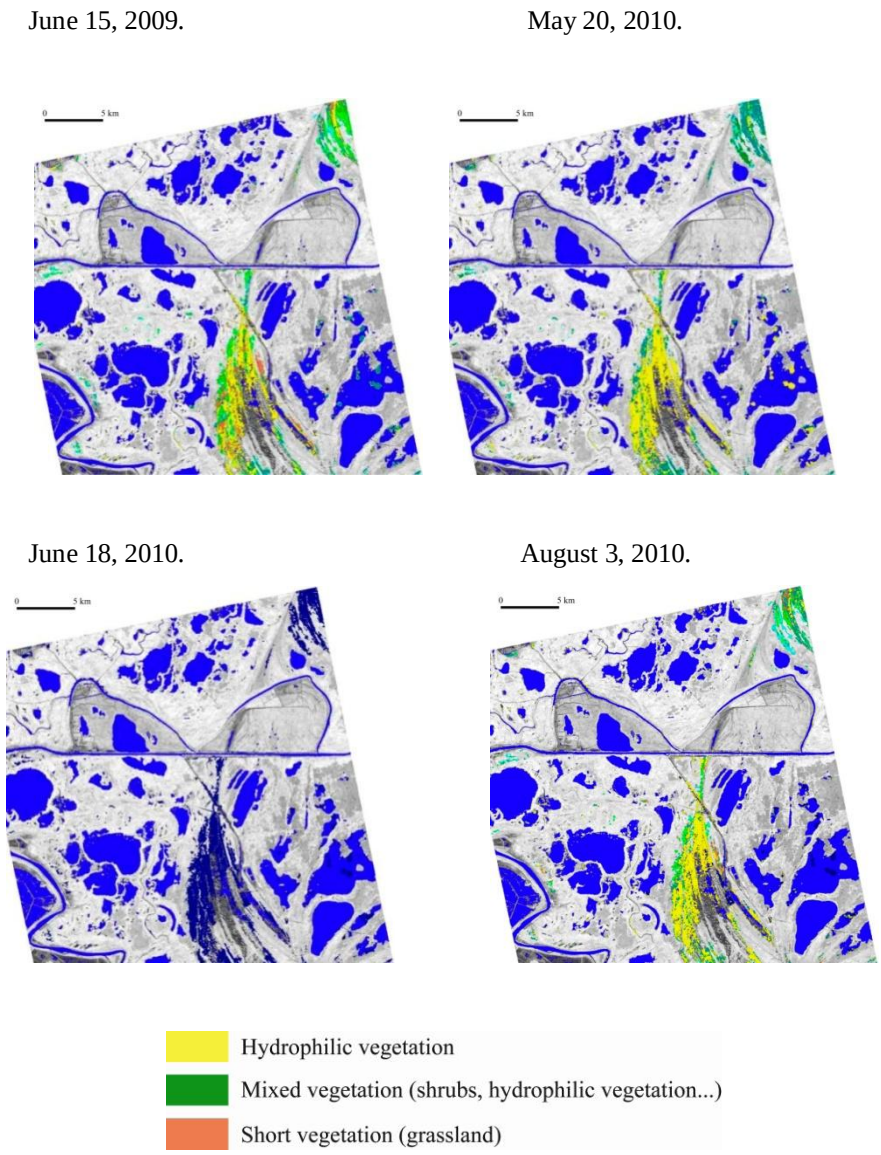


Figure 7. Diachronic evolution of the four dates: before, during and after flooding. The three images acquired prior to and after the floods represent the floodable classes according to the flooded thematic mask extracted from the image of June 18, 2010. The stable classes are in grey.

15

Examination of the different color composites permitted the definition of classes of interest for flooded and floodable class identification (table 2). We identified a forest class within the stable entropy classification classes, a mixed phragmites class that is also included in the stable classes, a less dense class represented by short vegetation included in the floodable class and classes that changed considerably between 15 June 2009 and 18 June 2010. Other land cover floodable classes were identified: the dunes and interdune depressions, hydrophilic vegetation (Typha, Carex), psamophilic vegetation (Euphorbia), halophilic vegetation (Salicornia). These classes were correlated with the flood classes and the floodable areas (table 3).

Table 3: Correspondence between classification results and flood classes

Classes	**Description**
Submerged: lakes, branches, canals	Permanent waters
Flooded	Flooded between May 20, 2010 and June 8, 2010
Floodable	Short vegetation (grassland), hydrophilic vegetation, psamophilic vegetation, mixed vegetation, mixed phragmites, forest, dominant phragmites, cut phragmites
Drained until August 3, 2010	Flooded on June 18, 2010 and drained until August 3, 2010

The creation of training and control point sets is the result of a random stratified sampling based on the whole set of areas of interest initially determined. 500 training points and 800 control points are set for each class. Results shown in table 4 reveal a very good performance of 95% for the water class, hence the interest for flood monitoring uses. We noted some overlap between short vegetation and mixed vegetation classes, where the performance does not exceed 82%. All the other classes show good performance levels (above 82%), the mean being 87%.

Table 4. Classification performance

Class	Accuracy (percent)
Forest	89
Lake	95
Mixed phragmites	94
Short vegetation	82
Typha_sp	87
Mixed vegetation	82
Cut phragmites	85

Dominant phragmites	84
Mean accuracy	87

Furthermore, work in progress leads us to believe that the performance attained by double polarization radar images in L band (PALSAR) in this study is superior to the performance obtained by optical imaging. The L band wave penetration capacity is particularly suitable for water level variation detection even underneath vegetation cover, such as reeds in the Danube delta. On the other hand, optical imaging only allows the detection of the response of the analyzed area and hence it is not directly sensitive to water level variation.

6. Conclusion and perspectives

This paper provides a methodology for detecting changes using temporal entropy applied to the evaluation of flooded and floodable areas during the Danube delta floods of 2010. Changes were examined at the shortest possible time intervals before and after a flood. These application examples suggest that change detection is possible and could play a fundamental role in the process of understanding, in a broad sense of the term, the temporal evolution of a flooded or floodable landscape via sequential analysis.

Temporal entropy calculated from polarimetric indices is an easy way of detecting changes in the areas under investigation and their comparison makes it possible to conduct a relevant qualitative analysis. Consequently, the analysis of the studied areas enabled us to detect the information most closely related to the changes of interest and least dependent on spatial variations (water level and plant cover). The use of SVM to classify change detection and to characterize the land cover, using respectively temporal entropy and polarimetric indices for each acquisition revealed the major contribution of the land cover based on floodable area classification. The use of polarimetric data helped in the detection of structural differences between the various hydrophilic plant covers and, thus, contributed to the mapping of these types of areas in relation to the water level. It also provided additional information on the spatial extent of the flooded areas.

As far as the future perspectives of our work are concerned, and in order to investigate wavelength influence for flood detection, we intend to conduct a multitemporal full polarization RADARSAT-2 data analysis. In addition, it may be interesting to draw a comparison between dual PALSAR wavelengths (L-band) and RADARSAT-2 wavelengths (C-band), and to document the complementarity of the two instruments/sensors used to detect the land cover of the floodable areas.

References

Burges C-J, 1998, *A tutorial on support vector machines for pattern recognition*, Data mining and knowledge discovery, U. Fayyad, Ed. Kluwer Academic, pp. 1-43.

Cameron, W.L., Youssef, N., and Leung, L.K., 1996, Simulated polarimetric signatures of primitive geometrical shapes, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 3, p. 793–803.

Cloude, S. R. and Pottier, E., 1997, A review of target decomposition theorems in radar polarimetry, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 1, pp. 68-78.

Cloude S., 2007, The Dual Polarization Entropy/Alpha Decomposition: A PALSAR Case Study. Proceedings : January 2007, Frascati, Italy.

Freeman, A. and Durden, S. L., 1998, A three-component scattering model for polarimetric SAR data, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, p. 963–973.

Fukuda, S. and Hirosawa, H., 2001, Support vector machine classification of land cover: application to polarimetric SAR data. *Geoscience and Remote Sensing Symposium*. 1, 187-189.

Gastescu P., Stiuca R., *Delta Dunarii. Rezervatie a biosferei*, Edition CD PRESS, 2008, 400p.

Rapport de synthèse GIEC, *Bilan 2007 des changements climatiques*, publié par le Groupe d'experts intergouvernemental sur l'évolution du climat.

Hanganu J., Dubyna D., Zhmud E., Grigoras¸ I., Menke U., Drost H., Sefan N., Sarbu, I., 2002, *Vegetation of the Biosphere Reserve "Danube Delta* - with Transboundary Vegetation Map on a 1:150,000 scale, Danube Delta National Institute, Romania; M.G. Kholodny - Institute of Botany & Danube Delta Biosphere Reserve, Ukraine and RIZA, The Netherlands. RIZA report 2002.049 2002, 89 pp.

Inglada J. and Mercier G., 2007, A New Statistical Similarity Measure for Change Detection in Multitemporal SAR Images and its Extension to Multiscale Change Analysis, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45 (5) pp 1432--1446.

Krogager, E., 1999, New decomposition of the radar target scattering matrix. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 18, p. 1525-1527.

Lee J. S., Grunes M. R, de Grandi G., 1999, Polarimetric SAR speckle filtering and its implication for classification. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, 2362-2373.

Lee, J.S. and Pottier, E., 2009, *Polarimetric radar imaging: from basics to applications*, Brian J. Thompson, New York.

Papathanassiou, K. P. and Cloude, S. R., 2001, Single-baseline polarimetric SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 11, p. 2352–2363.

Quin G., Pinel-Puysségur B, and Nicolas J.-M., 2012, Comparison of harmonic, geometric and arithmetic means for change detection in SAR time series, *EUSAR. 9th European Conference on Synthetic Aperture Radar*, 2012.

Touzi, R., 1997, Target scattering decomposition in terms of roll invariant target parameters, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 1, p. 73–84.

Touzi, R. and Charbonneau, F., 2002, Characterization of target symmetric scattering using polarimetric, SAR. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, p. 693-711.

Van Zyl, J.J., 1989, Unsupervised classification of scattering behavior using polarimetric data, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 27, no. 1, pp. 36-45.

Yajima, Y., Yamaguchi, Y. Sato, R. Yamada, H. and Boerner, W.-M., 2008, PALSAR Image Analysis of Wetlands Using a Modified Four-Component Scattering Power Decomposition, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 6, p. 1667-1673.

Vectorisation automatique des forêts dans les minutes de la carte d'Etat Major du 19ème siècle

Pierre-Alexis Herrault[1,2], David Sheeren[1], Mathieu Fauvel[1], Martin Paegelow[2]

[1]*Université de Toulouse, INP-ENSAT, UMR 1201 DYNAFOR*
Av. de l'Agrobiopôle, BP 32607, Auzeville Tolosane, 31326 Castanet Tolosan cedex
pierrealexis.herrault@ensat.fr;david.sheeren@ensat.fr;mathieu.fauvel@ensat.fr

[2]*Université de Toulouse, UTM, UMR 5602 GEODE*
5 Allées Antonio Machado, 31058 Toulouse cedex-France
paegelow@univ-tlse2.fr

RESUME. *Dans cet article, nous proposons une nouvelle méthode pour extraire automatiquement les forêts dans les cartes d'Etat-major du 19ème siècle numérisées. La démarche proposée est constituée de quatre étapes principales : filtrage de l'image, changement d'espace colorimétrique, identification des forêts à l'aide d'un détecteur non paramétrique (SVDD), post-traitement. La méthode est suffisamment robuste pour prendre en compte la diversité des représentations possibles des forêts dans ces cartes anciennes. Les résultats montrent des performances élevées avec une précision globale de détection obtenue de 95%. Cette approche ouvre de nouvelles perspectives pour les différentes études environnementales incluant une dimension historique.*

MOTS-CLÉS : *carte ancienne, CIELab, SVDD, morphologie mathématique.*

ABSTRACT. *In this article, we propose a new method to automatically extract forest features in historical 'Map of France 'from 19th century. The proposed method is based on four major steps: filtering of the image, color-space conversion, extraction of the forests based on a non-parametric detection algorithm (SVDD), post-processing. The method is sufficiently robust to take into account the diversity of the representation of the forests in historical map. Results show high performances with a global accuracy of 95%. This approach offers new perspectives in the different environmental studies including a historical dimension.*

KEYWORDS: *historical map, CIELab, SVDD, mathematical morphology.*

1. Introduction

Les cartes anciennes contiennent une riche quantité d'informations sur le passé concernant l'occupation du sol, leur structure, la morphologie des villes, les monuments, etc. L'extraction de ces informations présente un intérêt pour de nombreuses études qui portent par exemple sur les transformations des paysages sur le long terme, le développement urbain, l'évolution du trait de côte, et le changement global (Cousins, 2001 ; Bender et al., 2005 ; Gimmi et al., 2011 ; Smith and Cromley, 2012).

A l'heure actuelle, la vectorisation des éléments figurant sur ces cartes et leur intégration dans une base de données géographiques est généralement menée par digitalisation, avec une forte intervention humaine. Or, pour des raisons évidentes de temps et de coût, cette procédure est peu reproductible sur de vastes étendues ce qui limite la réalisation d'études historiques spatialisées.

Plusieurs travaux se sont intéressés au problème de vectorisation automatique de cartes topographiques, cartes géologiques ou documents plus anciens (Lichtner, 1985 ; Ansoult et al., 1990 ; Samet et Soffer, 1998 ; Wise, 1999, Leyk et al., 2006 ; Chiang, 2012 ; Shaw et al., 2011 ; Ivanowski et Kozak 2012 ; Mahmoud, 2012). La difficulté varie suivant le type de carte à traiter (couleur ou noir et blanc), sa qualité (déformation ou non du document numérisé) et le niveau de surcharge des différents éléments planimétriques (superposition des courbes de niveau, toponymes et réseaux divers sur les entités d'occupation du sol). Les méthodes développées sont souvent spécifiques à la carte considérée et peu reproductibles sur une autre source de données. Toutefois, il émerge des différents travaux une démarche commune pour vectoriser automatiquement ces cartes. Elle se compose de trois étapes principales : (1) une phase de « nettoyage », (2) une étape d'identification des éléments d'intérêt et (3) une étape de post- traitement.

Dans cet article, nous proposons une nouvelle méthode de vectorisation automatique des forêts dans les Minutes de la carte d'Etat-Major datant du milieu du 19ème siècle (Dupouey, 2007). Il n'existe actuellement aucune méthode de vectorisation dédiée à cette carte qui est pourtant largement exploitée dans différentes études, notamment en écologie historique (Andrieu et al., 2011). Fondée sur des techniques de traitement d'image, la procédure adopte la démarche générale de vectorisation en incluant un changement d'espace colorimétrique et l'application d'un détecteur supervisé non-paramétrique (Support Vector Data Description). Cette approche pourrait être également employée pour extraire d'autres régions colorées de cette même carte.

L'article est structuré de la manière suivante. Une brève description des caractéristiques de la carte d'Etat-major est donnée en section 2. Nous décrivons en section 3 la méthode proposée avec les principales étapes qui la compose. Nous présentons l'application de cette méthode sur différents extraits de minute et

discutons des résultats en section 4. Nous concluons et proposant différentes perspectives en section 5.

2. Les minutes de la carte d'Etat-Major

La méthode de vectorisation proposée est définie pour la reconnaissance des formations boisées dans les minutes de la carte d'Etat-Major de la France. Cette carte a été constituée au milieu du 19ème siècle, de 1818 à 1866, à partir de levés d'échelle variable suivant le lieu ainsi que des plans cadastraux napoléoniens (Dupouey, 2007). La carte générale d'Etat-Major couvre l'ensemble du territoire français au 1:80.000. Elle est éditée en noir et blanc. Les minutes de cette carte sont quant à elles en couleur et produites à l'échelle du 1:40.000. Leur qualité géométrique est réputée assez fiable mais très variable suivant le lieu. Le système de coordonnées planimétriques de la carte est basé sur la projection dite de Bonne avec comme ellipsoïde de référence celui de Plessis 1817. Il s'agit d'une projection équivalente dont les différents paramètres sont connus.

Les minutes de la carte utilisées dans cette étude ont été mises à disposition par l'IGN au format numérique. Les images fournies sont codées sur 3 octets (24 bits) correspondant aux 3 couleurs primaires (R, V, B). La résolution spatiale des images est de 4 mètres pour un scannage à 254 dpi. L'IGN fournit les minutes non pas dans le système géodésique initial mais dans le système légal de référence actuel (Lambert-93). Il est toutefois nécessaire de géoréférencer les minutes pour les rendre superposables aux données cartographiques d'aujourd'hui. Une procédure de géoréférencement adaptée aux cartes anciennes et fondée sur un modèle de régression à noyau a été proposée et appliquée dans ce sens (Herrault et al., 2013). Une illustration des minutes de la carte est fournie en section 4.

3. Méthode de vectorisation automatique

La méthode de vectorisation automatique des forêts que nous proposons est composée de différentes étapes (figure 1). Nous détaillons chacune de ces étapes dans les sections suivantes.

3.1. Pré-traitements pour le filtrage d'éléments en surcharge

Avant de procéder à la détection des forêts proprement dite, une étape de nettoyage est nécessaire pour éliminer des éléments planimétriques qui se superposent aux polygones représentant les forêts. Ces éléments concernent principalement certains symboles, la toponymie et les hachures rendant compte du relief. En plaine, ces hachures sont peu nombreuses. En revanche, leur densité est très importante en milieu de montagne ce qui rend le processus de filtrage délicat. La procédure proposée vise à éliminer ces hachures quel que soit le contexte orographique tout en préservant le contour des forêts. Trois étapes successives sont

appliquées : (1) dilatation morphologique, (2) filtrage de type médian, (3) filtrage de type passe-bas.

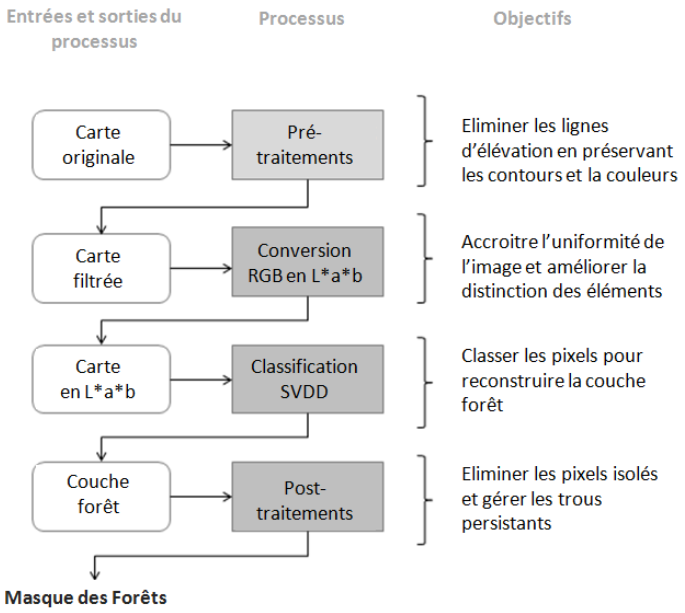


Figure 1.Chaine de traitement suivie pour l'extraction des forêts sur les minutes d'Etat-Major

La dilatation morphologique (étape 1 des pré-traitements) vise à combler la totalité des trous à l'intérieur des forêts formés par les symboles, les éléments textuels et les hachures. L'élément structurant choisi est un carré de taille 5x5. Cette taille d'élément permet d'éliminer efficacement ce type de trous sans impacter les trous résultant d'une perforation de la forêt. L'opérateur est appliqué sur chaque bande de l'image (R, V, B).

Le filtre médian est ensuite mis en œuvre à l'issue de la dilatation pour éliminer les hachures persistantes (étape 2 des pré-traitements). La taille de fenêtre du filtre est la même mais il est appliqué plusieurs fois jusqu'à obtenir une image suffisamment filtrée. Le nombre d'application du filtre varie en fonction de la densité des hachures présente sur la carte et donc du relief de la zone. Contrairement à un filtre moyen, ce filtre médian présente l'intérêt de respecter les contours des objets.

Un dernier filtre de type passe-bas et de taille 5x5 est finalement appliqué pour supprimer le bruit résiduel.

3.2. Transformation de l'espace colorimétrique : du RGB vers le CIELab

Les étapes précédentes ont permis de supprimer un certain nombre d'éléments se superposant aux forêts. Néanmoins, le filtrage a également altéré les couleurs de l'image. Les teintes ont été modifiées et ceci, au sein d'une même catégorie d'objet puisqu'en fonction du relief, la surcharge des hachures n'est pas la même. D'une manière plus générale, la couleur est rarement homogène dans les cartes historiques. L'intensité est susceptible de varier plus ou moins fortement suivant le lieu. Le changement d'espace colorimétrique est nécessaire afin de rendre plus robuste la reconnaissance automatique des objets sur base de leur couleur.

Le choix d'un espace colorimétrique adapté est une étape importante dans la vectorisation des cartes historiques en couleur. Il conditionne souvent la qualité du résultat final (Ivanowski et al., 2012). L'espace traditionnellement utilisé est l'espace RVB. Cet espace bien connu présente certaines limites pour la question qui nous occupe: la luminosité n'est pas uniforme en plus du fait qu'il est plus éloigné de la perception humaine, c'est-à-dire qu'une différence entre deux couleurs dans cet espace n'est pas proportionnel à la différence perçu par l'œil humain (Angulo et Serra, 2003). Les espaces colorimétriques cylindriques dans lesquels la luminosité est isolée sont mieux adaptés à la reconnaissance d'objets dont la luminosité n'est pas uniforme.

Dans notre approche, nous proposons de transformer l'image initiale dans l'espace colorimétrique CIELab (fig.2). Ce choix s'inspire de travaux précédents montrant que le CIELab est bien approprié aux images bruitées colorées (Angulo et Serra 2003, Ganesan et al. 2010).

Dans l'espace CIELab, la composante de clarté de la couleur est reportée sur l'axe L (*Luminosity*), variant de 0 à 100, et perpendiculaire à une pile de plan « ab ». Les composantes « a » et « b » sont définies par une gamme de valeurs variant respectivement du rouge au vert en passant par le gris, et du jaune au bleu en passant par le gris. Chacun de ces plans contient ainsi toutes les couleurs possibles pour une valeur de clarté. Dans notre cas, nous pouvons avoir plusieurs valeurs de clarté (i.e. plusieurs valeurs de L) pour une même couleur (ici le vert pour les forêts). La reconnaissance de ces objets est alors fondée sur la recherche des différentes partitions de vert existant pour chaque valeur possible de L. Cette recherche est effectuée à l'aide d'un algorithme de détection non paramétrique décrit ci-dessous.

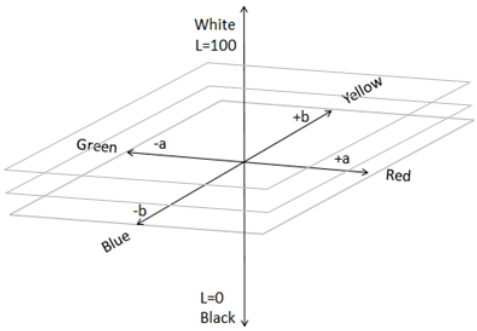


Figure 2.Modèle de représentation cartésien de l'espace CIELab

3.3 Extraction des forêts à partir du détecteur SVDD

La reconnaissance des forêts dans l'espace colorimétrique est réalisée au moyen de l'algorithme de détection supervisée SVDD (*Support Vector Description Descriptor*) (Sanchez-Hernandez et al. 2007, Fauvel et al. 2013). Deux raisons principales expliquent ce choix :

- contrairement à un algorithme de reconnaissance multi-classe, cette méthode supervisée ne requiert pas nécessairement de fournir des exemples d'apprentissage négatifs (i .e. des exemples correspondant aux autres classes que celle recherchée) ou du moins, pas de manière exhaustive (Tax et Duin, 2004). Cela rend la sélection des exemples d'apprentissage moins contraignante et plus rapide.

- contrairement à un algorithme de classification supervisée paramétrique (tel que le modèle de mélange gaussien), le détecteur SVDD ne nécessite pas l'estimation des distributions des classes à partir des exemples d'apprentissage. Cela rend l'approche également plus souple car la variabilité des représentations des forêts peut conduire à l'obtention d'une distribution multi-modale de la classe. Par ailleurs, la distribution des classes négatives (autres que forêt) peut être difficile à modéliser en raison d'une grande diversité des objets présents dans la carte. Cela nécessiterait un nombre important d'exemples d'apprentissage négatifs qui rendrait la méthode moins opérationnelle.

Concernant son fonctionnement, l'algorithme SVDD fait partie des méthodes d'apprentissage à noyau (Shawe-Taylor et Cristianini, 2004). Il consiste à trouver l'enveloppe d'une hypersphère de rayon minimal qui englobe les exemples de la classe d'intérêt en excluant les autres. Lorsque cette enveloppe optimale n'est pas linéaire, la prise en compte des non linéarités s'effectue par l'introduction de noyaux (noté k dans l'exemple qui suit) permettant de calculer une similarité entre deux variables dans un espace optimisé.

De manière plus formelle, appelons $S = \{x_i, y_i\}$ le jeu d'exemples d'apprentissage avec $x_i \in \mathbb{R}^d$ et y_i= 1 si x_i est positif et y_i = -1 si c'est un exemple négatif. Les paramètres $\hat{\beta}_i$ de l'hypersphère optimale, avec $i \in \{1, \dots, n\}$ sont déterminés en résolvant le problème d'optimisation suivant (pour les détails, voir Tax et Duin 2004) :

$$\max_{\beta_i} G(\beta) = \sum_{i=1}^{n} \beta_i \, y_i k(x_i, x_i) - \sum_{i,j=1}^{n} \beta_i \, \beta_j y_i y_j k(x_i, x_j) \quad (1)$$

Avec $C \geq \beta_i \geq 0$ et $\sum_{i=1}^{n} \beta_i \, y_i = 1$. (2)

Où C est le paramètre de régularisation compris entre 0 et 1 qui pénalise les erreurs durant l'optimisation. Ce paramètre permet de régler la proportion d'exemples d'apprentissage que l'on souhaite maintenir en dehors de l'hypersphère. Quand $C = 1$, aucune erreur n'est admise. Des valeurs de C inférieures permettent des erreurs pour une part des échantillons négatifs et positifs.

La valeur au carré $\mathcal{R}^2$ du rayon de l'hypersphère est donné par $G(\hat{\beta})$ (β à l'optimal). La distance euclidienne d_c au centre c de l'hypersphère pour un exemple z est donnée par :

$$d_c(z) = k(z, z) +$$

$$\sum_{i,j=1}^{n} \hat{\beta}_i \, \hat{\beta}_j y_i y_j k(x_i, x_j) \; - \; 2 \sum_{i=1}^{n} \hat{\beta}_i \, y_i k(z, x_i) \quad (3)$$

La détection est effectuée en évaluant si la distance d_c pour un exemple z est plus grande ou plus petite que le rayon de l'hypersphère $\mathcal{R}^2$. Dans le premier cas, il s'agit d'une classe négative (l'exemple a dépassé la frontière de l'hypersphère). Dans le second, il s'agit d'une classe positive.

Pour les séparations non linéaires, le choix d'un noyau adapté est un problème crucial dans l'utilisation de ces méthodes. Un noyau idéal permettrait d'englober les échantillons positifs dans une région de forme sphérique (ou au moins dans une région délimitée) au sein de l'espace des données d'apprentissage. Les noyaux doivent donc être à support compact. C'est la raison pour laquelle nous avons fait le choix d'utiliser le noyau gaussien (et non les noyaux polynomiaux par exemple), forme la plus compacte des fonctions à noyau. Le noyau est exprimé comme suit :

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right) \quad (4)$$

L'hyper paramètre σ est déterminé comme la moyenne des distances entre les échantillons positifs.

$$\sigma^2 = \frac{2}{(n_+ - 1)(n_+ - 1)} \sum_{\substack{i=1 \\ j=i+1}}^{n_+} \|x_i - x_j\| \quad (5)$$

3.4 Post-traitement

A l'issue de la détection des forêts effectuée à l'étape précédente, un post-traitement est réalisé sur le masque extrait (pixels identifiés comme « forêt » à la suite de la classification) pour corriger certains artéfacts. Un opérateur morphologique d'ouverture est d'abord appliqué pour supprimer les pixels isolés. Certains trous résiduels sont ensuite comblés par l'application d'une règle contextuelle : les groupes de pixels non forêt totalement bordés par des forêts sont reclassés en forêt. L'image originale ne contenant que les forêts est finalement reconstruite.

4. Résultats et discussion

La méthode proposée a été testée sur trois extraits différents des minutes d'Etat-major. Ils se différencient par leur contexte orographique, leur surcharge d'éléments planimétriques et la couleur qui est utilisée pour représenter les entités « forêt » (différents niveaux de vert) (Fig 3).

Figure 3. Les trois extraits sélectionnés pour l'application de la méthode proposée

L'étape de filtrage opérée en première partie du protocole de traitement a permis d'éliminer les éléments planimétriques se superposant aux polygones représentant les forêts. Premièrement, l'opérateur de dilatation a réduit efficacement le nombre

de pixels dans les basses fréquences. Ces pixels correspondant aux lignes d'élévation et à la toponymie ont été supprimé tout en préservant les contours d'objet.

Figure 4. Illustration du processus de filtrage pour l'extrait 2

Cette étape est apparue toutefois insuffisante pour procéder directement à la classification des pixels d'intérêt. Un nombre trop important d'éléments sombres est encore visible au sein des entités à extraire à la suite de ce processus. L'application successive d'un filtre médian a alors permis d'éliminer ces artefacts persistants en transférant les pixels sombres vers les hautes fréquences. Enfin, l'ultime opération de cette étape de filtrage a pour but de lisser le bruit de fond granuleux et améliorer l'uniformité de l'image. L'application du filtre passe-bas [1]a bien répondu à ce problème en conservant correctement les détails de l'image. L'image obtenue à la suite de cette procédure permet de procéder à l'extraction des pixels d'intérêt (fig 4).

La méthode de classification SVDD offre la possibilité d'introduire des éléments négatifs et de régulariser la proportion d'éléments d'apprentissage à maintenir en dehors de l'hyper sphère. Fixer le paramètre C à 1 revient à n'accepter aucune erreur dans le processus d'apprentissage. L'objectif étant d'aboutir à un compromis intéressant pour les trois extraits testés, nous décidons de fixer la valeur de C à 0.5.

[1] Le filtrage passe bas atténue les basses fréquences. Il a pour conséquence un adoucissement des détails et une suppression du bruit granuleux.

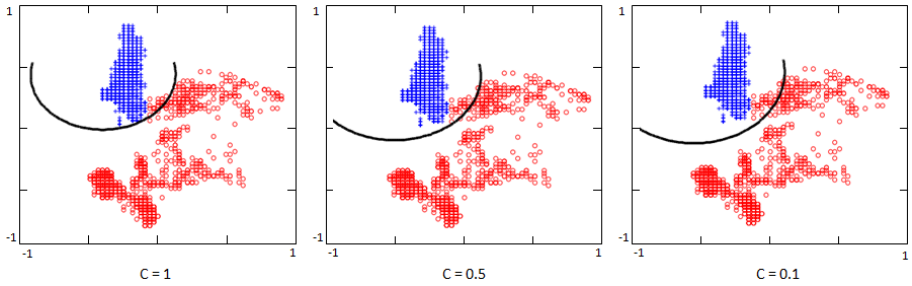


Figure 5. Impact du paramètre de régularisation C sur la procédure d'apprentissage pour l'extrait 1. Les pixels d'apprentissage de la classe positive sont représentés en bleu et les pixels de la classe négative sont représentés en rouge. La ligne noire représente la limite de décision trouvée par le détecteur SVDD.

La figure 5 illustre l'influence du paramètre de régularisation du détecteur SVDD sur la procédure d'apprentissage pour l'extrait 1. Pour cela, nous avons effectué trois phases d'apprentissage à l'aide du même jeu d'échantillon (négatifs et positifs) basées sur trois valeur différentes du paramètre C (C=0.1 ; C= 0.5 ; C=1). A noter que cette phase d'apprentissage et la procédure de classification qui en résulte sont à présent effectuées dans l'espace CIELab. Comme attendu, une faible valeur de C (ex : 0.1) conduit à accepter une plus grande part des échantillons négatifs dans l'hyper sphère. Détecter les forêts sur l'image entière avec cette valeur de C conduirait à classer trop de pixels non-forêt dans la classe forêt. Lorsque C est élevé (ex : 1), le détecteur ajuste parfaitement le nuage des échantillons positifs. Il y aura alors de très forts risques de surajustements[2] quand il s'agira de détecter la classe d'intérêt sur l'image entière. Dès lors que l'on classe les données réelles, un compromis doit être trouvé pour correctement apprendre les données tout en bénéficiant d'une bonne capacité de généralisation (Fauvel et al. 2013). Une valeur de C à 0.5 respecte ces deux critères. L'ensemble du nuage d'échantillons positifs est contenu dans l'hypersphère mais une part raisonnable de pixels négatifs est autorisé dans l'hypersphère afin d'éviter les problèmes de surajustement.

La chaine de traitement globale a été appliquée sur les trois extraits présentés en début de section. L'illustration des masques résultants est présentée dans la figure 6. La validation des résultats est effectuée par comparaison avec les couches d'extraction obtenue manuellement sur l'ensemble des extraits. Ils indiquent une précision globale importante (Kappa ~ 0.90).

[2] Manque de généralisation du modèle de détection malgré une grande précision sur l'ensemble d'apprentissage.

16

Tableau 1.Résultats de l'extraction opérée sur les trois extraits sélectionnés

	Extrait 1	Extrait 2	Extrait 3
Précision globale	96.3 %	95.83%	94.11%
Indice Kappa	0.91	0.90	0.89
Commission	3.96%	4.93%	4.21%
Omission	1.18%	1.00%	3.33%

Le tableau 1 répertorie les résultats d'extraction pour chacun des extraits. Si la proportion globale de pixels correctement classés avoisine les 95 % pour chacun des extraits, l'indice Kappa présente lui aussi de très bons scores avec des valeurs situées entre 0.89 et 0.91. Les résultats très proches sur chacune des images traitées montrent une bonne performance globale de la méthode proposée et une certaine robustesse. Quel que soit le type de représentation des forêts ou le niveau de surcharge de la carte, la démarche proposée fournit des résultats quasi similaires.

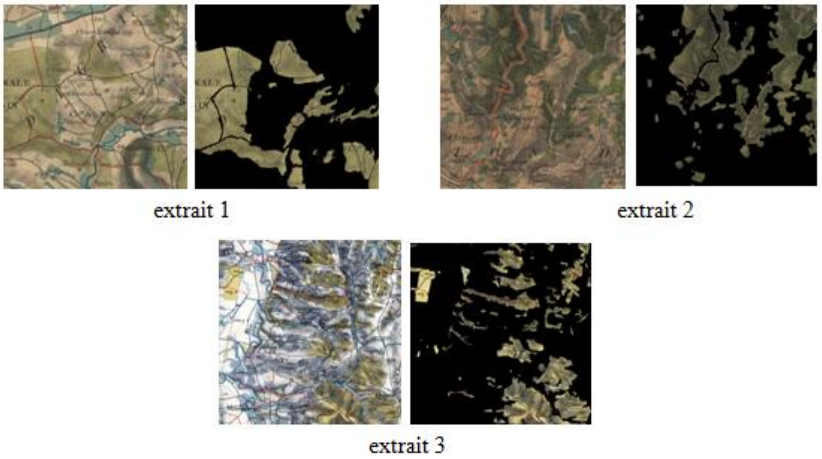


Figure 6.Masques des forêts obtenus après les post-traitements (à gauche, l'image originale ; à droite, le masque des forêts)

L'analyse détaillée des matrices de confusion montrent des valeurs de commission (éléments surdétectés) plus élevées que les valeurs d'omissions (éléments sous-détectés). La méthode a donc tendance à surestimer le masque des forêts. Ceci peut s'expliquer par deux éléments principaux :

- le compromis déterminé dans la mise en œuvre du détecteur SVDD conduit à accepter une part d'échantillons négatifs dans l'hypersphère afin d'éviter les problèmes de sur ajustement des échantillons positifs. Les paramètres retenus dans la phase d'apprentissage et appliqués ensuite pour extraire les pixels d'intérêt dans

l'image entière jouent un rôle déterminant. Certains pixels appartenant à d'autres classes ont été classifiées comme appartenant à la couche « forêt ».

-malgré une phase de post-traitement visant en partie à éliminer les pixels isolés, les paramètres de l'élément structurant utilisé pour l'ouverture morphologique (3*3) n'ont pu corriger l'ensemble des imperfections observées dans les masques résultants.

5. Conclusion

Dans cet article, une méthode d'extraction automatique des forêts sur les minutes de la carte d'Etat-major du 19e siècle est développée. La procédure montre des résultats performants avec une précision globale de 95 %. La première phase de filtrage a permis d'éliminer efficacement les objets se superposant aux forêts par une succession de filtres morphologiques et de convolution. En ce qui concerne la modification de l'espace colorimétrique de la carte, l'espace CIELab parait bien appropriée au traitement des cartes historiques. En accroissant l'uniformité de l'image, il offre la possibilité d'effectuer une classification des pixels d'intérêt de manière efficace. Quant au détecteur SVDD, il fournit des résultats d'extraction très précis. La possibilité d'introduire des échantillons négatifs sans pour autant échantillonner de manière exhaustive la classe négative rend plus facile l'identification de la classe d'intérêt. Dans cette étude le noyau gaussien est utilisé et les résultats expérimentaux montrent qu'il remplit correctement l'objectif d'extraction.

Les résultats pourraient être améliorés en considérant deux points majeurs. Premièrement, une recherche approfondie du paramètre de régularisation C du détecteur SVDD permettrait d'établir une limite de décision de manière plus fine. Deuxièmement, les post traitements doivent faire l'objet d'une attention particulière afin d'accroitre la précision des résultats issus de la procédure de classification. Les futures recherches dans ce travail portent quant à elles sur l'application de la démarche proposée à d'autres types de cartes présentant des problèmes similaires pour accroitre ses potentialités. Confronter cette approche à d'autres sources permettrait d'évaluer plus finement ses potentialités. Par ailleurs, ce travail s'inscrit également dans un objectif de suivi des trajectoires forestières depuis 1850 (Ladet et al., 2013). Relier ces informations avec des données taxonomiques d'aujourd'hui permettrait de vérifier une potentielle relation entre l'histoire des forêts et la configuration actuelle de la biodiversité.

18

Bibliographie

Andrieu E., Ladet S., Heintz W., Deconchat M. (2011). History and spatial complexity of deforestation and logging in small private forests, *Landscape and Urban Planning*, 103, 109-117

Angulo J. and Serra J. (2003) Mathematical morphology in color spaces applied to the analysis of cartographic images, S. Levachkine, J. Serra and M. Egenhoger (Eds.), Semantic Processing of Spatial Data, *in Proceedings of the GEOPRO 2003-International Workshop Semantic Processing of Spatial Data*, Mexico City, pp. 59-66

Ansoult M., Soille P., Loodts J. (1990) Mathematical morphology: a tool for automated GIS data acquisition from scanned thematic maps. *Photogrammetric Engineering and Remote Sensing*, vol. 56 (9), pp 1263-1271

Bender O., Boehmer H.J., Jens D., Schumacher K.P. (2005) Using GIS to analyse 200 years of cultural landscape change in Southern Germany, *Landscape and Urban Planning*, vol.70, pp 111-125.

Chiang Y.-Y., Leyk S, Knoblock C. A. (2012) Efficient and Robust Graphics Recognition from Historical Maps, in Graphics Recognition: Achievements, Challenges and Evolution, Selected Papers of the Ninth International Workshop on Graphics Recognition (GREC), *Lecture Notes in Computer Science*

Cousins S. (2001) Analysis of land-cover transitions based on 17th and 18th century cadastral maps and aerial photographs, *Landscape Ecology*, vol.16, pp 41-54

Dupouey J.L., Bachacou J., Cosserat R., Aberdam S., Vallauri D., Chappart G., Corvisier de Villèle, M.A. 2007. Vers la réalisation d'une carte géoréférencée des forêts anciennes de France. *Le Monde des Cartes*, 191, pp. 85-98

Fauvel M., Arbelot B., Benediktsson J.A., Sheeren D., Chanussot J. (2013). Detection of Hedges in a Rural Landscape Using a Local Orientation Feature: From Linear Opening to Path Opening, *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol.6, no.1, pp.15, 26

Ganesan P., Rajini V., Rajkumar I R. (2010) Segmentation and Edge Detection of Color Images Using CIELAB Color Space and Edge Detectors, *in Emerging Trends in Robotics and Communication Technologies (INTERACT)*, 2010 International Conference

Gimmi U., Lachat T, Burgi M. (2011) Reconstructing the collapse of wetland networks in the Swiss lowlands 1850–2000, *Landscape ecology* , vol.26, pp 1071-1083

Herrault P.-A., Sheeren D., Fauvel M., Monteil C., Paegelow M. (2013). A comparative study of geometric transformation models for the historical 'Map of France registration**,** *Geographia Technica*, no. 1/2013, pp. 34-46

Iwanowski M., Kozak J. (2012). Automatic detection of forest regions on scanned old maps, *Przeglad Elektrotechniczny (Electrical Review)*, ISSN 0033-2097, R. 88 NR 4

Ladet S., Lopez J., Grel A.,Valladares L., Brin A. (2013). Cartographie historique : outil clé pour caractériser la continuité temporelle. Application à des travaux en écologie forestière, *Actes du colloque SAGEO 2013*

Leyk S., Boesch R., Weibel R. (2006). Saliency and semantic processing: Extracting forest cover from historical topographic maps, *Pattern recognition*, vol. 39 (5), pp 953-968, doi: 10.1016/j.patcog.2005.10.018

Mahmoud M.I. (2012). *Information extraction from paper maps using oriented object analysis*, PhD thesis in Geo-science Management, Faculty of Geo Information Science and Earth Observation of the University of Twente

Samet H., Soffer A. (1998) Magellan: Map acquisition of geographic labels by legend analysis. *International journal on document analysis and recognition*, vol. 1 (2), pp 89-101, doi: 10.1007/s10032005

Sanchez-Hernandez C., Boyd D. S., Foody G. M. (2007). Oneclass classification for mapping a specific land-cover class: SVDD classification of fenland, *IEEE Trans Geosci Remote Sens*, vol. 45, no. 4, pp. 1061–1073

Shaw T., Bajcsy P. (2011) Automation of Digital Historical Map Analyses, *in IS&T/SPIE Electronic Imaging*, 7869-09, Session 3, Conference 7869: Computer Vision and Image Analysis of Art II, January 23-27; (oral presentation)

Shawe-Taylor J., Cristianini N. (2004). *Kernel Methods for Pattern Analysis*, Cambridge University Press

Smith M.J. and Cromley R.G. 2012, Measuring Historical Coastal Change using GIS and the Change Polygon Approach, *Transactions in GIS*, 16(1): 3–15

Tax D. M. J., R. Duin P. W. (2004). Support vector data description, *Machine Learning*, vol. 54, no. 1, pp. 45–66

Wise S. (1999) Extracting raster GIS data from scanned thematic maps, *Transactions in GIS*, 3(3): 221-237

Radar hyperfréquence embarqué appliqué à la cartographie du corridor fluvial

M. Jaud[1], R. Rouveure[1], P. Faure[1], M-O. Monod[1]

1. UR TSCF, Irstea Clermont-Ferrand
Campus des Cézeaux, 24 avenue des Landais, BP50085, 63172 Aubière cedex, France
marion.jaud@irstea.fr, raphael.rouveure@irstea.fr

RESUME. *La gestion des rivières est une composante essentielle d'un développement durable et de la préservation des ressources en eau. Dans ce contexte, des levés réguliers doivent être réalisés. Cet article présente le radar embarqué K2Pi pour cartographier le corridor fluvial. Afin d'en évaluer le potentiel, les cartes radar obtenues sur un tronçon de l'Allier (Auvergne, France) de 18km sont comparées à des images aériennes RGB et infra-rouges. De par sa souplesse d'utilisation et la faisabilité d'une acquisition sans GPS, ce système s'avère complémentaire des techniques classiques de suivi.*

ABSTRACT. *River monitoring plays a crucial role in sustainable management of water resources. Regular surveys have so to be performed. This paper will present the ship-borne radar K2Pi in a perspective of mapping the river corridor. A 18km long reach of the river Allier (Auvergne, France) has been scanned by the radar. The radar map is compared with RGB and infra-red aerial images, so as to examine the radar potential. The implementation of the K2Pi radar is relatively straightforward and do not need a GPS. This system appears to be complementary to conventional monitoring techniques.*

MOTS-CLES : *Cartographie radar, radar FMCW, Fourier-Mellin, ortho-images.*

KEYWORDS: *Radar mapping, FMCW radar, Fourier-Mellin, ortho-images.*

1. Introduction

De par leur attractivité et leur richesse écologique, les cours d'eau sont un objet géographique associé à d'importants enjeux environnementaux, sociétaux et économiques (Liverman et Cuesta, 2008). Depuis 2000, la France, comme tous les Etats membres de la Communauté Européenne, met en œuvre la Directive Cadre sur l'Eau (DCE) afin de parvenir en 2015 au « bon état » de tous les milieux aquatiques. Dans le cadre d'une gestion intégrée des cours d'eau, un suivi de leur morphodynamique s'impose (Malavoi et Bravard, 2010).

panoramique » centrée sur la position du radar. Cette image est corrigée des déplacements du véhicule et des effets de speckle.

2.2. Construction de la carte radar

L'association d'images radar panoramiques successives donne lieu à la construction d'une carte grâce à l'algorithme « R-SLAM » (Rouveure *et al*., 2008). Cet algorithme s'appuie sur la technique du SLAM (Simultaneous Localization And Mapping) qui permet à un véhicule mobile de construire une carte relative de l'environnement tout en se localisant dans cette carte 2D grâce à différents capteurs optiques (Dissanayake *et al*., 2001). Avec le radar K2Pi, la carte peut être construite de manière récursive, indépendamment de la dynamique de la plate-forme ou de la position du radar sur le véhicule.

Pour une image radar panoramique, cette méthode se décompose en deux étapes :

- un appariement de carte (via une corrélation croisée 3D) entre la n^{ième} image radar panoramique et la carte construite par itération jusqu'au rang n-1. Le maximum de corrélation donne une estimation des paramètres de déplacement et de rotation entre le scan d'ordre n-1 et le scan d'ordre n.
- une mise à jour de la carte globale par fusion de la n^{ième} image radar panoramique avec la carte construite jusqu'au rang n-1.

La carte radar finale est un raster en échelle de gris, avec une résolution spatiale de 20cm (figure 1c). Les calculs successifs des déplacements et rotations inter-scans permettent de reconstruire simultanément la trajectoire relative du radar, appelée « trajectoire R-SLAM ».

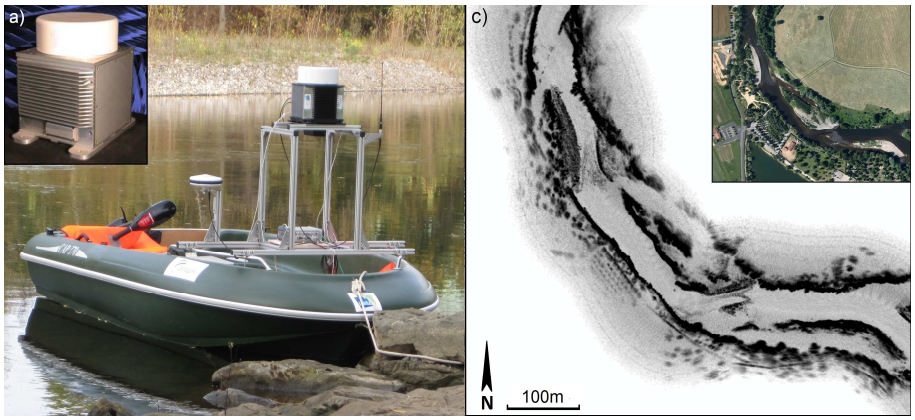


Figure 1: (a, b) Radar K2Pi embarqué sur un canot (photo Irstea©). (c) Extrait de la carte radar relative construite à partir de données radar acquises en Octobre 2012 sur l'Allier (Auvergne, France).

4

2.3. Géoréférencement de la carte radar

Une étape de géoréférencement vient compléter la construction de la carte radar. Afin de s'adapter au mieux au contexte d'étude et aux données disponibles, trois méthodes ont été développées (Jaud *et al*., sous presse) :

- géoréférencement par sélection de paires de points de contrôle sur la carte radar et sur une ortho-photographie ;
- géoréférencement par appariement de la trajectoire relative R-SLAM avec une trajectoire DGPS enregistrée durant l'acquisition
- géoréférencement image à image, via la méthode de Fourier-Mellin (Casasent et Psaltis, 1976) schématisée par la figure 2.

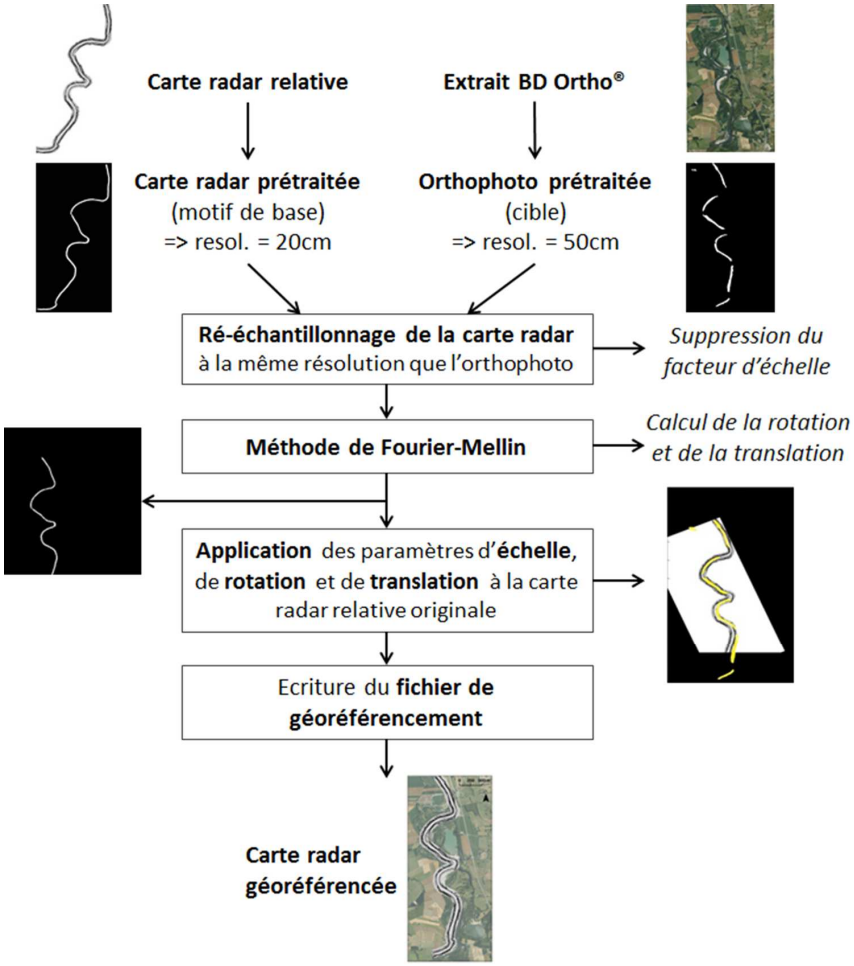


Figure 2:Différentes étapes de la procédure de géoréférencement basée sur la méthode de Fourier-Mellin pour le recalage d'une carte radar.

La méthode de Fourier-Mellin permet de déduire les paramètres de transformation en rotation et translation de la carte radar. Cette méthode se base sur la ressemblance globale entre la carte radar et une photographie aérienne. Toutefois, à une échelle plus fine, de nombreuses disparités apparaissent entre carte radar et photographie aérienne. Une étape de prétraitement des images s'avère nécessaire afin de ne conserver que les éléments caractéristiques communs. Une fois géoréférencée, la carte radar pourra être intégrée dans un SIG afin d'être exploitée. Cette méthode de géoréférencement image à image sera choisie préférentiellement car elle est entièrement automatique et ne nécessite aucun enregistrement GPS lors de l'acquisition radar. En contexte fluvial, le signal GPS peut en effet être masqué si la ripisylve est trop dense ou si le cours d'eau circule dans des gorges.

3. Acquisition des données

3.1. Zone d'étude

La zone d'étude est un tronçon de l'Allier (Auvergne), située entre Mirefleurs et Pont du Château (figure 3a). Cette zone a été subdivisée en deux parties :

- un premier sous-tronçon de 8 km entre Mirefleurs et Cournon ;
- un second sous-tronçon de 10 km entre Cournon et Pont du Château.

Ce secteur, de style fluvial sinueux, marque la frontière entre l'Allier montagnard en amont et l'Allier des plaines en aval (SAGE Allier Aval, 2007). Il présente ainsi des configurations très variées propices à une mise en situation du matériel : présence de zones urbanisées, de zones de falaises, de gravières, etc.

3.2. Moyens mis en œuvre

Les deux sous-tronçons ont été cartographiés par le radar K2Pi en Octobre 2012. Le canot, équipé d'un moteur électrique, a effectué le parcours de 18 km en un peu plus de 3 h, soit une vitesse moyenne d'environ 6 km/h. Cette vitesse ne doit pas être trop élevée afin de limiter l'écart entre deux images panoramiques radar successives. Au total, près de 12000 images panoramiques ont été acquises. La carte radar totale couvre près de 3,5 km² avec une résolution de 20 cm.

Parallèlement à l'acquisition radar, un survol de la zone a été effectué (figure 3b) afin d'acquérir des photographies en encodage couleur de type RGB ainsi que des images en InfraRouge (IR) couleur dans un rayon de 300 m de part et d'autre de la rivière (figure 3c). Ces images ont une résolution de 20cm, similaire à celle de la carte radar.

6

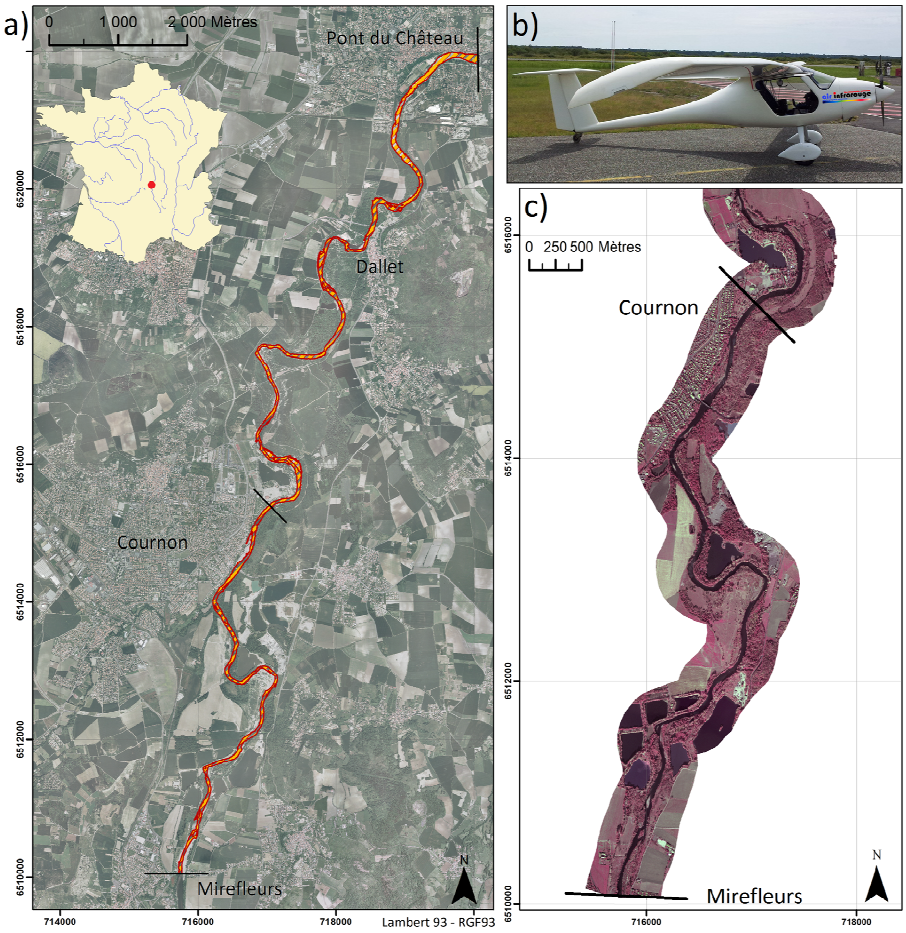


Figure 3 : a) Zone d'étude subdivisée en deux tronçons : Mirefleurs-Cournon (8km) et Cournon-Pont du Château (10km). b) Avion utilisé par la société Air Infrarouge pour le survol de la zone d'étude. c) Extrait de la mosaïque d'ortho-images IR.

4. Potentiel et limitations

Les atouts et limitations de la cartographie radar K2Pi par rapport aux images aériennes RGB et IR sont récapitulés dans le Tableau 1. Lors d'une acquisition depuis un canot, la portée du radar (100 m) est en pratique fréquemment réduite par la présence de talus de berges ou de falaises. Cet inconvénient limite l'étendue de la carte radar de part et d'autre du cours d'eau. On a donc très peu d'informations sur le proche bassin versant ou sur la nature des berges.

En imagerie aérienne, les résultats obtenus dépendent largement de la période durant laquelle le survol a été effectué. Cette date est fixée notamment en fonction des conditions météorologiques (vent faible, temps ensoleillé, absence de brumes à l'aplomb de la rivière, etc.). De plus, à l'automne, la fréquentation touristique est quasi-nulle et la ripisylve est moins dense, ce qui limite le masquage de l'interface eau/sol. En revanche, les effets d'ombre portée sont plus importants (figure 4a et 4b). La carte radar, quant à elle, n'est pas affectée par les conditions d'illuminations, les effets d'ombrages ou les conditions météorologiques (figure 4c). D'autre part, l'acquisition radar se déroulant depuis une embarcation, les effets de masquage par la ripisylve sont moins importants que pour une image aérienne.

Abstraction faite des effets d'ombrage, la localisation de l'interface eau/sol est facilitée par un fort contraste sur les images IR et sur la carte radar. La distinction de cette limite peut s'avérer plus difficile sur l'image RGB selon la couleur de l'eau et celle de la berge. De plus, la « couleur » des zones immergées varie sur les images RGB et IR (présence de végétation aquatique, sèches, bancs de sables, radiers, etc.), alors que l'eau apparaît globalement plus uniforme sur la carte radar. En effet, si la surface de l'eau est quasiment lisse, l'onde incidente se réfléchit de manière spéculaire, le radar ne reçoit donc pas d'écho retour. A l'inverse, sur les zones de remous une partie de l'onde incidente est réfléchie en direction du radar, permettant de mettre en évidence les obstacles émergés ainsi que les zones de rapides (figure 5).

Tableau 1. Potentiel et limitations de la cartographie radar par rapport aux images aériennes RGB et IR.

	Radar K2Pi	Photo RGB	Image IR
Mode d'acquisition	au sol	aéroporté	aéroporté
Géoréférencement des données	méthode de Fourier-Mellin	GPS + centrale inertielle	GPS + centrale inertielle
Zone visualisable	100m max. de part et d'autre du radar	fonction du nombre de lignes de vol	fonction du nombre de lignes de vol
Interprétabilité de l'image	faible (coeff. de rétro-diffusion)	forte (couleurs / textures)	forte (couleurs / textures)
Ombres portées + Influence éclairement	non	oui	oui
Masquage par la ripisylve	faible	oui	oui
« Couleur » de l'eau	uniforme	variable	quasi-uniforme
Localisation de rapides	oui	ambiguë	ambiguë

8

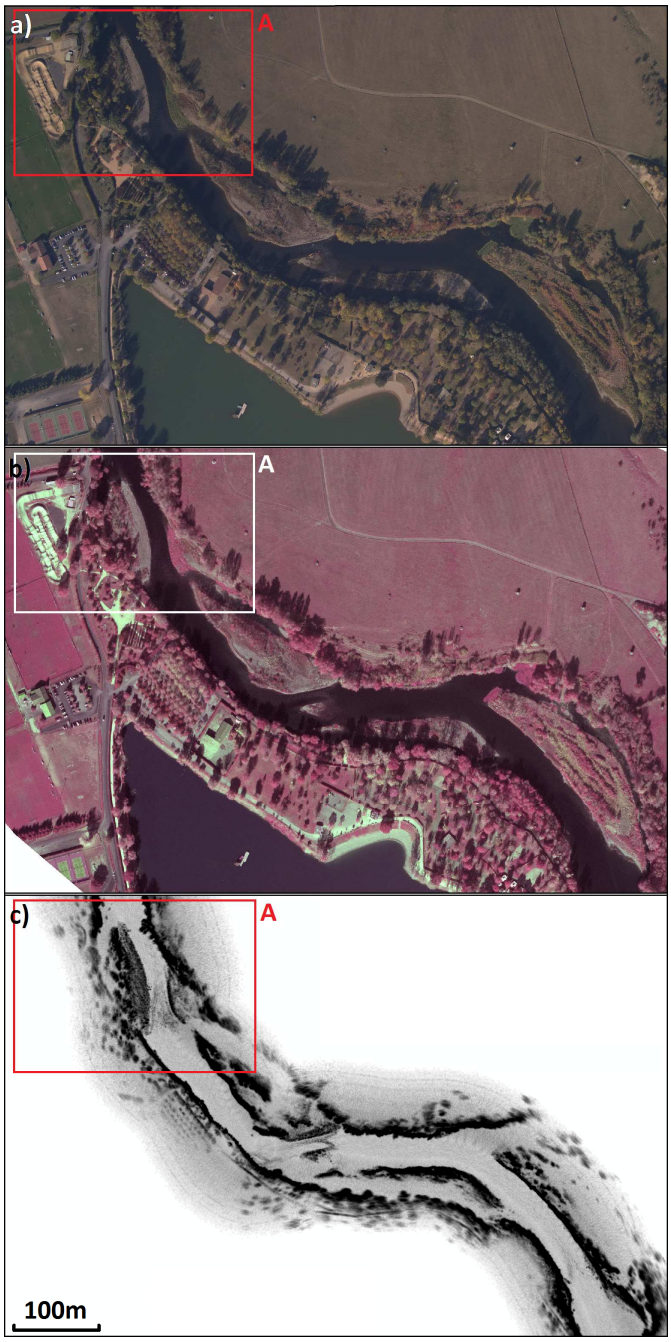


Figure 4 : Comparaison de (a) la photographie aérienne RGB, (b) l'image IR et (c) la carte radar au nord de Cournon. La zone A sera l'objet d'un zoom à la Figure 5.

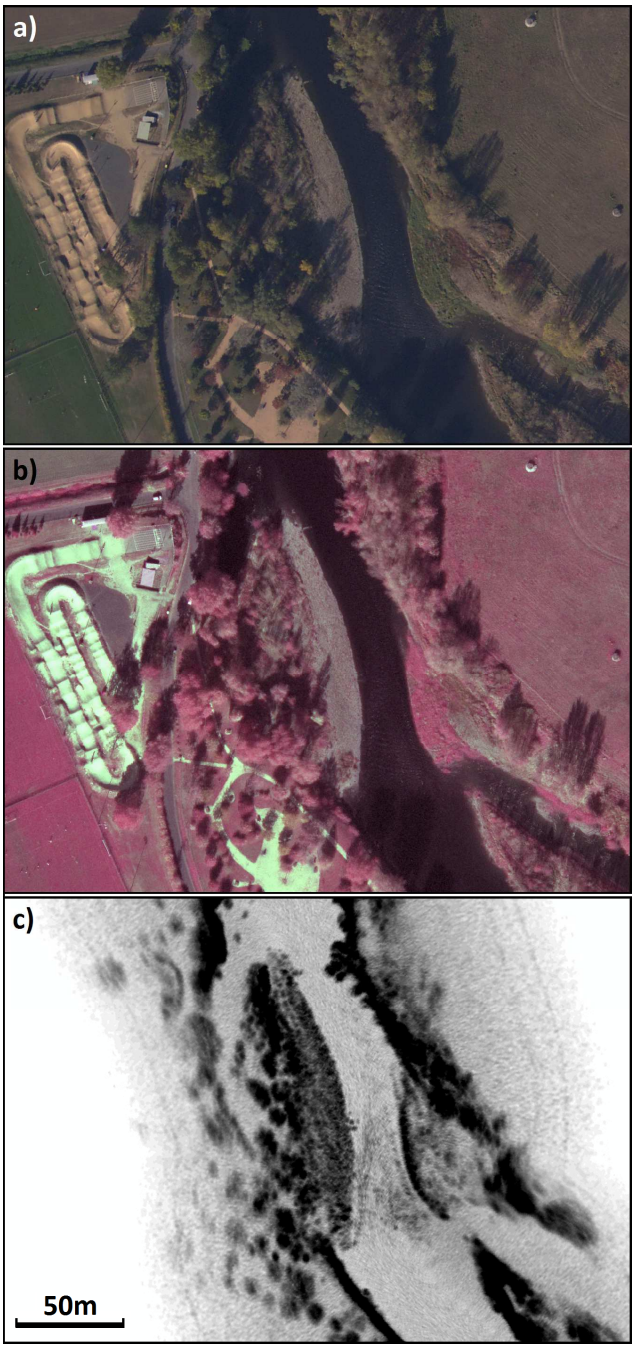


Figure 5 : Comparaison de (a) la photographie aérienne RGB, (b) l'image IR et (c) la carte radar – Zoom sur la zone A : rapide au nord de Cournon (cf. Figure 4).

5. Conclusion

Le radar cartographique K2Pi offre une grande souplesse d'utilisation puisque l'acquisition des données peut s'effectuer indépendamment des conditions météorologiques. Comparé aux techniques classiques de télédétection, l'acquisition n'est pas tributaire de capteurs annexes (GPS, centrale inertielle, etc.), ce qui en diminue le coût. Le radar K2Pi n'est donc pas soumis, contrairement à d'autres systèmes terrestres embarqués, tels que le LiDAR, aux phénomènes de masquages GPS, la carte étant géoréférencée a posteriori via la méthode de Fourier-Mellin.

Nous avons montré que la cartographie radar fait apparaître clairement la position de la limite eau/sol. Néanmoins, elle n'a pas vocation à se substituer aux méthodes classiques de télédétection, notamment du fait de sa couverture limitée et de sa difficulté d'interprétation. L'imagerie aérienne permet également de générer par stéréoscopie un Modèle Numérique de Terrain (MNT), ce que ne peut faire actuellement la cartographie radar. Plutôt adaptée à des études locales, comme par exemple les problématiques de localisation des bancs alluviaux (Malavoi et Bravard, 2010), elle s'avère une méthode rapide et relativement économique permettant de compléter les données de suivis aériens dont le pas de temps est parfois trop long par rapport à la dynamique de la zone.

Remerciements:

Ces travaux ont été réalisés dans le cadre du projet PELICAN N°ANR-2010-EMMA-033 financé par l'Agence Nationale de la Recherche. Ce projet a été labellisé par le pôle de compétitivité ViaMeca.

Bibliographie

Bird S., Hogan D., Schwab J. (2010). Photogrammetric monitoring of small streams under a riparian forest canopy. *Earth Surface Processes and Landforms*, vol. 35, n°8, p. 952-970.

Casasent, D., Psaltis, D. (1976). Position, rotation, and scale invariant optical correlation. *Applied Optics*, vol. 7, n°15, p. 1795–1799.

Dissanayake G., Newman P., Durrant-Whyte H.F., Clark S., Csobra M. (2001). A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics & Automation*, vol. 17, n°3, p. 229-241.

Heritage G., Hetherington D. (2007). Towards a protocol for laser scanning in fluvial geomorphology. *Earth Surface Processes and Landforms*, vol. 32, n°1, p. 66-74.

Higgitt D.L., Warburton J. (1999). Applications of differential GPS in upland fluvial geomorphology. *Geomorphology*, vol. 29, n°1-2, p. 121-134.

Horn J.D., Joeckel R.M., Fielding C.R. (2012). Progressive abandonment and planform changes of the central Platte River in Nebraska, central USA, over historical timeframes. *Geomorphology*, vol. 139, p. 372-383.

Jaud M., Rouveure R., Faure P., Monod M-O. (sous presse). Methods for FMCW radar maps georeferencing. *ISPRS Journal of Photogrammetry and Remote Sensing*.

Jung H.C., Hamski J., Durand M., Alsdorf D., Hossain F., Lee H., Hossain A., Hasan K., Khan A.S., Hoque A. (2010). Characterization of complex fluvial systems using remote sensing of spatial and temporal water level variations in the Amazon, Congo, and Brahmaputra Rivers. *Earth Surface Processes and Landforms*, vol. 35, n°3, p. 294-304.

Lane S.N., Widdison P.E., Thomas R.E., Ashworth P.J., Best J.L., Lunt I.A., Sambrook Smith G.H., Simpson C.J. (2010). Quantification of braided river channel change using archival digital image analysis. *Earth Surface Processes and Landforms*, vol. 35, n°8, p. 971-985.

Liverman D.M., Cuesta R.M.R. (2008). Human interactions with the Earth system: people and pixels revisited. *Earth Surface Processes and Landforms*, vol. 33, n°9, p. 1458-1471.

Malavoi J-R., Bravard J-P. (2010). *Éléments d'hydromorphologie fluviale*, Onema (Office national de l'eau et des milieux aquatiques), Vincennes.

Marcus W.A., Fonstad M.A. (2010). Remote sensing of rivers: the emergence of a subdiscipline in the river sciences. *Earth Surface Processes and Landforms*, vol. 35, n°15, p. 1867-1872.

Monod M-O. (1995). *Frequency modulated radar: a new sensor for natural environment and mobile robotics*. Thèse en Robotique, Université Paris 6.

Notebaert B., Verstraeten G., Govers G., Poesen J. (2009). Qualitative and quantitative applications of LiDAR imagery in fluvial geomorphology. *Earth Surface Processes and Landforms*, vol. 34, n°2, p. 217-231.

O'Neal M.A., Pizzuto J.E. (2011). The rates and spatial patterns of annual riverbank erosion revealed through terrestrial laser-scanner surveys of the South River, Virginia. *Earth Surface Processes and Landforms*, vol. 36, n°5, p. 695-701.

Peynot T., Scheding S., Terho S. (2010). The Marulan Data Sets: Multi-Sensor Perception in Natural Environment With Challenging Conditions. *The International Journal of Robotics Research*, vol. 29, n°13, p. 1602-1607.

Reina G., Underwood J., Brooker G., Durrant-Whyte H. (2011). Radar-based perception for autonomous outdoor vehicles. *Journal of Field Robotics*, vol. 28, n°6, p. 894-913.

Rouveure R., Marionneau A., Chanet M., Monod M-O. (2003). Méthodes dynamiques de caractérisation physique des sols agricole à l'échelle de la parcelle. *Ingénieries*, N° spécial, p. 7-21.

Rouveure R., Monod M-O., Faure P. (2008). Mapping of the environment with a high resolution ground-based radar imager, *Actes du colloque MELECON 2008*, Ajaccio, France.

SAGE (Schéma d'Aménagement et de Gestion des Eaux) Allier Aval (2007). *Complément et mise à jour des connaissances sur la dynamique fluviale de l'Allier entre Vieille Brioude et le Bec d'Allier*. Diagnostic Hydratec – Asconit Consultant.

Scherer S., Rehder J., Achar S., Cover H., Chambers A., Nuske S., Singh S. (2012). River mapping from a flying robot: state estimation, river detection, and obstacle mapping. *Autonomous Robots*, vol. 33, n°1-2, p. 189-214.

Skolnik, M.I. (1980). *Introduction to Radar Systems*, 2nd Ed., McGraw-Hill, New York.

Yang J., Rao D., Chung S-J., Hutchinson S. (2011). Monocular Vision based Navigation in GPS Denied Riverine Environments, *Actes du colloque AIAA Infotech at Aerospace Conference 2011*, St. Louis, USA.

Editeurs : F. GOURMELON, R. THIBAUD

Conception et réalisation : L. DAVID, A. MAULPOIX, M. TISSOT

http://sageo2013.sciencesconf.org/