Predicting sea surface salinity in a tidal estuary with machine learning

Nicolas Guillou, Georges Chapalain, Sébastien Petton

Oceanologia 65 (2023) 318—332

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.journals.elsevier.com/oceanologia

ORIGINAL RESEARCH ARTICLE

Predicting sea surface salinity in a tidal estuary with machine learning

Nicolas Guillou[a,*], **Georges Chapalain**[a], **Sébastien Petton**[b]

[a] *Cerema, DTecREM, HA, Technopôle Brest-Iroise, Plouzané, France*
[b] *Ifremer, University of Brest, CNRS, IRD, LEMAR, Argenton, France*

Received 14 April 2022; accepted 25 July 2022
Available online 10 August 2022

KEYWORDS
Multilayer perceptron;
Support vector regression;
Random forest;
River plume;
Numerical model;
Bay of Brest

Abstract As an indicator of exchanges between watersheds, rivers and coastal seas, salinity may provide valuable information about the exposure, ecological health and robustness of marine ecosystems, including especially estuaries. The temporal variations of salinity are traditionally approached with numerical models based on a physical description of hydrodynamic and hydrological processes. However, as these models require large computational resources, such an approach is, in practice, rarely considered for rapid turnaround predictions as requested by engineering and operational applications dealing with the ecological monitoring of estuaries. As an alternative efficient and rapid solution, we investigated here the potential of machine learning algorithms to mimic the non-linear complex relationships between salinity and a series of input parameters (such as tide-induced free-surface elevation, river discharges and wind velocity). Beyond regression methods, the attention was dedicated to popular machine learning approaches including MultiLayer Perceptron, Support Vector Regression and Random Forest. These algorithms were applied to six-year observations of sea surface salinity at the mouth of the Elorn estuary (bay of Brest, western Brittany, France) and compared to predictions from an advanced ecological numerical model. In spite of simple input data, machine learning algorithms reproduced the seasonal and semi-diurnal variations of sea surface salinity characterised by noticeable tide-induced modulations and low-salinity events during the winter period. Support Vector Regression provided the best estimations of surface salinity, improving especially predictions from the advanced numerical model during low-salinity

* Corresponding author at: Cerema, DTecREM, HA, 155 rue Pierre Bouguer, Technopôle Brest-Iroise, BP 5, 29280, Plouzané, France.
E-mail address: nicolas.guillou@cerema.fr (N. Guillou).
Peer review under the responsibility of the Institute of Oceanology of the Polish Academy of Sciences.

Production and hosting by Elsevier

https://doi.org/10.1016/j.oceano.2022.07.007

events. This promotes the exploitation of machine learning algorithms as a complementary tool to process-based physical models.

1. Introduction

At the interface between watershed, rivers and marine ecosystems, estuaries are important pathways for the exchange, transport and fate of materials (including suspended particles, dissolved nutriments, micro-plastics, or pollutants) between surrounding lands and coastal seas. As a marker of freshwater mixing, salinity is a valuable indicator of these exchanges, fluctuating under the combined influence of riverine inputs and run-offs, tidal intrusion and meteorological forcings. Thus, salinity is a key parameter for assessing the renewal capacity of an estuary by providing further insights into water quality, the health of habitats and biota (Choi and Lee, 2004; Dyer, 1973; Guo and Lordi, 2000). Given the sensitivity to meteorological and hydrodynamic conditions, salinity is also an indicator of the variability of extreme weather events (in occurrence and intensity) liable to impact coastal ecosystems. Reliable monitoring of this environmental parameter may therefore provide valuable information about the exposure, ecological state and robustness of an estuary. This includes especially in situ observations and real-time predictions.

As extensive observations are difficult to achieve (due to technical failure and maintenance operations), salinity is, most of the time, derived from process-based physical computer models liable to approach the interactions between fresh riverine water discharge, density-induced circulation, tide and surface wind forcings (Cruz et al., 2021; Robins et al., 2014; Zhang et al., 2021). However, the implementation of these numerical models requires important computational resources for approaching, at high spatial resolutions, the complex hydrodynamic interactions, exacerbated by increased bottom friction in shallow waters. These models rely furthermore on complex calibrations and an extensive amount of input data including, among others, the spatio-temporal distribution of surface forcings (e.g., wind velocity, atmospheric pressure) or the refined definition of water depth variations along the estuarine channel and bordering wetting-drying areas. For these reasons, whereas such advanced models enable a physical interpretation of processes, these numerical tools remain difficult to apply for rapid turnaround times predictions as requested in engineering and operational applications dealing with the ecological monitoring of the estuary.

However, with the development of Artificial Intelligence (AI) analysis techniques and methods, new solutions may be exploited to approach water quality parameters by including a limited number of input data and computational resources (Maier et al., 2010; Maier and Dandy, 1996). Thus, supervised learning approaches such as Artificial Neural Networks (ANN) are able to produce accurate predictions by learning and/or detecting the underlying patterns and complex relationships between a series of input data and a targeted parameter. MultiLayer Perceptrons (MLPs) refers to one of the most popular ANN models in water-engineering studies (Maier et al., 2010). The basic structure consists of a series of units, called neurons arranged in different hidden layers between (i) an input layer (with input feature) and (ii) an output layer (with the targeted variable). Each unit receives the input information with weight and transfers the output with non-linear activation functions. The different weights are determined during a training phase by error-minimization algorithms between ANN predictions of the targeted variables and the corresponding data.

Adapted to highly non-linear problems, ANNs were therefore exploited to approach the evolution of salinity in estuaries. Motivated by significant economic, ecological and social issues, numerous investigations were conducted in the San Francisco Bay and Sacramento-San Joaquin Delta estuary along the Pacific coast of California (USA) (Chen et al., 2018; Chung and Seneviratne, 2009; He et al., 2020; Rath et al., 2017). However, with the development of ANN data-driven approaches, complementary investigations were also conducted in broader estuarine environments and coastal bays connected to rivers including, among others, the river Murray (in South Australia) (Bowden et al., 2005), the Apalachicola River (Florida, USA) (Huang and Foo, 2002), the Danshui River estuarine system (northern Taiwan) (Chen et al., 2017) or the Hilo Bay (Hawaii) (Alizadeh et al., 2018). These different investigations exhibited the performance of ANN data-driven approaches for estimating salinity in these marine and estuarine environments.

Most investigations relied on MLP or similar ANN to approach salinity variations in response to multiple environmental forcing including freshwater input, water level, tide or wind (Maier et al., 2010). Thus, Huang and Foo (2002) implemented a three-layer ANN — varying the number of neurons in the range (9, 16, 33) in the hidden layer — to approach observed salinity at the mouth of the Apalachicola River system (Florida, USA) with a Root-Mean-Square error (RMSE) down to 1.6 ppt for a five days period. More recently, Chen et al. (2017) compared the exploitation of a three-layer ANN with a three-dimensional (3D) hydrodynamic model for approaching the sea surface salinity in the Danshui River (northern Taiwan). In spite of a tendency to underestimate peak salinity during flood tide and overpredict minimal salinity during ebb tide, the artificial networks considered were able to reproduce tide-induced variations while providing a better estimate than the hydrodynamic model with RMSE below 3.81 ppt between predictions and observations. In order to improve ANN predictive and structural validity, Rath et al. (2017) proposed a hybrid empirical-Bayesian neural network model for approaching salinity in the San-Francisco Bay-Delta estuary while accounting for uncertainties in model parameters.

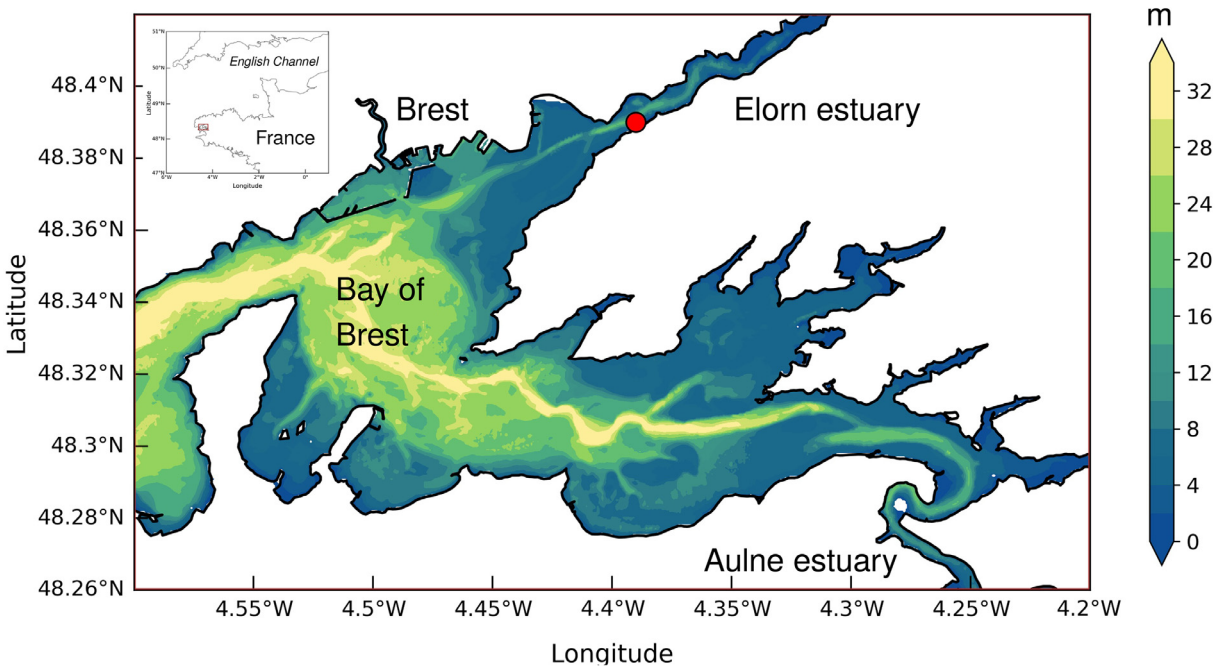

Figure 1 Mean water depth of the bay of Brest and the Elorn estuary. The red circle shows the location of the measurement station considered in the present investigation.

As MLPs and advanced ANNs can not treat the short- or long-term temporal dependencies between input and output time series, investigations were therefore extended to advanced Recurrent Neural Networks (RNNs). Thus, He et al. (2020) compared a series of neural network models, including MLPs and a widely-used RNN, the Long Short-Term Memory network (LSTM), for estimating the downstream boundary salinity in the Sacramento-San Joaquin Delta.

The present investigation complements these different applications of machine learning models for approaching the estuarine salinity under the combined influence of a tide, river freshwater input, precipitation, atmospheric pressure and surface wind. Thus, algorithms such as RNN or LSTM based on temporal dependencies between input and output data were disregarded. Extending the exploitation of ANNs, particular attention was devoted to the performances of a series of popular advanced machine learning (ML) algorithms, including MLP, Support Vector Regression (SVR) and Random Forest (RF). These advanced ML techniques were complemented by the simplest approaches including Multiple Linear Regression (MLR) and Multiple Polynomial Regression (MPR), this in order to assess the progress obtained with advanced ML models. Performances of ML algorithms were assessed against predictions of salinity derived from an ecological numerical model based on a 3D hydrodynamic approach. The application was conducted at the mouth of the Elorn estuary, in the bay of Brest (western Brittany, France), by exploiting a series of in-situ observations of sea surface salinity during a period of six years from 2015 to 2021 (Figure 1). Beyond extending the application of machine learning algorithms to salinity prediction in an estuary of north-western Europe, this study provided an extensive evaluation (not restricted to classical MLPs) about the suitability and capability of ML algorithms to predict the highly non-linear response of an environmental parameter to multiple coastal forcings.

The paper is organised as follows. Section 2 describes the site of application and the environmental conditions. Section 3.1 presents the in-situ observations of surface salinity exploited to train and assess performances of ML algorithms. Sections 3.2 and 3.3 successively describe the process-based physical model and deep-learning algorithms considered. Section 3.4 shows the dataset exploited for the extraction of input variables and the associated preprocessing. Section 4.1 assesses performances of the different ML algorithms. Section 4.2 compares results from ML with predictions from the 3D ecological model implemented in the bay of Brest. Section 4.3 finally discusses the sensitivity of results obtained from ML algorithms with respect to input data.

2. Study area

The site of application is located at the mouth of the Elorn estuary in the bay of Brest, a semi-enclosed basin of north-western Europe separated from the Atlantic Ocean by a 1.8 km wide strait (entitled the “Goulet de Brest”) (Figure 1). The bay is a rich ecosystem characterised by a diversity of marine species and macro-benthic communities which fosters the development of shellfish farming and professional fishing. Particular attention is therefore devoted to the ecological impact of surrounding agricultural, harbour and leisure activities (Chauvaud et al., 2000). Thus, as a result of intensive agriculture, the bay of Brest is receiving high nutrients load from freshwater inputs which increases eutrophic conditions (Le Pape et al., 1996). The bay is also subjected to harbour usage bringing together industrial, yachting, fishing and military activities. One of the

most recent major examples is the extension of the surface area of the harbour to welcome emerging activities from the marine energy sector.

More than 50% of the bay is shallower than 5 m and the maximum depth is around 50 m (Auffret, 1983). This coastal environment is subjected to dominant semi-diurnal tidal regimes with a spring tidal range exceeding 7 m that strongly influence the transport of water mass and suspended particles within the bay and exchanges with the Atlantic Ocean (Beudin et al., 2014; Frère et al., 2017; Petton et al., 2020; Salomon and Breton, 1991). Whereas the bay is characterised by important dispersal capacity directly influenced by strong tidal currents, reduced dispersal capacity is obtained over a long time scale. Thus, the averaged renewal capacity of water within the bay was estimated at three months (Agence de l'eau Loire Bretagne, 1997). This exhibited an increased sensitivity of the bay to substances remaining harmful after high dilution and/or whose degradation rate is low (e.g., metal salts, phytosanitary products, etc.).

Different rivers flow into the bay of Brest. However, the hydrology of the bay is mainly influenced by freshwater runoffs from the Aulne and Elorn rivers which account for around 63 and 15% of the total river input, respectively (Auffret, 1983). Whereas protected from north-western incoming Atlantic waves, this coastal environment may be subjected to local wind-generated surface gravity waves with significant wave heights up to 0.8 m in the northern part of the bay and within the Daoulas cove (Guillou, 2007; Petton, 2010).

Salinity in the Elorn estuary evolves mainly under the opposing contribution of freshwater and tidal flows. Stratification is thus liable to occur during neap tide and for high river discharges, fresh water dominating the upper part of the water column (Quéméneur et al., 1984). Such stratification conditions are liable to result in low salinity events with reduced values of surface salinity at the mouth of the Elorn estuary. But low salinity events may also occur under local and regional weather conditions as a result of the additional contribution of surface wind on salinity temporal variability (Poppeschi et al., 2021).

3. Material and methods

3.1. Observations

The investigation relied on in-situ observations of sea surface salinity for a six-year period (between 02/2015 and 02/2021) conducted at the mouth of the Elorn estuary (long. = 4.39°W, lat. = 48.39°N) (Figure 1). The instrumentation system, entitled BOCA (for "Bouée d'Observation Côtière Automatique multiparamètres") and implemented by the Cerema ("Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement") and its Laboratory of Coastal Engineering and Environment, consists of a multi-parameter YSI data probe attached to a buoy which automatically collects observations. Data, acquired with a time step of 1 s, were processed to obtain averaged values every 20 min. Salinity observations were characterised by different blank periods in relation to maintenance operations and system malfunction (Figure 2). However, over the six-year period (from 04/02/2015 to 01/02/2021), we obtained a series of 20,289 targeted variables, evenly distributed at an hourly time step. This corresponds nearly to more than two years of continuous observations of sea surface salinity at the mouth of the Elorn estuary. The recorded time series both captured (i) the seasonal evolution of sea surface salinity characterised by intense low salinity events during the winter period (with values below 20 ppt) and (ii) the semi-diurnal variations resulting from tidal advection and diffusion (Petton et al., 2020; Poppeschi et al., 2021). The considered dataset represented therefore a valuable source of information to investigate the temporal evolution of the salinity in the mouth of the Elorn estuary.

3.2. Process-based physical model

Performances of deep-learning algorithms were assessed against predictions from a high-resolution 3D hydrodynamic model implemented in the bay of Brest (Petton et al., 2020). Numerical simulations were conducted with the MARS model (Model for Application at Regional Scale) developed at Ifremer ("Institut Français de Recherche pour l'Exploitation de la Mer") (Lazure and Dumas, 2008). The model resolves (i) the continuity equation and the Reynolds-averaged momentum equations derived using the Boussinesq's approximations and the vertical hydrostatic equilibrium and (ii) the 3D transport equations of temperature and salinity. The horizontal turbulent viscosity was set constant equal to 0.5 m^2 s^{-1} whereas the vertical turbulent viscosity derives from a two-equation k-epsilon closure scheme. The computational domain covers the bay of Brest and extends in longitude from 4.09°W to 4.72°W and in latitude from 48.20°N to 48.44°N (Figure 3). This computational domain consists of a curvilinear grid with a horizontal spatial resolution of 50 m and 20-sigma vertical-grid cells. The model was driven by sea-surface elevation derived from a large-scale depth-averaged embedded model covering the western extent of Brittany (Le Roy and Simon, 2003). Atmospheric forcings (pressure, wind velocities, precipitation…) derived from the AROME model (Applications from Research to Operational MEsoscale) implemented by Météo-France (Ducrocq et al., 2005). Freshwater inputs from the different rivers of the bay were finally imposed by relying on hourly observations at upstream stations gathered in the database of Banque Hydro (2021). Further details about the model setup are available in Petton et al. (2020).

The model was assessed against a series of observations of hydrodynamic and environmental parameters including tidal sea level, current velocities, temperature and salinity (Petton et al., 2018, 2016). Predictions of salinity were compared with observations at two stations located at the entrance of the bay of Brest and in the south-eastern part of the bay (Poppeschi et al., 2021). In spite of a tendency to overestimate low salinity events, simulations reproduced the seasonal cycle of sea surface salinity. This model is therefore considered as a reference tool for assessing environmental and ecological issues within the bay, with the ability to capture the complex interactions between river plumes and tide- and wind-induced circulations as exhibited by the spatial distribution of sea surface salinity predicted during two contrasting events (Figure 3).

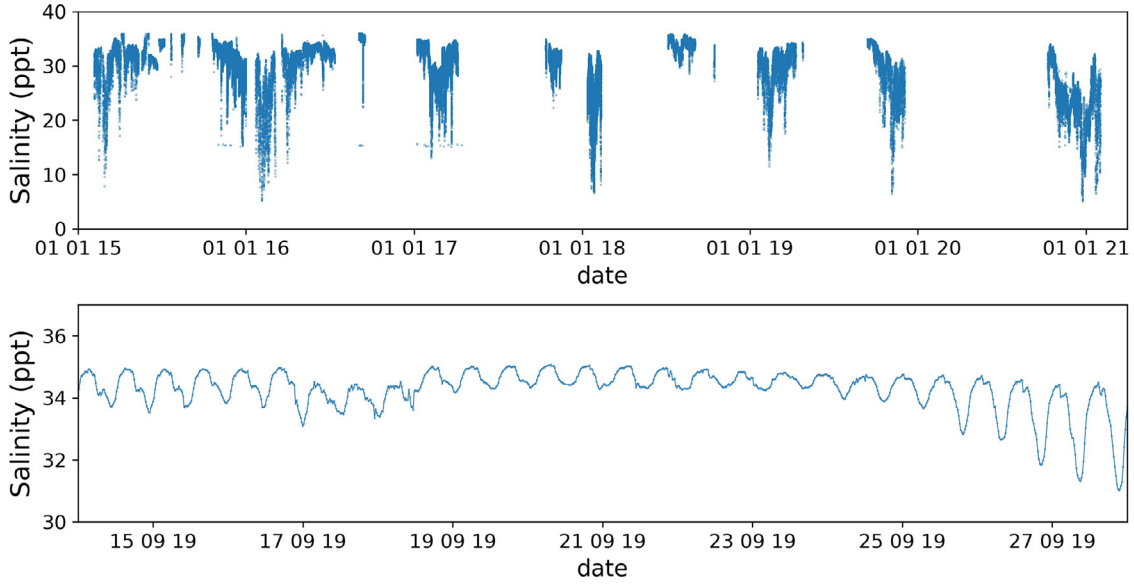

Figure 2 Time series of observed sea surface salinity at the mouth of the Elorn estuary (top) over the six-year period considered (between 2015 and 2021) with (bottom) a detailed view in tide-dominated conditions (between 15 and 27/09/2019).

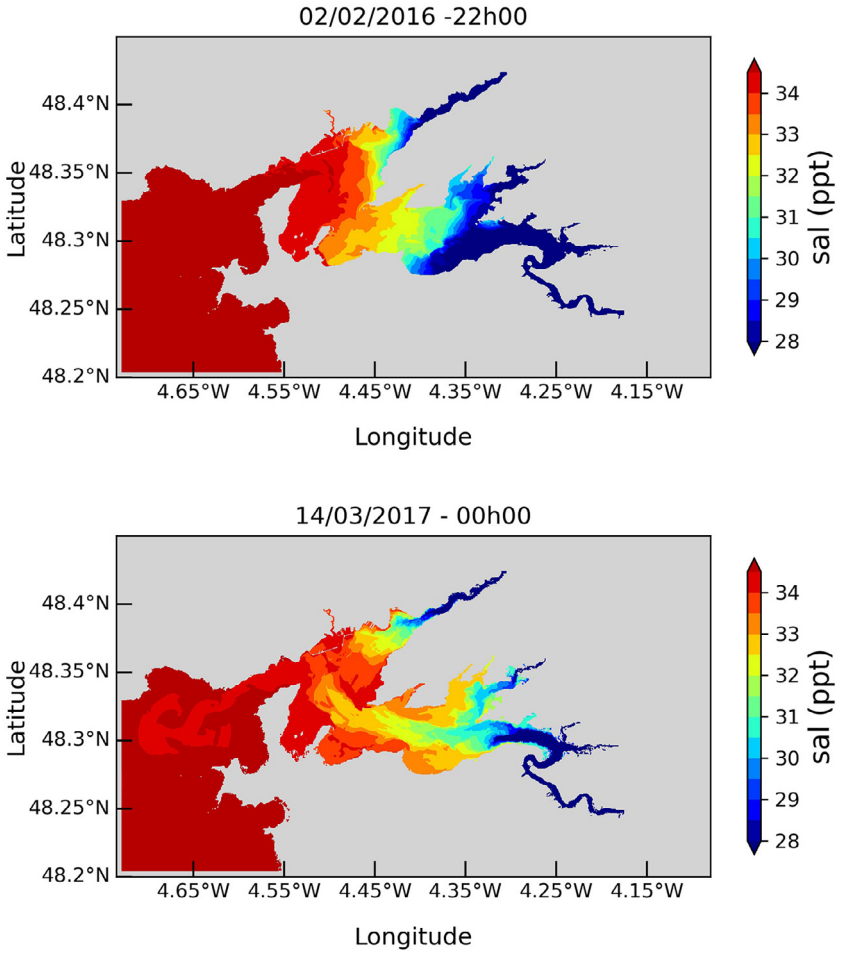

Figure 3 Spatial distribution of predicted sea surface salinity in the bay of Brest (top) on 02/02/2016 (22h00) with a strong effect of western wind on salinity intrusion within the bay and (bottom) on 14/03/2017 (00h00) with a noticeable interaction of salinity flow from Elorn and Aulne rivers. Note that these synoptic views show also the computational domain that covers the bay of Brest.

Table 1 Characteristics of machine learning algorithms retained.

Machine learning algorithms	Main characteristics
MLR and MPR	deg=2
MLP	1 hidden layer with 5 neurons, epochs=100, batch_size=10
SVR	RBF kernel, ϵ=0.1, C=100 and Γ=0.001
RF	n_estimators=1200, max_features=sqrt, max_depth=80, min_samples_split=3, min_samples_leaf=4 and bootstrap considered

Over the period of salinity observations at the mouth of the Elorn estuary, data from this numerical model were available from 2015 to 2018 with a time step of 15 min. The comparison with machine learning algorithms was therefore adapted to this dataset.

3.3. Machine learning algorithms

We propose here a brief description of multiple regression methods and machine learning algorithms considered in the present investigation. This includes simple Multiple Linear and Polynomial Regression methods (MLR and MPR), and more advanced MultiLayer Perceptron (MLP), Support Vector Regression (SVR) and Random Forest (RF). Further details about the parametrisation of these algorithms are provided in Table 1.

The performances and reliability of these different models were evaluated by relying on three statistical and scoring metrics including the Mean Absolute Error (MAE), the Root-Mean Square Error (RMSE), the Normalised Root-Mean Square Error (NRMSE) and the coefficient of determination R^2 between observations and predictions. The algorithms were implemented by relying on the Deep Learning Python libraries Scikit-learn and Keras (Keras, 2021). The random seed number was fixed to guarantee the reproductibility of results obtained.

3.3.1. MLR and MPR

MLR (Multiple Linear Regression) is one of the simplest supervised learning techniques, applied basically to determine the best linear trend lines between a series of input datasets and a targeted variable. The coefficients which weighted linearly the input values are determined by minimizing the sum of squared residuals between the estimated output and the targeted variable for all observations of the trained dataset. In comparison with MLR, MPR (Multiple Polynomial Regression) relies on a polynomial regression function. As the regression function includes non-linear terms, MPR are more adapted for approaching targeted observations subjected to non-linear response to input values (such as sea surface salinity). MPR depends naturally on the degrees of the polynomial regression function. However, preliminary estimations showed that increasing this degree diminished the performance of MPR. Thus, in the present investigation, we considered MPR with a polynomial regression function of degree two (Table 1).

3.3.2. MLP

MLP (Multilayer Perceptron) consists basically of three types of layers including (i) an input layer with a series of input features, (ii) hidden layers with a series of neurons (also called perceptrons) that receive the input values with weight and transfer it with a non-linear activation function, and (iii) the output layer with the final estimation of the targeted variable (Figure 4). Considering its capability for addressing the vanishing and exploding gradient problems in MLP, the Rectified Linear unit (ReLu) was retained for the activation function between hidden layers (Nair and Hinton, 2010). A linear function was considered for the output layer. Weights were updated by back-propagating the error from the output layer to the hidden and input layers with error-minimization algorithms. We relied here on the Adam optimization algorithm to optimize a mean squared error loss function between targeted variables and corresponding observations (Kingma and Ba, 2017). Further details about MLP are available, among others, in Azencott (2019). As increasing the depth of the network may increase the risk of over-fitting (therefore reducing the generalisation potential of the trained algorithm), we retained MLP with a reduced number of hidden layers and perceptrons per layer. Following the great part of salinity approaches based on MLP (Chen et al., 2017; Huang and Foo, 2002), we considered a three-layer ANN, thus restricting the algorithm to one hidden layer. For the case with one hidden layer, preliminary estimations showed that a slightly better approach of the observed salinity was obtained for five neurons in the hidden layer. The learning algorithm was finally implemented with a number of epochs (iteration of updated weights on batch samples) set to 100 and a batch size (number of subsamples of the trained dataset) set to 10 (Table 1).

3.3.3. SVR

Initially introduced by Vapnik (1995) and Cortes and Vapnik (1995), Support Vector Machine (SVM) is a kernel-based approach that provides a statistical model for distinguishing patterns of data. Thus, SVM relies on a hyperplane surface or a set of hyperplanes as a decision boundary to draw the line between different datasets (Figure 4). SVM was mainly considered for classification issues with Support Vector Classification (SVC). However, it was also adapted for regression problems (Drucker et al., 1997; Vapnik et al., 1996), thus resulting in Support Vector Regression (SVR). In SVR, the objective is to find the optimal surface that fits the data within a threshold value that defines how much error is acceptable in the model. This threshold value represents the distance between the hyperplane and the boundary line established by relying on data points closest to the hyperplane (data points also called Support Vectors). This method relies on a kernel function that transforms the data to a higher dimension and performs the separation. There are different types of kernel functions including linear, Gaussian, polyno-

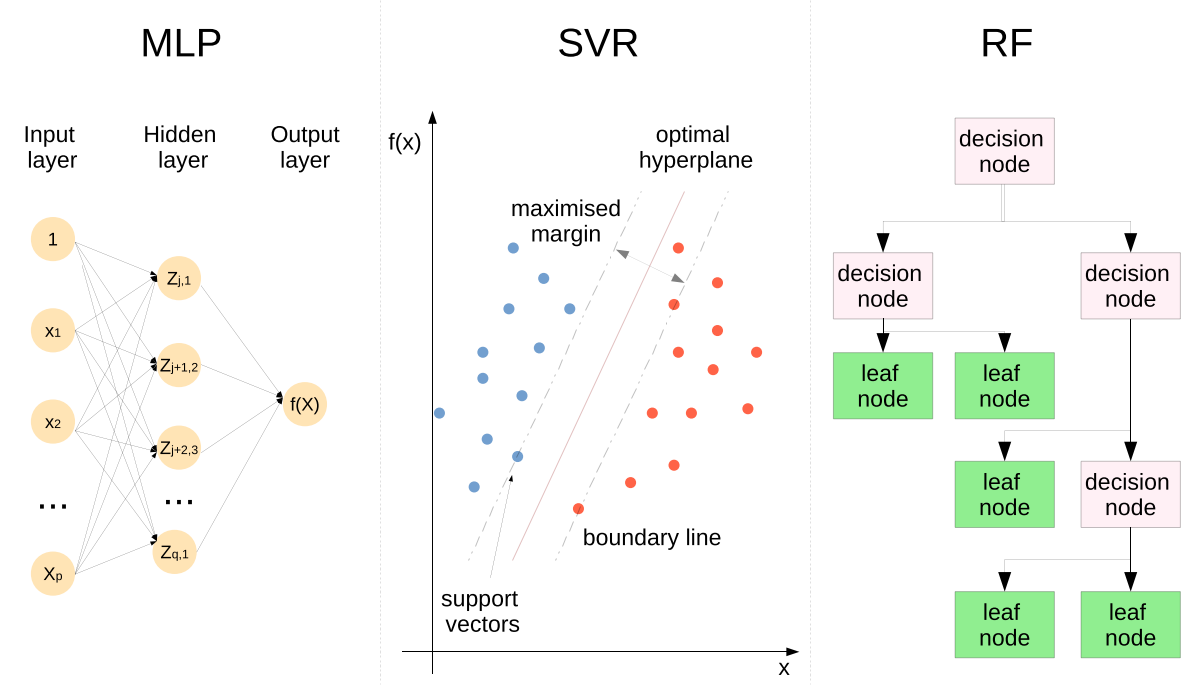

Figure 4 Schematic representation of (left) MLP (for a three-layer ANN), (middle) SVR (for a linear regression with one feature and one outcome) and (right) a decision tree of a RF.

mial or Radial Basis Function (RBF). In the present investigation, we relied on the RBF, considered as one of the most popular choices for a kernel type in SVRs (Hsu et al., 2010; Keerthi and Lin, 2003). In comparison with simple linear or multiple regression methods based on ordinary least squares, SVR offers therefore increased flexibility by defining an acceptable range of values for the model error via a hyperplane to fit the data. This enhances the generalised regression efficiency of SVR models.

However, three parameters have to be considered to establish the SVR model: (i) the loss function ϵ, (ii) the penalty parameter C and (iii) the slack parameter Γ. ϵ determines the region of insensitivity around the hyperplane. This term impacts tolerance for the error and the solution sparsity. However, in order to account for larger errors and integrate an increased number of data in the algorithm (thus improving its generalisation capability), slack variables C and Γ are also introduced. The penalty parameter C accounts for increased acceptable data points in the model. Low values of C will increase the tolerance for data points outside of ϵ as a reduced penalty is applied to these points whereas high values will heavily penalize these data points resulting in increased intolerance of the algorithm and a decision boundary more dependent on the individual data. In this latter situation, the trained algorithm may be overfitted. The slack parameter Γ defines finally the spread of the kernel considered (here the RBF kernel) and the decision region. Thus, low values of Γ will result in reduced curvature of the decision boundary with a broad decision region whereas high values will increase the curve of the decision boundary reducing the spread of the kernel with better coverage of data. However, high values of Γ tend also to increase the dependency between decision boundary and individual data points, resulting in overfitting of the algorithm. Further details about the description and implementation of SVR for regression issues of environmental parameters are available, among others, in Nguyen et al. (2021) and Su et al. (2015).

A tuning procedure was adopted to determine the three parameters which provided the best estimation of the targeted variable during the supervised learning. This evaluation was performed for a fixed ϵ of 0.1 with C in the range [0.001, 0.01, 0.1, 1, 10, 100] and Γ in the range [0.0001, 0.001, 0.01, 0.1] resulting, from preliminary computations, in the optimised parameters of C=100 and Γ=0.001 (Table 1).

3.3.4. RF

RF (Random Forest) is a popular machine learning algorithm that can be applied to both classification and regression. In comparison with other machine learning techniques, RF offers numerous advantages including stability, refined accuracy, applications to large datasets with heterogeneous feature types (e.g., categorical against numerical types). However, RFs may show limitations for predictions outside the range of training data. RF is an ensemble method that relies on a large number of small decision trees, called estimators, resulting in specific predictions of the targeted variables (Figure 4). Decision trees are flowchart-like structures designed to reach a final decision through a series of tests. Thus, decision trees are made of (i) nodes that correspond to tests, (ii) branches that account for outcomes of the tests, and (iii) leaf nodes that represent final decisions. RF relies on a series of hyperparameters including, for the most important, (i) the number of trees in the forest (n_estimators), (ii) the number of features considered for splitting at each leaf node (max_features), (iii) the maximum number of levels in

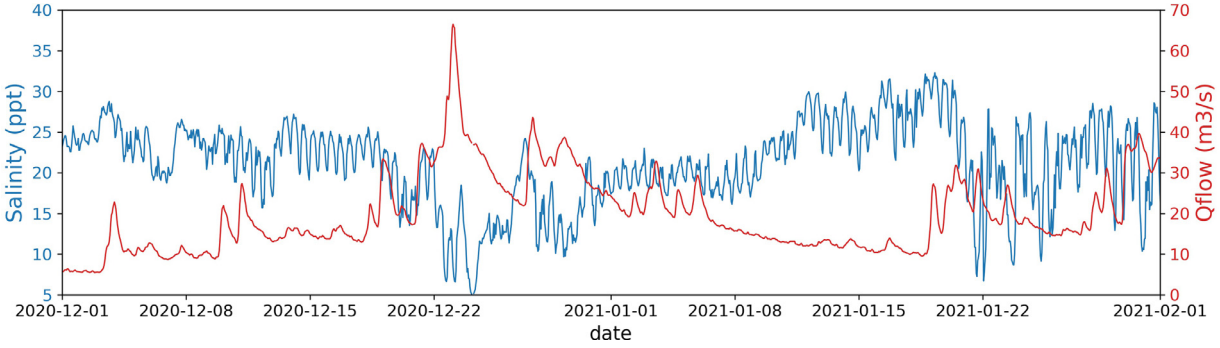

Figure 5 Time series of sea surface salinity observed at the mouth of the Elorn estuary and river outflow observed upstream of the Elorn river.

trees (max_depth), the minimum numbers of samples required to (iv) split a node (min_samples_split) and (v) at each leaf node (min_samples_leaf) and (vi) the method of selecting samples for training each tree (bootstrap or not).

Given the number of hyperparameters, the tuning procedure may be time-consuming in terms of computational resources. In the present investigation, we first relied on a K-fold cross-validation on trained dataset for a rough evaluation of the range of values of hyperparameters. This rough evaluation was then refined by directly specifying the values of hyperparameters to consider and retaining the parameters which provided the best estimation of the targeted variable (in a similar manner as for SVR — previous section). We obtained finally the following hyperparameters: n_estimators = 1200, max_features = square root of the number of features, max_depth = 80, min_samples_split = 3, min_samples_leaf = 4 and bootstrap considered (Table 1).

3.4. Output and input data

As described in Section 3.1, the targeted data was the sea surface salinity observed at the mouth of the Elorn estuary. Input data of machine learning algorithms were therefore selected in relation to their potential influence on salinity variation at this location. Thus, in order to represent the semi-diurnal variations associated with tide-induced river plume advection and dispersion as observed at the mouth of the Elorn estuary (Figures 2 and 3), we retained the variations of tide-induced free-surface elevation, FS_{tide}, as the first input parameter. Data were taken from tidal-gauge observations conducted by the French Navy SHOM ("Service Hydrographique et Océanographique de la Marine") in the harbour of Brest and available at a time step of one hour (SHOM, 2021). Given the close relationship with freshwater inputs exhibited in Figure 5, discharges from the river Elorn, $River_{Elorn}$, were also considered. Upstream river flows were extracted from hourly observations gathered in the database of Banque Hydro (2021). However, discharges from the Aulne river, which may also influence the salinity at the mouth of the Elorn estuary (Figure 3), were not considered as these data were highly correlated with discharges from the Elorn river. The detailed analysis of salinity variation performed by Poppeschi et al. (2021) exhibited furthermore the superimposed effect of meteorological conditions on extreme low salinity events in the bay of Brest, including especially the influence of surface wind. Thus, reduced salinity occurred not only after a peak in river discharge but also under favourable surface wind conditions liable to advect the river plume towards the centre of the bay. Meteorological observations of wind velocity magnitude and direction were therefore considered. However, the wind direction can not be characterised like its magnitude. Indeed, the value of 0 is similar to the value of 2Π. Thus, we selected the projection of the wind velocity along the orientation of the Elorn estuary, $Wind_{proj}$, (estimated at around 20° with respect to longitude) as an input parameter. Data were taken from international surface observations messages of the World Meteorological Organization for the city of Brest (WMO, 2021). The three input variables retained are listed in Table 2. The different input and output data considered were finally interpolated with a time step of one hour. Given the different range values of these input variables, these features were standardised by removing the mean and scaling to unit variance.

The application of machine learning algorithms was conducted by dividing the input and output datasets into two parts including (i) training for the supervised learning of data-driven approaches considered and (ii) validation for the comparison of these different models and their assessment with respect to the process-based physical model. Thus, the testing phase was ignored setting aside the evaluation of the generalisation error from the optimized model. Machine learning models were therefore trained and validated in the ratio 70:30% of the total observed dataset of sea surface salinity at the mouth of the Elorn estuary. However, predictions from the process-based physical model were available over the period 2015—2018, only. And the period of available observations extends from 04/02/2015 to 01/02/2021 (Section 3.1). Thus, in order to conduct the comparison of machine learning algorithms with predictions from the process-based physical model, the trained dataset was taken from the last 70% of input and output data whereas the validated dataset was taken from the first 30%. Thus, the training period extended from 25/03/2016 to 01/02/2021 whereas the validation period extended from 04/02/2015 to 25/03/2016.

Table 2 Description of input variables considered in machine learning algorithms.

Input variables	Description	Reference
FS_{tide}	Free-surface elevation at Brest harbour	SHOM (2021)
$River_{Elorn}$	River flow upstream of Elorn	Banque Hydro (2021)
$Wind_{proj}$	Wind velocity projection along the orientation of the Elorn estuary	WMO (2021)

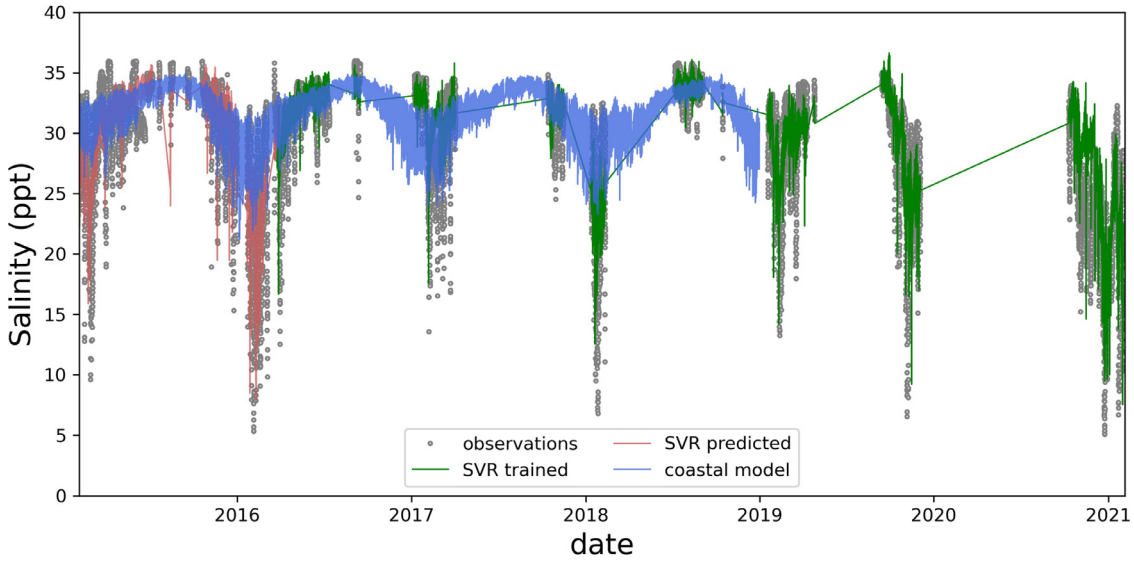

Figure 6 Time series of sea surface salinity observed at the mouth of the Elorn estuary, predicted from the coastal numerical model and obtained from optimised SVR during the training and validation periods from 04/02/2015 to 01/02/2021.

4. Results and discussion

4.1. Model selection

Results obtained from the five machine learning algorithms were very close to each other. Thus, the different models reproduced the seasonal cycle characterised by high salinity values (over 34 ppt) in summer and extreme low salinity events (with values below 20 ppt) in winter (Figures 6 and 7). Predictions obtained approached also the semi-diurnal modulations of salinity particularly noticeable at the mouth of the Elorn estuary where tidal currents predominantly influenced the mixing between salt water from the bay of Brest and fresh water from the Elorn river. However, despite very close results, the first classification of trained algorithms was established by relying on a series of statistical and scoring metrics (Table 3). MLR and RF resulted in the most important differences between predictions and in-situ observations. In spite of its simplicity of implementation, MPR provided slightly better predictions than more complex MLP. Indeed, as exhibited in the introduction, the response of sea surface salinity to external forcings (here tide, river outflow and wind) required an algorithm adapted to non-linear problems. Both MLP and MPR were able to capture these non-linearities, the first through nonlinear activation functions, the second with a multiple regression based on a polynomial function. This comparison exhibited furthermore that a great part of the non-linearities associated with the response of sea surface salinity was captured with a polynomial regression function of degree two which may explain the slight differences obtained between the two ML algorithms. Best performances were finally obtained with the optimised SVR which resulted in reduced differences and errors for MAE, RMSE and NRMSE between predictions and observations, and improved determination for R^2.

4.2. Deep-learning vs. physical model

In spite of reduced computational times, the machine learning model selected provided an approach to the temporal variations of sea surface salinity comparable to the numerical process-based physical model (Table 3). Thus, the five ML models considered resulted in lower RMSE (and NRMSE) than the numerical model. But the SVR model was the only one to provide slightly better MAE decreasing its values from 2.29 ppt to 2.26 ppt. These two values are very similar. However, most improvements were reached with the coefficient of determination R^2. Indeed, for the numerical model, this coefficient was negative exhibiting that predictions failed to fit observations whereas an estimation of 0.51 was obtained for the SVR model. These important differences were mainly associated with the approach of low-salinity events (Figures 6, 7 and 8). Indeed, as exhibited by Poppeschi et al. (2021), at the entrance of the bay, the numerical model overestimated surface salinity during these events with predicted minimum values of 25.5 ppt against observed minimum values of 23.5 ppt. These differences were here exhibited as the location considered (at the en-

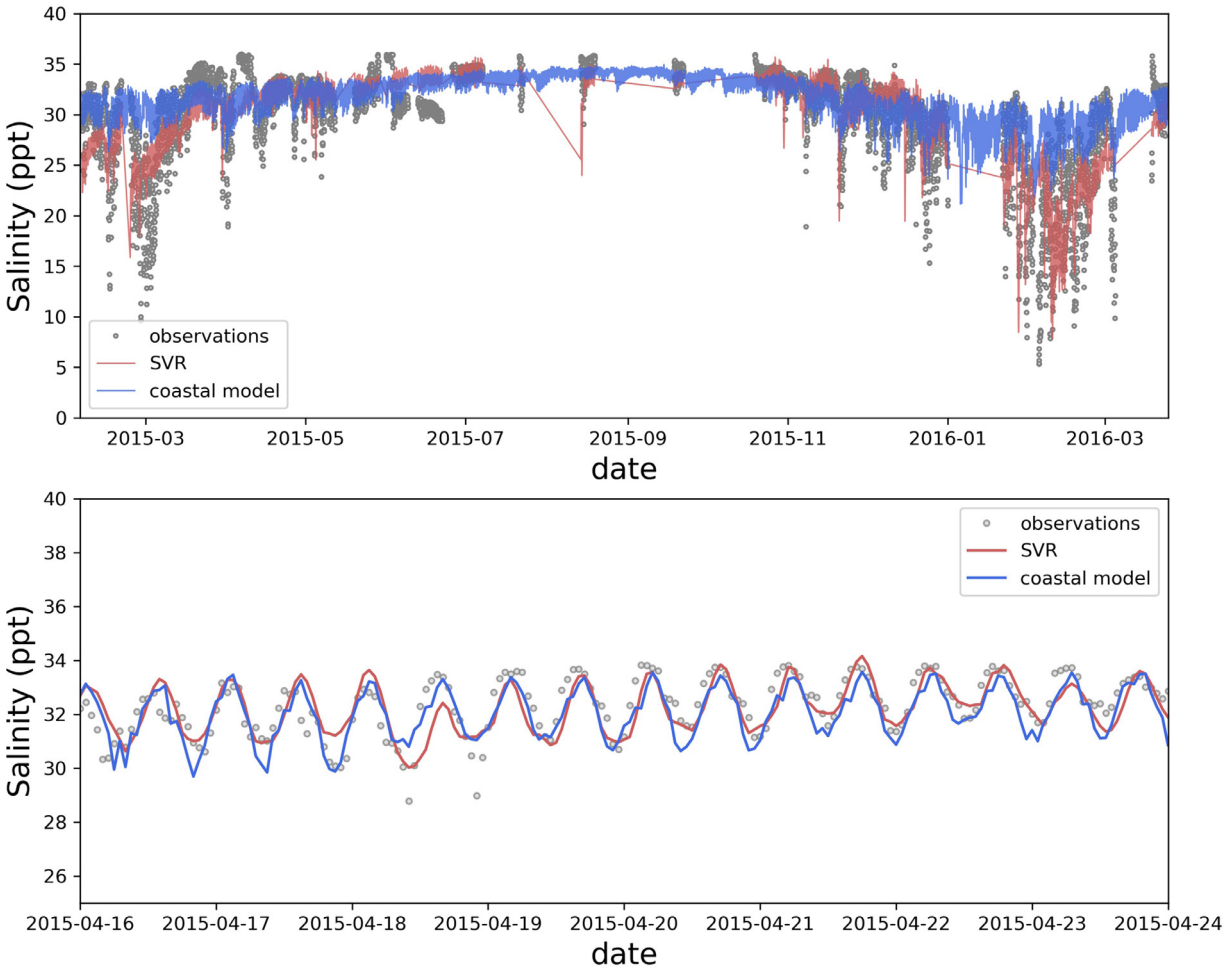

Figure 7 Time series of sea surface salinity observed at the mouth of the Elorn estuary and predicted by the coastal numerical model and the SVR algorithm (top) during the validation period (from 04/02/2015 to 25/03/2016) with (bottom) a detailed view in tide-dominated conditions (between 16 and 24/04/2015).

Table 3 Scoring for the evaluation of observed salinity for the validation dataset based on regression models (MLR and MPR), the three machine learning algorithms considered (MLP, SVR and RF) and the process-based physical model (MARS model). Statistical and scoring metrics considered include the Mean Absolute Error (MAE), the Root-Mean-Square Error (RMSE), the Normalised Root-Mean-Square Error (NRMSE) and the coefficient of determination R^2.

Deep-learning algorithms /Process-based physical model	MAE	RMSE	NRMSE	R^2
MLR	2.46 ppt	3.48 ppt	11.7%	0.29
MPR	2.33 ppt	3.14 ppt	10.5%	0.49
MLP	2.42 ppt	3.26 ppt	10.9%	0.48
SVR	**2.26 ppt**	**3.16 ppt**	**10.6%**	**0.51**
RF	2.44 ppt	3.32 ppt	11.1%	0.46
MARS model	2.29 ppt	3.73 ppt	12.5%	-2.52

trance of the Elorn estuary) was subjected to a stronger influence of fresh river discharges. Thus, during the validation period, observed minimum values reached 5 ppt while predictions from the coastal model remained over 22 ppt. These differences may be explained by the difficulty of the model to approach the transport of fresh waters from upstream river boundaries to the entrance of the estuary. A refined spatial numerical model may be implemented to approach the exchanges of water (and salinity) along the estuary, but this requires also an improved spatial distribution of the bathymetry (which is not currently available). In comparison, machine learning algorithms were able to capture a part of these low-salinity events. Thus, during the validation period, predictions from SVR resulted in minimum salinity of 5.4 ppt (against 5.3 ppt for observations). And these results were obtained with a limited number of input data, setting especially aside extensive measurement campaigns of water-depths spatial distribution.

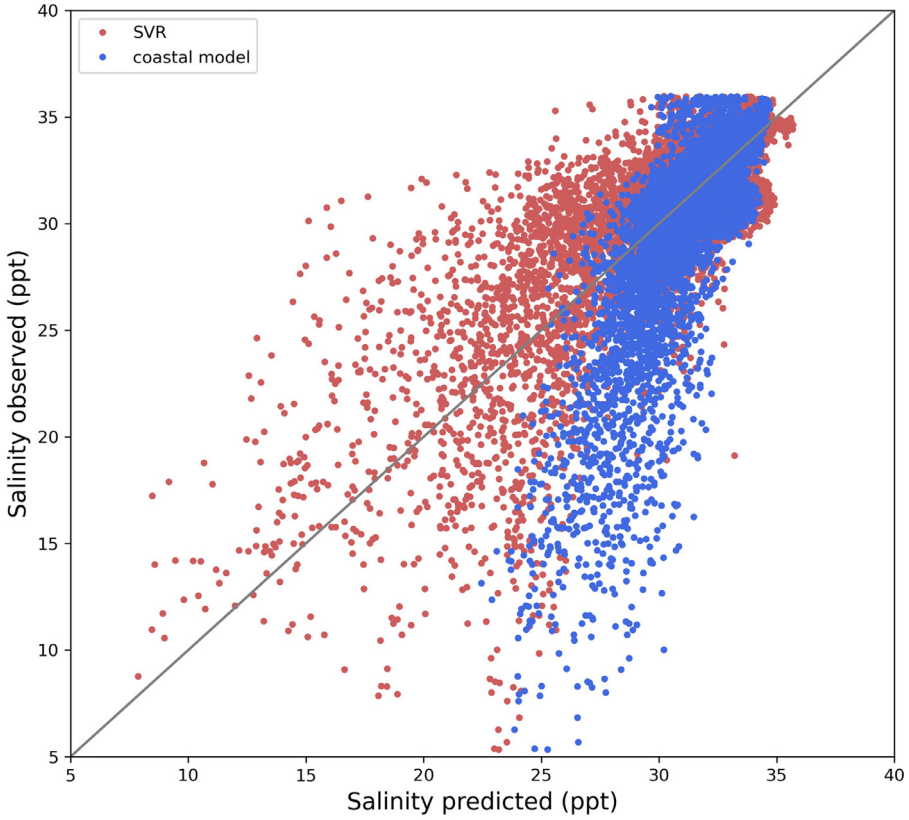

Figure 8 Correlation between sea surface salinity observed at the mouth of the Elorn estuary and predicted from SVR and the coastal numerical model during the validation period (from 04/02/2015 to 25/03/2016).

4.3. Sensitivity analysis

In the present investigation, the attention was dedicated to efficient and practical algorithms liable to approach environmental parameters such as sea surface salinity with limited computational resources and simple input data easily accessible. Thus, the investigation was conducted by relying on three input parameters: the tide-induced surface elevation in the nearest harbour, the upstream river flow and the wind velocity observed in the nearest station (Section 3.4). However, we may expect further improvements of ML algorithms by including input data more representative of processes driving the temporal variation of sea surface salinity at the mouth of the Elorn estuary (whereas these data are more difficult to access). Thus, tide-induced surface elevation may be replaced by tidal currents in the vicinity of the measurement point as a more refined parameter driving the tide-induced transport of salinity. We may also consider an extensive number of input data including the precipitation rate, P_{rate}, and the air temperature Temp. Indeed, the precipitation rate may be interesting to include to represent the impact of high rapid rainfall on surface salinity. Sea surface temperature may furthermore be exploited as a refined indicator of the seasonal variability between (i) the winter period characterised by an increased number of flood events and (ii) the summer period with reduced river outflow within the bay. The investigation of the influence of input data was conducted by relying on the SVR algorithm. The analysis was conducted in two steps. First, the model was exploited to investigate the interest of using tidal current instead of free surface elevation. Second, different estimations from the SVR model were compared varying the number of input data. These different applications were conducted by adopting the tuning procedure retained in Section 3.3.3, thus varying the values of hyperparameters with respect to the input dataset and parameters considered.

An estimation of sea surface salinity based on optimised SVR was conducted by replacing the free-surface elevation with the tidal current in the vicinity of the measurement location. Predictions from the process-based physical model were thus exploited to extract the horizontal components of the depth-averaged current velocities at the measurement point. As for wind velocity, we retained the projection of the current velocity along the orientation of the Elorn estuary as both horizontal components were highly correlated, and as the inclusion of the current direction required considering its orientation with respect to the estuary. The tuning procedure provided the optimised parameters of $C = 100$ and $\Gamma = 0.1$ (Table 4). The resulting optimised SVR, entitled SVR#2, resulted in statistical metrics comparable to values obtained with the five ML algorithms considered in Section 4.1 (Table 3). However, whereas the model approached the seasonal variability of sea surface salinity characterised by low-salinity events during the winter period, increased differences were obtained at the diurnal scales. Thus, SVR#2 resulted in lower tide-induced modulations of salinity than SVR#1, and this increased differences with observations. Indeed, the tidal current is a location-

Table 4 Scoring for the evaluation of observed salinity for the validation dataset based on optimised SVR with different input data.

Optimised SVR	Input variables	MAE	RMSE	NRMSE	R^2
SVR#1 (RBF kernel, ϵ=0.1, C=100 and Γ=0.001)	**FS_{tide}, $River_{Elorn}$, $Wind_{proj}$**	**2.26 ppt**	**3.16 ppt**	**10.6%**	**0.51**
SVR#2 (RBF kernel, ϵ=0.1, C=100 and Γ=0.1)	U_{proj}, $River_{Elorn}$, $Wind_{proj}$	2.44 ppt	3.41 ppt	11.4%	0.42
SVR#3 (RBF kernel, ϵ=0.1, C=10 and Γ=0.001)	FS_{tide}	3.26 ppt	4.83 ppt	16.2%	-38.1
SVR#4 (RBF kernel, ϵ=0.1, C=100 and Γ=0.001)	FS_{tide}, $River_{Elorn}$	2.28 ppt	3.20 ppt	10.7%	0.50
SVR#5 (RBF kernel, ϵ=0.1, C=100 and Γ=0.001)	FS_{tide}, $River_{Elorn}$, P_{rate}	2.26 ppt	3.14 ppt	10.5%	0.51
SVR#6 (RBF kernel, ϵ=0.1, C=100 and Γ=0.001)	FS_{tide}, $River_{Elorn}$, Temp	2.31 ppt	3.21 ppt	10.8%	0.50
SVR#7 (RBF kernel, ϵ=0.1, C=100 and Γ=0.001)	FS_{tide}, $River_{Elorn}$, $Wind_{proj}$, P_{rate}	2.25 ppt	3.13 ppt	10.5%	0.52
SVR#8 (RBF kernel, ϵ=0.1, C=100 and Γ=0.001)	FS_{tide}, $River_{Elorn}$, $Wind_{proj}$, P_{rate}, Temp	2.26 ppt	3.13 ppt	10.5%	0.52

specific characteristic of the hydrodynamic whereas the tide-induced free-surface elevation is more adapted to the global variation of the tidal cycle within the bay. Thus, free-surface elevation offers greater freedom than local tidal current to adapt the ML algorithm to targeted data. As an example, if the input current is highly rectilinear with two opposite directions between peak flood and ebb, we may expect rapid and frank variations of the predicted salinity while neglecting potential remote influences of salinity transport by rotary currents.

Taking into account the previous estimation, the sensitivity study to the number of input data was conducted by retaining the free-surface elevation to characterise the effect of the tide. As the wind velocity, meteorological data added (P_{rate} and Temp) were taken from observations messages of the World Meteorological Organization for Brest (WMO, 2021). Differences between SVR#3 (with FS_{tide}) and SVR#4 (with FS_{tide} and $River_{Elorn}$) confirmed the importance of both considering the tide-induced free-surface elevation and upstream river outflow to approach salinity variations at the mouth of the Elorn estuary (Table 4). However, reduced improvement was reached by including the third variable among the wind velocity, the precipitation rate and the air temperature. The inclusion of P_{rate} with FS_{tide} and $River_{Elorn}$ (SVR#5) appeared to provide slightly better estimations of sea surface salinity. But the three estimations from SVR#1, #5 and #6 were very close. An explanation is that meteorological conditions have an impact on sea salinity during localised events with a short period of time in comparison to the continuous and/or more frequent effect of tide and river discharges. For the precipitation rate, we may also refer to the nature and properties of watersheds of the bay of Brest. Thus, watersheds of the bay consist mainly of impermeable rocks and soils which increases the influence of precipitation on river discharges (Tréguer et al., 2014). The consequence is that floods and river discharges tend to mirror precipitation, especially during the winter period when soils are water-saturated. Reduced improvement was thus reached by including the precipitation rate.

However, whereas the coastal numerical model took into account past changes of hydrological conditions to predict salinity, ML algorithms (considered in the present investigation) neglected the previous evolution of dataset, setting especially aside the time delay between input data and the targeted parameter. And this time delay may be more important for the different input variables, including especially the river discharge and the precipitation rate. Thus, in the bay of Brest, sea salinity may be impacted by a peak in river discharges after a time lag of 10 days (Petton et al., 2020; Poppeschi et al., 2021). By analysing salinity observations at the entrance of the bay, Poppeschi et al. (2021) also noticed that low salinity events were always associated with a peak in precipitation between two and three days before these events. A detailed investigation of predictions from the numerical model confirmed furthermore the additional effects of surface wind, inputs from rivers (Aulne and Elorn) and tide-induced advection and diffusion on the duration and intensity of these low-salinity events. The inclusion of input parameters with a more refined definition may help to remove these uncertainties, but it may also be interesting to test algorithms liable to take into account previous states in the input parameters such as the LSTM.

5. Conclusion

A series of machine learning (ML) models were exploited to approach the temporal variations of sea surface salinity at the entrance of the Elorn estuary (bay of Brest, western Brittany, France). The attention was dedicated to regression models such as Multiple Linear Regression (MLR) and Multiple Polynomial Regression (MPR), and popular ML algorithms including MultiLayer Perceptron (MLP), Support Vector Regression (SVR) and Random Forest (RF). In order to assess the practical implementation of these algorithms in comparison to the more complex process-based physical model, we considered simple input data, easily accessible, in relation to their potential influence on sea surface salinity at the mouth of the Elorn estuary. This includes the observed free-surface elevation at the nearest harbour, the upstream river discharges and the wind velocity. A sensitivity study to input data considered additional parameters such as the tidal current, the precipitation rate and the air temperature. Performances of ML algorithms were evaluated with respect to observations gathered during a six-year period at the mouth of the estuary. Specific calibration studies were furthermore conducted for the different ML algorithms to establish optimised values of associated hyperparameters.

The present investigation exhibited (i) methodological conclusions associated with the implementation and intercomparison of ML algorithms and (ii) results associated with

the evolution of sea surface salinity at the mouth of the estuary.

From a methodological point of view, ML algorithms were found to provide estimations of observed sea surface salinity comparable to predictions from a process-based physical model, thus capturing the temporal variations from diurnal to seasonal time scales. However, whereas the process-based physical model reproduced the semi-diurnal variations of sea surface salinity, more important differences were obtained during low-salinity events with predicted minimum values over 22 ppt against observed minimum values of 5 ppt. This overestimation may be associated with the difficulty of the model to represent salinity transport from upstream river boundaries to the mouth of the estuary in relation to a coarse computational-grid resolution and definition of water-depths variations in this shallow-water environment. Instead, with a limited number of input data and reduced computational time compared to the 3D model, machine learning algorithms reproduced these low-salinity events. Both MLP and MPR were able to capture the non-linear nature of salinity variations to external forcings. Results obtained from MPR showed that a great part of these non-linearities was captured by a polynomial regression function of degree two. However, the best estimations were obtained for the Support Vector Regression. Whereas this algorithm required a tuning procedure of hyperparameters with additional computational time, this remained negligible in comparison to 3D numerical simulations of salinity transport in the bay of Brest. It is therefore suggested to rely on optimised SVR for approaching the evolution of salinity at the mouth of the Elorn estuary. Regarding input parameters, the inclusion of tidal currents may appear more relevant than free-surface elevation to account for salinity transport in the estuary. However, tidal currents were also highly variable at the scale of the bay and the selection of a rectilinear alternative component resulted in frank variations of salinity neglecting potential remote influences. Tide-induced free-surface elevation offered, in comparison, greater freedom to adapt the ML algorithms.

In an in-depth analysis of salinity variation, particular attention may be dedicated to the temporal variations and relation to input data. Thus, whereas semi-diurnal variations of sea surface salinity resulted from tide-induced advection and diffusion, low-salinity events at the mouth of the Elorn estuary appeared to be influenced by the intrusion of fresh waters from riverine inputs and an increased impact may be expected after a peak in precipitation. Thus, trained ML algorithms, treating river discharges as input variables, were able to capture these low-salinity events. The sensitivity study to input data confirmed furthermore the key role played by tide and river discharges on salinity variations in the estuary. And these major influences largely outweighed, in ML algorithms considered, the influence of other forcings such as surface wind shear stress, precipitation rate or air temperature whose impact may be mainly expected during isolated events of short duration. A refined definition of these input data over an extensive targeted dataset may help to remove these uncertainties, and refine the approach to salinity variations.

Trained ML algorithms may therefore be exploited to provide, with reduced computational time, a global evaluation of the temporal variation of a hydrological parameter such as sea surface salinity in an estuary under the combined complex influences of tide-induced transport and fresh river discharges. Results were, however, obtained by exploiting a six-year period of observations with a limited number of data in relation to blank periods due to maintenance operations and system malfunction. Thus, we may improve predictions with an extensive amount of data continuously acquired at the mouth of the Elorn estuary. Whereas the process-based physical model exhibited increased differences for approaching low-salinity events, it remains fundamental to encompass the physical mechanisms involved in the evolution of sea salinity. Thus, numerical modelling may be exploited as a complementary tool to ML algorithms (i) to provide further insights about parameters controlling the evolution of sea surface salinity and/or (ii) to produce new input data to train algorithms (hybrid approach). As the investigation was conducted in a single location in the bay of Brest, the potential of the ML algorithms may finally be evaluated by including broader observations, such as measurement points (i) at the entrance of the bay characterised by increasing mixing between fresh and marine waters and (ii) in the south-eastern part of the bay mainly influenced by freshwater inputs from the Aulne and Mignonne rivers. These extended observations may help to investigate the potential of ML algorithms to model salinity in locations impacted by contrasting remote effects of rivers discharges. Moreover, such an approach may serve broader applications in estuaries impacted by strong salinity variations to encompass, on an extended time scale, the potential effect of extreme weather events, especially storm surges and floods.

Acknowledgements

The present paper is a contribution to the research program INTERIMER (“INTEractions entre Rivière(s) et MER”) of the Laboratory of Coastal Engineering and Environment (Cerema, http://www.cerema.fr). The instrumentation system BOCA (“Bouée d’Observation Côtière Automatique multiparamètres”) at the mouth of the Elorn estuary was implemented as part of a collaboration between Cerema (LGCE, “Laboratoire de Génie Côtier et Environnement”), LEMAR (“Laboratoire des sciences de l’Environnement MARin”, “Institut Universitaire Européen de la Mer”) and the Urban Ecology Department of Brest Métropole. The authors thank Eric Duvieilbourg (LEMAR) and Antoine Douchin (Cerema) for their technical support in setting up the instrumentation system. The authors acknowledge also the support of Béatrice Quéau from Brest Métropole for the maintenance, with Antoine Douchin, of the BOCA instrumentation system and the exchange about observed data acquired. The numerical model exploited here was implemented on computer facilities DATARMOR of “Pôle de Calcul et de Données pour la Mer” (PCDM) (http://www.ifremer.fr/pcdm).

CRediT authorship contribution statement

Nicolas Guillou: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing — original draft, Writing — review & editing, Visualization, Supervision, Project administration. **Georges Chapalain:** Writing

— review & editing, Data Curation, Project administration. **Sébastien Petton:** Writing — review & editing, Software.

References

Agence de l'eau Loire Bretagne, 1997. Contrat de baie — La Rade de Brest et son bassin versant : Etat des lieā. URL: https://www.documentation.eauetbiodiversite.fr/notice/00000000015df039937aacbc62f2d250 (accessed 11.5.21).

Alizadeh, M.J., Kavianpour, M.R., Danesh, M., Adolf, J., Shamshirband, S., Chau, K.-W., 2018. Effect of river flow on the quality of estuarine and coastal waters using machine learning models. Engineering Appl. Comput. Fluid Mech. 12, 810—823. https://doi.org/10.1080/19942060.2018.1528480

Auffret, G., 1983. Dynamique sédimentaire de la marge continentale celtique - Evolution Cénozoïque - Spécificité du Pleistocène supérieur et de l'Holocène. Semantic Scholar [WWW Document], URL https://www.semanticscholar.org/paper/Dynamique-s%C3%A9dimentaire-la-marge-continentale-du-Auffret/1b555139d5867c69dbedecd6f4455e1d7a7e2094 (accessed 11.3.21).

Azencott, C.A., 2019. Introduction au Machine Learning, Dunod, Cambridge, UK., 227 pp.

Banque, Hydro, 2021. Banque Hydro. http://hydro.eaufrance.fr/indexs.php (accessed on 05/2021).

Beudin, A., Chapalain, G., Guillou, N., 2014. Modelling dynamics and exchanges of fine sediments in the bay of Brest. La Houille Blanche 47—53. https://doi.org/10.1051/lhb/2014062

Bowden, G.J., Maier, H.R., Dandy, G.C., 2005. Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. J. Hydrol. 301, 93—107. https://doi.org/10.1016/j.jhydrol.2004.06.020

Chauvaud, L., Jean, F., Ragueneau, O., Thouzeau, G., 2000. Long-term variation of the Bay of Brest ecosystem: benthic-pelagic coupling revisited. Mar. Ecol. Prog. Ser. 200, 35—48. https://doi.org/10.3354/MEPS200035

Chen, L., Roy, S.B., Hutton, P.H., 2018. Emulation of a process-based estuarine hydrodynamic model. Hydrol. Sci. J. 63, 783—802. https://doi.org/10.1080/02626667.2018.1447112

Chen, W., Liu, W., Huang, W., Liu, H., 2017. Prediction of Salinity Variations in a Tidal Estuary Using Artificial Neural Network and Three-Dimensional Hydrodynamic Models Comp. Water Energy Environ. Eng. 6 (1), 107—108.

Choi, K.W., Lee, J.H.W., 2004. Numerical determination of flushing time for stratified water bodies. J. Marine Syst. 50, 3—4. https://doi.org/10.1016/J.JMARSYS.2004.04.005

Chung, F.I., Seneviratne, S.A., 2009. Developing Artificial Neural Networks to Represent Salinity Intrusions in the Delta, in: World Environmental and Water Resources Congress 2009. In: Presented at the World Environmental and Water Resources Congress 2009. American Society of Civil Engineers, Kansas City, Missouri, United States, 1—10. https://doi.org/10.1061/41036(342)483

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273—297. https://doi.org/10.1007/BF00994018

Cruz, E.R., Nolasco, R., Padin, X.A., Gilcoto, M., Babarro, J.M.F., Dubert, J., Pérez, F.F., 2021. A High-Resolution Modeling Study of the Circulation Patterns at a Coastal Embayment: Ría de Pontevedra (NW Spain) Under Upwelling and Downwelling Conditions. Front. Mar. Sci. 8, 661250. https://doi.org/10.3389/fmars.2021.661250

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V., 1997. Support Vector Regression Machines. Neural Information Processing Systems 9, MIT Press, 155—161.

Ducrocq, V., Bouttier, F., Malardel, S., Montmerle, T., Seity, Y., 2005. Le projet AROME. La Houille Blanche 91 (2), 39—43. https://doi.org/10.1051/lhb:200502004

Dyer, K.R., 1973. Estuaries: A physical introduction DYER, K. R. 1973. Wiley-Interscience, New York, London, xv + 140 pp.

Frère, L., Paul-Pont, I., Rinnert, E., Petton, S., Jaffré, J., Bihannic, I., Soudant, P., Lambert, C., Huvet, A., 2017. Influence of environmental and anthropogenic factors on the composition, concentration and spatial distribution of microplastics: A case study of the Bay of Brest (Brittany, France). Environ. Pollut. 225, 211—222. https://doi.org/10.1016/j.envpol.2017.03.023

Guillou, N., 2007. Rôles de l'hétérogénéité des sédiments de fond et des interactions houle-courant sur l'hydrodynamique et la dynamique sédimentaire en zone subtidale — applications en Manche orientale et à la pointe de la Bretagne [WWW Document]. URL: https://www.calameo.com/books/001058329dee68c4a2d96 (accessed 11.3.21).

Guo, Q., Lordi, G.P., 2000. Method for quantifying freshwater input and flushing time in estuaries. J. Environ. Eng. 126, 675—683.

He, M., Zhong, L., Sandhu, P., Zhou, Y., 2020. Emulation of a Process-Based Salinity Generator for the Sacramento—San Joaquin Delta of California via Deep Learning. Water 12, 2088. https://doi.org/10.3390/w12082088

Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2010. A Practical Guide to Support Vector Classification, National Taiwan University Papers, Taipei, 16 pp.

Huang, W., Foo, S., 2002. Neural network modeling of salinity variation in Apalachicola River. Water Res. 36, 356—362. https://doi.org/10.1016/S0043-1354(01)00195-6

Keerthi, S.S., Lin, C.-J., 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. Neural Comput 15, 1667—1689. https://doi.org/10.1162/089976603321891855

Keras, 2021. Keras, Simple. Flexible. Powerful. https://keras.io (accessed on 09/2021).

Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs].

Lazure, P., Dumas, F., 2008. An external—internal mode coupling for a 3D hydrodynamical model for applications at regional scale (MARS). Adv. Water Resour. 31, 233—250. https://doi.org/10.1016/j.advwatres.2007.06.010

Le Pape, O., Del Amo, Y., Menesguen, A., Aminot, A., Quequiner, B., Tréguer, P., 1996. Resistance of a coastal ecosystem to increasing eutrophic conditions: the Bay of Brest (France), a semi-enclosed zone of Western Europe. Cont. Shelf Res. 16, 1885—1907.

Le Roy, R., Simon, B., 2003. Réalisation et validation d'un modèle de marée en Manche et dans le Golfe de Gascogne. Application à la réalisation d'un nouveau programme de réduction des sondages bathymétriques. (No. Rapport n°002/03). SHOM.

Maier, H.R., Dandy, G.C., 1996. The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters. Water Resour. Res. 32, 1013—1022. https://doi.org/10.1029/96WR03529

Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environ. Modell. Softw. 25, 891—909. https://doi.org/10.1016/j.envsoft.2010.02.003

Nair, V., Hinton, G.E., 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10).

Nguyen, T.G., Tran, N.A., Vu, P.L., Nguyen, Q.-H., Nguyen, H.D., Bui, Q.-T., 2021. Salinity intrusion prediction using remote sensing and machine learning in data-limited regions: A case study in Vietnam's Mekong Delta. Geoderma Regional 27, e00424. https://doi.org/10.1016/j.geodrs.2021.e00424

Petton, S., 2010. Etude des processus hydrodynamiques et hydro-sédimentaires affectant un estran de type marais salé de la rade de Brest (anse de Penfoul) colonisé par l'espèce invasive spartine (Spartina Alterniflora Loisel). In: Centre d'Etudes Techniques Maritimes et Fluviales, 37 pp.

Petton, S., Le Berre, D., Haurie, A., Pouvreau, S., 2016. HOMER Campaign : Mooring time series. https://doi.org/10.17882/43082

Petton, S., Le Roy, V., Bellec, G., Queau, I., Le Souchu, P., Pouvreau, S., 2018. Marine environmental station database of Daoulas bay. https://doi.org/10.17882/42493

Petton, S., Pouvreau, S., Dumas, F., 2020. Intensive use of Lagrangian trajectories to quantify coastal area dispersion. Ocean Dynam. 70, 541—559. https://doi.org/10.1007/s10236-019-01343-6

Poppeschi, C., Charria, G., Goberville, E., Rimmelin-Maury, P., Barrier, N., Petton, S., Unterberger, M., Grossteffan, E., Repecaud, M., Quéméner, L., Theetten, S., Le Roux, J.-F., Tréguer, P., 2021. Unraveling Salinity Extreme Events in Coastal Environments: A Winter Focus on the Bay of Brest. Front. Mar. Sci. 8, 966. https://doi.org/10.3389/fmars.2021.705403

Quéméneur, M., Kerouel, R., Aminot, A., 1984. Cycle de la matière organique dans l'estuaire de l'Elorn et relations avec les bactéries. Ifremer.

Rath, J.S., Hutton, P.H., Chen, L., Roy, S.B., 2017. A hybrid empirical-Bayesian artificial neural network model of salinity in the San Francisco Bay-Delta estuary. Environ. Modell. Softw. 93, 193—208. https://doi.org/10.1016/j.envsoft.2017.03.022

Robins, P.E., Lewis, M.J., Simpson, J.H., Howlett, E.R., Malham, S.K., 2014. Future variability of solute transport in a macrotidal estuary. Estuar. Coast. Shelf Sci. 151, 88—99. https://doi.org/10.1016/j.ecss.2014.09.019

Salomon, J.C., Breton, M., 1991. Numerical study of the dispersive capacity of the Bay of Brest, France, towards dissolved substances. Environ. Hydraul. 459—464.

SHOM, 2021. https://www.data.shom.fr (accessed on 05/2021).

Su, H., Wu, X., Yan, X.-H., Kidwell, A., 2015. Estimation of subsurface temperature anomaly in the Indian Ocean during recent global surface warming hiatus from satellite measurements: A support vector machine approach. Remote Sens. Environ. 160, 63—71. https://doi.org/10.1016/j.rse.2015.01.001

Tréguer, P., Goberville, E., Barrier, N., L'Helguen, S., Morin, P., Bozec, Y., Rimmelin-Maury, P., Czamanski, M., Grossteffan, E., Cariou, T., Répécaud, M., Quéméner, L., 2014. Large and local-scale influences on physical and chemical characteristics of coastal waters of Western Europe during winter. J. Marine Syst. 139, 79—90.

Vapnik, V., Golowich, S.E., Smola, A., 1996. Support vector method for function approximation, regression estimation and signal processing. In: Proceedings of the 9th International Conference on Neural Information Processing Systems, Guide Proceedings [WWW Document]. URL: https://dl.acm.org/doi/abs/10.5555/2998981.2999021 (accessed 3.7.22).

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York, USA, 189 pp.

WMO, 2021. World Meteorological Organization. Données d'observation des principales stations météorologiques URL: https://www.data.gouv.fr/fr/datasets/donnees-d-observation-des-principales-stations-meteorologiques (accessed on 05/2021).

Zhang, H., Shen, Y., Tang, J., 2021. Hydrodynamics and water renewal in the Pearl River Estuary, China: A numerical study from the perspective of water age. Ocean Eng. 237, 109639. https://doi.org/10.1016/j.oceaneng.2021.109639