Defining food sampling strategy for chemical risk assessment

Nathalie Wesolek, Alain-Claude Roudot

Defining food sampling strategies for chemical risk assessment

Nathalie Wesolek, Alain-Claude Roudot

ABSTRACT

Collection of accurate and reliable data is a prerequisite for informed risk assessment and risk management. For chemical contaminants in food, contamination assessments enable consumer protection and exposure assessments. And yet, the accuracy of a contamination assessment depends on both chemical analysis and sampling plan performance. A sampling plan is always used when the contamination level of a food lot is evaluated, due to the fact that the whole lot can not be analysed, but only samples, which are drawn from the lot. An efficient sampling plan enables to take samples from a food lot, following a given protocol, with a relatively low risk of misestimating the true mean concentration of the food lot after analysis of the food samples. Sampling plan performance testing is achieved thanks to mathematical validation methods. The best fit sampling plan is the one that gives the best compromise between the lowering of the risk of misestimating the true lot concentration and the practical feasibility (not too much time consuming nor money consuming). This chapter presents two sampling plan validation strategies: a parametric method developed by Whitaker and co-workers from 1972 and a non parametric method set up by Schatzki *et al.* (Schatzki, 1995; Campbell *et al.* 2003). To our knowledge, these are the only two methods sufficiently evolved for having been applied to real situation cases for food sampling validation. These statistical methods are first explained from a theoretical point of view. Then, each one is illustrated by a practical application to a sampling plan validation for a specific chemical risk in a food commodity, thanks to workable contamination data gathered in the literature. According to us, in its general mathematical principle, the non parametric method is more appropriate to cases with contaminants distributed heterogeneously in a food lot. However, due to its ease of use, the parametric method applies best to cases where the distribution of the contaminant is homogeneous. A food contaminant is homogeneously distributed in a contaminated food lot when the contamination incidence rate for individual food items taken from the contaminated lot is high, and when the concentration levels in each food item are rather alike. Otherwise, when the contamination incidence rate is low, and when the concentrations differ greatly in each food item, this means that the contaminant is heterogeneously distributed within the food lot. For these reasons, the first sampling plan validation technique (parametric method) is applied, in this chapter, to phycotoxin contamination in shellfish lots at the cultivation zone, as it is considered as being a homogeneously distributed contaminant case. For the heterogeneously distributed contaminant case, mycotoxin contamination data for pistachios at retail stage are exploited in order to put into practice the non-parametric sampling plan validation method. Both phycotoxins and mycotoxins are natural toxins that are unsafe for human. Limits of contaminations are set by national and international safety agencies, but sampling strategies have a great influence on the detected results in food lots. The chapter will show that an optimal sampling strategy can be obtained in each of the two cases, but that they require different mathematical approaches in order to obtain reliable Operating Characteristics (OC) curves showing:

- the consumer risk (risk of accepting lots at a true concentration above the contaminant's concentration threshold);
- the producer risk (risk of rejecting lots at a true concentration under the contaminant's concentration threshold).

INTRODUCTION

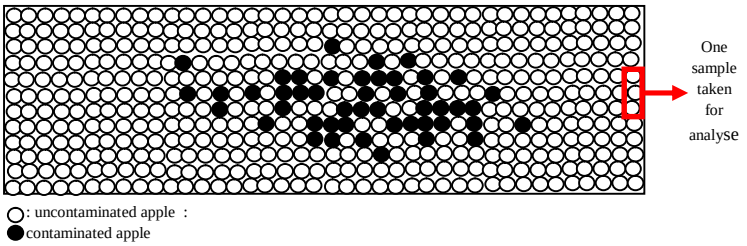
Chemical contaminant analyses in food are of utmost importance, as these contaminants can trigger dangerous health effects on food consumers, especially for people at risk like pregnant women or young children. Some contaminants are metabolised in the body and then eliminated, but may still be dangerous above a threshold concentration due to acute exposure. Other chemical contaminants aren't eliminated and can accumulate in the body at a level increasing along with time and consumption frequency of contaminated food. By this way, besides acute exposure, they generate chronic exposure due to very high body burdens.

Knowing consumption levels and dangerous doses levels in the body thanks to toxicological studies, legislators set maximum contamination levels in food for human consumption. Food having levels above these thresholds are not authorized for sale. Furthermore, legislators set sampling plans, which means they define the way samples must be taken from a food lot, as function of the food lot size. A food lot of a given food type can be a food consignment, an industrial batch, food having the same origin, etc., and is supposed to have a contamination level as homogeneous as possible.

In case European legislators don't define a sampling plan in detail for a food type and contaminant type, they state at least that a sampling plan, enabling the detection of any overshooting of the levels, must be set. A sampling plan's objective is to obtain samples having a contamination level representative of the mean contamination level of the food lot. A very simple imaginary example of an inappropriate sampling plan can be given with an apple truck consignment. This inappropriate sampling plan (Fig. 1) involves taking

one sample of three apples from the rear of the truck whereas the contaminated apples, represented in black, are located in the centre. In this example, even in case of rigorous laboratory analysis in an accredited laboratory, the level of contamination obtained for the sample is not reflecting the mean contamination level of the whole consignment.

Fig.1. An inappropriate sampling plan example with one sample taken in an apple truck consignment



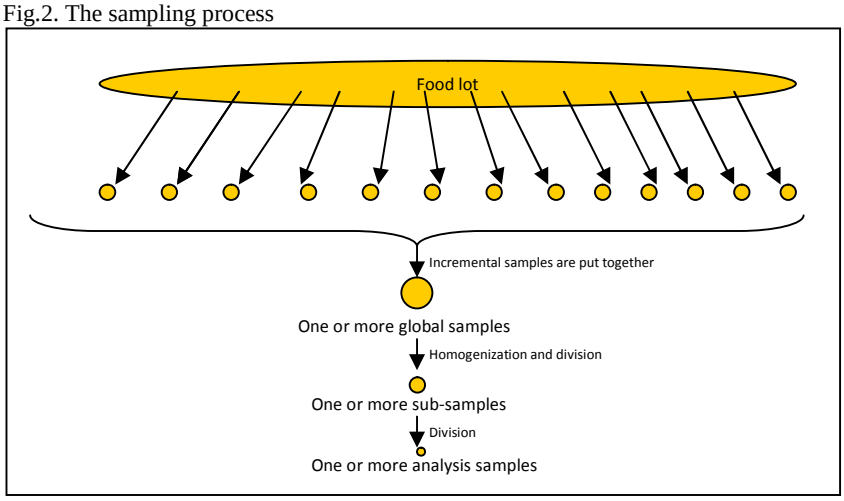
The sampling plan defines many criteria:

- for the lot (type and size) and
- for the sampling process.

The Sampling Process

A sampling process comprises many steps involving various kinds of samples as described in fig. 2.

Fig.2. The sampling process



Incremental samples are taken from the food lot and are aggregated to form global samples. These global samples are usually simply called "samples". They are homogenized and may be divided in sub-samples, before being further divided in analysis samples. The analysis samples are the ones submitted to chemical analyse.

A sampling process is specific to a food type and a contaminant type, due to a sampling complexity increasing along with:

- the heterogeneity of the product (solid versus liquid products);
- the heterogeneity of contamination (contaminant spread unevenly in the food commodity, see the abstract part for more details);
- the low maximum concentration level of the contaminant (i.e. 100 µg/kg is harder to detect than 100 mg/kg) (for more details see Blanc, 2006).

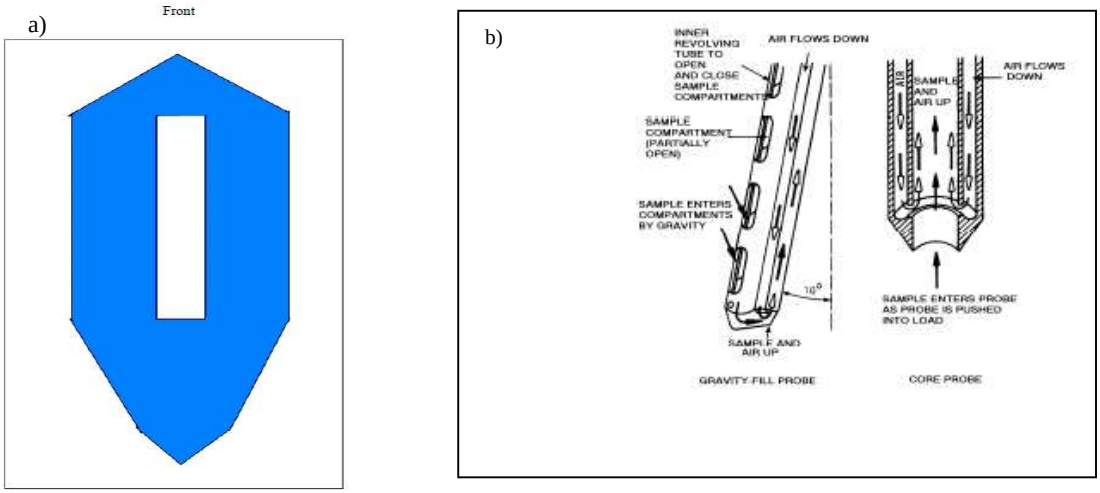
A complex sampling process for a contaminant spread heterogeneously and at a low concentration level in a granular food product, for a reliable concentration result, consists of:

- taking a sufficient number of incremental samples of an appropriate size from the lot with an adapted device;
- selecting proper locations to take incremental samples that are representative of the lot mean concentration level;
- mixing the incremental samples in a sufficient number of global samples;
- using a proper homogenisation method.

The complexity of the sampling process vouches for an appropriate accuracy of the inspection results.

An example of a complex sampling process is given in the practical procedure for grain handlers (USDA, 2009) for grain sampling in order to determine the aflatoxin level. In this procedure, samples must be representative and sufficiently large (recommended size 4.5 kg) to compensate for the uneven distribution of the contaminant. Each sample is taken with a probe and by this way, consists of many incremental samples (Fig. 3b). For stationary lots of grain in trucks and in other open-top carriers, at least two probe samples from any point in the shaded area (Fig. 3a), for lots that are 21,120 L or less, are taken. For larger lots, at least three probe samples are drawn.

Fig.3. a) Sampling Pattern for Sampling Trucks, Trailers, and Wagons (Courtesy: Charles R. Hurburgh, Jr., Iowa State University); b) Mechanical probes used to sample stationary lots of grain in trucks and in other open-top carriers (USDA, 2009)



Lot Type and Size

The selection of the food lot type and size is part of the sampling plan too. It is a complex selection that involves a choice as function of:

- the final objective that might be identifying bad lots in order to track the contamination back to the source or assessing the population s exposure;
- the contamination pattern.

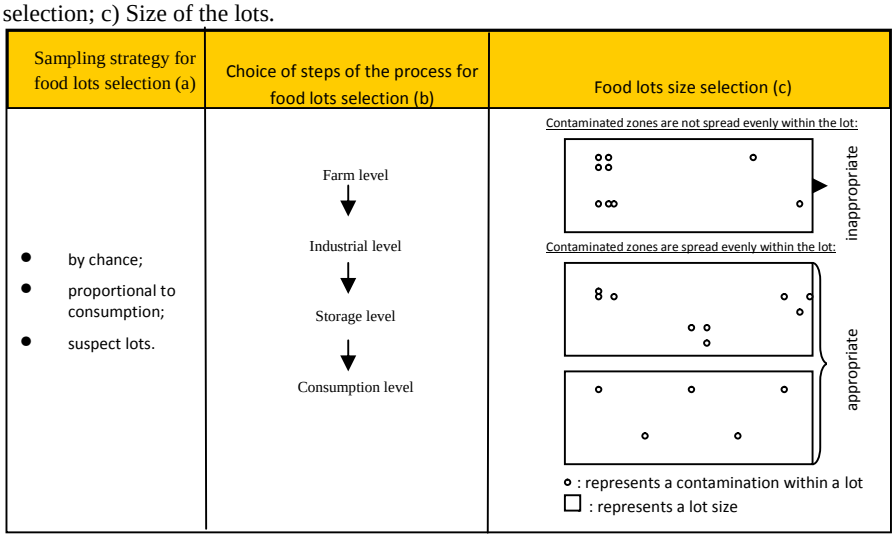
These two criteria of choice are going to be further explained hereunder and in fig. 4.

As chemical analyses are very expensive, systematic analyses are very seldom achieved on each food lot. Instead, for a given food (Fig. 4a):

- at a given sampling frequency, some food lots are selected by chance or food brands are selected in order to be proportional to the consumption levels of each brand, or
- food lots that are suspected of being contaminated are selected.

The most relevant steps of the process from farm to consumption (flow diagram in Fig. 4b) must be chosen for sampling a lot: farm level, industrial level, transport level, storage level, consumption level Indeed, contamination might occur or develop at a step and not at another. This is due to specific contamination patterns for each contaminant. The objective is to eliminate bad quality lots as soon as possible in the process to ensure consumer s safety and to lower the global mean contamination level. Moreover, still due to these patterns, chemical contaminants can be widely distributed or locally distributed at each process step and then generate plant and animal product widespread or local contaminations. More precisely, the contaminated zone within a farming area, fishing zone, storage area etc. can be either vast or very small. For these reasons, the choice of the lot size to be sampled must be judicious. A proper food lot size choice must be guided by the fact that a food lot must have a globally constant contamination level within the lot (Fig. 4c) even in the cases when contamination patchiness pattern drives to located contaminated zones within the lot. It says that these contaminated zones must be quite evenly located within a lot. This enables to avoid mixed up contamination levels within the lot as much as possible.

Fig.4. Lot selection for sampling: a) Sampling strategy for food lots selection; b) Choice of the steps of the process for food lots selection; c) Size of the lots.



Sampling Plan Validation

Finally, an accurate sampling plan ensures that lots chosen for sampling are properly defined and selected and that samples in the whole sampling process are properly obtained. The final goal is to state that the contamination level of the analysis samples is strictly the same as the contamination mean level of the whole lot tested with the lowest error as possible.
In order to enable the scientific validation (mathematical validation) of a given sampling plan, both the sampling process and the lot selection process must be studied. Two sampling plan validation methods are going to be discussed: a parametric method developed by Whitaker and co-workers from 1972 and a non parametric method set up by Schatzki *et al.* (Schatzki, 1995; Campbell *et al.* 2003). To our knowledge, these are the only two methods sufficiently evolved for having been applied to real situation cases.

ACCEPT PROBABILITIES OBTENTION FOR SAMPLING PLAN VALIDATION

Parametric Method from Whitaker: Principle and Application to Phycotoxins

The sampling plan validation method developed by Whitaker consists of calculation steps achieved from data of contaminant concentration of samples taken from a lot. The samples concentrations from a lot are adjusted to a theoretical distribution thanks to a goodness of fit test. This operation is made for a few lots. Furthermore, the variability between samples concentrations from a lot is studied in order to predict this variability for any lot mean concentration within a given range of concentrations. Both the theoretical distribution and the prediction of concentrations variability between samples of the same lot are used to calculate the accept probabilities of lots for the sampling plan tested. All these steps are further explained in the following sections and applied to phycotoxins with the evaluation of sampling plans to detect the contaminant okadaic acid in mussels. Okadaic acid is a phycotoxin and part of the diarrhetic shellfish poison family. Bivalve molluscan shellfish feed on phytoplankton (micro-algae) which might potentially contain biotoxins like okadaic acid. In order to ensure food safety in Europe, the level of okadaic acid equivalents in live bivalve molluscs must not exceed 160 µg per kg (Regulation 853/2004/EC). Therefore, sampling plans must be set by official authorities, in each European country, in order to monitor production areas to check the presence of okadaic acid in shellfish, knowing that mussel may be used as an indicator species (Regulation 854/2004/EC). It is a good point to take samples at the production areas, due to the fact that the phycotoxin concentration in shellfish, is not going to vary in the next steps of the process (transport, shelf life) as many shellfish are sold raw. The only process step influencing the okadaic acid concentration level is the cooking step (McCarron *et al.*, 2008), as concentration increases in steamed mussel meat in comparison to raw mussel meat. During a contamination event, although the contaminant is spread quite homogeneously in all the mussels, there is a high inter-individual variability as regards the contamination levels between mussels, even for mussels taken at the same sampling point and at the same time (Edebo *et al.,* 1988; Duinker *et al.,* 2007). For this reason, setting an appropriate sampling plan is difficult and must be made on a scientific basis better than suggested by an approximate rule of thumb based on suppositions. However there has never been a probabilistic evaluation of the impact that the sample size and the number of samples taken, have on the analysis results and the corresponding decision for the shellfish lot.

Theoretical Distribution
In order to estimate the probabilities associated with sampling lots of a food product for a contaminant analysis, the suitability of a theoretical distribution is studied. This means that the observed distribution of the contaminant's concentration in samples (of the

same size) drawn from one lot is compared to a theoretical distribution. This operation is repeated for several lots with the use, at each time, of a statistical test to validate the goodness of fit.
First of all, for the selection of a theoretical distribution to test, the observed probability density functions are drawn (one per lot). In order to do this, the data are spread into classes after determination of the number of classes as shown hereafter. The Sturges method can be used to determine the number of classes:

$$\text{Number of classes} = \sup(\log_2(n_l) + 1)$$

With n_l the number of samples within the lot and sup(): round the number to the superior integer. In order to set the class interval, the following formula can be used:

$$\text{Class interval} = \frac{\text{Maximum value} - \text{minimum value}}{\text{Number of classes}}$$

Then, the observed probability density functions are drawn as a histogram of the proportion of samples as function of the contamination class intervals. The observed probability density functions shape characterises a distribution type. It might be symmetric (normal), when the 2 "halves" of the histogram appear as mirror-images of one another. If it is not the case, then the shape is called skewed (non-symmetric). For skewed distributions, it is quite common to have one tail of the distribution considerably longer or drawn out relative to the other tail. A "skewed right" distribution is one in which the tail is on the right side. A "skewed left" distribution is one in which the tail is one the left side.

Results of practical application to okadaic acid:

Data used:
For the application to okadaic acid in mussels, the data used are from Pr. Arne Duinker who has supplied us (personal communication) with raw data on individual mussels contaminated with okadaic acid equivalents obtained during experiments that lead to a publication in 2007 (Duinker *et al.*). These data consist of contamination levels in mussels contaminated on collectors of rather high density in a stratified fjord. Four different lots were sampled, and all the samples from a lot were taken at the same sampling point, at the same time. For each lot: 29 or 30 samples were taken, each sample consisting of one mussel. Then each individual sample was submitted to chemical analysis. Given the Regulation 853/2004/EC, the data, expressed in concentrations in steamed mussels, must be converted to concentrations in raw mussels. This conversion is done thanks to the publication of McCarron *et al.* (2008). Indeed, they published a theoretical conversion value: $\text{Concentration}_{\text{in raw meat}} = \text{Concentration}_{\text{in steamed meat}} / 1.2667$.

Results:
The probability density functions obtained for each of the four lots show distributions highly skewed to the right, which says that the right tail is longer, the mass of the distribution is concentrated on the left of the figure. An example of a right-skewed distribution obtained is given in Fig. 5.

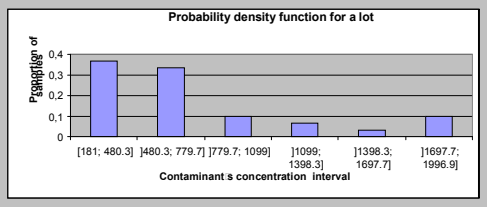


Fig.5. Histogram of observed probability density function for a lot with concentrations in µg/kg

The choice of a theoretical distribution type must be made between continuous data distributions: Normal distribution; Gamma family distributions with Exponential, Beta, Pareto, Weibull and Gumbel distributions. For right-skewed distributions, Exponential, Beta, Pareto, Weibull, or Gumbel might fit; whether for left-skewed distributions, Beta or Gauchy distributions could be appropriate. Furthermore, the slope of the tail, as well as a common practice in the field of application may guide a theoretical distribution choice.
Secondly, the cumulative distribution functions of the sensed theoretical distributions and of the observed distribution are compared to each other. In order to draw the observed and theoretical cumulative frequency distributions, the concentration values of the samples, named x, are ranked within each lot into ascending order with x_r the value of the concentration in the rank. For the calculation of the parameters of each theoretical distribution, the mean and variance are calculated for each of the lots on the sample results (observed data). Then, the method of moments is used to calculate each theoretical distribution's parameters. Afterwards, the observed and theoretical frequency distributions must be plotted on graphs in order to enable a visual comparison. In order to achieve this goal, the ordinates of the cumulative frequency distributions are calculated as follows for the observed and theoretical distributions:

- For the observed distribution, the proportion of samples corresponding to each x_r value must be recorded within each lot. Furthermore, there are two graphs per lot that are plotted using the following ordinates:

$F_{1\ observed}(x_r)$: $P(x < x_r)$: sum of the proportion of individuals until the rank r-1.
$F_{2\ observed}(x_r)$: $P(x \leq x_r)$: sum of the proportion of individuals until the rank r.

-For each theoretical distribution, there is only one graph per lot that is plotted using the following ordinate:

$F_{theoritical}(x_r) = P(x \leq x_r)$

Following of the practical application to okadaic acid:

As it is used for continuous data and can simulate highly positively skewed probability density functions, the theoretical distribution tested is the lognormal distribution. Furthermore, the lognormal distribution parameters can be easily calculated thanks to the method of moments (Calculations in Annex 1). The comparison of the observed and lognormal cumulative frequency distributions for each lot shows a good visual fit (theoretical ordinates can be computed as shown in Annex 2).
An illustration of cumulative distribution functions comparisons is given in Fig. 6.

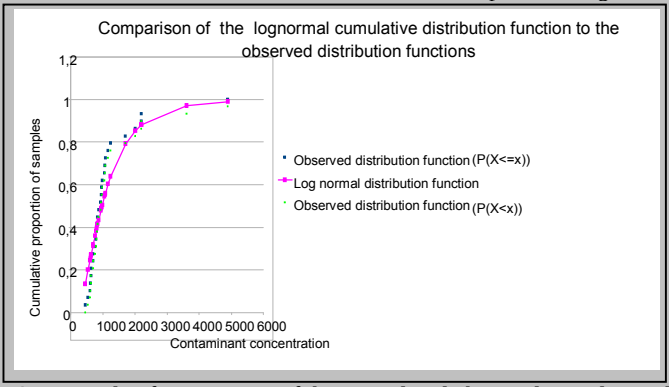


Fig.6. An example of comparison of theoretical and observed cumulative distribution functions with concentration in µg/kg

Thirdly, the statistical goodness of fit test must be chosen. The goodness of fit of the observed data to the theoretical distributions can be tested thanks to the Kolmogorov-Smirnov and the Anderson-Darling statistical tests. Both tests are goodness of fit tests of an empirical distribution of a sample of a random variable to a theoretical distribution of this random variable. These tests must be performed on cumulative frequency distributions in order to compare the observed distribution to a theoretical distribution. Both tests measure the differences between the theoretical and observed probabilities for each contaminant concentration within one lot. However, the Kolmogorov-Smirnov test involves finding the maximum vertical distance between the cumulative frequency distributions, whereas the Anderson-Darling test is based upon a weighted square of the vertical distance between these distributions. The main difference between these tests is that the Anderson-Darling test is more sensitive to deviations in the tails of the distribution than is the older Kolmogorov-Smirnov test. For the Kolmogorov-Smirnov test, if the parameters of the theoretical cumulative frequency distribution are calculated from the lot that is tested, which is the case here, the results must be taken with caution, because the test has got a tendency towards validating the distribution tested in all cases.
For each Kolmogorov-Smirnov goodness of fit test:

Test hypothesis:
H_0: The observed distribution conforms to the theoretical distribution.
H_1: The observed distribution doesn't conform to the theoretical distribution.

At the desired risk level, H_0 can not be rejected if the variable tested (D_{calc}) is inferior to a critical value found in a table for the corresponding number of samples in the lot This means that the adjustment of the observed data to the theoretical distribution test can not be rejected at the risk level chosen. Furthermore, the p-value is the probability of observing the D_{calc} value under H_0.
For each Anderson-Darling goodness of fit test:

This test is similar to the Kolmogorov-Smirnov test. The test hypothesis is the same, but it uses a different test statistic: A^2_n. The critical values and the test statistic depend on the specific distribution that is being tested.
Anderson-Darling can be applied to any distribution, but finding tables of critical values isn't so easy. If the estimated statistic exceeds the critical value at a particular significance level the null hypothesis can be rejected.

Following of the practical application to okadaic acid:

The conformity of the observed distribution to the lognormal distribution is further tested thanks to the two goodness of fit statistical tests. The theoretical calculations, as well as the way they can be put into practice are given in Annex 3 for the Kolmogorov-Smirnov test and in Annex 4 for the Anderson-Darling test. The results of these tests are given in Table 1.

Table 1. Goodness of fit test results

	Kolmogorov-Smirnov test			Anderson-Darling test		
	D_{calc}	Critical value	p-value	A^2_n	Critical value	p-value
Lot n°1	0.1393	0.2457	57.91%	1.0639	2.49	32.48%
Lot n°2	0.1829	0.2457	25.39%	1.01	2.49	35.14%
Lot n°3	0.2503	0.2417	3.84%	6.704	2.49	0.05%
Lot n°4	0.1547	0.2417	42.66%	1.2052	2.49	26.52%

Critical values and p-values are obtained for the risk level of 5%.

For lots n°1, 2 and 4, for both tests, the test statistics (D_{calc} and A^2_n) are inferior to the critical values, which means that the hypothesis H_0 can not be rejected at a 5% risk level. Moreover, the p-values are above 5%, which means too that H_0 is not rejected. So, at a 5% risk level, the lognormal distribution of the population can not be rejected. There is a discrepancy for lot n°3, as it shows opposite results. However, we consider that three lots validated out of four by the two goodness of fit tests is enough to consider that when samples from any lot are drawn, their contamination levels fit the lognormal distribution. Moreover, the samples distribution is still considered to be lognormal, even if the samples taken are of a bigger size than one mussel per sample.

Variability

Associated to the theoretical distribution that fits the distribution of samples from a lot, the calculation of the mean and the variance from samples of a lot enable to compute the probability that one sample taken by chance from this lot has a given concentration. This is why mean and variance data between samples must be calculated, knowing that the variance is a measure of how far the concentrations of samples are spread out from each other for one lot. So variance data are studied in order to enable a prediction of the contaminant concentration variability between samples.

The variability, more precisely the variance, between sample concentrations is due to the sampling process (see Fig. 2) and more specifically to the following points:

- taking incremental samples and aggregating them in a sample;
- homogenizing and dividing the sample into sub-samples;
- dividing the sub-sample into analysis samples and achieving a chemical analysis.

The first point generates variability due to the intrinsic inter-individual concentration variability. The variability at the second point is mostly caused by the grinding method in case the particles obtained are not small enough to ensure that after division in sub-samples, the concentrations in each sub-sample are strictly the same. The latest point triggers differences in concentration levels mainly because of the errors induced by the analysis method.

These observations are used to form a mathematical equation. We know that total variance is the sum of variance components, due to the fact that variance components are additive in case they are due to independent sources of random error. Consequently, for estimating components of variance, there must not be systematic error sources, stemming per example from a deficient instrument, a wrong use of the instrument by the experimenter, etc. Furthermore, we approximate that the inter-sample variability, the inter-subsample variability and the inter-analysis sample variability have not link one with the other. So, we assume, according to Whitaker *et al.*, that total variance (Variance $_{Total}$) is the sum of sampling variance (Variance $_{Sampling}$), sub-sampling variance (Variance $_{Sub-sampling}$) and analysis variance (Variance $_{Analysis}$):

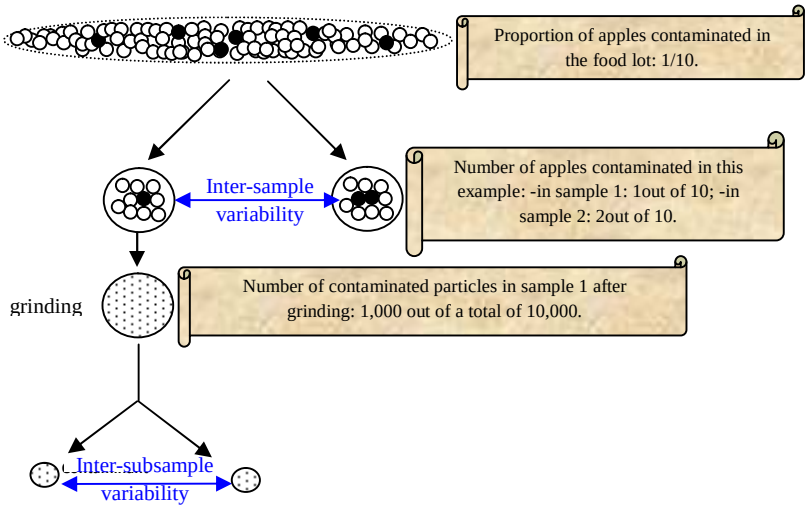
$$\text{Variance}_{Total} = \text{Variance}_{Sampling} + \text{Variance}_{Sub\text{-}sampling} + \text{Variance}_{Analysis}$$

However, sampling variance is not of the same order of magnitude at all as sub-sampling variance and analysis variance.

In Whitaker□s method, total variance, sub-sampling variance, and analysis variance are accurately quantified. But, given the fact that, when working on experimental data, Whitaker *et al.* always found that sub-sampling variance, and analysis variance were negligible in comparison to total variance we assume that we will not calculate the negligible variances. This is a slight modification of Whitaker□s method that makes the method much easier to achieve. Moreover, we can demonstrate, in an intuitive approach, that theses variances are always negligible. In the following, the orders of magnitudes of the variance components are going to be further defined. Indeed, while the laboratory analysis management today is outstanding thanks to the validated and efficient detection methods and procedures available for quality assurance in laboratories (accreditation), this is not necessarily true of the sampling operation, which seems to be the weak link in the sanitary control system for agricultural products. The sampling operation is often the main source of error when assessing the sanitary quality of a lot of agricultural commodities for pesticides, mycotoxins or heavy metals contaminations (Blanc, 2006). It can be added that it seems quite logical that the variability is very much lower between sub-samples

than between samples of the same origin, because the sample consists of aggregated whole individuals while the sub-sample consists of aggregated particles after grinding. This is particularly true for individuals of big size. Per example, for a contaminant spread in a lot of apples, on one individual apple out of 10, with the contaminant spread homogeneously into the flesh of the apple, consider taking two samples of 10 apples and dividing each sample after grinding into two sub-samples (Fig. 7). The inter-sample variability is the variability between the contamination level of at least two samples taken from the same lot. The samples might have contained zero, one, two or even more contaminated apples with given probabilities. So the inter-sample variability is high. Consider, per example, that the sample contained one contaminated apple. For this sample, a grinding process, dividing each apple into 1,000 particles, would trigger a sample contamination level of 1000 contaminated apple particles out of a total of 10,000 particles in the sample. The inter-subsample variability is the variability between the contamination level of at least two sub-samples coming from the same sample. Each subsample contamination level might consist of a certain amount of contaminated particles with a given probability, so the inter-subsample variability is rather low.

Fig.7. An example of sampling and sub-sampling variance components for a contaminated lot of apples with the contaminated apples or particles in black



When assuming that sub-sampling variance and analysis variance are negligible in comparison to the sampling variance, the following approximation can be made:

$$\text{Variance}_{\text{Total}} \approx \text{Variance}_{\text{Sampling}}$$

This approximation is now used to mention sampling variance instead of total variance.

In order to better comprehend which type of experimental data must be used, it is useful to further define the sampling variance. When two samples haven't got the same mean and when they come from the same population, then the difference between their means is simply due to sampling error. Sampling error is not the same for all the samples and all the populations. It can be small or big. Two factors determine its magnitude: the population variance, and the number of individuals in the sample:

1. The variance of the population: the bigger the population variance, the bigger the sampling error.
2. The number of individuals pooled together in each sample: the bigger the number of individuals, the smaller the sampling error. This principle is called the law of large numbers.

The latest factor requires further explanations:

The variability between samples consisting of pools of individuals is the variability between means. Indeed, we can consider that the concentration of a pool is equal to the mean of the concentrations of the individuals in the pool. The standard error of the mean is the standard deviation of the sample mean estimate of a population mean. It is usually estimated by the sample estimate of the population standard deviation divided by the square root of the sample size (assuming statistical independence of the values in the sample):

$$\text{Standard error}_{\text{mean}} = \text{Standard deviation} / (n^{0.5})$$

Knowing that the standard deviation is the square root of the variance, we can deduce that:

$$\text{Variance}_{\text{for pools}} \times \text{number}_{\text{of individuals in a pool}} = \text{Variance}_{\text{for individuals}}$$

Finally, we consider that:

$$\text{Variance}_{\text{for individuals}} = \text{Variance}_{\text{sampling}} = \text{Variance}_{\text{Total}}$$

Variance experimental data must be plotted against their respective mean concentration levels: Variance $_{sampling}$ = f(mean concentration). Each dot in the graph corresponds to data obtained for one lot with sampling variance and mean concentration calculated from the samples taken from the lot. Then, a regression curve is obtained in order to have an equation of the variance as function of the concentration. This equation is going to be useful to compute the accept probabilities.

Results of practical application to okadaic acid:
Data used:
An illustration with real data is given for okadaic acid in mussels with sampling variance and mean concentration data gathered from the literature. Total variance data (roughly equal to sampling variance) were compiled from 11 publications from various countries (Sweden, Italy, Germany, Ireland, Hong Kong, Tunisia) and one master thesis (Sweden) with the references in Annex 5. However, the literature review involved a much higher number of publications about okadaic acid levels in mussel, but all the publications were not relevant for variance data gathering. More concretely, for total variance data to be relevant, there had to be at least two separate mussel samples analyzed for one sampling point (means a specific area in the water), the minimum of two samples being taken at the same time. Furthermore, total variance for concentrations on individual shellfish were used directly as they were, whether total variance for concentrations on pools of shellfish were transformed following the formula:
Variance $_{for\ pools\ of\ mussels}$ x number $_{of\ shellfish\ in\ a\ pool}$ = Variance $_{for\ individual\ mussels}$
The concentration data in hepatopancreas given in a publication had to be transformed in whole flesh data, as the European regulation 853/2004 states that the okadaic acid results must be given per kg of whole flesh. In order to achieve this transformation, the following formula has been used:
(Concentration in hepatopancreas) / 6 = Concentration in whole flesh
The corresponding total variance data had to be modified too, as it had to be divided by 6^2, which means divided by 36. Moreover, it was decided to take into account total variance data for as much countries as possible in order to ensure that the sampling plan validation would not be country specific, but would on the contrary represent a global validation. This is a good way to check that the variability is not climate specific either. So, total variances data recorded in scientific publications were gathered and adapted in order to enable the development of a regression equation to predict the total variance between individual mussels in whole flesh as function of okadaic acid concentration (Fig. 8).

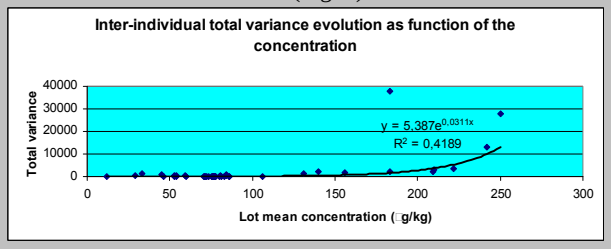


Fig. 8. Development of an exponential regression equation

Accept Probabilities

For a lot of a given mean concentration, the accept probability is computed as the probability that a sample consisting of a pool of individuals taken from the lot, has got a concentration level inferior or equal to the threshold (usually the legal food safety threshold). This probability is calculated thanks to the theoretical distribution and the total variance equation obtained in the previous sections.

Results of practical application to okadaic acid:

Table 2. An example of calculation of the accept probabilities

Lot mean concentration (µg/kg)	Population Variance	variance for a pool of 30 shellfish	mu for the lognormal distribution	sigma for the lognormal distribution	Accept probability
10	7.35211953	0.24507065	2.301361234	0.04947432	1
20	10.0340935	0.334469783	2.99531436	0.02891064	1
30	13.6944226	0.456480755	3.40094385	0.02251826	1

The lot mean concentration is an input. The population variance is the total variance calculated with the following formula: variance = 5.387* exp(0.0311*concentration). The variance for a pool of 30 shellfish is equal to the population variance divided by 30 (the number of individuals). The parameters of the lognormal distribution: mu and sigma are calculated thanks to the method of moments (Calculations in Annex 1). The accept probability is computed as the ordinate of the lognormal theoretical cumulative frequency distribution.

Non-Parametric Method from Schatzki: Principle and Application to Mycotoxins

In this section, for the whole method description, the notations are different from Whitaker's method. Due to the complexity of Schatzki's method, the theoretical description is given directly for aflatoxin contamination in pistachios. However, the particular case of retail pistachios is given as a practical application of the method for retail data computed from a few publications.

For granular materials and heterogeneously distributed contaminants, this methods appears as better adapted and we are going to see the reasons for this.

Many granular materials, which contain chemical contaminants, frequently have such contaminants distributed among the granules so that while the individual granules are uniform, the contaminant concentration varies widely among the granules. A small fraction of granules may contain the bulk of the contamination. This is why such type of contamination is called a heterogeneously distributed contamination. The contamination level of a lot is obtained by taking one or more samples of a predetermined size, measuring the level of contamination in the sample and assuming the level(s) measured are representative of the lot concentration. However, such measurements are subject to analytical, subsampling, and sampling errors. Of these, the sampling errors are by far the largest, so that the other errors may be ignored to a first approximation (Schatzki and Pan, 1996). Due to the low incidence of contamination (per example, only one granule out of 10,000 might be contaminated), a non-parametric method is more adapted than a parametric method due to the difficulty of getting a distribution close to the reality. Indeed, sources of error are minimized because, contrarily to the other method, there is no adjustment to a theoretical distribution and no variance data required.

The steps of Schatzki□s method are the following:

- Step 1: Calculating bins and their geometrical means
- Step 2: Distributions for individuals calculated from samples results and for many lots of various mean concentrations;
- Step 3: Calculating a merged master distribution
- Step 4: Determining mean lot concentrations thanks to the merged master distribution after calculating bins for the concentrations of individual granules
- Step 5: Calculations for the distribution of simulated samples
- Step 6: Accept probabilities

The method, primarily developed by Schatzki on pistachios, was applied to retail data. To our knowledge, this is the first time that sampling plans for retail data have been evaluated. Pistachios might potentially contain aflatoxin B1 (AFB1), which is a mycotoxin produced by the fungus *Aspergillus*. In order to ensure food safety, the level of AFB1 in ready-to- eat pistachios must not exceed 8 ppb (Regulation 165/2010/EC). Therefore, sampling plans must be set by official authorities in order to monitor retail pistachios to check the presence of AFB1. The incidence of contamination is very low, as a very low number of individuals within a lot might be contaminated; and the contamination level sustainable by a single nut is as high as 10^6 ppb (Schatzki and Pan, 1996). This study is a probabilistic evaluation of the impact that the sample size and the number of samples taken, have on the analysis results reliability.

Binning Sample Concentrations

Throughout this method, upper case will be used to refer to things related to C, the n-sized sample concentration. The corresponding lower case is used for things related to c, the concentration in a single nut.

The relation between sample size, n nuts, sample concentration, C, and the probability, $P_i(n)$, that this concentration falls within a predetermined range [bin$_i$l of C is based on the underlying probability distribution function (pdf), f(c), of aflatoxin concentration, c, in individual nuts. For this reason, a number J of bins of C must be set in order to bin sample concentration results for lots, obtained from real testing. The number of levels J which need to be considered is determined by the number of distinguishable levels of C. This number depends on the dynamic range and precision of C.

Results of practical application:

The dynamic range of C is limited by the minimum detectable level C_o of aflatoxin (approximately 0.1 ppb) and the maximum level sustainable by a single nut (which appears to be about l0^6 ppb or a little more) or approximately 7 decades: $10^6 / 10^{-1} = 10^7$.

The precision was indicated by Schatzki thanks to experimental data to be approximately 25%, from which it follows that a half-decade in C (approximately a factor of 3) covers ± 2 SD (1 ± 2 * 25% or 0.5-1.5):

Confidence Interval $_{95\%}$ □C □2*0.25*C

= C (1 ± 0.5)

= C * [0.5 ; 1.5]

And 1.5/0.5 = 3 □$10^{0.5}$

Thus, the experimental data can be expressed as J = 7/0.5 = 14 independent probabilities $P_i(n)$, corresponding to J logarithmic bins B_i of fixed size □ = $\log_{10} C_i^+/C_i^- = 0.5$, where C_i^+ and $C_i^- = C_{i-1}^+$ are the limits of B_i.

Bin limits for B_1, $C_1^- = 0.1$ (Limit of detection) and C_1^+ is calculated as follows:

$\text{Log}_{10}(C_1^+) = 0.5 + \log_{10}(0.1)$, so $C_1^+ = 10 \wedge (0.5 + \log_{10}(0.1))$

This calculation must be continued for the following bins. Then, for each bin, the geometric midpoint of the bin: C_i is calculated.

$C_i = (C_i^+ * C_i^-)^{0.5}$

The bins obtained (in ppb) are in Table 3.

Table 3: Concentration bins in ppb

	C-	C+	C
Bin1	0,1	0,31622777	0,17782794
Bin2	0,31622777	1	0,56234133
Bin3	1	3,16227766	1,7827941
Bin4	3,16227766	10	5,62341325
Bin5	10	31,6227766	17,7827941
Bin6	31,6227766	100	56,2341325
Bin7	100	316,227766	177,827941
Bin8	316,227766	1000	562,341325
Bin9	1000	3162,27766	1778,27941
Bin10	3162,27766	10000	5623,41325
Bin11	10000	31622,7766	17782,7941
Bin12	31622,7766	100000	56234,1325
Bin13	100000	316227,766	177827,941
Bin14	316227,766	1000000	562341,325

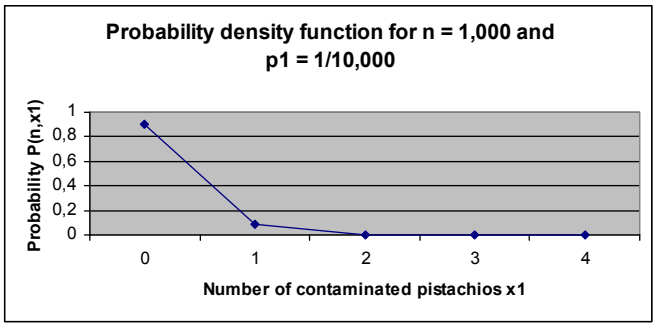
Distributions for Individuals

Two points are going to be discussed here. The first point is that a theoretical reasoning enables to determine the sample probability from the individual nut probability. The second point is that a sparse approximation is required to do the reverse: estimate the individual nut probability from the sample probability. Due to the fact that the contamination incidence is low, it would not be possible to have probability data on individuals directly via experiments. This is the reason why these probability data on individuals must be computed from probability data on samples.

- Determining the sample probability from the individual probability:

First, lets simplify the sample distribution by considering that all contaminated nuts have the same aflatoxin level, c_1, and occur with a probability p_l, a fraction $p_o = 1 - p_1$ being noncontaminated. This distribution may be approximated by a Poisson distribution (Schatzki, 1995 for further details). The Poisson distribution is a discrete probability distribution that fits rare events. In practice, an event is rare if its probability of occurrence is less than 0.05 and if n is at least equal to 50 (Scherrer, 1984).
An example of a Poisson probability density function that is obtained if n (the number of pistachios in the sample) is equal to 1,000 pistachios and if p_1 is equal to 1/10,000 is given in Fig. 8.

Fig.8. An example of probability density function



To be closer to the reality, for a given mean contamination of the lot, the above concepts need to be generalized to more than a single level of contamination c_1, knowing that to each level of contamination c_i corresponds a probability p_i. The lot distribution is modeled as a collection of nuts with a fraction p_i having aflatoxin concentration c_i, i = 1, ..., J and a fraction p_o uncontaminated. For each concentration level, the distribution of the number of contaminated pistachios in a sample, is the Poisson distribution. So, there are as many Poisson distributions as there are concentration levels. More theoretical details on the sample probability calculated from individual probability are in Annex 6.

- Determining the individual probability from the sample probability:

A sparse approximation is required in order to calculate the individual nut probability from the sample probability given by experiments (More details in Schatzki, 1995).
Experiments giving contamination levels of samples are used to calculate c_i and p_i. The sample size n is chosen to be small compared to $1/p_i$ for all i except i = 0, then all $n*p_i << 1$. Moreover, if the sample size is small enough, then a single nut is the sole nut that is contaminated in the sample. Sample concentrations from one lot that fall within a concentration bin are approximated by the geometric midpoint of the bin: C_i. The proportion of samples from a lot that fall within C_i is called P_i. From the observed P_i and C_i, the c_i and p_i can be calculated:

$$c_i = n * C_i \quad \text{and} \quad p_i = P_i / n$$

Then, for each lot, a distribution can be drawn with p_i as function of c_i.

Results of practical application:

Data used:
Sample results from publications for pistachios taken at retail stage (supermarket, grocery, market, etc.) are used. It must be possible to approximate the number of nuts in each sample (sample size n). Sample size must be small enough in order to consider, with the smallest risk of error as possible, that at most one nut is contaminated per sample.
Each publication taken into account gives aflatoxin sample results aiming to assess the population's exposure. So, in this particular case, a lot is considered to be a snapshot of what the consumers could have consumed. The samples taken from this lot must be representative of the average consumers habits via an appropriate sharing out between pistachio brands, packaging sizes, places where they are bought. All the sample data from one publication are considered to come from one single lot. As the next step consists of obtaining one probability density function per lot, then solely publications comprising at least 2 contaminated samples at 2 different concentration levels were taken into account.
The publications are the following: Sarhang Pour *et al.* (2010), Ghali *et al.* (2009), Set *et al.* (2010), Thuvander *et al.* (2001), Fernane *et al.* (2010).

Calculations and results:
For each publication found (that gives retail pistachios analysis data with a known sample mass) the probability P_i for each level C_i is assigned. Then the P_i and C_i are transformed into p_i and c_i, which means that the probabilities and concentrations for the samples are transformed into probabilities and concentrations for individual nuts. An example of results obtained is given for the data from Thuvander *et al.* in Table 4.

Table 4: Example of results obtained for the distribution of individual nuts (from Thuvander *et al.* raw data)

Sample			Individual nut	
Ci	Pi	n	ci	pi
0,56234	0,23809524	393	220,99962	0,00060584
177,83	0,04761905	393	69887,19	0,00012117
1778,28	0,04761905	393	698864,04	0,00012117

Concentrations are given in ppb.
For each lot, the p_i are plotted against the c_i for the lot distribution.

Merged Master Distribution for Individuals

Source: Schatzki 1999:
Schatzki (1995a) noted that the different lots showed similar distributions, differing mainly by a constant multiplier of p_i, $i > 0$. Accordingly, all the experimental distributions were shifted vertically on a log p plot to coincide as closely as possible to obtain a merged master distribution (Schatzki, 1999). This distribution is going to be used afterwards to compute p_i for different mean concentrations (m) of the lot. Indeed, per example, the probability p_i for an individual nut contaminated at a c_i of 25 ppb is lower for a slightly contaminated lot than for a highly contaminated lot.

Results of practical application:

The lot distributions are plotted on a graph (full log scale) and regression curves are drawn for each lot. These regression curves take the form: y = k * x^z, as shown in Fig. 9.

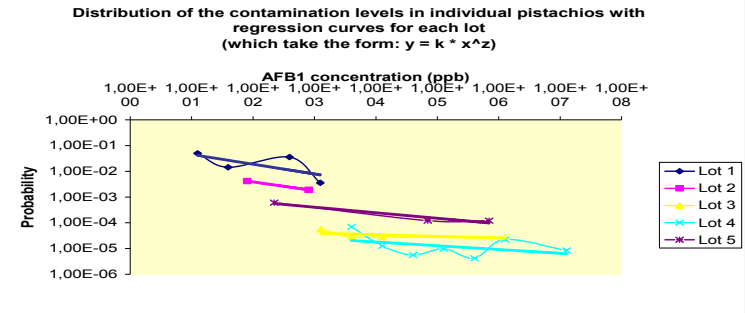


Fig.9. Lot distributions with calculated data represented by dots and their corresponding regression curves

In order to calculate a merged master distribution (for each lot mean concentration level), the relationship between k and z is studied: z is plotted as function of k and a regression curve is obtained (Fig. 10).

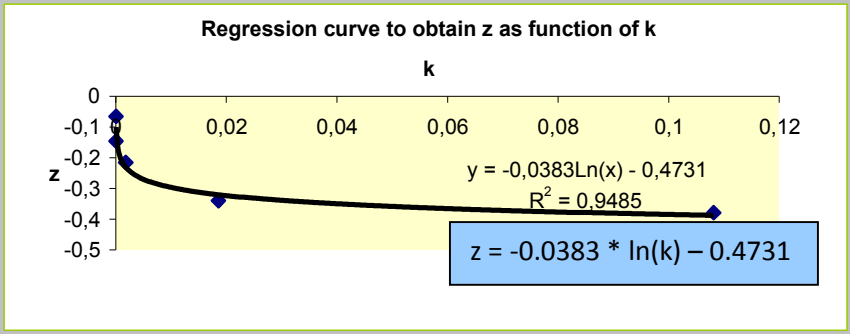


Fig.10. Graph in order to obtain the equation of z as function of k

Determining Mean Lot Concentrations

Mean lot concentrations are determined thanks to the merged master distribution after calculating bins for the concentrations of individual granules.

The quasi-continuous distribution of aflatoxin among the individual nuts of a lot is approximated as a discrete distribution, binned into half-decile bins. Here c_i is the geometric midpoint of bin i. All concentrations falling into bin i are approximated by c_i. Theoretical c_i are determined as in Schatzki 1999. The distribution for any hypothetical lot of the mean aflatoxin concentration m is computed thanks to the p_i evaluated from the merged master distribution equation of p_i as function of c_i. For each simulated lot, m (mean contamination level) is calculated:

$m = \square p_i * c_i$

Results of practical application:

Simulations of two different lot mean contamination levels m are shown in Table 5.

Table5.Calculations in order to obtain two m levels from two k levels.

ci	k0	k0.5	z0	z0.5	p0	p0.5	ci x p0	ci x p0.5
25	0.00001	0.00002	-0.03215495	-0.05870249	9.01673E-06	1.65565E-05	0.000225418	0.000413911
79	0.00001	0.00002	-0.03215496	-0.05870250	8.689244E-06	1.54751E-05	0.00068645	0.001222536
250	0.00001	0.00002	-0.03215497	-0.05870251	8.37326E-06	1.44632E-05	0.002093314	0.003615805
791	0.00001	0.00002	-0.03215498	-0.05870252	8.06881E-06	1.35176E-05	0.006382425	0.010692427
2500	0.00001	0.00002	-0.03215499	-0.05870253	7.7757E-06	1.26346E-05	0.019439245	0.031586569
7906	0.00001	0.00002	-0.03215500	-0.05870254	7.4931E-06	1.18089E-05	0.059240428	0.09336133
25000	0.00001	0.00002	-0.03215501	-0.05870255	7.22079E-06	1.10372E-05	0.180519635	0.275930665
79057	0.00001	0.00002	-0.03215502	-0.05870256	6.95836E-06	1.03159E-05	0.550107138	0.815547056
250000	0.00001	0.00002	-0.03215503	-0.05870257	6.70547E-06	9.64179E-06	1.676368503	2.410446383
						m =	2.495062557	3.642816682

Here c_i (in ppb) are inputs, k_0 is set at 0.00001 and $k_{0.5}$ is set at 0.00002 (per example). Then the corresponding z_0 and $z_{0.5}$ are calculated thanks to the following equation: $z = -0.0383 * \ln(k) \square 0.4731$. The p_0 is obtained, knowing that $p_0 = k_0 * c_i \wedge z_0$ and $p_{0.5}$ is obtained too. Afterwards, the $c_i * p_0$ are calculated for each c_i level, as well as the $c_i * p_{0.5}$. It follows that the concentration mean level of the first lot can be computed as the sum of all the $c_i * p_0$ and is equal to 2.495 ppb. The concentration mean level of the second lot is 3.643 ppb. This process is repeated for a wide range of m levels thanks to different k inputs in order to achieve the next steps.

Calculations for the Distribution of Simulated Samples

A simulated sample is considered as a set of small samples in order to enable a mathematical calculation of the distribution of this simulated sample with the method. More precisely, a large simulated sample (per example 10 kg) can be thought of as a set of *N* small samples, each adequately small, which are ground and analyzed separately and from which the large sample aflatoxin concentration is derived by arithmetically averaging the results. While this would not be as efficient as blending the ground samples before analysis, the results would be the same. Yet each of the small samples would not contain more than a single significantly contaminated nut, enabling the use of the calculations detailed above.

In order to assure that the probability of obtaining a small sample with 2 contaminated nuts is less than 5% of that of a single nut, but that the chance of getting at least some contamination in the range of concentration of interest is not much less than 10%:

$$n * p_{i\ max} \square 0.1$$

For each m, the following statements are taken into account:

- n must be inferior or equal to: $0.1/p_{i\,max}$ (with $_{pi\,max}$ the probability of contamination $c_i = 25$).
- N (the number of small samples) is close to: $(n*N)/(0.1/_{pi\,max})$

In order to define n and N:

- N is chosen as the smallest rounded integer of $(n*N)/(0.1/_{pi\,max})$ enabling n inferior or equal to $0.1/p_{i\,max.}$
- n is calculated as: n=(n*N)/N

Per example, in order to simulate a sampling plan consisting of taking one 10kg sample of pistachios: n*N = 7140 = the number of nuts in 10kg.

Then C_i and P_i for "the small samples" are calculated:

$C_i = c_i/n$

$P_i = (c_i*p_i)*n/c_i$

And the probability of a "small sample" being noncontaminated is $P_0 = 1- \sum P_i$

Results of practical application:

Table 6. Simulation of sample distribution for one 10 kg pistachio sample (7,140 nuts) thanks to P_i and C_i of "small samples"

m	2.49506256	3.64281668
pi max	9.0167E-06	1.6556E-05
n inferior or equal to:	11090.4891	6039.93919
N close to	0.64379487	1.18213111
N round integer	1	2
n recalculated	7140	3570
ci	ci x p0	ci x p0,5
25	0.00022542	0.00041391
79	0.00068645	0.00122254
250	0.00209331	0.0036158
791	0.00638243	0.01069243
2500	0.01943925	0.03158657
7906	0.05924043	0.09336133
25000	0.18051964	0.27593066
79057	0.55010714	0.81554706
250000	1.6763685	2.41044638
ci	Ci	Ci
25	0.0035014	0.0070028
79	0.01106443	0.02212885
250	0.03501401	0.07002801
791	0.11078431	0.22156863
2500	0.35014006	0.70028011
7906	1.10728291	2.21456583
25000	3.50140056	7.00280112
79057	11.072409	22.1448179
250000	35.0140056	70.0280112
ci	Pi	Pi
25	0.06437949	0.05910656
79	0.06204119	0.05524625
250	0.05978504	0.05163369
791	0.05761127	0.04825786
2500	0.05551848	0.04510562
7906	0.05350072	0.04215785
25000	0.05155641	0.0394029
79057	0.0496827	0.0368279
250000	0.04787708	0.03442117
Sum of Pi	0.50195238	0.41215979
P0	0.49804766	0.58784021

For the 2 lot mean concentration levels m calculated before, the corresponding $p_{i\ max}$ are reported with $p_{i\ max}$ being the p_i for $c_i = 25$ ppb.

After some calculations, we obtain $n_{\ recalculated} = n$ and $N_{\ round\ integer} = N$.

The $c_i * p_0$ and $c_i * p_{0.5}$ calculated in previous section are reported here.

$C_i = c_i/n$

$P_i = (c_i*p_i)*n/c_i$

Sum of $P_i = \sum P_i$

$P_0 = 1- \sum P_i$

Accept Probabilities

The sample distribution is established by Monte Carlo. A computer program is developed. This program computes a big number of iterations of samples for each lot. Each sample iteration is calculated as the mean of "small samples" results. Small samples are drawn at random, considering their calculated probabilities of appearance. The accept probabilities are computed as they are the fraction of the sample distribution that fall below the threshold concentration. The lot mean concentration m is recalculated.

Results of practical application:

Programming is achieved under the R software (8,000 iterations). For the mean concentration level m = 3.64 ppb, per example, given the N, C_i and P_i calculated before, the inputs are the following:

```
x<-c(0, 0.0070028, 0.02212885, 0.07002801, 0.22156863, 0.70028011, 2.21456583, 7.00280112, 22.1448179, 70.0280112)

y<-c(0.58784021, 0.05910656, 0.05524625, 0.05163369, 0.04825786, 0.04510562, 0.04215785, 0.0394029, 0.0368279, 0.03442117)

resamples<-lapply(1:8000,function(i)sample(x,size=2,replace=T,prob=y))

r.mean<-sapply(resamples,mean)

sort(r.mean)

mean(r.mean)
```

With: - the variable x being the C_i, the variable y being the P_i (without forgetting the probability of noncontamination)
- in the lapply function, size=N
- the accept probabilities are calculated as the proportion of samples having concentrations under the threshold. The concentrations in ascending order are given by the sort() function.
- m is recalculated thanks to mean(r.mean)

TESTS AND SELECTION OF THE BEST FIT SAMPLING PLAN

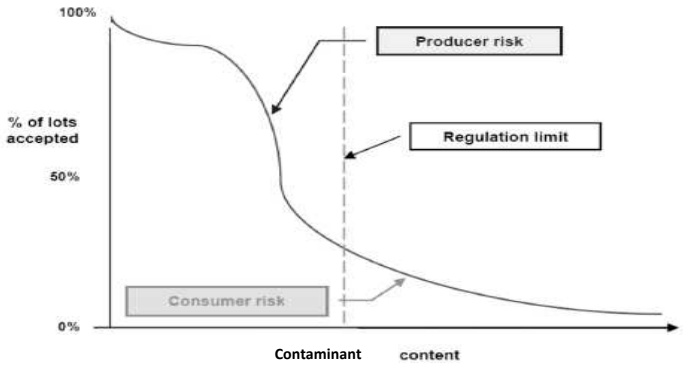
Operating Characteristics (OC) Curve Principle

Operating Characteristics (OC) curves enable to calculate the probability of mistake in determining the average contaminant concentration level for various sampling schemes (Fig. 11).

They show the risk of:

- accepting lots at a true concentration above the threshold (consumer risk);
- rejecting lots at a true concentration under the threshold (producer risk).

Fig.11. Operating Characteristics (OC) curves principle (Blanc, 2006)



A sampling plan must be selected considering the technical and economic feasibility as well as the consumer risk and producer risk levels. It is therefore crucial to assess these two risks in order to select a best fit sampling plan. An ideal sampling plan would lead to the acceptance of all the lots with contaminant content below the allowed limits and to the rejection of those with contaminant content higher than these limits, thus reducing the risk to consumer and producer to zero. In reality however, these risks can not be totally eliminated, but only reduced as much as possible, keeping in mind the feasibility. When this is achieved, then a best fit sampling plan is selected.

OC Curves Obtention for Phycotoxins and Selection of a Sampling Plan

For Whitaker□s method, the accept probabilities that have been calculated, correspond to a sampling plan with a single sample taken and compute the probability that this sample is inferior or equal to the threshold concentration, given the lot mean concentration. In order to obtain an OC curve, the accept probabilities must be plotted against the lot mean concentrations.

Sampling Plans for One Sample of Various Size

The corresponding OC curves are plotted without further calculations.

Results of practical application for phycotoxins (Fig. 12):

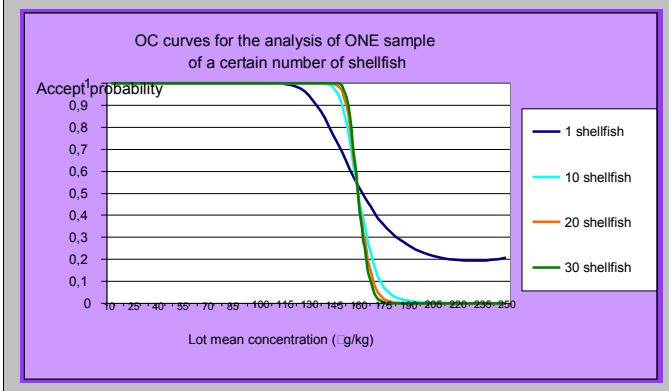


Fig.12. OC curves for sampling plans with one sample taken

For a 30 shellfish sample:
Accept probability = 95% for a concentration of 152 µg/kg
Accept probability = 5% for a concentration of 169 µg/kg
Increasing the number of shellfish per sample decreases both consumer and producer risks.

Multiple Samples Sampling Plans

When the sampling plan consists of taking a number q of samples, and if the lot is accepted only if each of the q samples tests under the threshold concentration, then:
Accept probability = $(P_a)^q$
With P_a the accept probability for a sample of a given number of shellfish in a pool.

Results of practical application for phycotoxins (Fig. 13):

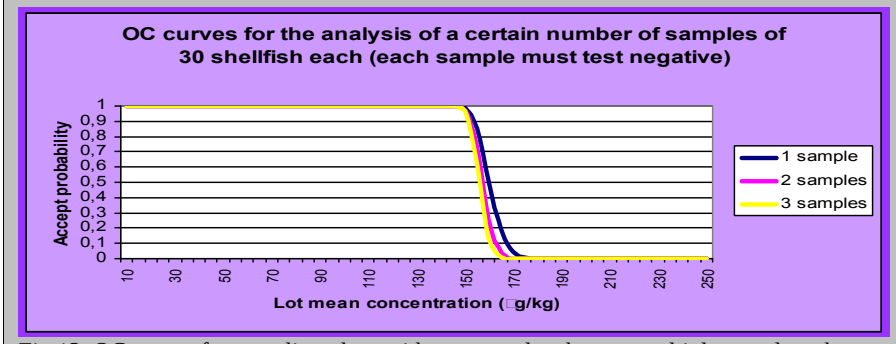


Fig.13. OC curves for sampling plans with one sample taken, or multiple samples taken

For 2 samples of 30 shellfish:
Accept probability = 95% for a concentration of 151 µg/kg
Accept probability = 5% for a concentration of 164 µg/kg
Increasing the number of samples (all samples must test at ≤ 160 µg/kg for the lot to be accepted) decreases consumer risk but increases producer risk.

Best Fit for Phycotoxins: Two Samples Sampling Plan

Results of practical application for phycotoxins (Fig. 14):

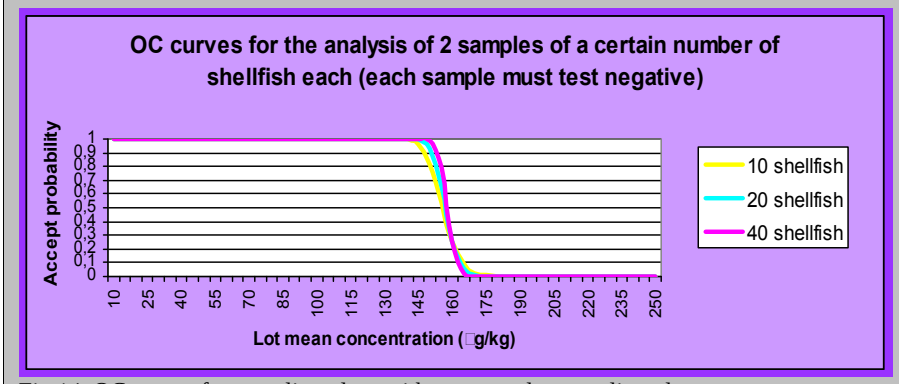


Fig.14. OC curves for sampling plans with two samples sampling plans

For 2 samples of 40 shellfish:
Accept probability = 95% for a concentration of 152µg/kg
Accept probability = 5% for a concentration of 163.5µg/kg

Increasing the number shellfish per sample for 2 samples (all samples must test at 160 µg/kg for the lot to be accepted) decreases consumer and producer risks.

The best fit sampling plan is a 2 samples sampling plan with a sample size of 30-40 individual mussels.

OC curves Obtention for Mycotoxins and Selection of a Sampling Plan

For Schatzki's method, for a given sample size, the accept probabilities P_a that are calculated thanks to the computer program correspond to a sampling plan involving that one single sample is taken. The accept probabilities for a given sampling plan, after the requested calculations, must be plotted against the lot mean concentrations.

Sampling Plans for Various Sample Size and Number of Samples Taken

When the sampling plan consists of taking a number q of samples, and if the lot is accepted only if each of the q samples tests under the threshold concentration, then:
Accept probability = $(P_a)^q$

Results of practical application:

For the sampling plans tested, each of the samples must test under 8 ppb for the lot to be accepted (Fig. 15).
The first sampling plan involves taking one 10 kg sample, whereas the second sampling plan consists of taking one 30 kg sample. For the first sampling plan, the producer risk is low, but the consumer risk is really high. It can be observed that when the sample weight is higher, the accept probabilities are lower for lots having a lot mean concentration under 8 ppb (good lots) as well as for lots having a lot mean concentration above 8 ppb (bad lots). So, the producer risk is higher, but the consumer risk is lower.
The third sampling plan consists of taking three 10 kg samples and the fourth sampling plan, consists of three 30 kg samples. For these two sampling plans, the producer risk is high. So, even if the consumer risk is low, such a sampling plan strategy can not be validated, as too many lots of good quality would be rejected.

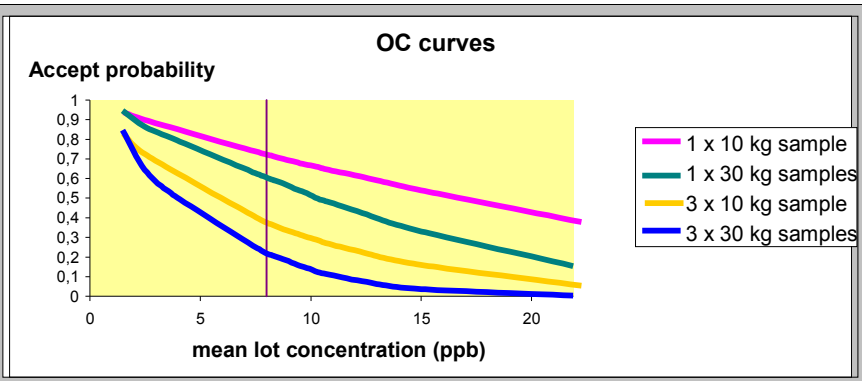


Fig.15. OC curves for sampling plans with 10 kg samples or 30 kg samples and involving taking either one or three samples

All these results show either a too high accept probability for bad lots or a too low accept probability for good lots. And yet, the objective is to define a sampling plan that would have a high accept probability for good lots and a low accept probability for bad lots.

Sequential Sampling Plans

A sequential sampling plan consists of taking a first sample and testing it:
- If the concentration of the sample is under a low concentration, then the lot is accepted immediately;
- If the sample concentration is above a high concentration, the lot is rejected without further testing;
- In the other cases, it says when the sample concentration is between the low and the high concentrations cited above, then a second sample is required.

So, a second sample is tested only if necessary. The same kind of criteria of acceptance and rejection as well as the criteria for further testing can be used for the second sample. And this operation can be repeated for a third and even a fourth sample, depending on the maximum number of samples that it has been decided to test in each specific sequential sampling plan.

Results of practical application:

A sequential sampling plan with one up to three samples tested is compared to the sampling plan described before and which consists of taking three 10 kg samples and accepting the lot if each of the three samples tests under 8 ppb (Fig. 16).
The sequential sampling plan tested is the following:
A first sample is taken and:

- If the 10 kg sample tests □2ppb, the lot is accepted,
- Otherwise, if the 10 kg sample tests □6ppb and > 2 ppb, the analysis of another 10 kg sample is requested,
- If the 10 kg sample > 6 ppb the lot is refused.

The same criteria are applied for the second sample.
If the third sample tests □2ppb, the lot is accepted, otherwise, it is rejected.

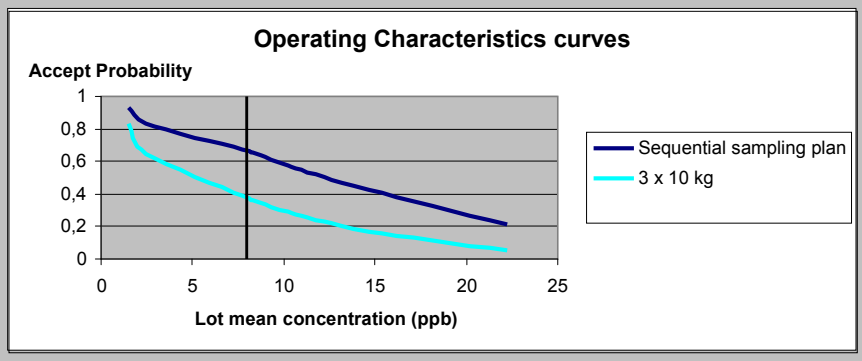


Fig.16. OC curves for a sampling plan with three 10 kg samples and a sequential sampling plan

It can be observed, when comparing the two OC curves obtained in Fig. 13 that, even though the limits (2 ppb and 6 ppb) for the sequential sampling plan are set below the regulation limit (8 ppb), this sampling plan does not give good results. Indeed, the consumer risk is too high. However, the producer risk is lower, than for the other 3 x 10 kg sampling plan.

Finally, the fact that the lot might have been accepted after the analysis of only one sample (sequential sampling plan) triggers a high consumer risk due to the low incidence of contamination.

Best Fit for Retail Pistachios: Non Sequential Sampling Plan but with Varying Thresholds

Results of practical application:

This best fit type of sampling plan involves testing each of the samples in all cases in order to enable a low consumer risk. However, in order to reduce the producer risk, the thresholds are variable. Various strategies are tested, and the one that gives the lowers both producer and consumer risks is reported in Fig. 17 under the name: "Special sampling plan".

This sampling plan that gives the most valuable results (best fit special sampling plan) consists of testing 4 samples of 10 kg each and accepting the lot if:
Each of the 4 samples test ≤12 ppb
OR 3 of the samples test ≤2 ppb and 2 ppb < 1 sample ≤34 ppb
OR 2 samples ≤0.1 ppb and 0.1ppb < 1 sample ≤2 ppb and 1 sample > 2 ppb.

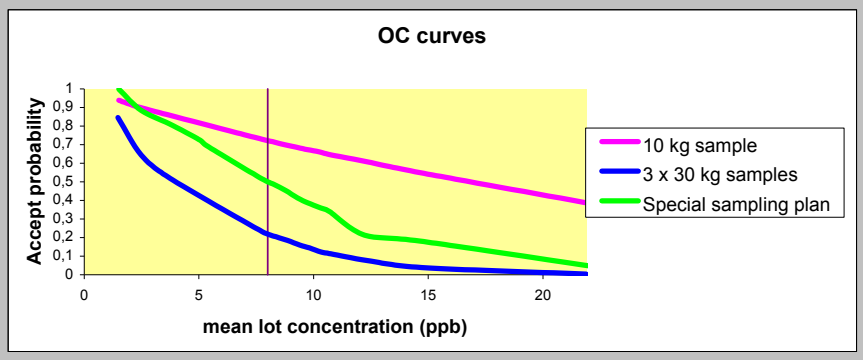


Fig.17. OC curves for a sampling plan with three 30 kg samples and another sampling plan with one 10 kg sample compared to a special best fit sampling plan

The best fit sampling plans protects both the consumer and the producer. It is a well balanced sampling plan.

CONCLUSIONS

We can conclude that an optimal sampling plan for homogeneously distributed contaminants is obtained thanks to OC curves plotting accept probabilities calculated with Whitaker's method. These accept probabilities are obtained thanks to the contaminant's distribution and thanks to an equation of the total variance as function of the concentration. For heterogeneously distributed contaminants, Whitaker's method does not appear as best adapted, contrarily to Schatzki's method. This method computes accept probabilities for a sample considered as a set of "small samples" (containing at most one contaminated individual). The contamination probabilities for "small samples" are computed from Poisson distributions of individuals. Two best fit sampling plans are proposed: one for phycotoxins in shellfish and another for mycotoxins in pistachios.

ACKNOWLEDGEMENTS

This project was financed by the BASELINE European research project, which is part of the 7th Framework Programme (grant agreement 222738). This project is titled "Selection and improving of fit-for-purpose sampling procedures for specific foods and risks" and spans from 2009 to 2013.

REFERENCES

Blanc, M. (2006). Sampling: the weak link in the sanitary quality control system of agricultural products. *Mol. Nutr. Food Res.,* 50, p.473 – 479

Campbell, B.C., Molyneux, R.J., Schatzki, T.F. (2003). Current research on reducing pre- and post harvest aflatoxin contamination of U.S. almond, pistachio and walnut. *Toxin reviews*, vol. 22, n°2 and 3, p.225-266

Carmody, E. P., James, K. J., Kelly, S. S. (1996). Dinophysistoxin-2: the predominant diarrhoetic shellfish toxin in Ireland. *Toxicon*, vol. 34, p. 351-359.

D'Agostino and Stephens, *Goodness-Of-Fit Techniques***,** Marcel-Dekker, New York, 1986, Table 4.7, p.123. All of Chapter 4, pp.97-193, deals with goodness-of-fit tests based on empirical distribution function (EDF) statistics.)

Duinker, A., Bergslien, M., Strand, Ø, Olseng, C. D., Svardal, A. (2007). The effect of size and age on depuration rates of diarrhetic shellfish toxins (DST) in mussels (*Mytilus edulis* L.). *Harmful Algae*, vol. 6, p. 288-300.

Edebo, L., Lange, S., Li, X.P., Allenmark, K., Lindgren, K., Thompson, R. (1988). Seasonal, geographic and individual variation of okadaic acid content in cultivated mussels in Sweden. *APMIS*, vol. 96, issue 7-12, p. 1036-1042.

Fernane, F., Cano-Sancho, G., Sanchis, V., Marin, S., Ramos, A.J. (2010). Aflatoxins and ochratoxin A in pistachios sampled in Spain: occurrence and presence of mycotoxigenic fungi. *Food additives and contaminants. Part B, Surveillan*ce, vol. 3, n°3, p. 185-192.

Ghali, R., Belouaer, I., Hdiri, S., Ghorbel, H., Maaroufi, K., Hedilli, A. (2009). Simultaneous HPLC determination of aflatoxins B1, B2, G1 and G2 in Tunisian sorghum and pistachios. Journal of food composition and analysis, 22, p. 751-755.

Godhe, A., Svensson, S., Rehnstam-Holm, A.-S. (2002). Oceanographic settings explain fluctuations in Dinophysis spp. and concentrations of diarrhetic shellfish toxin in the plankton community within a mussel farm area on the Swedish west coast. *Marine Ecology Progress Series*, vol. 240, p. 71-83.

Kacem, I., Hajjem, B., Bouaïcha, N. (2009). First evidence of okadaic acid in *Mytilus galloprovincialis* mussels, collected in a Mediterranean lagoon, Tunisia. *Bull Environ Contam Toxicol*, vol. 82, p. 660-664.

Klöpper, S., Scharek, R., Gerdts, G. (2003). Diarrhetic shellfish toxicity in relation to the abundance of *Dinophysis spp.* in the German bight near Helgoland. *Marine Ecology Progress Series*, vol. 259, p. 93-102.

Lindegarth, S., Torgersen, T., Lundve, B., Sandvik, M. (2009). Differential retention of okadaic acid (OA) group toxins and pectenotoxins (PTX) in the blue mussel, *Mytilus edulis* (L.), and European flat oyster, *Ostrea edulis* (L.). *Journal of Shellfish Research*, vol. 28, p. 313-323.

Mak, C.Y., Yu, H., Choi, M.C., Shen, X., Lam, M.H.W., Martin, M., Wu, R.S.S., Wong, P.S., Richardson, B.J., Lam, P.K.S. (2005). Okadaic acid, a causative toxin of diarrhetic shellfish poisoning, in green-lipped mussels *Perna viridus* from Hong Kong fish culture zones: Method development and monitoring. *Marine Pollution Bulletin*, vol. 51, p. 1010-1017.

McCarron, P., Kilcoyne, J., Hess, P. (2008). Effects of cooking and heat treatment on concentration and tissue distribution of okadaic acid and dinophysistoxin-2 in mussels (*Mytilus edulis*). *Toxicon*, vol. 51, issue 6, p. 1081-1089.

Millot, G. (2009). *Comprendre et r□aliser les tests statistiques □l□aide de R, Manuel pour les d□butants.* Ed. de boeck.

Reizopoulou, S., Strogyloudi, E., Giannakourou, A., Pagou, K., Hatzianestis, I., Pyrgaki, C. Granéli, E. (2008). Okadaic acid accumulation in macrofilter feeders subjected to natural blooms of *Dinophysis acuminata*. *Harmful Algae*, vol. 7, p. 228-234.

Sarhang Pour, R., Rasti, M., Zighamian, H., Daraei Garmakhani, A. (2010). Occurrence of aflatoxins in pistachio nuts in Esfahan province of Iran. *Journal of food safety*, vol. 30, issue 2, p. 330-340.

Schatzki, T.F. (1999). Distribution of Aflatoxin in Pistachios. 6. Seller□s and Buyer□s Risk*J. Agric. Food Chem.*, 47, 3771-3775 3771.

Schatzki, T.F. (1995). Distribution of Aflatoxin in Pistachios. 1. Lot Distributions. *J. Agric. Food Chem.* 1995, 43, 1561 -1 565.

Schatzki, T.F., Pan, J.L. (1996). Distribution of aflatoxin in pistachios. 3. Distribution in pistachio process streams. *J. Agric. Food Chem*. 44, p. 1076-1084.

Scherrer, B. (1984). *Biostatistiques.* Ed. Chicoutimi: Gaëtan Morin.

Set, E., Erkmen, O. (2010). The aflatoxin contamination of ground red pepper and pistachio nuts sold in Turkey. *Food and chemical toxicology*. Vol. 48, issues 8-9, p. 2535-2537.

Sidari, L., Nichetto, P., Cok, S., Sosa, S., Tubaro, S., Honsell, G., Della Loggia, R. (1998). Phytoplankton selection by mussels, and diarrhetic shellfish poisoning. *Marine Biology*, vol. 131, p. 103-111.

Svensson, S., Förlin, L. (2004). Analysis of the importance of lipid breakdown for elimination of okadaic acid (diarrhetic shellfish toxin) in mussels, *Mytilus edulis*: results from a field study and a laboratory experiment. *Aquatic Toxicology*, vol. 66, p. 405-418.

Thuvander, A., Möller, T., Enghardt Barbieri, H., Jansson, A., Salomonsson, A.C., Olsen, M. (2001). Dietary intake of some important mycotoxins by the Swedish population. Food additives and contaminants. Vol. 18, n° 8, p. 696-706.

USDA, 2009. Practical procedures for grain handlers:
Archive.gipsa.usda.gov/pubs/primer.pdf

Whitaker, T.B., Dickens, J.W., Monroe, R.J., Wiser, E.H. (1972). Comparison of the observed distribution of aflatoxin in shelled peanuts to the negative binomial distribution. *Journal of the American Oil Chemists Society*, 49, p. 590-593.

Wrange, A.-L. (2008). Investigating mechanisms behind species-specific differences in uptake of diarrhetic shellfish toxins (DST) between oysters (Ostre edulis) and mussels (Mytilus edulis). Master thesis.

APPENDIXES

Annex 1

As an example, the calculation of the parameters mu and sigma of the Lognormal distribution thanks to the method of moments is the following:

$$\mu = \ln(\overline{x}) - 0.5 \ln\left(\frac{s^2}{\overline{x}^2} + 1\right) \quad ;$$

$$\sigma = \left(\ln\left(\frac{s^2}{\overline{x}^2} + 1\right)\right)^{0.5} .$$

with s^2 the variance and $\overline{x}$ the mean on the sample results

Annex 2

For the lognormal distribution, the ordinates can be computed thanks to the following Excel formula:
LOGNORMDIST (x_r ; □; □)

Annex 3

Kolmogorov-Smirnov test:

The critical value found in a table (per example, table of critical values on: www.apprendre-en-ligne.net/random/t-KS.pdf) must be compared to the variable tested named D_{calc}. The calculations of D_{calc} under Excel are shown in the following, as well as the calculation of the p-value. However, the R software may be used as well, if one is more used to it.

- Using Excel:

There are 2n differences calculated, n being the number of samples in the lot tested. So, for each contaminant concentration recorded, there are two difference values calculated:

$$d_1 = F_{theoritical}(x_r) - F_{1\ observed}(x_r)$$
$$d_2 = F_{2\ observed}(x_r) - F_{theoritical}(x_r)$$

x_r: value of the contaminant concentration in the rank
$F_{1\ observed}(x_r)$: $P(x < x_r)$
$F_{2\ observed}(x_r)$: $P(x \leq x_r)$
With x being the contaminant concentration value.

The variable that is tested (D_{calc}) is the biggest value among the 2n differences calculated.
The p-value formula is:

$$\text{p-value} = 2 \cdot D_{calc} \sum_{f=0}^{pe[n(1-Dcalc)]} \left[C_n^f \cdot (1 - D_{calc} - f/n)^{n-f} \cdot (D_{calc} + f/n)^{f-1}\right]$$

f: counter taking integer values from 0 to pe[1 □D_{calc}]
pe[ ]: integer part of n(1 □D_{calc}). Consists of suppressing the decimals
C_n^f: combination of n objects from a set of f objects (from Millot, 2009)

This formula is only valid if there are no identical values among the observed values.

- Using the R software:

In practice, when using R software for the adjustment to a lognormal distribution, the R function is the following:
>ks.test(x_r, "plnorm", meanlog = μ, sdlog = σ^2)
When using this function, μ and σ^2 must be replaced by their numerical values and x_r must correspond to the variable in a data frame created with an appropriate formula under R. The software gives D_{calc} and the p-value as an output.

Annex 4

Anderson-Darling test:
The critical values at the risk level of 5% can be found in the book from D'Agostino and Stephens, 1986.

Theory:
The test statistic is:

$$A^2_n = n \int_{-\infty}^{+\infty} \frac{[F_{2observed}(x_r) - F_{theoritical}(x_r)]^2}{F_{theoritical}(x_r)(1 - F_{theoritical}(x_r))} d F_{theoritical}(x_r)$$

For the normal and lognormal distributions, the test statistic, A^2_n is calculated from:

$$A^2_n = -n - (1/n) \sum_{r=1}^{n} (2r-1) . [\ln(G_{theoritical}((x_r - m)/s)) + \ln(1-G_{theoritical}((x_{n-r+1} - m)/s))]$$

where n is the number of samples, and G is the standard normal cumulative distribution function with m the mean and s the standard deviation.
This formula needs to be modified for a small number of samples (n < 30):

$$A^2_{small} = A^2_n (1+(0.75/n)+(2.25/n^2))$$

Using the R software:
The R function to be used under the R software for this test, after loading the "ADGofTest" package, for the goodness of fit to the lognormal distribution, is the following:
>ad.test (x_r, plnorm, meanlog = μ, sdlog = σ^2)
This formula must be applied in the same way as the Kolmogorov-Smirnov test formula.

Annex 5

The publications which data are used are the following: Godhe *et al.* (2002); Svensson *et al.* (2004); Wrange *et al.* (2008); Sidari *et al.* (1998); Lindegarth *et al.* (2009); Klöpper *et al.* (2003); Carmody *et al.* (1996); Duinker *et al.* (2007); Mak *et al.* (2005); Kacem *et al.* (2009); Reizopoulou *et al.* (2008); Edebo *et al.* (1988).

Annex 6

To obtain $P_i(n)$:

$$P_i(n) = \int ... \int_{C_i^- < C < C_i^+} P(n; ..., x_1, x_2, ..x_j) I$$

This expression is the expression of the $P_i(n)$ in terms of n and $\{p_i, c_i\}$. This sample probability, while laborious to compute, will only depend on the set of np_j and the bin limits. The approach is then exactly the same as for the single-level contamination case. One estimates $\{P_i(n)\}$ from the fraction of samples in each bin B_i. Next one computes $\{p_i, c_i\}$ as described below, which, in turn, allows the computation of the $\{P_i(n)\}$ at any other n.