Sensitivity analysis for high quantiles of ochratoxin A exposure distribution

Isabelle Albert, Jean-Pierre Gauchi

ELSEVIER

International Journal of Food Microbiology 75 (2002) 143–155

INTERNATIONAL JOURNAL OF
Food Microbiology

www.elsevier.com/locate/ijfoodmicro

Sensitivity analysis for high quantiles of ochratoxin A exposure distribution

I. Albert*, J.-P. Gauchi

Food Risk Analysis Group, Biometrics Unit, National Institute for Agricultural Research (INRA), Domaine de Vilvert, 78352 Jouy-en-Josas cedex, France

Received 28 March 2001; received in revised form 29 October 2001; accepted 17 November 2001

Abstract

Using available data from a consumption survey and contamination data on ochratoxin A (OA) in food, a sensitivity analysis (SA) for high quantiles (95th and 99th quantiles) of OA exposure distribution was carried out, obtained by a Monte Carlo simulation in French children. Exposure assessment for food contaminants is important to control the risk of foodborne diseases. Risk assessors are interested in high quantiles of contaminant exposure distributions. As these exposure distributions are generally very asymmetrical, it is difficult to obtain relevant and stable high quantiles in such a context. Determining OA exposure distribution is complex because it is based on the sum of elementary exposure distributions (eight foodstuffs are analysed here), and each one of these is the product of a consumption distribution and a contamination distribution. The SA enables us to quantify the influences of the parameter variability of the consumption and contamination probability density functions (pdf) which have been fitted to the data, our simulation model inputs, on the 95th and 99th quantiles of the output exposure distribution. After some preliminary trials, we have postulated a quadratic polynomial regression model for the quantiles of OA exposure distribution in view of undertaking this SA. This regression model comprises 32 main factors, their 496 two-factor interactions and their 32 quadratic terms. The 32 factors are the parameters of the fitted pdf: 16 parameters of Gamma distributions relative to the eight consumed foods and 16 parameters of Gamma distributions relative to the eight food OA contaminations. For an optimal parameter estimation of such a large model, we used an experimental design approach depending on a resolution-V fractional factorial design of 6561 experiments. The factor ranges are established by a preliminary study of bootstrap sampling. From the bootstrap samples, the factor ranges are obtained taking into account the correlation between the two parameters of the fitted Gamma pdf. A full exposure distribution is simulated for each of the 6561 experiments. The consumption dependencies are taken into account by the Iman and Conover method. On the basis of this analysis, validated and useful models for each desired quantile are obtained showing a major influence of the parameters of ''Cereals'' (consumption and contamination) and slightly less so for parameter of ''Pork'' consumption in the sensitivity of the quantiles.

Keywords: Sensitivity analysis; Exposure assessment; Monte Carlo; High quantiles; Regression; Design of experiments; Food safety

* Corresponding author. Tel.: +33-1-3465-2843; fax: +33-1-3465-2217.
E-mail addresses: Isabelle.Albert@jouy.inra.fr (I. Albert), Jean-Pierre.Gauchi@jouy.inra.fr (J.-P. Gauchi).

0168-1605/02/$ - see front matter
PII: S0168-1605(01)00747-4

1. Introduction

Ochratoxin A (OA) is a mycotoxin produced by fungi of the genera *Aspergillus* and *Penicillium*. This mycotoxin may be a contaminant of grain stored in poor conditions, especially in Europe, and through the food chain also contaminates other foodstuffs, especially pork and poultry meat. Several toxicological studies (e.g. Pfohl-Leszkowick, 1999; JECFA, 1995; CCFAC, 1999) have shown this mycotoxin is an acute kidney toxin in humans. For these reasons, this mycotoxin remains under the close surveillance of the «Conseil Supérieur d'Hygiène Publique» in France and thus evaluating human exposure to this mycotoxin is an important public health question.

A consumption survey (ASPCC, 1999)—was conducted from June 1993 to June 1994 on 1161 French individuals (children, women and men) to estimate risk assessment exposure to OA in food. During 7 days, the participants were questioned on their consumption of several types of food. Our sensitivity analysis (SA) was restricted to children (individuals younger than 15) and to eight types of food, given in Table 1, keeping in mind that certain cover several foods, consumed by children. Table 2 gives descriptive statistics of these data for child consumers and non-consumers, and for men and women descriptive statistics are given in Verger (1999) and Gauchi (in press).

From consumption and contamination data, several exposure statistics have been computed by a non-parametric method—the NP–NP method—and a parametric method—the MP–P method (Gauchi, in press). In this paper, we use the MP–P method to evaluate the high quantiles of the OA exposure distribution in the SA. Indeed, the parametric method (the mixed parametric/parametric method, named the MP–P method in the text) seems to be very suitable for estimating relevant and stable high quantiles. This MP–P method, based on a Monte Carlo simulation, depends on the random samplings from fitted pdf (mixed parametric pdf for consumption and parametric pdf for contamination). The 95th and 99th quantiles of the exposure distribution are the model outputs (the responses) of our SA. The inputs are the parameters of the pdf fitted to the data. The SA enables us to quantify the influences of these parameters (and their interactions) on the responses via a quadratic polynomial model postulated for modelling the variation of these responses. For an optimal estimation of the parameters of the polynomial model, we used an experimental design approach depending on a resolution-V fractional factorial design of 6561 experiments. The factors of this design are the fitted pdf parameters. To simulate a meaningful range for each factor, we determine it by a parametric bootstrap approach (Efron and Tibshirani, 1993) carried out on the consumption and contamination samples. In

Table 1
Definition of the eight types of food analysed

Cereals: rice, wheat, corn starch, bread, pasta, flour, crispbread, breakfast cereals, crackers, biscuits, pastries, croissants and the like, popcorn, cornflakes, cooked sweet corn
Raisins: dried raisins
Other dried fruits: apricots, bananas, dates, figs, prunes
Pork: kidneys, cutlets, fillet, roast, ribs, loin, black pudding, ham pâté, salami, rillettes, andouillettes, bacon, sausages, foie gras
Poultry: chicken liver, chicken, duck, goose
Fruit juices: fruit juices
Wines: several different wines
Coffee: dry coffee (correction by a dilution coefficient of 17)

Table 2
Consumption in grams over an average day for all children (French children) and only those who consume a particular type of food

Food	*n*	Mean	Standard deviation	Minimum	Maximum
All the children					
Cereals	232	165.36	89.87	26.14	619
Raisins	232	0.25	1.19	0	11.71
Other dried fruit	232	0.41	2	0	17.14
Pork	232	39.12	25.98	0	140
Poultry	232	16.38	16.78	0	95.71
Fruit juices	232	81.67	149.72	0	1497.14
Wines	232	0.53	2.47	0	23.86
Coffee	232	0.34	1.47	0	12.57
The children who consume:					
Cereals	232	195.36	89.87	26.14	619
Raisins	17	3.39	3	0.43	11.71
Other dried fruit	15	6.29	5.12	1.29	17.14
Pork	219	41.44	24.87	0.57	140
Poultry	167	22.76	15.68	1.43	95.71
Fruit juices	157	120.69	168.68	1.43	1497.14
Wines	25	4.94	6	0.43	23.86
Coffee	24	3.33	3.36	0.14	12.57

order to consider the correlation between the pdf parameters, three experimental domains of variation were designed inside a 95% concentration ellipse.

This type of SA should be distinguished from marginal analysis: what is the effect of infinitely small changes (perturbations)? The effects in the SA are measured by the parameter vector β of the postulated quadratic polynomial regression model.

In Section 2, we present how we obtain the OA exposure distribution and the methodology of our SA. In Section 3, we give the results, and we conclude in Section 4 by a discussion.

2. Methods

The contamination data ($\mu g\ kg^{-1}$ OA in food) come from elements independent of the consumption survey: they originate from SCOOP task 3.2.7. (see detailed reference). Table 3 gives descriptive statistics of OA contamination in the eight types of food. For poultry contamination data, the only information available was 31 values between 0 and 0.03, and 62 values between 0.03 and 0.18: the mean of 0.075 (Table 3) is calculated by the weighted arithmetic mean of the two means calculated from the two uniform distributions on [0; 0.03] and [0.03; 0.018]. On the other hand, the standard deviation of the unknown data is estimated by the Bienaymé–Tchebysheff inequality by supposing that the range [0; 0.18] encompasses 100% of the data. This inequality indicates that 75% of the data are contained in the interval $[\mu \pm 2\sigma]$, which, with an estimation of 0.075 for μ, leads here to an estimation of 0.0342 for σ. In the following, poultry contamination data are simulated by a Gamma distribution with mean = 0.075 and standard deviation = 0.0342 corresponding to a shape parameter = 4.81 and a scale parameter = 64.12. Histograms of consumption and contamination data are given in Figs. 1, 2 and 3.

Table 3
Descriptive statistics of the contamination data ($\mu g\ kg^{-1}$) for each food

Food	n	Mean	Standard deviation	Minimum	Maximum
Cereals	183	0.65	0.98	0.20	7.20
Raisins	13	0.71	1.16	0.10	4.30
Other dried fruit	33	0.28	0.29	0.10	1.60
Pork	1011	0.30	0.29	0.10	6.10
Poultry	93	0.075	0.0342	0.00	0.18
Fruit juices	19	0.25	0.78	0.02	3.45
Wines	104	0.16	0.26	0.00	1.64
Coffee	155	1.12	1.70	0.02	3.10

Several methods and tools are needed to carry out this SA. First, we present the MP–P method used for estimating an exposure distribution and the quantiles that we are interested in. We refer the reader to previous work (Gauchi, in press) for more detailed information about this method. Second, for the SA, we expose the four-step methodology used to attain our objectives: (i) postulation of a polynomial regression model; (ii) construction of the appropriate experimental design; (iii) determination of the factor ranges; (iv) validation of the model.

2.1. Construction of the exposure distribution

2.1.1. Definition of exposure

Determining OA exposure distribution is rather complicated because it is based on the sum of eight elementary exposure distributions (one per type of food analysed), and each one of these is the product of a consumption distribution and a contamination distribution. Let us give a rigorous definition of the exposure in our context. Consider the consumption C_j of one food j, divided by the individual weight, i.e. a normalized consumption, and the contamination T_j in OA of this food. Then, for p foods consumed per day, we define the normalized global exposure E as the sum of the p normalized products consumption×contamination: $E = \sum_{j=1}^{p} C_j T_j$, expressed in nanograms per kg of body weight per day, i.e. in $ng \times bw^{-1} \times day^{-1}$ units.

2.1.2. The MP–P method

The MP–P method, described in detail previously (Gauchi, in press), is essentially parametric in the sense that a mixed pdf is adjusted on each food consumption and a parametric pdf is adjusted on each food contamination. For consumption, apart from cereals, the typical histogram that can be plotted for each foodstuff is shown in Fig. 4, where two zones appear corresponding to the two sub-classes: non-consumers (sub-class of proportion h, represented by

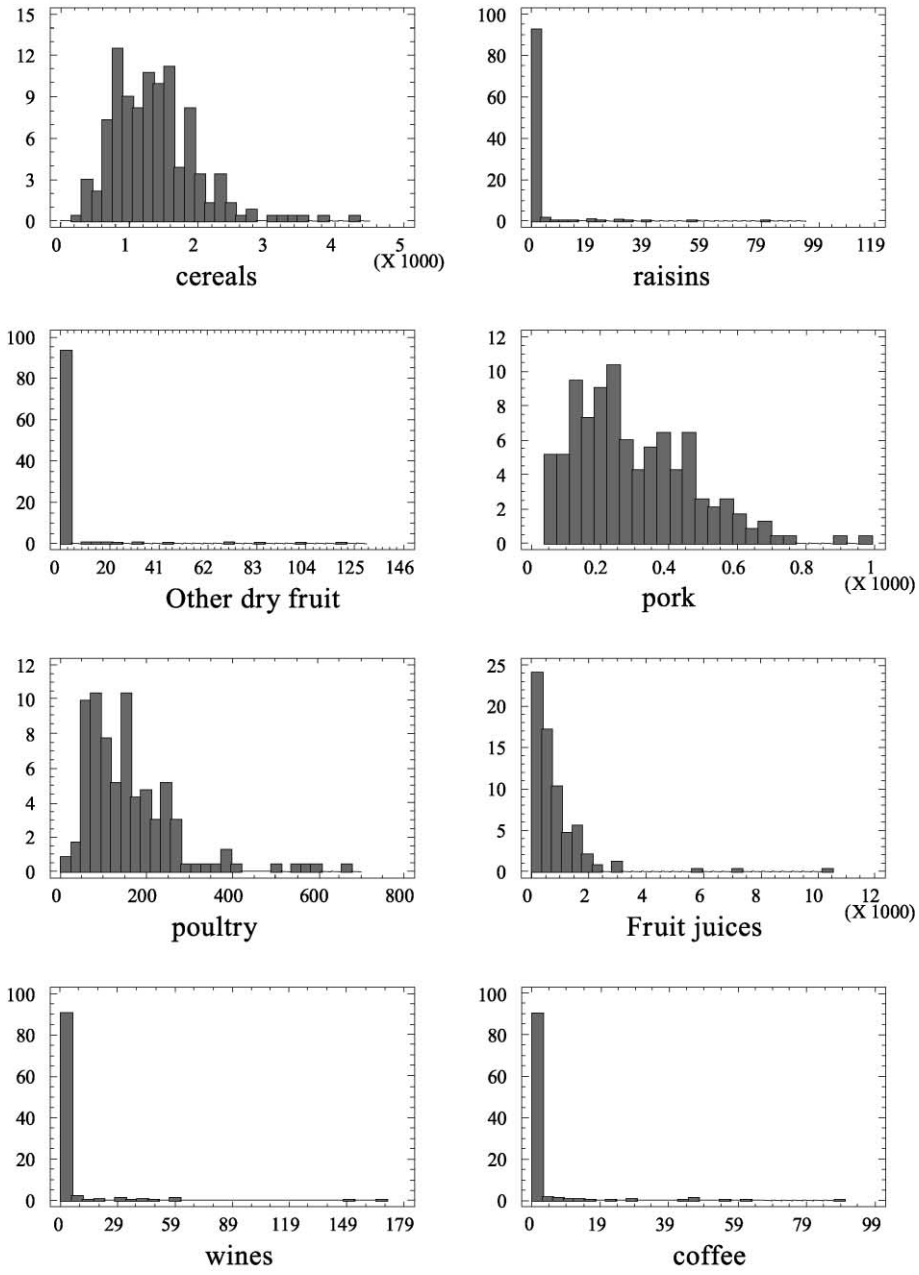

Fig. 1. Consumption histograms (relative frequencies) for children studied (French children).

the tall bar on the left) and consumers (sub-class of proportion $(1-h)$, represented by the asymmetrical part). In this situation, we propose for each consumed foodstuff j to fit a mixed distribution defined by a continuous uniform distribution for the non-consumers and a Gamma distribution for the consumers. An example of such a mixed pdf has been given by Vose (1996). The Gamma distribution has been validated by means of the Anderson–Darling statistic (D'Agostino and Stephens, 1986) among several asymmetrical

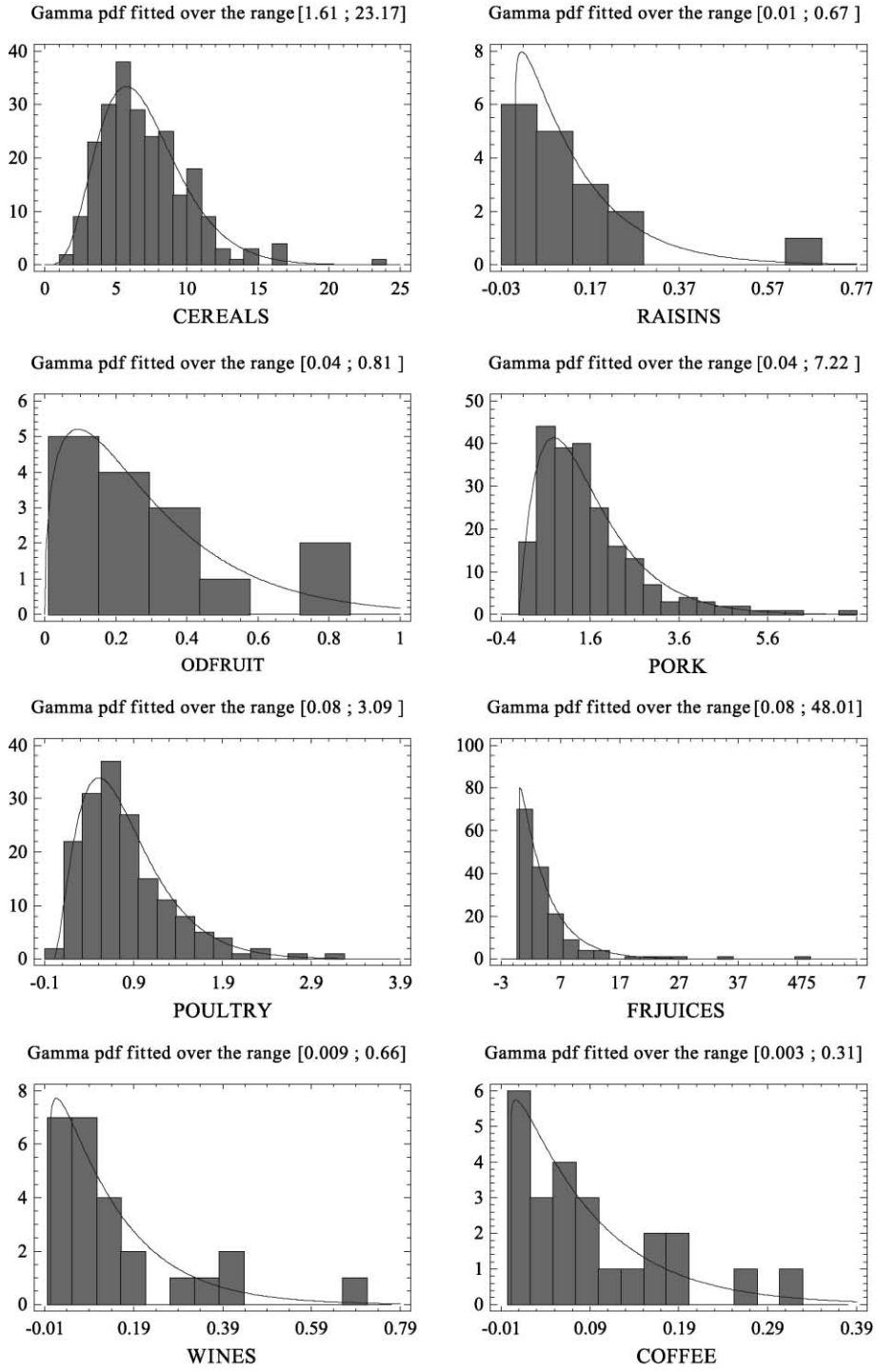

Fig. 2. Normalized consumption histograms (relative frequencies) and fitted Gamma pdf for children consumers (French children).

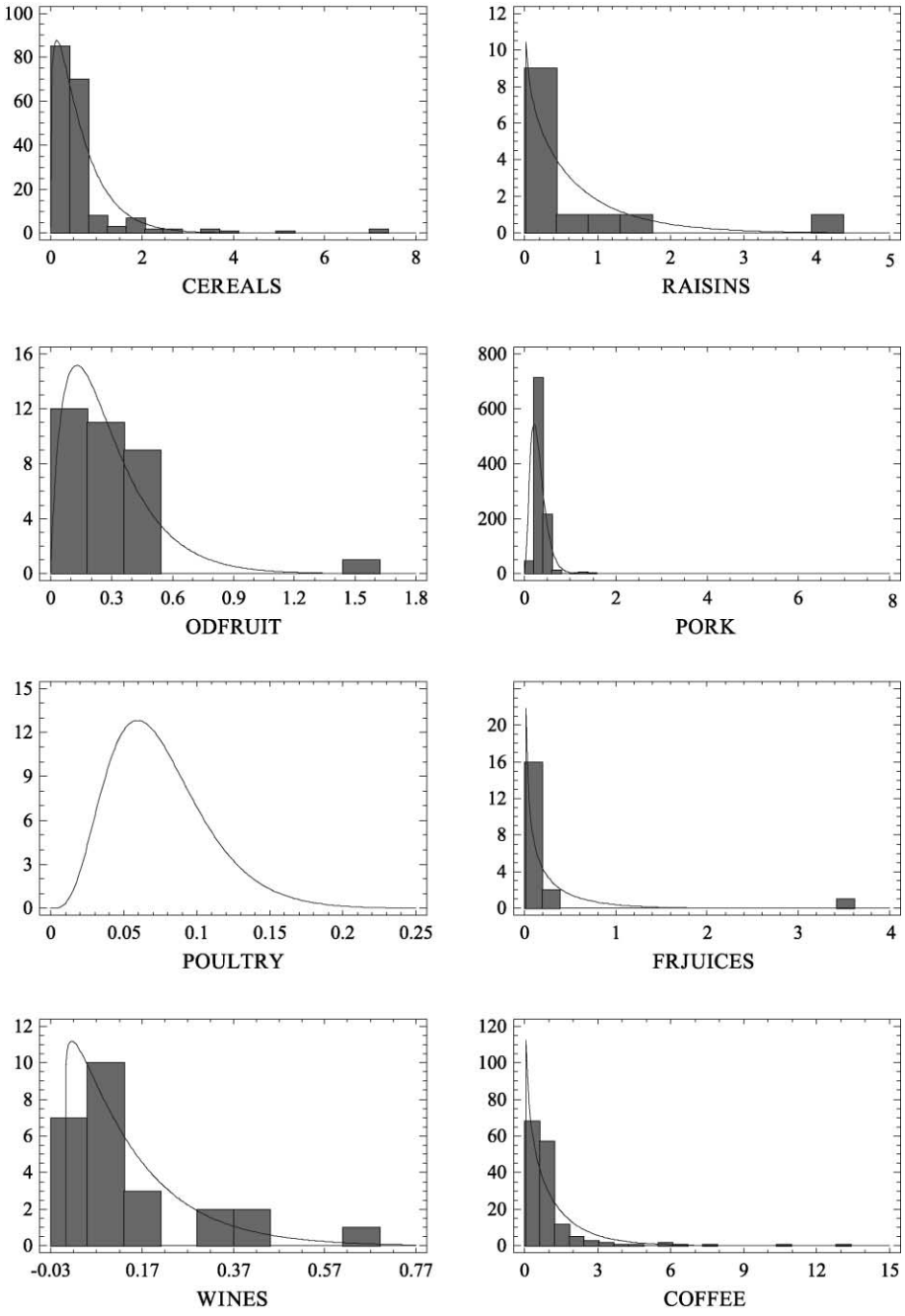

Fig. 3. Contamination histograms (relative frequencies) and the fitted Gamma pdf.

distributions. In the same way, this statistic has been used to decide Gamma distributions for contamination data.

The Gamma pdf for a continuous random variable X defined in $[0, +\infty>]$ $\left[\text{is: } f(x,r,\lambda) = [1/\Gamma(r)]\lambda^{-1}(\lambda^{-1}(x-\theta))^{r-1}e^{-\frac{(x-\theta)}{\lambda}}\right]$, where r, λ, θ are the shape,

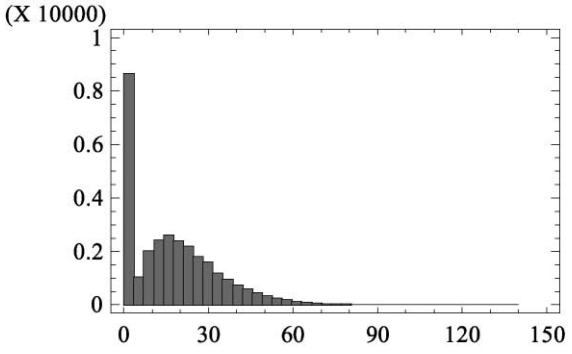

Fig. 4. Typical consumption histogram.

scale and threshold parameters, respectively, and $\Gamma(r)$ is the usual Euler's integral. The mean and variance of the Gamma distribution are connected to r and λ in the following way:

$$E(X) = r\lambda \text{ and } V(X) = r\lambda^2. \quad (1)$$

The estimated parameters of the corresponding pdf are given in Table 4. The fitted Gamma distributions are presented in the histograms Figs. 2 and 3.

An exposure value is calculated as follows:

$$\hat{E}_i^{(S)} = \sum_{j=1}^{p} \tilde{c}_{i,j}\tilde{t}_{i,j}, \quad (2)$$

where:

- $\tilde{c}_{i,j}$ is a random normalized consumption for the foodstuff j, drawn from the corresponding cumulative density function (cdf) of the above-mentioned mixed distribution, whose density parameters are given in Table 4;
- $\tilde{t}_{i,j}$ is a random contamination for the foodstuff j, drawn from the fitted Gamma cdf, whose density parameters are given in Table 4.

Then, with $N = 10{,}000$ outputs of form (2), we obtain an exposure simulation set S. The eight vectors of the N consumption random deviates have been arranged by means of the Iman and Conover (1982) method to be correlated with the aim of respecting the consumption dependencies. The N outputs lead to calculate various usual statistics and notably the 95th and 99th quantiles. Fig. 5 shows the simulated output exposure histogram.

2.1.3. Calculation of quantiles

The αth theoretical exposure quantile is here referred to as Q_α and defined by: $F_E(Q_\alpha) = \alpha$, where F_E is the theoretical cdf of the exposure of the population studied. The αth empirical exposure quantile $\hat{Q}_\alpha^{(S)}$ is defined by:

$$\#[\hat{E}_i^{(S)} \le \hat{Q}_\alpha^{(S)}]/N = \alpha, \quad (3)$$

where the notation $\#[x_i \le K]$ means the number of x_i less than or equal to K. For example, if $N = 10{,}000$, the estimate of the 0.95th quantile is the 9500th largest value of all the $\hat{E}_i^{(S)}$. If αN, is not an integer, the

Table 4
Parameters of the Gamma pdf fitted to the normalized consumption data (only to those who consume) and parameters of the Gamma pdf fitted to the contamination data

Food	Parameters of the fitted Gamma pdf	
	Consumption	Contamination ($\theta \hat{=} 0$)
Cereals	$\hat{r} = 2.594$; $\hat{\lambda} = 0.480$; $\hat{\theta} = 1.610$	$\hat{r} = 0.439$; $\hat{\lambda} = 0.674$
Raisins	$\hat{r} = 0.701$; $\hat{\lambda} = 5.484$; $\hat{\theta} = 0.010$	$\hat{r} = 0.375$; $\hat{\lambda} = 0.530$
Other dried fruit	$\hat{r} = 0.582$; $\hat{\lambda} = 2.286$; $\hat{\theta} = 0.043$	$\hat{r} = 0.998$; $\hat{\lambda} = 3.503$
Pork	$\hat{r} = 1.714$; $\hat{\lambda} = 1.127$; $\hat{\theta} = 0.041$	$\hat{r} = 1.085$; $\hat{\lambda} = 3.591$
Poultry	$\hat{r} = 1.813$; $\hat{\lambda} = 2.499$; $\hat{\theta} = 0.080$	$\hat{r} = 4.810$; $\hat{\lambda} = 64.12$
Fruit juices	$\hat{r} = 0.939$; $\hat{\lambda} = 0.209$; $\hat{\theta} = 0.083$	$\hat{r} = 0.108$; $\hat{\lambda} = 0.423$
Wines	$\hat{r} = 0.764$; $\hat{\lambda} = 5.594$; $\hat{\theta} = 0.092$	$\hat{r} = 0.413$; $\hat{\lambda} = 2.404$
Coffee	$\hat{r} = 0.784$; $\hat{\lambda} = 8.961$; $\hat{\theta} = 0.033$	$\hat{r} = 0.431$; $\hat{\lambda} = 0.386$

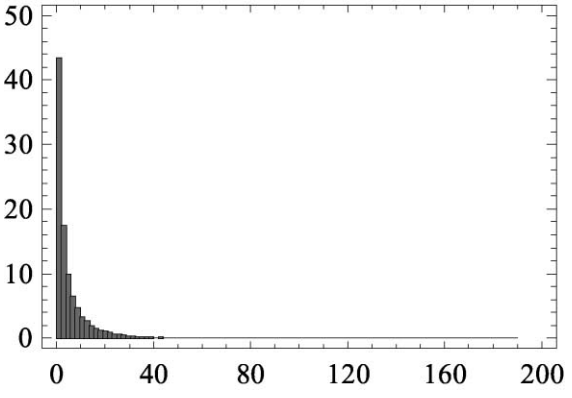

Fig. 5. Exposure output histogram (relative frequencies) with the MP–P method.

following procedure can be used. Assuming $\alpha \leq 0.5$, let $k = [(N+1)\alpha]$, the largest integer $\leq (N+1)\alpha$. Then, we define the empirical αth and $(1-\alpha)$th quantiles by the kth largest and $(N+1-k)$th largest values of all the $\hat{E}_i^{(S)}$, respectively. These quantiles are known as the empirical quantiles.

2.2. Methodology of the SA

The SA used is based on two main joint tools: a quadratic polynomial model and a simulation design which is optimal relative to this model. Our aim is to determine 95th and 99th quantile models for clearly expressing the influence of the factors chosen and, secondly, to have at our disposal a simulation tool to easily evaluate the amplitude of a quantile variation when new inputs are given.

2.2.1. Polynomial model

After some preliminary trials, we assumed a full quadratic polynomial regression model for the responses (outputs) studied. This regression model is based on 32 main factors, their 496 two-factor interactions and their 32 quadratic terms. The 32 factors are the pdf parameters: 16 parameters of the Gamma distributions relative to the eight consumed foods (each of these distributions is described by three parameters; the influence of the threshold parameter is not studied because it merely demarcates the lower position of the distribution of the consumers to consider the non-consumers), and 16 parameters of the two parameter Gamma distributions relative to the eight food OA contaminations. The general form of the model is:

$$y = \beta_0 + \sum_{j=1}^{32} \beta_j z_j + \sum_{j=1}^{31} \sum_{k=j+1}^{32} \beta_{jk} z_j z_k + \sum_{j=1}^{32} \beta_{jj} z_j^2 + \varepsilon, \quad (4)$$

where the β are unknown parameters, the factors z are the consumption and contamination pdf parameters, and ε is an error term for which no particular probability distribution is assumed and we only suppose its expectation is zero. Model (4) is a second degree (quadratic) polynomial in z. In the following, model (4) is successively applied to the 95th and 99th empirical quantiles, $\hat{Q}_{0.95}$ and $\hat{Q}_{0.99}$ respectively, defined by Eq. (3).

2.2.2. Experimental design

In order to estimate optimally the β parameters of model (4), an appropriate simulation design is chosen. As the three-factor interactions and over are assumed negligible, a full factorial design of 3^{32} experiments was not necessary, especially as it would have taken an extremely long time. Instead of this unfeasible design, we built a resolution-V fractional factorial design (Box et al., 1978) of 6561 experiments by means of the Factex procedure of the SAS software (SAS/QC® Institute). A resolution-V design allows us to estimate independently the parameters of the main effects of the factors (the z_j), the parameters of the 496 two-factor interactions (the $z_j z_k$), and the parameters of the 32 quadratic effects (the z_j^2) assuming that three-factor and over interactions are negligible. The factor levels were coded as -1, 0 and $+1$.

2.2.3. Factor ranges

The z factors, i.e. the parameters of the consumption and contamination pdf, are the inputs of the SA. The aim is to make the inputs vary in a reasonable range but large enough to observe how the response outputs vary. Because we had only one data set for consumption and contamination, no hypothesis of the factor variations could be made, so no supplementary surveys or contamination analyses were available. Thus, to simulate a meaningful range for each factor, we decided to determine it by a parametric bootstrap approach (Efron and Tibshirani, 1993). We considered the correlation between the shape and scale parameter of the pdf. For each consumption and contamination sample (16 in all), 10,000 bootstrap samples were obtained. A parametric bootstrap sample is obtained by sampling n times [n is the size of the data sample; see Table 2 (consumers) and Table 3 for the values of n] in the distributions fitted to the data. Table 4 gives the parameters of those distributions. From the mean and the standard deviation of each bootstrap sample, we calculated the scale and shape parameters from relation (1). As the bootstrap means were biased due to the sparseness of the data (see histograms Figs. 1–3), all the bootstrap scale and shape parameters were re-centred on the fitted param-

eter values given in Table 4. As the scatter plot of the 10,000 parameter pairs (scale and shape) showed an elliptical appearance for each food consumption and food contamination, we determined a variation domain by calculating a 95% concentration ellipse, one in each zone. In each ellipse, three squared variation domains were designed for the shape and scale parameters: the first one in the lower part of the ellipse (zone I), the second one in the middle part (zone II) and the third one in the upper part (zone III). Thus, in this approach, three polynomial models were determined for the 95th and 99th quantiles. Fig. 6 gives an example of the three zones for the consumption of cereals. For example, Tables 5, 6 and 7 indicate the three factor levels −1, 0 and +1 for the lower, middle and upper zones, respectively.

2.2.4. Model validation

To check whether the estimated regression model was a valid approximation, we proceeded as follows:

2.2.4.1. Calculation of the adjusted squared multiple correlation coefficient.

$$R^2_{\text{adj}} = 1 - (1 - R^2)\frac{(n-1)}{(n-p)},$$

where R is the usual multiple correlation coefficient, n is the number of observations (here $n = 6561$) and p is the number of parameters in the model (in Eq. (4), $p = 1+32+496+32 = 561$).

2.2.4.2. Examination of the residuals.

We considered the DFFITS$_i$ statistic defined as:

$$\text{DFFITS}_i = \frac{(\hat{y}_i - \hat{y}_{(i)})}{s_{(i)}\sqrt{h_i}},$$

where $\hat{y}_i$ is the prediction for the ith observation and $\hat{y}_{(i)}$ is the ith prediction but without the ith observation, $s_{(i)}$ is the estimated standard deviation after deleting the ith

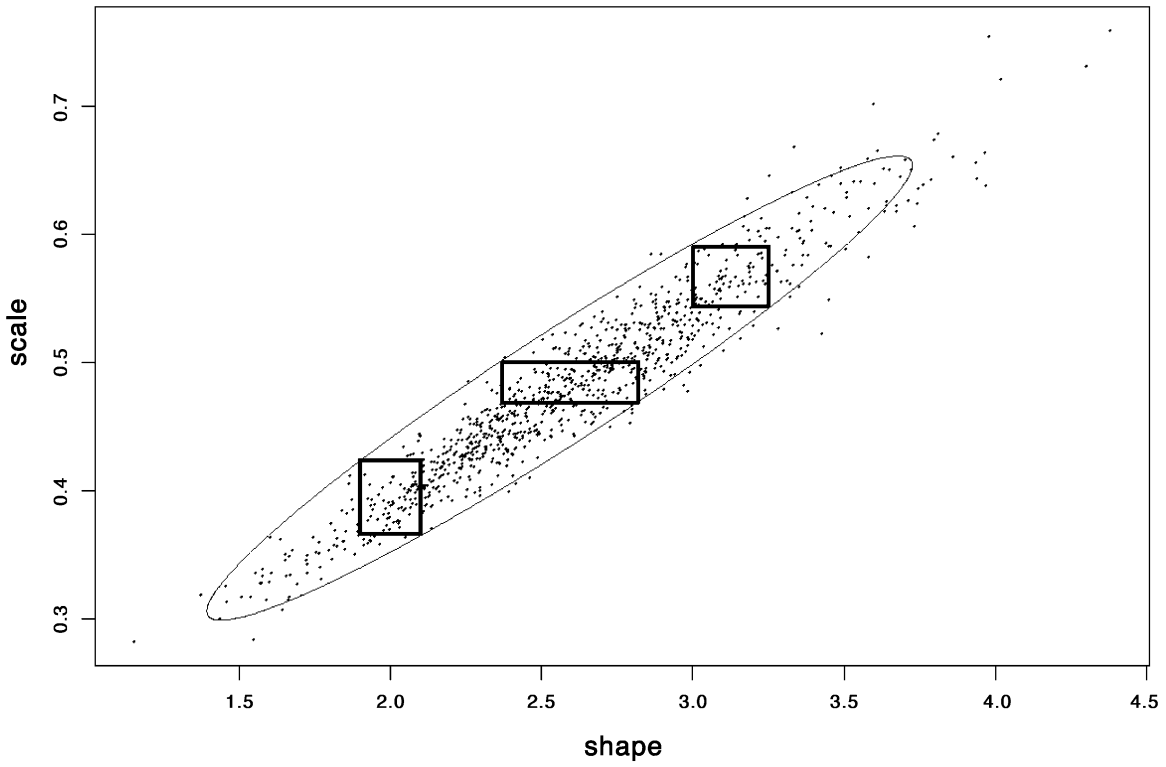

Fig. 6. Scatter plot of the shape and scale parameters for cereals obtained by bootstrap showing the three experimental domains.

Table 5
Levels of the inputs of the sensitivity analysis for the lower part (zone I)

Food		Factor range					
		Consumption			Contamination		
		−1	0	1	−1	0	1
Cereals	r	2.148	2.215	2.282	0.257	0.292	0.327
	λ	0.395	0.404	0.412	0.306	0.359	0.412
Raisins	r	0.0524	0.165	0.278	0.2822	0.375	0.468
	λ	0.492	1.574	2.656	0.3055	0.5305	0.755
Other dried fruit	r	0.1691	0.328	0.487	0.402	0.582	0.762
	λ	0.500	1.116	1.732	1.189	1.685	2.180
Pork	r	1.381	1.455	1.529	0.547	0.577	0.607
	λ	0.855	0.930	1.005	1.722	1.808	1.894
Poultry	r	1.470	1.535	1.600	3.705	3.891	4.077
	λ	1.859	2.050	2.241	46.784	51.17	55.56
Fruit juices	r	0.5618	0.6466	0.7314	0.0693	0.1078	0.1464
	λ	0.1051	0.1254	0.1457	0.2486	0.4233	0.598
Wines	r	0.323	0.485	0.647	0.294	0.3161	0.338
	λ	0.777	2.525	4.273	1.416	1.753	2.090
Coffee	r	0.342	0.508	0.673	0.271	0.2918	0.3125
	λ	4.122	5.888	7.654	0.1796	0.1928	0.2059

observation, and h_i is the ith diagonal of the projection matrix for the predictor space (see Belsley et al., 1980 for more details). As the distribution of the model residuals shows an approximately normal distribution, a possible threshold is the size-adjusted cut-off recommended by Belsley et al. (1980) equal to $2\sqrt{p/n}$. Therefore, a value of DFFITS_i above this cut-off indicates an influential residual e_i. The DFFITS_i was calculated by SAS software (Proc REG).

Table 6
Levels of the inputs of the sensitivity analysis for the middle part (zone II)

Food		Factor range					
		Consumption			Contamination		
		−1	0	1	−1	0	1
Cereals	r	2.3690	2.5944	2.8198	0.3987	0.4395	0.4802
	λ	0.4688	0.4796	0.4904	0.6210	0.6740	0.7260
Raisins	r	0.5122	0.7012	0.8902	0.2822	0.3750	0.4680
	λ	3.6110	5.4840	7.3570	0.3055	0.5305	0.7550
Other dried fruit	r	0.4042	0.5818	0.7594	0.8160	0.9980	1.1790
	λ	1.5492	2.2862	3.0230	3.0400	3.5020	3.9640
Pork	r	1.6320	1.7140	1.7950	1.0240	1.0850	1.1450
	λ	0.5635	1.1269	1.6903	3.5477	3.5907	3.6340
Poultry	r	1.7590	1.8310	1.9030	4.6150	4.8100	5.0000
	λ	2.2890	2.4950	2.7020	59.0400	64.1200	69.2000
Fruit juices	r	0.8450	0.9388	1.0330	0.0693	0.1078	0.1464
	λ	0.1862	0.2087	0.2312	0.2486	0.4233	0.5980
Wines	r	0.5895	0.7645	0.9390	0.3657	0.4127	0.4596
	λ	3.5740	5.5940	7.6140	1.7800	2.4040	3.0300
Coffee	r	0.5862	0.7842	0.9820	0.3861	0.4310	0.4760
	λ	7.0300	8.9610	10.8920	0.3421	0.3860	0.4300

Table 7
Levels of the inputs of the sensitivity analysis for the upper part (zone III)

Food		Factor range					
		Consumption			Contamination		
		−1	0	1	−1	0	1
Cereals	r	2.8838	2.9738	3.0637	0.552	0.587	0.622
	λ	0.544	0.5552	0.566	0.9363	0.9893	1.0425
Raisins	r	1.124	1.237	1.350	1.305	1.404	1.502
	λ	8.312	9.394	10.476	3.095	3.339	3.583
Other dried fruit	r	0.676	0.835	0.994	1.233	1.413	1.593
	λ	2.84	3.456	4.072	4.825	5.3202	5.815
Pork	r	1.899	1.973	2.047	1.5625	1.5925	1.6225
	λ	1.2488	1.3241	1.399	5.287	5.373	5.459
Poultry	r	2.062	2.127	2.192	5.543	5.729	5.915
	λ	2.749	2.940	3.131	72.68	77.07	81.44
Fruit juices	r	1.146	1.231	1.316	0.238	0.2689	0.299
	λ	0.272	0.292	0.312	0.997	1.141	1.285
Wines	r	0.882	1.044	1.206	0.487	0.509	0.531
	λ	6.915	8.663	10.41	2.718	3.055	3.392
Coffee	r	0.895	1.060	1.225	0.549	0.570	0.591
	λ	10.26	12.03	13.796	0.565	0.578	0.591

2.2.4.3. Test experiments. A final test was performed to evaluate the model validation: the comparison between the predicted outputs for K, typically 10,000, new input combinations randomly sampled in the three squared experimental domains and the corresponding K simulated outputs. Then, we check if the K residuals are Gaussian and if their variation ranges are reasonable. We calculate the percentages δe_i:

$$\delta e_i = \frac{e_i}{y_i} \times 100 = \frac{\hat{y}_i - y_i}{y_i} \times 100, \qquad (5)$$

where y_i is the observation and $\hat{y}_i$ the value predicted by the model to see the percentages of discrepancy between predicted and observed values for the test experiments.

3. Results

We fitted the model (4) to the 6561 experiments for the two responses $\hat{Q}_{0.95}$ and $\hat{Q}_{0.99}$ in the three zones as described in Section 2.2.3, after the factors z_j had been

Table 8
R^2_{adj} statistic for main models (95th quantile and zone II)

Model	Number of parameters	R^2_{adj}
Model 1: $y = \beta_0 + \sum_{j=1}^{32} \beta_j z_j + \sum_{j=1}^{31} \sum_{k=j+1}^{32} \beta_{jk} z_j z_k + \sum_{j=1}^{32} \beta_{jj} Z_j^2 + \varepsilon$	560	0.9715
Model 2: $y = \beta_0 + \sum_{j=1}^{32} \beta_j z_j + \sum_{j=1}^{32} \beta_{jj} z_j^2 + \varepsilon$	64	0.9682
Model 3: $y = \beta_0 + \sum_{j=1}^{32} \beta_j z_j + \sum_{j=1}^{31} \sum_{k=j+1}^{32} \beta_{jk} z_j z_k + \varepsilon$	528	0.9695
Model 4: $y = \beta_0 + \sum_{j=1}^{32} \beta_j z_j + \varepsilon$	32	0.9664

appropriately coded (use of orthogonal polynomials and scaling). Table 8 gives the values of the $R_{adj}{}^2$ of the four main models in the middle part for the 95th quantile. According to these $R_{adj}{}^2$, the model with only the main effects was chosen in the first step (model 4 in Table 8). Indeed, the model has just 32 terms and its value of $R_{adj}{}^2$ is still very high ($R_{adj}{}^2 = 0.97$). The results being equivalent for the 99th quantile and the $R_{adj}{}^2$ values being slightly smaller (about 3% less), the model with only the 32 main effects was also chosen for the 99th quantile. For both quantiles, the results were similar in the other two zones.

In the second step, with a view to parsimony, for each zone, variables were selected considering their explained sum of squares. The variables with the higher sum of squares were kept in the final model. Table 9 gives the final models with their parameter estimations for the 95th and 99th quantiles and for the three zones. The model for the 95th quantile includes only four terms (the main effects) in the lower and upper zones and five terms in the middle zone. The results are similar for the 99th quantile with a slightly smaller percentage of explained variability for each model. The results show a major influence of the parameters of the distributions on cereals (contamination and consumption). In the middle part, a slight influence of the pork consumption scale parameter appears. We noted that both the orthogonality of the simulation designs and the above-mentioned coding enable us to compare rigorously the influences of all the food distribution parameters. We obtained the marginal effect of each parameter and all of the parameters could be compared with each other because they have the same scale. On the other hand, the constant terms of the models of Table 9 represent the quantile means inside the cuboidal zones: for the 95th quantile they are 28.1, 20.1 and 16.7, and for the

Table 9
Polynomial models of the 95th and 99th quantiles in the three zones of the 95% concentration ellipse with z_1 = cereals contamination scale parameter, z_2 = cereals contamination shape parameter, z_3 = cereals consumption shape parameter, z_4 = cereals consumption scale parameter, and z_5 = pork consumption scale parameter

95% concentration elliptical zone	Percentile	Final model	R^2
I (Lower part)	95th	$\hat{y} = 28.1 - 8.2z_1 + 5.4z_2 + 1.2z_3 - z_4$	0.97
	99th	$\hat{y} = 61.5 - 18.2z_1 + 7.3z_2 + 2.7z_3 - 2.2z_4$	0.96
II (Middle part)	95th	$\hat{y} = 20.1 - 3.1z_1 + 2.6z_2 + 2.6z_3 - 0.7z_4 - 0.5z_5$	0.97
	99th	$\hat{y} = 39.1 - 6z_1 + 3.6z_2 + 4.9z_3 - 1.5z_4 - 0.5z_5$	0.94
III (Upper part)	95th	$\hat{y} = 16.7 - 1.7z_1 + 1.3z_2 + 0.7z_3 - 0.5z_4$	0.94
	99th	$\hat{y} = 30.4 - 3.2z_1 + 1.9z_2 + 1.3z_3 - z_4$	0.89

99th quantile they are 61.5, 39.1 and 30.4 according to the upper, middle and lower zone, respectively. We can compare these values to the value given in Gauchi (in press): for the 95th and 99th quantiles, the following 95% confidence intervals were calculated [14.07; 29.56] and [28.08; 76.92], respectively. We noted that the quantile means obtained with this SA lie inside the 95% confidence intervals. Thus, we assume that the SA outputs are reasonable values.

The distributions of the residuals of the different models did not indicate outliers. Only a few DFFITS measures slightly exceeded the threshold as defined in Section 2.2.4.2. Ten thousand new combinations of inputs have been randomly sampled inside the three experimental domains. In the three zones, the distributions of the residuals of each model are approximately Gaussian and have moderate variation ranges. For example, in the middle part for the 95th quantile, the estimated standard deviation of the model residuals is estimated at 2.92 (for the 10,000 test experiments). It is higher than the estimated standard deviation (0.37) of the residuals on the 6561 experiments from which the model has been chosen, but their variation range is reasonable because 95% of these residuals have percentage, as defined in Eq. (5), between $\pm 28\%$. The discrepancy between the predicted and observed values for the test experiments is acceptable. Fig. 7 represents the histogram of the 10,000 residuals for the 95th quantile in the middle zone.

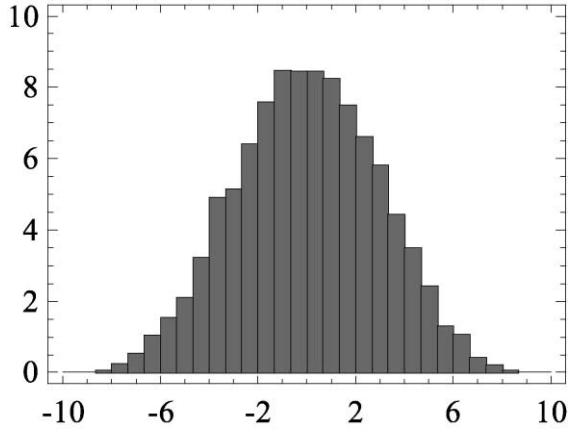

Fig. 7. Histogram of the 10,000 residuals for the 95th quantile for the 10,000 new observations in the middle part.

4. Discussion

On the basis of the observation, well known to risk analysts, that to obtain a good estimation of the high quantiles of exposure is a difficult operation when it is based on data which are both sparse and characterised by very asymmetrical distributions, we undertook a sensitivity analysis of the 95th and 99th quantiles of exposure to OA. We conducted an analysis of their sensitivity to probability density parameters adjusted to consumption and contamination data. That is, we measured the influence of the shapes of the consumption and contamination distributions of the studied foods to highlight those foods that possibly require particular investigation in order to better estimate the high exposure quantiles.

One of the original ideas of this work was to generate a realistic variability of data available by a parametric bootstrap method. This enabled us to construct variation intervals of the density parameters concerned, intervals presumed to be representative of the child population. Another original idea was to take into account the natural correlation of the parameters of the Gamma laws using concentration ellipses. However, we have not explored all the variation domains of the inputs: we studied the sensitivity of the 95th and 99th quantiles in three subdomains of cuboidal symmetry distributed judiciously in the variation domain (lower, middle and upper zone). The advantage of these subdomains is that they make it possible to set up orthogonal experimental designs and to take into account a large variation of the quantiles (see the constant terms of the models in the three zones). On these subdomains, it appears very clearly that a small number of parameters have a great influence on estimating these quantiles compared with the very large number of parameters envisaged at the beginning. These include mainly the parameters of the laws of consumption and contamination of the category cereals. Cereals are known by epidemiologists and toxicologists to play a major role in health problems concerning OA in Europe.

It is difficult to take into account all the volume of the variation domain of the inputs (for example, because of the non-convexity of this domain). The necessary tools are derived from the theory of optimum experimental designs. This could possibly improve our work. However, it would be appropriate beforehand to

guarantee the pertinence of our domain of variation (obtained by bootstrap sampling) ideally with data from further surveys.

Finally, it is difficult to compare our results with other results because, to our knowledge, no other research group has conducted a study on contamination by OA. The other estimations of the OA exposure quantiles found elsewhere are very empirical as indicated previously in Gauchi (in press) and the sensitivity analysis presented here is the first SA on OA exposure quantiles.

Acknowledgements

We thank P. Verger and J.C. Leblanc (INRA, Scientific Directorate for Human Nutrition and Food Safety, Paris) for their important preliminary data handling and providing the consumption survey data and the contamination raw data.

References

ASPCC, 1999. Observatoire des Consommations Alimentaires: Etude des distributions statistiques des consommations de huit groupes d'aliments susceptibles de contenir de l'Ochratoxine A. Internal Technical Report of CREDOC, for INRA, based on data from ASPCC survey, April 1999.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics John Wiley & Sons, New York.

Box, G.E.P., Hunter, W.G., Hunter, J.S., 1978. Statistics For Experimenters.

Codex Committee on Food Additives and Contaminants (CCFAC), 1999. Position paper on ochratoxin A. Codex Committee on Food Additives and Contaminants, Thirty-first Session, The Hague, The Netherlands, 22–26 March 1999.

D'Agostino, R.B., Stephens, M.A., 1986. Goodness-of-Fit Techniques. Marcel Dekker, New York.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap Chapman & Hall, London.

Gauchi, J.P., 2002. Quantitative assessment of exposure to the mycotoxin ochratoxin A in food. Risk Anal., in press.

Iman, R.L., Conover, W.J., 1982. A distribution-free approach to inducing rank correlation among input variables. Commun. Stat., Simul. Comput. 11, 311–334.

JECFA (Joint FAO/WHO Expert Committee on Food Additives), 1995. Ochratoxin A Report: TRS 859-JECFA 44/35 Tox monograph: FAS 35-JECFA 44/363. World Health Organization, Geneva.

Pfohl-Leszkowick, A., 1999. Les Mycotoxines dans L'alimentation: Évaluation et Gestion du Risque. Tec&Doc Editor, Paris, France Chap. 3.

SCOOP task 3.2.7. Report of the European Communauty, French contribution, in press. Origin of data: Direction Générale de l'Alimentation (Ministère de l'Agriculture et de la Pêche), Direction Gènèrale de la Concurrence, de la Consommation et de la Répression des Fraudes (Ministère de l'Economie, des Finances et de l'Industrie).

Verger, P., 1999. Exposure assessment to certain contaminants, working group «Contaminants». Internal Technical Report, March 1999. INRA, Paris.

Vose, D., 1996. Quantitative Risk Analysis—A Guide to Monte Carlo Simulation Modelling Wiley, Chichester, UK.