Electronic Journal of Applied Statistical Analysis
EJASA (2011), Electron. J. App. Stat. Anal., Vol. 4, Issue 2, 255 – 264
e-ISSN 2070-5948, DOI 10.1285/i20705948v4n2p255

GLOBAL SENSITIVITY ANALYSIS APPLIED TO FOOD SAFETY RISK ASSESSMENT

Jean-Christophe Augustin*

Ecole Nationale Vétérinaire d'Alfort, France

Received 02 October 2010; Accepted 08 July 2011
Available online 14 October 2011

***Abstract**: This study proposes an example of the implementation of the global sensitivity analysis developed by Saltelli in the frame of the contamination assessment of Listeria monocytogenes in smoked salmon at consumption. This method allowed identifying subsets of relevant parameters among the input variables of the model characterized by a great variability and/or uncertainty. This example illustrates the significance to perform these studies for food safety risk assessment models in order to identify the most influential parameters justifying management options, to identify parameters that need to be further studied, and to simplify models when complexity is not justified.*

***Keywords**: Sensitivity analysis, microbial risk assessment, Listeria monocytogenes*

1. Introduction

Sensitivity analysis (SA) aims to study how the variation in the output of a model can be apportioned, qualitatively or quantitatively, to different sources of variation in the model input [16]. All the scientific fields are concerned with this approach as soon as a modeling process is involved. The purposes of SA can be grouped into three categories: i) assessment of the quality and of the relevance of a model, ii) identification of the key factors to establish research priorities, to identify regions for which the model output variation is maximum, or to identify factors interacting with each other, iii) identification of factors or assumptions having little impact on the output variation to simplify the model.

* E-mail: jcaugustin@vet-alfort.fr

Many methods were developed to perform SA. Three types of methods were identified [16]: i) screening methods generally used to identify subset of important factors among hundreds of input factors. The most famous methods in this field are one-factor-at-a-time (OAT) experiments, iterated fractional factorial design or sequential bifurcation methods, ii) local SA using partial derivatives, and iii) global SA.

Global SA focuses on the output variation over the entire range of variation of the input factors. One important property of global methods is their capacity to estimate the sensitivity to individual factors while all other are varied as well. Four kinds of methods were developed to perform global SA: i) scatter plots, ii) regression analysis, input-output correlation (linear or rank correlation and partial correlation coefficients), iii) ANOVA and response surface method, iv) variance-based methods (Fourier Amplitude Sensitivity Test – FAST, Sobol' method, extended FAST, Saltelli method). Scatter plots are non-quantitative graphical tools, regression analysis and input-output correlation methods rely on the assumptions that the output and input factors are linearly or monotonically related, in ANOVA, it is assumed that the output is normally distributed and the response surface method is not adapted to models involving many input factors. The SA methods based on the conditional variances and then on the variance decomposition allow quantifying the effect of the variability of the input variables and of their interactions on the variability of the output variables without any assumptions on the model structure. They were first developed in the 70s with the work of Cukier and collaborators [2, 3] who proposed the FAST method allowing to calculate indices reflecting the sensitivity of an output variable to the input variables (first order indices). The Russian mathematician Sobol' proved the variance decomposition of the output variable as the sum of conditional variances linked to each input variables and to their interactions and proposed a calculation method of all the indices based on the Monte Carlo method [17]. Saltelli and collaborators [7, 15] proposed the concept of total indices more adapted to the study of models involving many input variables and developed simple methods allowing the estimation of the sensitivity indices. More recently, moment-independent methods based on uncertainty importance measures that look the entire output distribution without referring to one of its moments were proposed [1].

Quantitative microbiological risk assessment (QMRA) is increasingly used in food safety to numerically estimate the probability and the severity of adverse health effects resulting from the exposure to microbiological foodborne hazards. These studies are developed in the framework of risk analysis which has emerged over the past decade as the internationally recognized model for improving food control systems with the objectives of producing safer food, reducing the number of foodborne illnesses and facilitating the international trade of foods. QMRA should help risk managers to implement appropriate management options. The QMRA models generally involve many steps depending on the length of the food chain under consideration. For instance, processes like microbiological contamination of raw food ingredients, killing of microorganisms during processing such as pasteurization, multiplication of microorganisms during storage steps must be modeled and generally numerous modeling assumptions are required to describe these phenomena. Furthermore, the input factors are mostly biological ones and are characterized by a great natural variability as well as a great uncertainty by lack of extensive knowledge about these biological factors. For example, the variability of the minimal temperature allowing the cell multiplication of a bacterial species can be very large depending on the strains under consideration and the accurate description of the distribution describing this biological variability requires the study of numerous strains. It must also be pointed out that in the field of QMRA the validation of modeling hypothesis or the accurate characterization of input factors often requires

time consuming, painful and expensive laboratory studies. SA has then naturally emerged relatively early in this field to determine the main risk-determining phenomena. The first paper referring to SA in the field of food safety risk assessment was then published in 2000 by Zwietering and van Gerwen [20] and the approach was illustrated by highlighting the steps influencing the survival and the growth of *Salmonella* during the processing of chicken meat. Methods initially used to perform SA in QMRA consisted essentially in OAT-like experiments [20] and input-output correlation, mostly Spearman rank correlation coefficients [4, 8, 9, 12, 20] and sometimes Pearson correlation coefficient [18]. In 2002 and 2004, Frey and Patil [6, 11] proposed extensive reviews of SA methods usable for QMRA models and they show that ANOVA was a valuable method. Since these publications, ANOVA is increasingly used in QMRA studies [10, 13, 14]. These methods are also frequently extended with "what-if" scenarios consisting in assessing the impact on risk estimates of management options concerning input factors set to specific values or range of values. SA is then essentially used in QMRA to identify key management options but model simplification, which is another important goal of SA, could also be useful for QMRA. This concern is topical in QMRA and, recently, Zwietering [19] worried about the increasing complexity of models and raised the issue of the usefulness of this complexity.

The aim of this study was to investigate the usefulness of global SA and more especially the method proposed by Saltelli to identify key risk management options and to simplify models used in QMRA that are often characterized by a strong non-additivity and nonlinearity. The Saltelli method was implemented to study a model describing the contamination of smoked salmon by *Listeria monocytogenes* at the moment of the consumption.

2. Materials and methods

The contamination assessment model used in this study is taken from [5]. Only the main aspects of this model will be exposed in this section with a first part concerning the principles of the Saltelli method to perform SA.

2.1 The Saltelli method

The Saltelli method is based on a numerical procedure for computing the full set of first order sensitivity indices, S_i, and total effect indices, St_i, for all the input factors X_i (input factors are assumed not correlated) of the model f with an output Y:

$$Y = f(X_1, X_2, ..., X_k) \tag{1}$$

The first order indices of the k input factors are:

$$S_i = \frac{V[E(Y \mid X_i)]}{V(Y)} \tag{2}$$

where:

$$V[E(Y \mid X_i)] = V(Y) - E[V(Y \mid X_i)] \tag{3}$$

The second element is the expectation of the conditional variance of Y knowing X_i, which will decrease with increasing impact of X_i on the variance of Y.
These first order indices are then included between 0 and 1 and quantify the effect of the variability of each input factor on the total variance of the output. For an additive model, the sum of first order indices is equal to 1.
The total effect indices of the k input factors are:

$$St_i = 1 - \frac{V[E(Y \mid X_{\sim i})]}{V(Y)} \quad (4)$$

where $V[E(Y \mid X_{\sim i})]$ is the total contribution of input factors different from X_i to the variance of Y. St_i quantifies the sum of the first order effect with the effects linked to the interactions of X_i with the other input factors.
The numerical procedure for computing these indices proposed by Saltelli consists in generating two matrices A and B of N lines (N simulation runs) and k columns (k studied factors) of pseudo-random numbers provided by a space-filling design. Latin hypercube sampling (LHS) was used in this study with respect to the range of variation of each input factors. Then k matrices C_i containing all the columns of B but the i^{th} column of A are generated. The model is then run for each row of the k+2 matrices providing k+2 output vectors of length N, y_A, y_B, and y_{Ci}. First order and total effect indices are then calculated with the following formulae:

$$S_i = \frac{y_A \cdot y_{Ci} - g_0}{y_A \cdot y_A - f_0^2} \quad (5)$$

$$St_i = 1 - \frac{y_B \cdot y_{Ci} - f_0^2}{y_A \cdot y_A - f_0^2} \quad (6)$$

with:

$$f_0 = \frac{1}{N}\sum_{u=1}^{N} y_A \text{ and } g_0 = \frac{1}{N}\sum_{u=1}^{N} y_A \cdot y_B \quad (7)$$

2.2 Sensitivity analysis of the model assessing the contamination of smoked salmon by Listeria monocytogenes

The contamination assessment model for *L. monocytogenes* in cold smoked vacuum packed salmon is fully described in [5]. The model evaluates the microbiological contamination of portions at the end of the food shelf life assuming an initial contamination of portions with few microbial cells. Briefly, the final contamination depends on the evolution of the initial contamination that is dependent on numerous factors like the biological characteristics of the microorganism, the physico-chemical characteristics of the salmon, the storage conditions during the distribution chain (Figure 1).

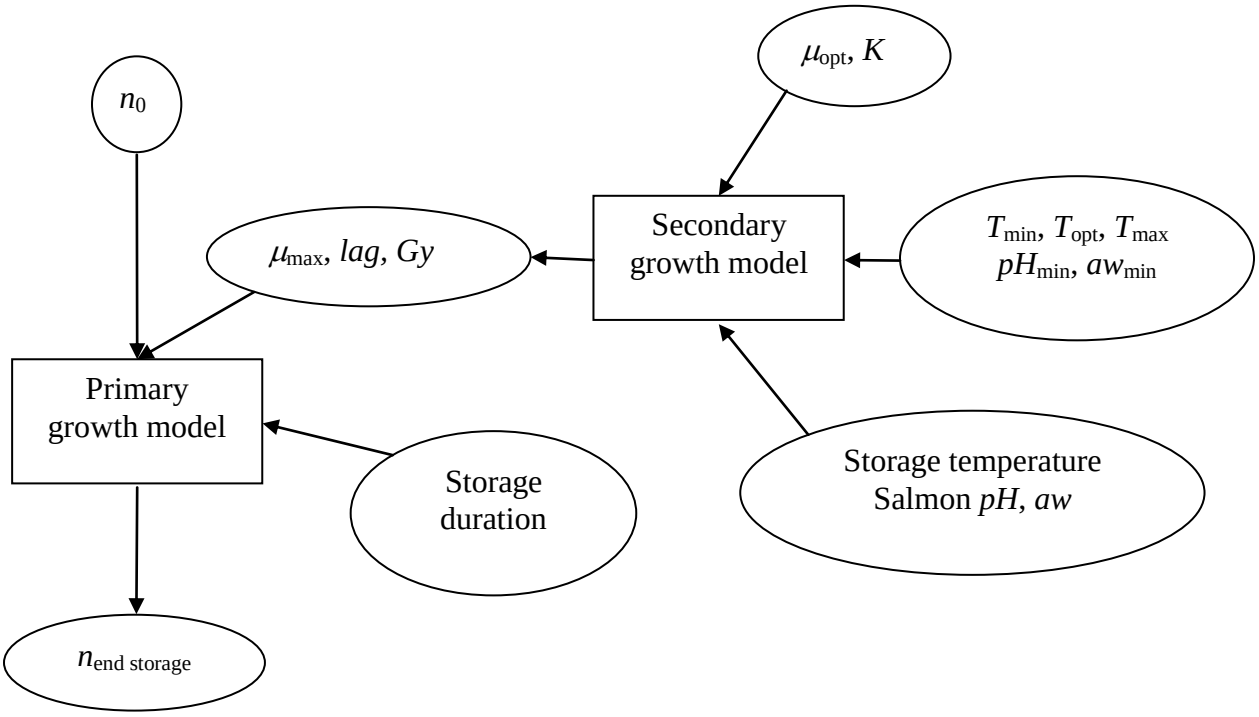

Figure 1. Flow diagram of the stochastic model used to estimate the *Listeria monocytogenes* contamination of salmon after a storage step of the chain distribution.

Predictive microbiology models allow describing the behavior of microorganisms as a function of these factors. Primary models describe the kinetics of bacterial populations as a function of primary parameters: lag time (initial delay before growth begins) and growth rate. Secondary models describe the evolution of primary parameters with environmental (temperature, food characteristics) and biological factors (cardinal values, initial physiological state of contaminating cells). All these factors are characterized by a great natural variability and/or uncertainty. The 23 microbiological, food, and distribution chain independent input factors of the model are presented in Table 1.

Two SA were performed, the first by assuming no knowledge about the variation distribution of input factors apart form their ranges of variation, the second by assuming informative probability density functions for input factors characterized by approximately the same range of variation as the uniform distributions used in the first case (Table 1). In the first case, the values for input factors in Saltelli matrices were obtained from the LHS by assuming uniform distributions defined on the ranges of variation. In the second case, the values were deduced from the LHS by inverting the cumulative distribution functions of input factors. $N = 5 \times 10^4$ simulation runs were performed and final bacterial density expressed in logarithm was calculated with the Matlab software (The MathWorks Inc., Natick, MA, USA). The simulations were performed with an Intel Core i7 2.66 GHz and the computation time was approximately 8 h for each case. Eq. (5), (6) and (7) were then used to calculate sensitivity indices.

Table 1. Identification of the input factors of the contamination assessment model and their distributions.

Input factors	Description	Range	Distribution[a]
Microbiological parameters			
K	Initial physiological state of *L. monocytogenes*	0 – 36	exp(N(0.76, 1.23))
n_0	Initial number of *L. monocytogenes* cells	1 – 20	P(10)
Gy (log)	Growth yield of the *L. monocytogenes* population	4 – 6	N(5, 0.43)
μ_{opt} (h^{-1})	Optimum specific growth rate of *L. monocytogenes* in cold smoked salmon	0.25 – 1.55	N(0.924, 0.286)
T_{min} (°C)	*L. monocytogenes* minimum temperature for growth	-2.8 – 0.6	N(-1.08, 0.72)
T_{opt} (°C)	*L. monocytogenes* optimum temperature for growth	36.4 – 40.0	N(38.2, 0.76)
T_{max} (°C)	*L. monocytogenes* maximum temperature for growth	40.5 – 46.1	N(43.3, 1.2)
pH_{min}	*L. monocytogenes* minimum pH for growth	3.9 – 4.5	N(4.19, 0.12)
aw_{min}	*L. monocytogenes* minimum water activity for growth	0.901 – 0.943	N(0.922, 0.009)
Food characteristics			
pH	Cold smoked salmon pH	5.6 – 6.1	N(5.84, 0.1)
aw	Cold smoked salmon water activity	0.935 – 0.985	N(0.960, 0.010)
Distribution chain parameters			
$dTEP$ (h)	Duration of transport and storage	0 –55	E(12)
$TTEP$ (°C)	Temperature of transport and storage	-2 – 7	N(2.5, 2)
dC (h)	Duration of storage in cold room	0 – 44	E(10)
TC (°C)	Temperature of storage in cold room	1.5 – 6.5	N(4, 1)
dM (h)	Duration of storage at retail	0 – 270	E(60)
TM (°C)	Temperature of storage at retail	2 – 9	N(5.5, 1.5)
dV (h)	Duration of the journey back home	0 – 2.3	E(0.5)
TV (°C)	Temperature of the journey back home	6 – 20	N(13, 3)
dR (h)	Duration of the storage in the refrigerator	0 – 414	E(90)
TR (°C)	Temperature of the storage in the refrigerator	2 – 12	N(7, 2)
dD (h)	Duration of unrefrigerated storage before consumption	0 – 24	E(5)
TD (°C)	Temperature of unrefrigerated storage before consumption	10 – 25	N(17.5, 3)

[a] $N(a,b)$ is the normal distribution with expected value a and standard deviation b, $P(a)$ is the Poisson distribution with expectation a, $E(a)$ is the exponential distribution with expectation a.

3. Results and discussion

The first order and total effect sensitivity indices are presented in Table 2. The difference between first order and total effect indices illustrates that most of input factors influenced the output variance in interaction with other factors. When assuming uniform distributions for input factors (case #1), the sum of first order indices was equal to 0.52, the variance of input factors explained thus only 52% of the total variance of the output. The total effect indices were then used to identify the key input factors instead of only the first order indices.

Table 2. – Estimates of the first order (*S*) and total effect (*St*) indices of the sensitivity analysis.

	Case #1 (uniform distributions)		Case #2 (informative distributions)	
Factors	***St***	***S***	***St***	***S***
K	0.58	0.21	0.31	0.16
μ_{opt}	0.35	0.05	0.18	0.07
aw	0.29	0.06	0.11	0.07
TR	0.28	0.06	0.14	0.06
dR	0.24	0.05	0.39	0.25
aw_{min}	0.11	0.00	0.03	0.01
T_{min}	0.10	0.01	0.03	0.02
n_0	0.10	0.06	0.02	0.02
TM	0.10	0.00	0.05	0.00
dM	0.09	0.00	0.09	0.05
pH	0.06	0.00	0.01	0.00
T_{opt}	0.08	0.01	0.01	0.00
Gy	0.06	0.00	0.01	0.00
T_{max}	0.06	0.00	0.00	0.00
dD	0.06	0.01	0.04	0.02
TD	0.06	0.00	0.01	0.00
pHmin	0.05	0.00	0.01	0.00
TTEP	0.05	0.00	0.00	0.00
TC	0.05	0.00	0.01	0.00
dC	0.05	0.00	0.01	0.00
dTEP	0.05	0.00	0.01	0.00
TV	0.05	0.00	0.00	0.00
dV	0.05	0.00	0.00	0.00

In case #1 (uniform distributions for input), the most important factors influencing the variance of the log of the bacterial contamination ($St > 0.1$) were the number of contaminating cells (n_0) and biological parameters linked to the capacity of bacterial cells to rapidly enter active multiplication phase (*K*) with a high growth rate (μ_{opt}) and to grow when conditions are not optimal (T_{min}, aw_{min}). We also observed that the water activity of the salmon (*aw*) was an important factor and, less surprisingly, that the storage conditions in the domestic refrigerator (*TR*, *dR*) and that the retail temperature (*TM*) were also significant. The impact of other input factors was less significant ($St < 0.1$) and they can therefore be set to a fixed value in their range of variation without notably affecting the final log contamination in *L. monocytogenes*. The cut-

off value of 0.1 was empirically chosen and corresponded to a break between a few very significant input factors and a larger group of factors with *St*-values comprised between 0.05 and 0.10.

When using informative probability distribution functions (case #2), the ranking of most influential factors was slightly altered. The subset of important factors was reduced in comparison with the case #1. Only the domestic storage conditions (*TR*, *dR*), the physiological state (*K*) of contaminating cells and their growth rate (μ_{opt}), and the water activity of salmon (*aw*) were recognized as important (Table 2). The global SA allowed thus to identify key input factors for management options: manufacturers could be encouraged to better control the water activity of their product and communication programmes could be implemented to make the consumers aware of the importance of fridge storage conditions. Other key input factors were identified and required additional research for a better characterization, this was the case of the initial physiological state of contaminating cells controlling their lag phase. In comparison with case #1, no new important input factor was identified showing that an uninformative approach is useful in first intention to simplify the model and to restrict the acquisition of data only for eventually significant input factors.

4. Conclusions

This application of global SA to the assessment of the bacterial contamination of a food shows the relevance of such approach to study models used in QMRA. The Saltelli method is relatively easy to perform and gives important information to scientists performing modelling as well as to risk managers. Comprehensible quantitative indices taking into account all interactions between factors are calculated with this method with acceptable computation time. These indices allow the identification of key input factors but also the factors that are not significant to simplify the model. It can be very useful to perform SA from the design of a model even if no extensive knowledge is available for the values of input factors to only focus on those needing additional research and an accurate description.

Acknowledgement

This study was carried out in the framework of the PhD thesis of Mariem ELLOUZE in collaboration with Jean-Pierre GAUCHI. The author wishes to gratefully thank them for this work. This work was supported by a CIFRE financial grant (n°624/2006) from the company CRYOLOG and the French Association Nationale de la Recherche et de la Technologie (ANRT). I would also like to thank the Sym'Previus consortium for providing the cardinal values of *L. monocytogenes* and the CEMAGREF and ANIA for providing food storage thermal profiles.

References

[1]. Borgonovo, E. (2006). Measuring uncertainty importance: investigation and comparison of alternative approaches. *Risk Analysis*, 26, 1349-1361.

[2]. Cukier, R.I., Fortuin, C.M., Schuler, K.E., Petschek, A.G., Schaibly, J.H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. *The Journal of Chemical Physics*, 59, 3873-3878.

[3]. Cukier, R.I., Levine, H., Schuler, K. (1978). Non linear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26, 1-42.

[4]. Delignette-Muller, M.L., Rosso, L. (2000). Biological variability and exposure assessment. *International Journal of Food Microbiology*, 58, 203-212.

[5]. Ellouze, M., Gauchi, J.-P., Augustin, J-C. (2010). Global sensitivity analysis applied to a contamination assessment model of *Listeria monocytogenes* in cold smoked salmon at consumption. *Risk analysis*, 30, 841-852.

[6]. Frey, H.C., Patil, S.R. (2002). Identification and review of sensitivity analysis methods. *Risk Analysis*, 22, 553-578.

[7]. Homma, T., Saltelli, A. (1996). Importance measures in global sensitivity analysis of non linear models. *Reliability Engineering and System Safety*, 52, 1-17.

[8]. Koutsoumanis, K., Pavlis, A., Nychas, G.-J.E., Xanthiakos, K. (2010). Probabilistic model for *Listeria monocytogenes* growth during distribution, retail storage, and domestic storage of pasteurized milk. *Applied and Environmental Microbiology*, 76, 2181-2191.

[9]. Mataragas, M., Zwietering, M.H., Skandamis, P.N., Drosinos, E.H. (2010). Quantitative microbiological risk assessment as a tool to obtain useful information for risk managers - Specific application to *Listeria monocytogenes* and ready-to-eat meat products. *International Journal of Food Microbiology*, 141, S170-S179.

[10]. Membré, J.-M., Kan-King-Yu, D., Blackburn, C.W. (2008). Use of sensitivity analysis to aid interpretation of a probabilistic *Bacillus cereus* spore lag time model applied to heat-treated chilled foods (REPFEDs). *International Journal of Food Microbiology*, 128, 23-33.

[11]. Patil, S.R., Frey, H.C. (2004). Comparison of sensitivity analysis methods based on applications to a food safety risk assessment model. *Risk Analysis*, 24, 573-585.

[12]. Pérez-Rodríguez, F., Todd, E.C.D., Valero, A., Carrasco, E., García, R.M., Zurera, G. (2006). Linking quantitative exposure assessment and risk management using the food safety objective concept: an example with *Listeria monocytogenes* in different cross-contamination scenarios. *Journal of Food Protection*, 69, 2384-2394.

[13]. Pouillot, R., Miconnet, N., Afchain, A.L., Delignette-Muller, M.L., Beaufort, A., Rosso, L., Denis, J.-B., Cornu, M. (2007). Quantitative risk assessment of *Listeria monocytogenes* in French cold-smoked salmon: I. Quantitative exposure assessment. *Risk Analysis*, 27, 683-700.

[14]. Pouillot, R., Goulet, V., Delignette-Muller, M.L., Mahé, A., Cornu, M. (2009). Quantitative risk assessment of *Listeria monocytogenes* in French cold-smoked salmon: II. Risk characterization. *Risk Analysis*, 29, 806-819.

[15]. Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145, 280-297.

[16]. Saltelli, A., Chan, K., Scott, E.M. (2000). *Sensitivity Analysis*. Chichester, John Wiley.

[17]. Sobol', I.M. (1990). Sensitivity estimates for nonlinear mathematical models. *Mathematicheskoe Modelirovanie*, 2, 112-118 [Translated as Sensitivity analysis for nonlinear mathematical models, *Math. Modeling Comput. Experiment* (1993), 1, 407-414.]
[18]. Yang, H., Mokhtari, A., Jaykus, L.-A., Morales, R.A., Cates, S.C., Cowen, P. (2006). Consumer phase risk assessment for *Listeria monocytogenes* in deli meats. *Risk Analysis*, 26, 89-103.
[19]. Zwietering, M.H. (2009). Quantitative risk assessment: Is more complex always better? Simple is not stupid and complex is not always more correct. *International Journal of Food Microbiology*, 134, 57-62.
[20]. Zwietering, M.H., van Gerwen, S.J.C. (2000). Sensitivity analysis in quantitative microbial risk assessment. *International Journal of Food Microbiology*, 58, 213-221.