Research Article

Camille Béchaux*, Amélie Crépet and Stéphan Clémençon

Improving Dietary Exposure Models by Imputing Biomonitoring Data through ABC Methods

Abstract: New data are available in the field of risk assessment: the biomonitoring data which is measurement of the chemical dose in a human tissue (e.g. blood or urine). These data are original because they represent direct measurements of the dose of chemical substances really taken up from the environment, whereas exposure is usually assessed from contamination levels of the different exposure media (e.g. food, air, water, etc.) and statistical models. However, considered alone, these data provide little help from the perspective of Public Health guidance. The objective of this paper is to propose a method to exploit the information provided by human biomonitoring in order to improve the modeling of exposure. This method is based on the Kinetic Dietary Exposure Model which takes into account the pharmacokinetic elimination and the accumulation phenomenon inside the human body. This model is corrected to account for any possible temporal evolution in exposure by adding a scaling function which describes this evolution. Approximate Bayesian Computation is used to fit this exposure model from the biomonitoring data available. Specific summary statistics and appropriate distances between simulated and observed statistical distributions are proposed and discussed in the light of risk assessment. The promoted method is then applied to measurements of blood concentration of dioxins in a group of French fishermen families. The outputs of the model are an estimation of the body burden distribution from observed dietary intakes and the evolution of dietary exposure to dioxins in France between 1930 and today. This model successfully fit to dioxins data can also be used with other biomonitoring data to improve the risk assessment to many other contaminants.

Keywords: chemical risk assessment, approximate Bayesian computation, biomonitoring data, kinetic model, exposure model

DOI 10.1515/ijb-2013-0062

1 Introduction

Many chemicals are known to accumulate in the environment and to be very slowly eliminated from the human body such as polychlorinated biphenyls, dioxins, organochlorine pesticides, chlorobenzenes, and polybrominated diphenyl ethers. They are of great concern for the public health and are listed by the Stockholm convention as *Persistent Organic Pollutants* [1]. To assess the risk related to these chemicals, the use of the body burden is now considered as the best way [2]. Recently, new data allow for a direct assessment of the body burden: the biomonitoring data. However, as highlighted in Lyons et al. [3], understanding the effects on public health of exposure to environmental chemicals requires establishing relationships among events along an exposure–health evaluation–risk assessment continuum. Therefore, biomonitoring data need to be linked to external exposure to be interpreted in terms of source of exposure

***Corresponding author: Camille Béchaux,** Anses – DER, 27 Avenue du général Leclerc, Maisons-Alfort 94700, France, E-mail: camille.bechaux@anses.fr; camille.bechaux@gmail.com

Amélie Crépet, Anses – DER, 27 Avenue du général Leclerc, Maisons-Alfort 94700, France, E-mail: amelie.crepet@anses.fr

Stéphan Clémençon, Telecom ParisTech – LTCI UMR, Paris, France, E-mail: stephan.clemencon@telecom-paristech.fr

and for risk management purposes. External exposure is usually assessed from statistical models considering current contamination levels of the different exposure media (e.g. food, air, water, etc.) and behaviors of individuals (consumption habits, activity pattern, etc.). However, these classical exposure models do not take the accumulation in the body into account. Some reverse-dosimetry or steady-state approaches aim at taking this elimination time into account through toxicokinetic models [4]. However, they do not account for the level of the past exposures and especially its evolution. While because of the long half-life of persistent chemicals, body burden measured today also reflects the past exposure of the individuals. It is well highlighted that exposure to environmental contaminants has significantly changed over the last decades due to changes in production, in use, and in successive regulations [5] but data on past exposure are rarely available. Therefore, to be able to interpret biomonitoring data when assessing the risk related to persistent chemical, a comprehensive method that takes lifelong exposure and accumulation into account is needed. It is precisely the objective of this paper to develop a method to link the biomonitoring data with the external exposure taking the accumulation and the past exposure into account.

Different statistical points are developed and combined in this paper to propose a comprehensive approach. First, we modified the *Kinetic Dietary Exposure Model* (KDEM) which was introduced in the context of dietary risk assessment by Bertail et al. [6]. This dynamic model takes into account the accumulation phenomenon of a given chemical in the body and a single-compartment pharmacokinetics model governing its elimination. A simple pharmacokinetics model is suitable for many environmental contaminants that accumulate mainly in lipids (e.g. MeHg, polychlorinated biphenyls, polybrominated diphenyl ethers, dioxins, etc.). The KDEM approach was used to assess exposure to dietary methylmercury (MeHg) in the French adult female population [7]. However, these results have never been validated with measured body burdens provided by biomonitoring data [8]. Moreover, the model does not take the trend of exposure into account. That is why, a modified model is developed which integrates a scaling function λ to describe the evolution of the exposure. Data on past contamination or past exposure are rarely available but it is proposed here to use biomonitoring data to estimate the parameters of this function. Given the very high complexity of the likelihood function of the new model and the computational difficulties faced when trying to optimize it, Approximate Bayesian Computation (ABC in short, Beaumont et al. [9]) is proposed to fit the numerical model from the biomonitoring data available. This approach bypasses the exact evaluation of the likelihood function to estimate the posterior distributions by means of simulations and well-chosen summary statistics. ABC has recently become popular for statistical inference in complex models such as those considered in biosciences, in population-genetics [10], and in mathematical epidemiology [11] for instance. In order to perform model selection simultaneously with the estimation of the parameter, an original two-stage hierarchical ABC-MCMC algorithm is implemented following in the footsteps of Toni et al. [12]. Specific summary statistics, especially relevant for risk assessment, are proposed based on Cramer–Von Mises statistic.

The paper is structured as follows: Section 2 outlines the key ingredients governing the modified KDEM which accounts for the possible temporal evolution in exposure. Section 3 specifies the principles of the ABC method used to fit the new dynamic model of exposure from punctual biomonitoring data. Relevant specific summary statistics for risk assessment are proposed, and the model selection for the parametric modeling of the function λ is described. Section 4 is devoted to the application of the methodology to blood concentration in dioxins measured in a population of French fishermen families.

2 Dynamic modeling of dietary exposure

The KDEM was introduced by Bertail et al. [6] in order to describe the temporal evolution of the total body burden of a chemical present in a variety of foods. Its dynamic is ruled by two components, a *marked point process* (MPP) which models dietary behavior and governs the accumulation of the chemical in the human body and a linear differential equation accounting for its physiological elimination.

The accumulation due to successive dietary intakes is described by the MPP $(T_n, U_n)_{n\in N}$, where the T_n's are the successive times when an intake of value U_n occurs (with $T_0 = 0$ by convention). The U_n's are independent from the T_n's and the inter-intake times $\Delta T_n = T_n - T_{n-1}$, $n \geq 1$.

The elimination process between two successive intakes is described by a *single-compartment* pharmacokinetic model. The total body burden of the chemical $x(t)$ decreases with time t according to the linear differential equation:

$$\frac{dx}{dt}(t) = -\gamma \times x(t), \tag{1}$$

where $\gamma > 0$ denotes the elimination rate. The parameter γ is related to the biological half-life of the chemical in the body $t_{1/2} = \log(2)/\gamma$, namely the time needed for the total body burden x to decrease by half in the absence of any additional intake.

In order to integrate the evolution of the exposure into the model, a scaling function of time $\lambda(t)$ is introduced in the KDEM model. This function describes the possible evolution of the dietary exposure through time and is assumed to be of a parametric form, fully determined by a vector θ. The shape of the exposure evolution is unknown. Therefore, the function λ must be represented in a flexible manner. That is why, a linear combination of functions from a B-spline basis of degree $m \geq 1$: $\lambda(t) = \sum_{i=0}^{m} \beta_i B_i^m(t)$ was proposed.

Finally, exposure is described by a piecewise deterministic process $X(t)$ of initial value x_0 and evolving according to eq. (1) between intake times. The total body burden values $X_n = X(T_n)$ at intake times T_n are defined through the recurrence equation:

$$\begin{cases} X_{n+1} = X_n e^{-\gamma \Delta T_{n+1}} + \lambda_\theta(T_{n+1}) U_{n+1} \\ X_0 \quad = x_0 \end{cases}. \tag{2}$$

3 An ABC method to fit the model to biomonitoring data

We propose to implement an ABC method [9] to fit the parameter θ of the scaling function λ by comparing the distribution of the measured body burdens from data to values simulated from KDEM.

The principle of standard ABC relies on the basic rejection algorithm: a candidate value θ^* is generated from a prior distribution $\Pi(\theta)$. Based on the numerical model defined by θ^*, a simulated data set X^* is generated. It is next reduced to a vector of summary statistics s^*. One then computes a "distance", $\rho(s^*, s)$, between the simulated and the observed summary statistics and compares it to a tolerance level δ: if $\rho(s^*, s) \leq \delta$, θ^* is accepted. After a chosen number of iterations, the accepted θ^* s form the posterior distribution of θ.

3.1 Specific summary statistics and distances for risk assessment

The choice of the vector of summary statistics s and the corresponding distance ρ is paramount in ABC methods. This aspect has been widely studied in the case of population genetic data [9, 13, 14]. Regarding biomonitoring data, specific summary statistics must be defined. Here, the summary statistics are built from distances between the observed and the simulated distributions of body burdens of different age classes of individuals. Individuals with ages in the same range, $[a_k, a_{k+1}]$, at time t are grouped, in order to form $K \geq 1$ (disjoint) groups of comparable sizes. Denote by $I_k \subset \{1, \ldots, n\}$ the corresponding subsets of indexes and then form the empirical distributions of the corresponding exposures: $F_k = (1/\#I_k) \sum_{i\in I_k} \delta_{X_i(t-\tau_i)}$, denoting by $\#I_k$ the cardinality of I_k, $1 \leq k \leq K$. The choice of the number K of age classes is critical since it determines the dimensionality of the vector of summary statistics

$s = (F_1, \ldots, F_K)$. Indeed, when the summary statistic is not sufficient, there is no guarantee that the posterior distribution based on the latter equals to the exact posterior distribution, i.e. based on the likelihood function) [15]. However, increasing the number K of summary statistics, thereby increasing the amount of information available may reduce the accuracy of the procedure insofar as the probability of accepting a value decreases exponentially as dimensionality increases [9].

For each age class k in $\{1, \ldots, K\}$, the empirical distribution $\tilde{F}_k(dx)$ of the observed body burdens, $\{\tilde{X}_i(t-\tau_i),\ \ i \in I_k\}$, is compared to the simulated body burdens $\{X_i^*(t-\tau_i),\ \ i \in I_k\}$ with the model defined by the parameter value θ^*, $F_k^*(dx)$. Various quantities can be considered to evaluate the discrepancy between two distributions [16]. A very popular measure of global discrepancy is the two-sample Cramer–Von Mises statistic which is widely used in nonparametric hypothesis testing [17]. In the present context, it corresponds to the quantity:

$$\rho_{VM}(\tilde{F}_k,\ F_k^*) = \frac{\#I_k}{2}\int_{-\infty}^{+\infty} \left(\tilde{F}_k(x) - F_k^*(x)\right)^2 F_k(dx), \tag{3}$$

where $\tilde{F}_k(x) = (1/\#I_k)\sum_{i\in I_k} I\{\tilde{X}_i(t-\tau_i) \le x\}$, $F_k^*(x) = (1/\#I_k)\sum_{i\in I_k} I\{X_i^*(t-\tau_i) \le x\}$, and $F_k = (F_k^* + \tilde{F}_k)/2$.

In the field of risk assessment, one generally focuses on the right tail of the exposure distribution. Thus, it may be relevant to consider the truncated version:

$$\tilde{\rho}_q(\tilde{F}_k,\ F_k^*) = \frac{\#I_k}{2}\int_{q}^{+\infty} \left(\tilde{F}_k(x) - F_k^*(x)\right)^2 F_k(dx), \tag{4}$$

choosing q as a specific quantile of the distribution $\tilde{F}_k$. We consider the following dissimilarity measure between the distribution of observed data $\tilde{X}$ and that of simulations X^*:

$$\rho(\tilde{X},\ X^*) = \sum_{k=1}^{K} \tilde{\rho}_q(\tilde{F}_k,\ F_k^*). \tag{5}$$

3.2 A two-stage hierarchical ABC-MCMC algorithm for model selection

Many algorithms can generate observations from a posterior distribution without computing likelihoods explicitly, including basic rejection algorithms. In order to reduce the number of simulations required to estimate the posterior distribution, an ABC-MCMC algorithm is used [15]. Thereby, a path of a Markov chain whose equilibrium distribution is the approximate posterior distribution $\Pi(\theta|\rho(\tilde{X}, X^*) \le \delta)$ is simulated. In this way, the values of the parameter θ providing simulations close to the observed data are visited preferentially.

Here, a model selection is performed simultaneously with the estimation of the parameter θ of the scaling function λ. A two-stage hierarchical algorithm is proposed to consider m as an additional parameter [12]. It makes it possible to choose the best form for the scaling function λ among a finite number of models, indexed by $m \in \{1, \ldots, M\}$ and sorted by increasing complexity. The two-stage hierarchical algorithm is based on an ABC-MCMC procedure and implemented as follows.

1. If now at m, propose a move to m^* according to a transition kernel $q(m \to m^*)$.
   (i) if now at θ_m^*, propose a move to $\theta_{m^*}^*$ according to a transition kernel $q'(\theta_{m^*} \to \theta_{m^*}^*)$
   (ii) Generate X^* from model m^* and $\theta_{m^*}^*$
   (iii) Compute $\alpha' = \min\left\{1, \frac{\Pi(\theta^*|m^*)q'\left(\theta_{m^*}^* \to \theta_{m^*}\right)}{\Pi(\theta_{m^*})q'\left(\theta_{m^*} \to \theta_{m^*}^*\right)} 1\left(\rho(X^*, \tilde{X}) \le \delta\right)\right\}$
   (iv) Accept $\theta_{m^*}^*$ with probability α' and otherwise stay at θ_{m^*}, then return to **1.**
2. Compute $\alpha = \min\left\{1, \frac{\Pi(m^*)q(m\to m^*)}{\Pi(m)q(m\to m^*)} 1\left(\rho(X^*, \tilde{X}) \le \delta\right)\right\}$
3. Accept m^* with probability α, stay at m otherwise and then return to **1.**

This algorithm was implemented in the software R, and the code is available in the supplementary material. It can be noticed that the higher the posterior probability, the greater the number of accepted θ. This

ensures a good estimate of the posterior distribution for the selected model. However, for the models which are poorly represented in the marginal posterior distribution of m, the small number of accepted θ does not provide an accurate estimation of the posterior distribution of θ[12]. It should be noticed that this model selection implicitly penalizes the models with the highest numbers of parameters since the probability for θ to be accepted decreases with the dimension.

4 Application to dioxin biomonitoring data in France

4.1 Data and exposure assessment

In the French National study on dioxin blood levels in French consumers of freshwater fish (ICAR study, Anses and InVS [18]) conducted in 2009, the dioxin concentrations in blood (pg/g fat) were recorded for 606 adults between 18 and 75 years of age. All individuals were from fishermen families. In order to compare these measurements with the estimated body burdens (pg/kg body weight), the concentrations were converted into total body burdens using the equation proposed by Deurenberg et al. [19]. This conversion assumes that the balance between all lipid compartments in the body and the fat content of the human body varies with age, sex, weight, and height. Consumed quantities of 115 food items covering the main diet of each individual taking part in the study were collected by means of a food frequency questionnaire and a photograph manual of portions' size.

Contamination data in food are provided by three French studies. Mean concentrations in edible parts of the 26 most consumed species of freshwater fish are provided by the ICAR study. Contamination data of seafood come from the CALIPSO study [20]. This study provides the mean contamination in 30 species of marine fish, 17 species of molluscs and crustaceans, and 14 other products. The mean concentration of other foods known to contribute to the exposure to dioxins (24 food items including the different kinds of meat, cooked meat, ready meals, dairy products, eggs, and fat such as butter and oil) comes from the Second French Total Diet Study [21].

The reference distribution of the dioxins intake U in 2009 in adulthood is estimated by combining the distribution of the reference contamination of foods with consumption provided by the food questionnaire of the ICAR study. Since the consumption is known to be subject to wide fluctuations during childhood, reference intakes U_n at age under 18 are simulated by combining the consumption distribution of 1,455 French children aged 3–17 years recorded in the INCA2 study [22] with the distribution of the reference contamination vector.

4.2 Modeling λ for dioxin exposure

The purpose of the scaling function λ is to correct the intake according to the date on which it occurred. It thus describes the historic evolution of the exposure over time. Dioxins are naturally found in the environment [23], but the emissions due to human activities are suspected of having significantly changed since the 1930s [24]. Thereby we set $\lambda(t \leq 1930) = 0.1$ the level of exposure related to the background contamination of the environment, as calculated in Van Der Molen et al. [24]. Since the data were collected in 2009, individual reference intake U corresponds to this date and the correction factor is equal to 1 in 2009, i.e. $\lambda(2009) = 1$. The scaling function λ is assumed to be a continuous function in the interval $[1930; 2009]$. Therefore, the linear combination of functions from the B-spline basis of degree $m \geq 1$ has only two knots (1930 and 2009 namely). In this situation, the number of parameters of the vector θ to be estimated corresponds to the number $m+1$ of basis functions. In the present study, $M = 4$ models are tested, each model $m \in \{1, \ldots, , 4\}$ corresponding to a B-spline basis of degree m.

The evolution of environmental contamination by dioxins has already been studied in certain countries by Lorber [4], Van Der Molen et al. [24], and Aylward and Hays [25]. Their results were used to set

appropriate prior distributions for the parameters governing the different candidate model. Typical scaling functions λ obtained from these prior distributions for θ are drawn in red in Figure 2.

4.3 ABC implementation

Summary statistics of Section 3 were calculated using $K = 5$ age classes. This number of classes guarantees a sufficient number of individuals in each class (Table 1). Observed body burdens and their distributions by age classes are depicted in Figure 1. In risk assessment, a focus is made on the most exposed individuals. Therefore, the threshold q of the distance ρ (eq. (4)) is set to 0.5. Thereby, the 50% of the most exposed individuals of each age class are used to fit the model.

Table 1 Comparison of observed and simulated body burden (mean and 95% credibility interval) distributions for each age class

Age (years)			Observed and predicted body burden (pg\kg bw)		
	Size		25th p	Mean	97.5th p
18–37	124	Obs.	1,022	4,097	8,695
		Sim.	959 [854;1,152]	4,501 [4,018;5,387]	9,072 [8,463;11,615]
38–46	123	Obs.	3,253	6,358	13,881
		Sim.	1,571 [1,396;1,905]	6231 [5515;7567]	15,408 [13,662;18,685]
47–54	130	Obs.	4,254	11,709	22,327
		Sim.	1,388 [1,220;1,697]	11,397 [9,998;13,951]	25,881 [21,048;32,601]
55–61	112	Obs.	3,539	14,916	30,932
		Sim.	1,806 [2,076;2,550]	21,760 [18,945;26,729]	32,742 [26,084;45,422]
62–75	117	Obs.	9,958	20,213	43,491
		Sim.	4,686 [4,038;5,814]	22,092 [19,110;27,325]	56,144 [42,070;68,530]

The choice of tolerance δ corresponds to a trade-off between the bias and the variance of the estimate [9] and, in practice, between computability and accuracy. The tolerance is often set to be a quantile P_δ of the empirical distribution function of the distance between observed and simulated data [9, 11, 26]. In this work, $P_\delta = 0.05$ which means that the 5% of simulated X that are closest to $\tilde{X}$ considering all the models together are kept.

As proposed in Wegmann et al. [26], a series of 10,000 simulations have been preliminarily performed, where values for the parameter θ are each time randomly drawn from the prior distribution $\Pi(d\theta)$. This calibration step allows to conveniently define the tolerance δ such that $P_\delta = 0.05$. Moreover any of these simulations for which the condition $\rho(\tilde{X}, X^*) \leq \delta$ is true can be used as a starting point for the chain. This removes the issue of the large number of iterations that are necessary to first satisfy the condition $\rho(\tilde{X}, X^*) \leq \delta$. These preliminary simulations are also used to adjust the proposal range by choosing an appropriate transition kernel in order to ensure a thorough exploration of the whole parameter space.

The θ which satisfy $\rho(\tilde{X}, X^*) \leq \delta$ are used as a starting point for the chains of the ABC-MCMC algorithm. Thus, there is no initial burn-in period. The convergence of the chains is checked by testing the goodness-of-fit between successive intermediate posterior distributions of θ for each parameter, using a two-sample Kolmogorov–Smirnov test [27]. This test indicates that the chains have converged after 150,000 iterations. The acceptance rate is close to 38% which ensure a thorough exploration of the parameter space. The posterior distributions of each parameter are built from the accepted θ of the 150,000 iterations of two different chains. The function λ obtained from these posterior distributions for the four models are shown in black in Figure 2.

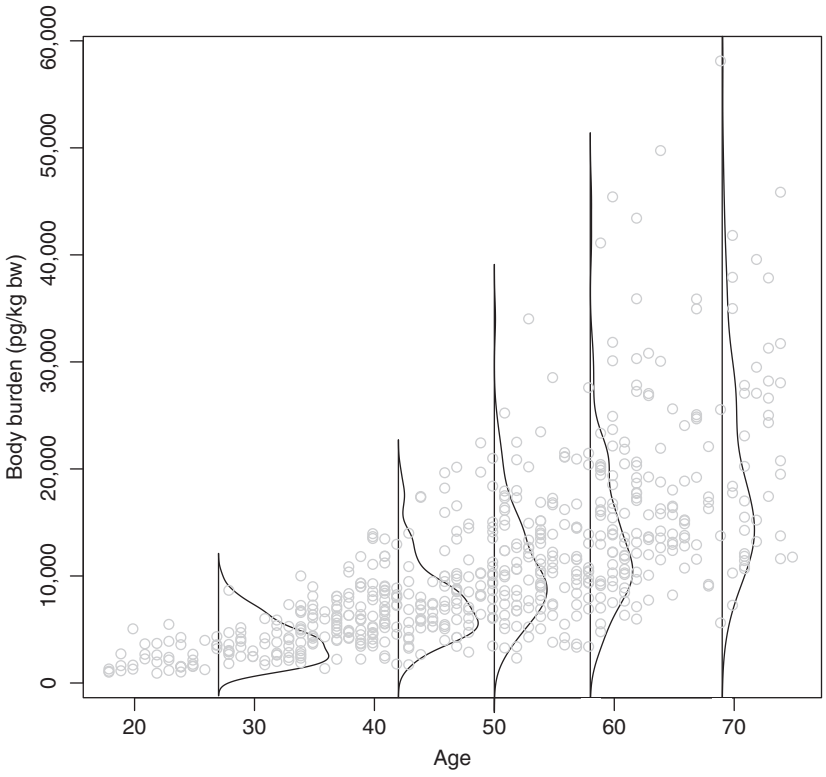

Figure 1 Observed body burdens $\tilde{X}$ and their distributions by age classes ($K = 5$)

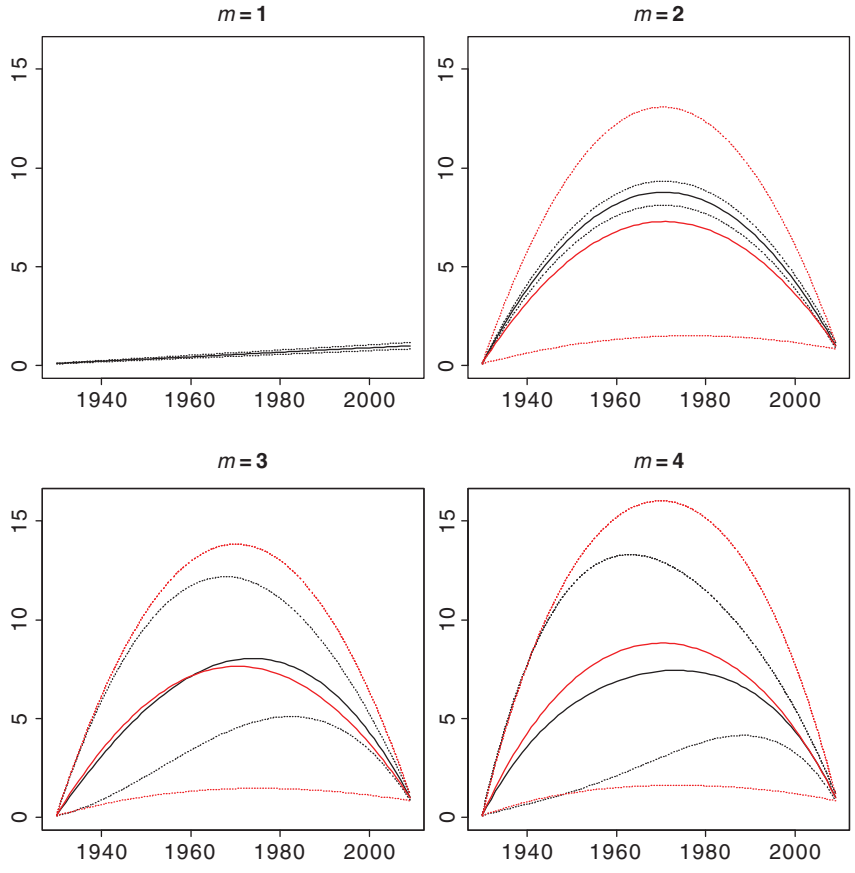

Figure 2 Function λ obtained from the prior distributions of the parameters (p5, mean, p95) in red and from posterior distributions (p5, mean, p95) in black according to the four models

4.4 Results and model validation

For the model $m = 2$, one reaches a rate of 41% of accepted θ (against 0% for model 1, 35% for model 3, and 24% for model 4). This model is thus considered as the best regarding the selection method proposed in Section 3. To confirm this result, the Root Mean Square Error of Prediction (RMSEP) of each model is calculated. The data set is randomly divided into three subsets respecting the proportion of individuals in each age class. Then the model is fitted from two-thirds of the data set, and the RMSEP is calculated on the remaining third using the posterior distribution of θ for each model. This procedure is repeated three times, and the mean values of the RMSEP are: 13,599 for $m = 2$, 13,535 for $m = 3$, and 13,452 for $m = 4$. Since, no value of θ was accepted for $m = 1$, no RMSEP is calculated. The RMSEP decreases when the number of parameters increases. However, the values obtained remain quite comparable, and the gain of prediction provided by an increasing number of parameters does not seem significant enough. This confirms the selection of model $m = 2$.

The scaling functions provided by each of the four tested models are shown in black in Figure 2. Since, no value of θ was accepted for $m = 1$, there is no scaling function associated with this model. The scaling function provided by the models $m = 2$, $m = 3$, and $m = 4$ are closed regarding the mean. However, the interval around the mean is much more bigger for models $m = 3$ and $m = 4$. The previous selection model concluded that model $m = 2$ was the best model. The scaling function provided by this model shows that exposure has increased since the introduction of dioxins in the 1930s to a peak in the 1970s corresponding to a seven times higher exposure than the current exposure. Then, the exposure has decreased to the current exposure values.

From the obtained scaling function, the body burdens of the individuals were predicted. Thereby, the goodness-of-fit of the model is checked by comparing the distributions of the predicted body burdens and the observed body burdens. Figure 3 shows the distributions of the observed body burdens, the body burdens predicted from KDEM without scaling function and the body burdens predicted from the corrected KDEM with the scaling function considering the mean estimation of the $m = 2$ model parameters. This comparison shows a much better fit of the corrected KDEM model to the biomonitoring data regarding the whole population.

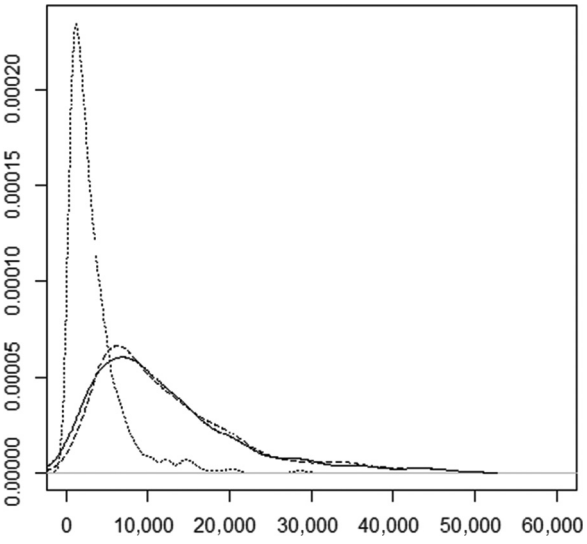

Figure 3 Distributions of the observed body burden (solid line), the body burdens predicted by KDEM (dotted line), and the body burdens predicted by the corrected KDEM $m = 2$ (dashed line) for the ICAR population (from the mean estimation of the model parameters)

In order to be relevant to risk assessment, the model also needs good prediction abilities. As made for the model selection, the quality of the model prediction is checked by performing a cross-validation. Body burdens for one-third of individuals are simulated with the model fitted from the remaining two-thirds. The distributions of these simulated body burdens are summarized and compared with the observed body burdens of the corresponding individuals. Results of these comparisons are given in Table 1. A distribution is considered as well predicted if the observed percentiles are included in the credibility interval of the predicted percentiles. Table 1 shows that the mean and the high percentiles are well estimated for each age class whereas for the low percentiles intervals do not contain the observed percentiles. This can be explained by the choice of specific summary statistics that only used the 50% that represent the most exposed individuals to fit the model.

5 Discussion

This paper proposes an original method based on an ABC approach to calibrate a new dynamic exposure model using biomonitoring data. This method improves chemical exposure and risk assessment as well as risk management in many ways. First, the combination of biomarker measurements with contamination levels usually exploited separately makes it possible to refine exposure assessment. The comparison of KDEM outputs with measured body burdens highlighted that the current contamination and consumption data even used in a dynamic exposure model were not sufficient to explain the actual accumulation of dioxins in the body. Indeed, environmental contamination by dioxins in France has changed with time. Individuals were thus exposed to higher intakes in the past. Unfortunately, the value of these intakes are unknown in France since no past data are available. However, the biomonitoring data from fishermen families and the modified KDEM made it possible to estimate the scaling function parameters without past data. The obtained scaling function highlights that exposure to dioxins in France peaked in the 1970s. This is fully consistent with observed evolutions in the emission of dioxin-like PCBs in the environment [28].

In France, regarding the general population, no biomonitoring data on dioxins are available. Risk assessment for the general population is thus conducted using dietary intakes provided by the Second French Total Diet Study [21]. This risk estimate does not take into account the exposure evolution and the accumulation phenomenon. This could lead to underestimate the dioxin body burdens, as observed for the Fishermen families. The KDEM corrected by the obtained scaling function can be applied to the general population intakes to predict body burdens. Therefore, assuming the exposure evolution to be the same between the fishermen population and the general population, the risk related to dioxins can be refined. The method also makes it possible to propose new risk indicators which depend on time [29]. Indeed, as individual exposure over life is described as a process $X(t)$, the maximum exposure over a given period of time: $\max_{t\in[0,T]}X(t)$ or the time spent above a reference threshold u: $\int_{t=0}^{T} I_{X(t)\geq u}dt$ can be calculated. A new time dimension can therefore be introduced in risk assessment.

The proposed method also improves risk management. The weakness of biomonitoring data lies in its inability to identify exposure sources and their related part to the overall exposure. Indeed, these data alone do not provide any information on the exposure route (inhalation, ingestion, etc.) or for example on the food the most involved in the body burden. Therefore, linking dietary exposure with blood concentration data constitutes therefore a useful tool for the interpretation of biomonitoring data by the risk managers. The promoted method can be used to interpret biomonitoring data from cohort studies of the US National Health and Nutrition Examination Survey [30] or the French National survey on nutrition and health [31] for instance. This can be done for dioxins but also for other substances like polychlorinated biphenyls, polybrominated diphenyl ethers, arsenic, cadmium, mercury, and pesticides.

An ABC was proposed to fit the corrected KDEM to the biomonitoring data. This application confirms that ABC methods are particularly suitable in case of complex likelihood. Indeed, the computational difficulties faced when trying to optimize the likelihood are bypassed in this method. However, ABC

methods require the choice of sufficient distances and summary statistics as highlighted by Joyce and Marjoram [14] in case of high-dimensional data in genetics. This choice can be tricky, and a wrong choice can lead to fail to fit the model [15]. The distance and the summary statistics defined here were proven able to correctly predict the body burdens of the population. They were specifically defined for risk assessment and can be used for other chemicals. An original two-stage hierarchical ABC-MCMC algorithm is proposed to perform the ABC method. By considering the model as an additional parameter, the parameter estimation for each model is performed simultaneously with the model selection. However, as highlighted by Toni et al. [12] this selection procedure may penalize/favor some models because of the choices made for the prior distribution (which is also inherent in the standard Bayesian model selection) and the tolerance. For this reason, the proposed model selection was validated by a usual criterion, the RMSEP. In our application, the RMSEP confirmed the model selection made from the hierarchical algorithm.

Using the ABC-MCMC algorithm proposed by Marjoram [15], the number of proposed values to obtain a sufficient rate of accepted θ seems reasonable since after 150,000 iterations, around 57,000 θ were accepted. Otherwise when a more complicated model of exposure is used or if it is necessary to be more restrictive regarding the tolerance δ, then it would be relevant to consider alternative algorithms. Toni et al. [12] showed that the sequential Monte Carlo method, first developed by Sisson et al. [32] and derived from a sequential importance sampling algorithm from Del Moral et al. [33], may yield much better computational performance.

Regarding dioxins, a single-compartment pharmacokinetic model, determined by a single parameter, i.e. the half-life, has been proved relevant to describe the elimination process by the human body [4, 24]. This makes it possible to link exposure with biomonitoring data from a simple model with reliable statistical properties, as described by Bertail et al. [8]. Such pharmacokinetic models are suitable for many other chemicals like polychlorinated biphenyl, polybrominated diphenyl ethers, and methyl mercury, which allows for a direct use of the method proposed in this paper with these chemicals. However, for certain other chemicals, a model which more effectively describes the anatomical, physiological, physical, and chemical phenomena involved in the absorption, distribution, metabolism, and excretion of chemicals in the human body might be necessary. Physiologically based pharmacokinetic models, providing a multi-compartment description, may be a relevant choice in these situations [34].

References

1. Jones K, Voogt P. Persistent organic pollutants (pops): state of the science. Environ Pollut 1999;100:209–21.
2. Van Leeuwen F, Younes M. Consultation on assessment of the health risk of dioxins: re-evaluation of the tolerable daily intake (TDI): executive summary. Food Addit Contam 2000;17:223–40.
3. Lyons M, Yang R, Mayeno A, Reisfeld B. Computational toxicology of chloroform: reverse dosimetry using Bayesian inference, Markov chain Monte Carlo simulation, and human biomonitoring data. Environ Health Perspect 2008;116:1040–6.
4. Lorber M. A pharmacokinetic model for estimating exposure of americans to dioxin-like compounds in the past, present, and future. Sci Total Environ 2002;288:81–95.
5. Hays S, Becker R, Leung H, Aylward L, Pyatt D. Biomonitoring equivalents: a screening approach for interpreting. Regul Toxicol Pharmacol 2007;47:96–109.
6. Bertail P, Clémençon S, Tressou J. A storage model with random release rate for modeling exposure to food contaminants. Math Biosci Eng 2008;35:35–60.
7. Verger P, Tressou J, Clémençon S. Integration of time as a description parameter in risk characterisation: application to methyl mercury. Regul Toxicol Pharmacol 2007;49:25–30.
8. Bertail P, Clémençon S, Tressou J. Statistical analysis of a dynamic model for dietary contaminant exposure. J Biol Dyn 2010;4:212–34.
9. Beaumont M, Zhang W, Balding D. Approximate Bayesian computation in population genetics. Genetics 2002;162: 2025–35.
10. Csilléry K, Blum M, Gaggiotti O, François O. Approximate Bayesian computation (ABC) in practice. Trends Ecol Evol 2010;25:490–1.

11. Blum M, Chi Tran V. HIV with contact tracing: a case study in approximate Bayesian computation. Biostatistics 2010;11:644–60.
12. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J Royal Soc Interface 2009;6:187–202.
13. Blum M. Approximate Bayesian computation: a nonparametric perspective. J Am Stat Assoc 2010;105:1178–87.
14. Joyce P, Marjoram P. Approximately sufficient statistics and Bayesian computation. Stat Appl Genet Mol Biol 2008;7:1544–6115.
15. Marjoram P. Markov Chain Monte Carlo without likelihoods. Proc Natl Acad Sci USA 2003;100:15324–8.
16. Rachev S. Probability metrics and the stability of stochastic models, 1991.
17. Lehmann E, Romano J. Testing statistical hypotheses, 2005.
18. Anses and InVS. Etude Nationale d'Imprégnation Aux Polychlorobiphényles des Consommateurs de Poissons d'Eau Douce (ICAR-PCB). Rapport d'Étude Scientifique et Synthèse, 2011.
19. Deurenberg P, Weststrate J, Seidell J. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. Br J Nutr 1991;65:105–14.
20. Leblanc J. CALIPSO, fish and seafood consumption study and biomarker of exposure to trace elements, pollutants and omega3, 2006.
21. Sirot V, Tard A, Venisseau A, Brosseaud A, Marchand P, Le Bizec B. Dietary exposure to polychlorinated dibenzo-p-dioxins, polychlorinated dibenzofurans and polychlorinated biphenyls of the French population: results of the second French total diet study. Chemosphere 2012;88:492–500.
22. Lioret S, Dubuisson C, Dufour A, Touvier M, Calamassi-Tran G, Maire B, Volatier J, Lafay L. Trends in food intake in french children from 1999 to 2007: results from the INCA (Étude Individuelle Nationale Des Consommations Alimentaires) dietary surveys. Br J Nutr 2010;103:585–601.
23. Sinkkonen S, Paasivirta J. Degradation half-life times of PCDDs, PCDFs and PCBs for environmental fate modeling. Chemosphere 2000;40:943–9.
24. Van Der Molen G, Kooijman S, Slob W. A generic toxicokinetic model for persistent lipophilic compounds in humans: an application to TCDD. Fundam Appl Toxicol 1996;31:83–94.
25. Aylward L, Hays S. Temporal trends in human TCDD body burden: decreases over three decades and implications for exposure levels. J Expo Anal Environ Epidemiol 2002;12:319–28.
26. Wegmann D, Leuenberger C, Excoffier L. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics 2009;182:1207–18.
27. Robert G, Gelman A, Gilks W. Weak convergence and optimal scaling of random walk metropolis algorithms. Ann Appl Probability 1997;7:110–20.
28. INERIS. Données technico économiques sur les substances chimiques en France: les polychlorobiphenyles (PCB), 2011. Available at: http://www.ineris.fr/substances/fr/.
29. Clémençon S, Tressou J. Exposition aux risques alimentaires et processus stochastiques: le cas des contaminants chimiques. J De La Société Française De Statistique 2009;150:3–29.
30. Centers for disease Control & Prevention. National Health and Nutrition Examination Survey, 2000. Available at: http://www.cdc.gov/nchs/about/major/nhanes/nhanes99_00.htmS.
31. Falqa G, Zeghnouna A, Pascala M, Vernayb M, Le Stratc Y, Garnierd R, Olichone D, Bretina P, Castetbonb K, Fréry N. Blood lead levels in the adult population living in France the French nutrition and health survey (ENNS 2006–2007). Environ Int 2011;37:565–71.
32. Sisson S, Fan Y, Tanaka M. Sequential Monte Carlo without likelihoods. Proc Natl Acad Sci USA 2007;104:1760–5.
33. Del Moral P, Doucet A, Jasra A. Sequential Monte Carlo samplers. J Royal Stat Soc Ser B 2006;68:411–32.
34. Reddy M, Yang R, Andersen M, Clewell H. Physiologically based pharmacokinetic modeling: science and applications. Hoboken, New Jersey: John Wiley and Sons, 2005.

Note: Supplemental Material: The online version of this article (DOI: 10.1515/ijb-2013-0062) offers supplementary material, available to authorized users.