Recommendations to address uncertainties in environmental risk assessment using toxicokinetics-toxicodynamics models

Virgile Baudrot[1*] and Sandrine Charles[1]

[1]Univ Lyon, Université Lyon 1, UMR CNRS 5558, Laboratoire de Biométrie et Biologie Évolutive, F-69100 Villeurbanne, France

Abstract

Providing reliable environmental quality standards (EQS) is a challenging issue for environmental risk assessment (ERA). These EQS are derived from toxicity endpoints estimated from dose-response models to identify and characterize the environmental hazard of chemical compounds as those released by human activities. The classical toxicity endpoints are the x% effect/lethal concentrations at a specific time t ($EC/LC(x,t)$), or the multiplication factors applied to environmental exposure profiles leading to x% of effect reduction at a specific time t ($MF(x,t)$). However, classical dose-response models used to estimate the toxicity endpoints have some weaknesses such as their dependency on observation time-points which are likely to differ between species. Also, real exposure profiles are hardly ever constant over time, what makes impossible the use of classical dose-response models and compromises the derivation of $MF(x,t)$, actually designed to tackle time-variable exposure profiles. When dealing with survival or immobility toxicity test data, these issues can be overcome with the use of the General Unified Threshold model of Survival (GUTS), a toxicokinetics-toxicodynamics (TKTD) model, providing an explicit framework to analyse both time and concentration-dependent data sets, as well as a mechanistic derivation of $EC/LC(x,t)$ and $MF(x,t)$ whatever x and at any time t of interest. In addition, the assessment of a risk is inherently built upon probability distributions, so that the next critical step for ERA is to characterize uncertainties of toxicity endpoints, and sequentially of EQS. The innovative approach investigated in our paper is the use of the Bayesian framework to deal with uncertainties raising in the calibration process and propagated all along the successive prediction steps until the $LC(x,t)$ and $MF(x,t)$ derivations. We also explored the mathematical properties of $LC(x,t)$ and $MF(x,t)$ as well as the impact of different experimental designs in order to provide some recommendations for a robust derivation of toxicity endpoints leading to reliable EQS.

Keywords. Survival models ; Dose-Response ; GUTS ; Lethal Concentration ; Multiplication Factor ; Margin of safety ; Environmental Risk Assessment

1. Introduction

Assessing the environmental risk of chemical compounds requires environmental quality standards (EQS) such as PNECs, RACs and MAC-EQS under the ECHA, EFSA PPR and WFP regulatory

frameworks respectively (EFSA PPR, 2013; ECHA, 2017). Derivation of EQS results from the combination of assessment factors with toxicity endpoints mainly derivated from estimated or measured exposure response of a set of target species to that chemical compound (EFSA PPR, 2013; Isigonis et al., 2015; Syberg and Hansen, 2016; ECHA, 2017). Deriving reliable toxicity endpoints is challenging and the subject matter is very controversial (Laskowski, 1995; Jager, 2011). Today, Environmental Risk Assessment (ERA) rests on fitting classical dose-response models to quantitative toxicity test data. For acute effect assessment, such data are collected from standard toxicity tests, from which the 50% lethal or effective concentration (LC_{50} or EC_{50}) is generally estimated at the end of the exposure duration, meaning that the monitoring of observations over time is not fully exploited. In addition, classical dose-response models implicitly assume that the exposure concentration remains constant all along the experiment, what makes difficult to extrapolate the results to more realistic scenarios with time-variable exposure profiles combining different heights, widths and frequencies of contaminant pulses (Reinert et al., 2002; Brock, 2009; Jager, 2011; Ashauer et al., 2013). To overcome this gap at the organism level, the use mechanistic models, such as toxicokinetics-toxicodynamics (TKTD) models, is now promoted in order to describe the effects of a substance of interest by integrating the dynamics of the exposure (Jager et al., 2011; EFSA PPR, 2013; Hommen et al., 2016). Indeed, TKTD models appear highly advantageous in terms of mechanistic understanding of the chemical mode of action, of deriving time-independent parameters, of interpreting time-varying exposure and of making predictions under untested situations (Jager et al., 2011; Ashauer et al., 2013). Another of their advantages for ERA is the possible calculation of any x% lethal $LC(x,t)$ or effective $EC(x,t)$ whatever x and at any given exposure duration t. Also, from time-variable concentration profiles as observed in the environment, it is possible to estimate a margin of safety such as the exposure multiplication factor, $MF(x,t)$, leading to any x% of effect reduction due to the contaminant at any time t (Ashauer et al., 2013). When focusing on survival rate of individuals, a General Unified Threshold model of Survival (GUTS) has been proposed to unify the majority of TKTD survival models (Jager et al., 2011). In the present paper, we consider the two most used derivations named Stochastic Death (GUTS-RED-SD) and Individual Tolerance (GUTS-RED-IT). GUTS-RED-SD model assumes that all individuals are identically sensitive to the chemical substance by sharing a common internal threshold concentration and that mortality is a stochastic process once this threshold is reached. On the contrary, GUTS-RED-IT model is based on the Critical Body Residues (CBR) approach, which assumes that individuals differ in their threshold, following a probability distribution, and die as soon as the internal concentration reaches the individual-specific threshold (Jager et al., 2011). The robustness of GUTS models for calibration and prediction has been widely demonstrated in previous studies, with little differences between both GUTS-RED-SD and GUTS-RED-IT models in terms of calibration and prediction (Ashauer et al., 2013; Baudrot et al., 2018c; Jager and Ashauer, 2018). Sensitivity analysis

of toxicity endpoints derivated from GUTS, such as $LC(x,t)$ and $MF(x,t)$, have also been investigated (Ashauer et al., 2013; Baudrot et al., 2018c), but the question of how uncertainties are propagated is still under-studied.

Quantifying uncertainties or level of confidence associated with toxicity endpoints is undoubtedly a way to improve trust in risk predictors and to avoid decision that could increase, rather than decrease, the risk (Gray and Cohen, 2012; Beck et al., 2016). The Bayesian framework has many advantages to deal with uncertainties since the distribution of parameters, and so their uncertainties, is embedded in the inference process. While the construction of priors on model parameters can be seen as a carrier of subjectivity (Ferson, 2005), there is a proved added-value by taking advantage of information from the experimental design (Delignette-Muller et al., 2017; Baudrot et al., 2018c). Consequently, coupling TKTD models with Bayesian inference allows to estimate the probability distribution of toxicity endpoints, and any other predictions coming from the mechanistic (TKTD) model, by taking into account all the constraints resulting from the experimental design. Moreover, Bayesian inference, which revealed particularly efficient with GUTS models (Delignette-Muller et al., 2017; Baudrot et al., 2018c), can also be used to optimize the experimental design by quantifying the gain of knowledge from priors to posteriors (Albert et al., 2012). At last, Bayesian inference is also tailored for decision making as it confronts the assessors with a range of values, rather than just a point, which is particularly valuable for risk assessment (Ferson, 2005; Gray and Cohen, 2012).

In the present study, we explore how scrutinizing uncertainties helps to provide recommendations on the experimental design and the characteristics of toxicity endpoints used for EQS, while maximizing their reliability. We first give an overview of TKTD models with a focus on GUTS (Jager et al., 2011). Handling GUTS models within the R-package *morse* (Baudrot et al., 2018a) is then illustrated with five example data sets. Then, we explore how a variety of experimental designs influence the uncertainties in derived $LC(x,t)$ and $MF(x,t)$. Finally, we provide a set of recommendations on the use of TKTD models for ERA, based on their added-value and the way the uncertainty may be handled under the Bayesian framework.

2. Material and methods

2.1. Data from experimental toxicity tests

We used experimental toxicity data sets, detailled in (Ashauer et al., 2011; Nyman et al., 2012; Ashauer et al., 2016) testing all together five chemical compounds (carbendazim, cypermethrin, dimethoate, malathion and propiconazole) on the survival rate of the amphipod crustacean *Gammarus pulex*. Two experiments were performed for each compound, one exposing *G. pulex* to constant concentrations, and the other exposing *G. pulex* to time-variable concentrations (see Table 1). In constant exposure experiments, *G. pulex* was exposed to eight concentrations for four days. In time-variable exposure

experiments, *G. pulex* was exposed to two different pulse profiles, consisting in two one-day exposure pulses with short and longer interval between them.

Table 1: Characteristics of data sets used in the manuscript. The "Profile type" is the type of exposure profiles (constant or time-variable), "Data points" refers to the number of data points in the data set, "Nbr profiles" is the number of profiles in the dataset "N_{init}" is the initial number of individuals in profile, "Nbr days" is the number of days for each experiment, and "Time points per profile" is the number of time points for each time series (each constant profiles consisted in 5 time-points).

Product	Profile type	Data points	Nbr profiles	N_{init}	Nbr days	Time points per profile
Carbendazim	constant	40	8	20	4	5
Cypermethrin	constant	40	8	20	4	5
Dimethoate	constant	40	8	20	4	5
Malathion	constant	40	8	20	4	5
Propiconazole	constant	40	8	20	4	5
Carbendazim	variable	51	4	80	10	$[8, 14, 16, 13]$
Cypermethrin	variable	61	4	80	10	$[10, 18, 18, 15]$
Dimethoate	variable	58	4	80	10	$[10, 16, 17, 15]$
Malathion	variable	70	2	70	22	$[35, 35]$
Propiconazole	variable	74	4	70	10	$[11, 21, 21, 21]$

2.2. GUTS modelling

In the following, we detail the mathematical equations of GUTS models describing the survival rate over time for organisms exposed to a profile of concentrations of a single chemical product. All other possible derivations of GUTS models are fully described in (Jager et al., 2011; Jager and Ashauer, 2018). We provide here a summary of GUTS-RED-SD and GUTS-RED-IT reduced models in order to introduce notations and equations relevant for mathematical derivation of explicit formulations of the x% Lethal Concentration at time t, denoted $LC(x,t)$, and of the Multiplication Factor leading to x% mortality at time t, denoted $MF(x,t)$.

2.2.1. Toxicokinetics

We denote $C_w(t)$ the external concentration of a chemical product which can be variable over time. As there is no measure of internal concentration, we use the scaled internal concentration, denoted $D_w(t)$, which is therefore a latent variable as described by the toxicokinetics part of the model as follows:

$$\frac{dD_w(t)}{dt} = k_d(C_w(t) - D_w(t)) \tag{1}$$

where k_d $[time^{-1}]$ is the dominant rate constant, corresponding to the slowest compensating process dominating the overall dynamics of toxicity.

As we assume that the internal concentration equal 0 at $t = 0$, the explicit formulation for constant

concentration profiles is given by:

$$D_w(t) = C_w \left(1 - e^{-k_d t}\right) \tag{2}$$

An explicit expression for time-variable exposure profiles is provided in Supplementary Material as it can be usefull for implementation, but not for mathematical calculus below. The GUTS-RED-SD and GUTS-RED-IT models are based on the same model for the scaled internal concentration. They do not differ in the TK part, but in the TD part describing the death mechanism.

From the toxicokinetics equation (2), we can easily compute the $x\%$ depuration time, that is the period of time after a pulse leading to $x\%$ of reduction in the scaled internal concentration:

$$DRT_x = \frac{-\log(x\%)}{k_d} \tag{3}$$

2.2.2. Toxicodynamics

Model GUTS-RED-SD:. The GUTS-RED-SD model supposes that all the organisms have the same internal threshold concentration, denoted z $[mol.L^{-1}]$, and that, once exceeded, the instantaneous probability to die, named $h(t)$, increases linearly with the internal concentration. The mathematical equation is:

$$h(t) = b_w \max_{0 \leq \tau \leq t} (D_w(t) - z, 0) + h_b \tag{4}$$

where b_w $[L.mol.time^{-1}]$ is the killing rate and h_b $[time^{-1}]$ is the background mortality rate.

Then, the survival probability along time under GUTS-RED-SD model is given by:

$$S_{SD}(t) = \exp\left(-\int_0^t h(t)\, dt\right) \tag{5}$$

Model GUTS-RED-IT:. The GUTS-RED-IT model supposes that the threshold concentration is distributed among organisms, and that the death is immediate as soon as this threshold is reached. The probability to die at the maximal internal concentration with background mortality h_b is given by:

$$S_{IT}(t) = \exp(-h_b t)(1 - F(\max_{0 < \tau < t} (D_w(\tau)))) \tag{6}$$

Assuming a log-logistic function, we get $F(x) = \dfrac{1}{1 + (x/m_w)^{-\beta}}$, with m_w $[mol.L^{-1}]$ the median and β the shape of the threshold distribution, what gives:

$$S_{IT}(t) = \exp(-h_b t) \left(1 - 1/\left(1 + \left(\frac{\max_{0 \leq \tau \leq t} (D_w(\tau))}{m_w}\right)^{-\beta}\right)\right) \tag{7}$$

2.3. Implementation and Bayesian inference

GUTS models were implemented within a Bayesian framework through *JAGS* (Plummer, 2016) by using the R-package *morse* (Baudrot et al., 2018a). The Bayesian inference methods, choice of priors and parameterisation of MCMC process have previously been fully explained (Delignette-Muller et al., 2017; Baudrot et al., 2018c,a). The joint posterior distribution of parameters was used to predict survival curve under tested and untested exposure profiles, for the calculation of $LC(x,t)$ and $MF(x,t)$, and for the computing of goodness-of-fit measures (see hereinafter). The use of the joint posterior distribution allow to quantify the uncertainty around all these predictions, and therefore the computing of their median and their 95% credible intervals as follow: under a specific exposure profile, we simulated the survival rate over time for every joint posterior parameter set ; then at each time point of the time series, we computed 0.5, 0.025 and 0.975 quantiles, thus providing median and 95% limits.

2.4. Measures of model robustness

Modelling is always associated with testing its robustness: the robustness in fitting data used for calibration, but also the robustness for predictions on new data (Grimm and Berger, 2016). To evaluate the robustness of estimations and predictions with the two GUTS models, we calculated their statistical properties by means of the Normalized Root Mean Square Error (NRMSE), the Posterior Predictive Check (PPC), the Watanabe-Akaike Information Criterion and the Leave-One-Out Cross-Validation (LOO-CV) (Gelman et al., 2013).

2.4.1. Normalized Root Mean Square Error

The Root Mean Square Error (RMSE) allows to charaterize the difference between observations and predictions from the posterior distribution. With N observations and $y_{i,obs}$ the observed number of individuals ($i \in \{1,\dots,N\}$), then for each estimation $y_{.,j}$ of the markov chain of size M ($j \in \{1,\dots,M\}$) resulting from the Bayesian inference, we can define the $RMSE_j$ such as:

$$RMSE_j = \sqrt{\frac{1}{N}\sum_{i}^{N}(y_{i,j} - y_{i,obs})^2} \quad \Rightarrow \quad NRMSE_j = \frac{RMSE_j}{\overline{y_{obs}}} \tag{8}$$

Where Normalized RMSE (NRMSE) is given by dividing RMSE with the mean of the observations denoted $\overline{y_{obs}}$. We then have the distribution of the NRMSE from which we returned the median and the 95% credible interval in Table 2.

2.4.2. Posterior Predictive Check (PPC)

The Posterior Predictive Check consists in comparing replicated data drawn from the joint posterior distribution to observed data. A measure of goodness-of-fit is the percentage of observed data lying within the 95% predicted credible intervals (Gelman et al., 2013).

2.4.3. WAIC and LOO-CV

Information criteria as WAIC and LOO-CV are common measures of predictive precision also used to compare models. The WAIC is the sum of the log predictive density computed for every points, to which a bias is added to take into account the number of parameters. The LOO-CV use the log predictive density estimated from a training subset and applied it on another one (Gelman et al., 2013). Both WAIC and LOO-CV were computed with the R-package *bayesplot* (Gabry and Mahr, 2017).

2.5. Mathematical definition and properties of $LC(x,t)$

The $LC(x,t)$ makes sense only in the situation of constant exposure profiles (i.e., whatever time t, $C_w(t)$ is constant). In such situations, we can provide an explicit formulation of the survival rate over time considering both models GUTS-RED-SD and GUTS-RED-IT. Many software provides an implementation of GUTS models what facilitate the possibility to compute the $LC(x,t)$ at any time and any x% (Jager and Ashauer, 2018). Our Bayesian implementation of GUTS models using the R language is one of them (Baudrot et al., 2018a).

Let $LC(x,t)$ be the lethal concentration for x% of organisms at any time t, and $S(C,t)$ be the survival rate at constant concentration C and time t. Then, the $LC(x,t)$ is defined as:

$$S(LC(x,t),t) = S(0,t)\left(1 - \frac{x}{100}\right) \tag{9}$$

where $S(0,t)$ is the survival rate at time t when there is no contaminant, which reflects the background mortality.

2.5.1. GUTS-RED-SD model

The lethal concentration $LC_{SD}(x,t)$ is given by:

$$LC_{SD}(x,t) = \frac{-k_d \ln\left(1 - \frac{x}{100}\right)}{b_w\left(k_d(t-t_z) - e^{-k_d t_z} + e^{-k_d t}\right)} + \frac{k_d z(t-t_z)}{k_d(t-t_z) - e^{-k_d t_z} + e^{-k_d t}} \tag{10}$$

As mention in Supplementary Material, under time-variable exposure, t_z is also variable with time, while in the case of constant exposure, t_z is exactly $-1/k_d \ln(1 - z/C_w)$. When time increase, the $LC_{SD}(x,t)$ curve become a vertical line at point z, and when time tends to infinity, the convergence is:

$$\lim_{t\to+\infty} LC_{SD}(x,t) = z \quad , \quad \text{with} \quad t_z = \frac{-1}{k_d}\ln\left(1 - \frac{z}{LC_{SD}(x,t)}\right) \tag{11}$$

193 *2.5.2. GUTS-RED-IT model*

194 The lethal concentration $LC_{IT}(x,t)$ is given by:

$$LC_{IT}(x,t) = \frac{m_w}{(1-e^{-k_d t})} \sqrt[\beta]{\frac{x}{100-x}} \tag{12}$$

195 It is then straightfoward to see that when t increases, the $LC_{IT}(x,t)$ converges to:

$$\lim_{t\to+\infty} LC_{IT}(x,t) = m_w \sqrt[\beta]{\frac{x}{100-x}} \tag{13}$$

196 In the specific case of $x = 50\%$, we get: $\lim_{t\to+\infty} LC(50,t) = m_w$.

197 *2.5.3. Calculation of the density distribution of $LC(x,t)$*

198 The calculation of $LC(x,t)$ is based on equation (9). Then, using the GUTS models and the
199 estimates of parameters from the calibration processes, we compute the survival rate without contam-
200 ination (i.e., the background mortality denoted $S(0,t)$) and a set of prediction of the survival rate over
201 a range of concentrations (i.e., $S(C,t)$). This process provides the distribution of the $LC(x,t)$ using
202 equation (9).

203 *2.6. Mathematical definition and properties of the multiplication factor $MF(x,t)$*

204 Contrary to the lethal concentration $LC(x,t)$ used in situations of constant exposure profiles, the
205 multiplication factor, $MF(x,t)$ can be computed for both constant and time-variable exposure profiles.
206 With the exposure profile $C_w(\tau)$, with τ running from 0 to t, the $MF(x,t)$ is defined as:

$$S\left(MF(x,t) \times C_w(\tau), t\right) = S\left(0,t\right)\left(1 - \frac{x}{100}\right) \tag{14}$$

207 In the Supplementary Material, we show that the internal damage $D_w(t)$ is linearly related to
208 the multiplication factor since whatever the exposure profile (constant or time-variable), we get the
209 following relation:

$$D_w^{MF}(t) = MF(x,t) \times D_w(t) \tag{15}$$

210 where $D_w^{MF}(t)$ is the internal damage when the exposure profile is multipied by $MF(x,t)$.

211 *2.6.1. GUTS-RED-SD model*

212 The multiplication factor $MF_{SD}(x,t)$ is given by:

$$MF_{SD}(x,t) = \frac{-\frac{1}{b_w}\ln\left(1-\frac{x}{100}\right) + \int_0^t \max_{0<\tau<t}\left(D_w(\tau) - z\,,\,0\right)d\tau}{\int_0^t \max_{0<\tau<t}\left(D_w(\tau) - \frac{z}{MF(x,t)}\,,\,0\right)d\tau} \tag{16}$$

When the external concentration is constant, we can use the explicit expression of $D_w(t)$ for $C_w(t) = C_w$, and get:

$$MF_{SD}(x,t) = \frac{-\frac{1}{b_w}\ln\left(1-\frac{x}{100}\right) + \frac{C_w}{k_d}\left(e^{-k_d t} - e^{-k_d t_z}\right) + (C_w - z)(t - t_z)}{\frac{C_w}{k_d}\left(e^{-k_d t} - e^{-k_d t_{z,MF}}\right) + \left(C_w - \frac{z}{MF(x,t)}\right)(t - t_{z,MF})} \tag{17}$$

where t_z has been previously defined and $t_{z,MF} = \frac{-1}{k_d}\ln\left(1 - \frac{z}{MF(x,t)C_w}\right)$. As for the $LC_{SD}(x,t)$, the expression of $t_{z,MF}$ prevents to have a whole explicit formulation of $MF_{SD}(x,t)$.

2.6.2. GUTS-RED-IT model

The multiplication factor $MF_{IT}(x,t)$ is given by:

$$MF_{IT}(x,t) = \sqrt[\beta]{\frac{100 + x\left(\frac{\max\limits_{0<\tau<t}(D_w(\tau))}{m_w}\right)^{-\beta}}{100 - x}} \tag{18}$$

Therefore, from a GUTS-RED-IT model, solving the toxicokinetics part giving $\max\limits_{0<\tau<t}(D_w(\tau))$ is enough to find any multiplication factor for any x at any t. When the external concentration is constant, this maximum is $C_w\left(1 - e^{-k_d t}\right)$.

3. Results

3.1. Goodness-of-fit of GUTS-RED-SD and GUTS-RED-IT

For all compounds, Table 2 shows that fitting on constant exposure profiles provide better fit than for time-variable exposure profiles (see also graphics of Posterior Predictive Check in Supplementary Material), whatever the measure of goodness-of-fit (except with the NRMSE measure of GUTS-RED-IT on dimethoate). This result could be expected since, as pointed by Table 1, there are always more time series in data sets with constant exposure profiles. But also, since there are explicit solutions of differential equations with constant exposure profiles for both models GUTS-RED-SD and GUTS-RED-IT, the computing process is easier contrary to time-variable exposure profile requiring the use of a numerical integrator.

For validation, whatever the measure of goodness-of-fit, predictions are always better when parameters are calibrated on data sets with variable exposure profiles to then predict on data set under constant exposure profiles, than the other way round.

Based on Table 2, it is hard to differentiate GUTS-RED-SD from GUTS-RED-IT with the quality of their fits. At least, we can notice that GUTS-RED-IT model is particularly bad for Carbendazim and Dimethoate under time-variable exposure profiles. Still under variable exposure profiles, for Malathion

and Propiconazole data sets, we can observed a large 95% credible interval for GUTS-RED-IT (see figures in Supplementary Material). While NRMSE and %PPC tend to better qualified GUTS-RED-IT, the uncertainty is penalized with Bayesian measures WAIC and LOO-CV. In fact, the percentage of recovery extracted from a PPC is totally blind to point large credible interval, since it increases when the credible interval increases.

Table 2: Results of calibration and validation of models GUTS-RED-SD and GUTS-RED-IT for the five chemical compounds: Carbendazim (car), Cypermethrin (cyp), Dimethoate (Dim), Malathion (mal) and Propiconazole (prz). Profiles of exposure concentration are either constant, denoted *cst*, or variable, denoted *var*. The notation $cst \rightarrow var$ means that calibration was done on data set of constant exposure and validation was done on data set oftime-variable exposure profile (see data set in Table 1). The measures NRMSE, %PPC, WAIC and LOO-CV assess the goodness-of-fit and are fully explained in section 2.4.

		GUTS SD				**GUTS IT**			
Product	Profile	NRMSE	%PPC	WAIC	LOO-CV	NRMSE	%PPC	WAIC	LOO-CV
Calibration									
car	cst	0.112	100	402.41	403.27	0.124	100	420.11	422.09
cyp	cst	0.095	100	196.37	206.78	0.092	100	188.07	189.09
dim	cst	0.122	97.5	308.94	309.41	0.171	90.0	357.38	358.74
mal	cst	0.090	100	248.87	249.59	0.112	92.5	273.01	273.54
prz	cst	0.102	100	282.03	285.57	0.118	80.0	308.03	314.93
car	var	0.159	82.1	1006.0	1012.1	0.499	32.1	1222.4	1216.4
cyp	var	0.196	91.7	829.04	833.48	0.116	97.2	793.95	801.23
dim	var	0.129	83.3	1224.8	1226.8	0.161	55.6	1357.2	1344.7
mal	var	0.196	97.8	762.58	766.76	0.148	100	908.56	934.80
prz	var	0.164	95.5	951.10	894.02	0.038	97.7	3262.8	1436.2
Validation: data used for parameter calibration $\rightarrow$ data for prediction and goodness-of-fit									
car	cst $\rightarrow$ var	0.159	42.9	17709	4578.4	0.148	50.0	12800	4541.0
cyp	cst $\rightarrow$ var	0.196	91.7	1760.5	1423.5	0.183	88.9	1283.4	1141.0
dim	cst $\rightarrow$ var	0.129	83.3	1845.7	1685.3	0.199	63.9	1708.7	1628.9
mal	cst $\rightarrow$ var	0.196	67.4	10162	2610.7	0.169	63.0	1258.5	1286.1
prz	cst $\rightarrow$ var	0.164	95.5	940.54	900.90	0.176	90.9	894.41	940.74
car	var $\rightarrow$ cst	0.164	67.5	537.14	537.79	0.228	90.0	437.01	437.01
cyp	var $\rightarrow$ cst	0.071	82.5	537.62	488.90	0.051	87.5	453.65	378.89
dim	var $\rightarrow$ cst	0.013	97.5	302.24	302.30	0.157	87.5	389.32	393.68
mal	var $\rightarrow$ cst	0.053	80.0	470.28	512.86	0.049	90.0	869.45	732.94
prz	var $\rightarrow$ cst	0.040	77.5	797.60	660.09	0.041	80.0	1661.3	1107.8

3.2. Comparison of $LC(x,t)$ with GUTS-RED-SD and GUTS-RED-IT

There is no obvious difference between GUTS-RED-SD and GUTS-RED-IT in their goodness-of-fit nor in the calculation of $LC(x,t)$ along time t or percentage of affected population, x.

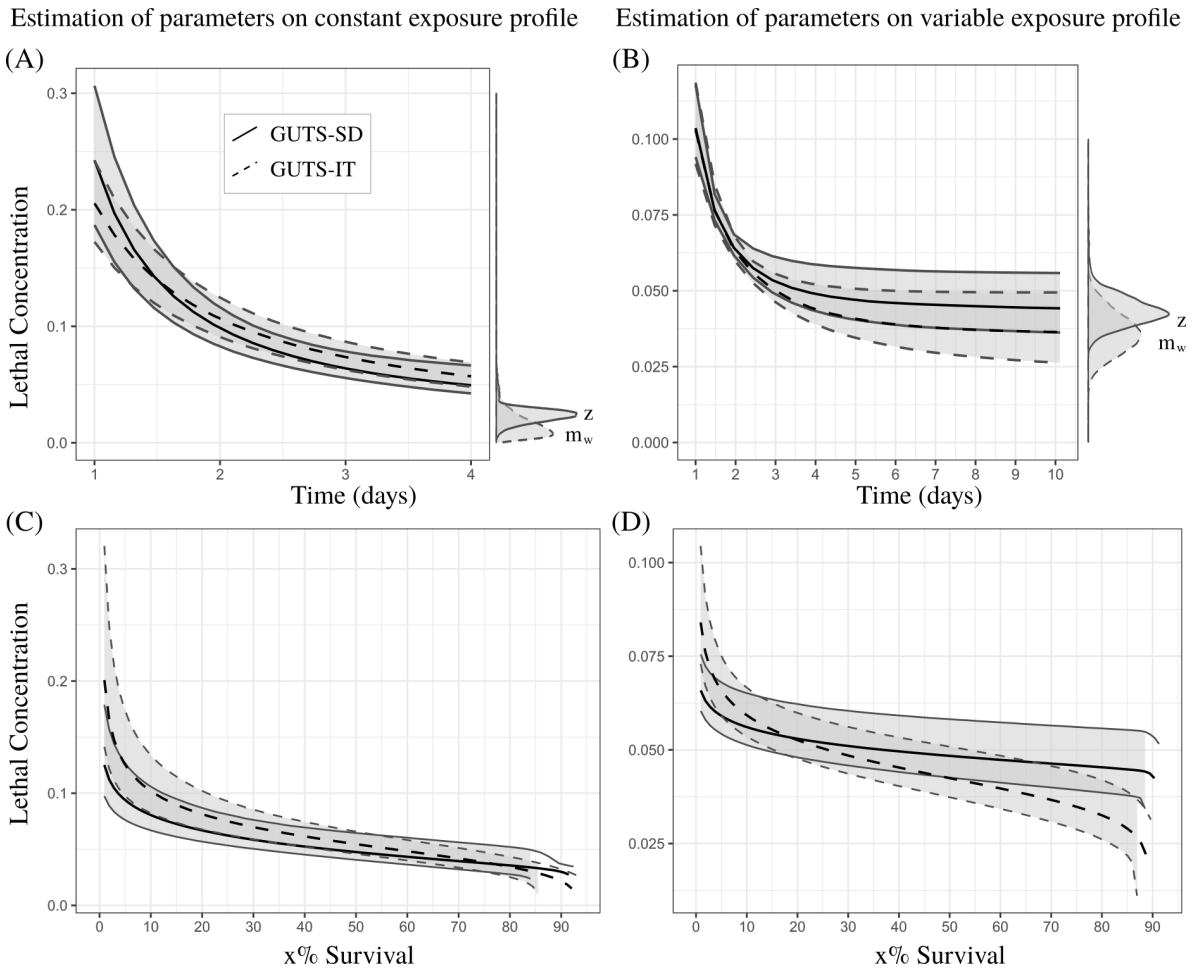

Figure 1: Comparison of $LC(x,t)$ for GUTS-RED-SD, solid lines, and GUTS-RED-IT, dashed lines, for Cypermethrin (see Supplementary Material for other compounds). Parameters are estimated on data under constant (A, C) and variable (B, D) concentration profiles. Black lines are median and grey zones are 95 % credible bands. (A, B) Lethal concentration for 50% of the organisms ($LC(50,t)$) from day 1 to the end of the experiment. (C, D) Lethal concentration at the end of experiment (4 and 10 days respectivelly) along the percentage of population affected.

3.2.1. $LC(x,t)$ as a function of time t

As expected, from Figures 1-(A,B) and Supplementary Material, we see that $LC(x,t)$ decreases with time. Rarely pointed is the shape of this decrease which is exponential and converges toward different values according to the model. This asymptotic behavior is known as the incipient $LC(x,t)$ (Jager et al., 2006). A direct consequence for risk assessors is that evaluation of $LC(x,t)$ at early time induces higher uncertainty than at a later time (specific time being relative to the species and the compound). In other words, the sensitivity of $LC(x,t)$ to time t decreases as long as t increases. For instance, we see on Figures 1-(A,B) that a little change in time around day 2 leads to a greater change

in the estimation of $LC(x,t)$ than around day 4. However, we have to note that the uncertainty of $LC(x,t)$ does not always decreases when time increases. For instance, in Figure 1-(B), the uncertainty at day 6 and afterwards is greater than around day 3.

When t increases to infinity, the $LC(x,t)$ convergences towards the distribution of parameter z for GUTS-RED-SD (see equation (11)) and $m_w \sqrt[\beta]{\frac{x}{100-x}}$ for GUTS-RED-IT (see equation (13)). The specific $LC_{50,t}$ tends to z for GUTS-RED-SD and to m_w for GUTS-RED-IT (see equations (11) and (13)). The recommendation for risk assessors would be to use the advantages of TKTD models in order to extrapolate the $LC(x,t)$ on a longer period than the duration of the experiment in order to visualize the uncertainties around the incipient $LC(x,t)$ defined by the asymptote. At least, we recommend to look at the $LC(x,t)$ at the last time of the experiment, what is in line with the common procedure in ERA.

3.2.2. $LC(x,t)$ as a function of percentage of affected population, x

From Figure 1-(C,D), we can see that the uncertainty of $LC(x,t)$ is greater at low x, that is when the effect of the contaminant is weak. While computing $LC(x,t)$ at $x > 50\%$ is never used for ERA, we can also see that the uncertainty of $LC(x,t)$ increases when x tends to 100%. As a consequence, while the uncertainty is not always minimal at the standard value of $x = 50\%$, it seems to be always smaller around this value than around $x = 10\%$ another classical value used in ERA. Consequently, for risk assessors, while TKTD models allow to compute the $LC(x,t)$ whatever x, if only one value has to be chosen, we recommend to keep the standard of $x = 50\%$. On the other hand, the risk assessor has to keep in mind that 50% is not the optimal threshold in term of reduction of uncertainty, depending on the data set, the model (GUTS-RED-SD or GUTS-RED-IT) and the parameter estimates.

3.3. Comparison of $MF(x,t)$ with GUTS-RED-SD and GUTS-RED-IT

3.3.1. $MF(x,t)$ as a function of time t

As expected, Figures 2-(D-F) show that the multiplication factor is decreasing when the time at which the survival rate is checked increases. In other words, the later the survival rate is assessed, the lower is the multiplication factor. Also, these graphics reveal there is no typical pattern of curves of multiplication factors over time t of exposure. Under a constant exposure profile, the curve shows an exponential decreasing pattern, while under pulsed exposure, we observe a constant phase and, surrounding peaks, an sudden decrease of the multiplication factor. The multiplication factor is obviously highly variable around a pulse in the concentration of the chemical product. Therefore, a recommendation would be to wait for some times (e.g., several days) after a peak before computing a multiplication factor. More generally, the multiplication factor is designed to be compared with the assessment factor (AF) classically used in concert with the effect/lethal concentration value based on realistic time-variable exposure profiles to derive an EQS. As a consequence, when using $MF(x,t)$

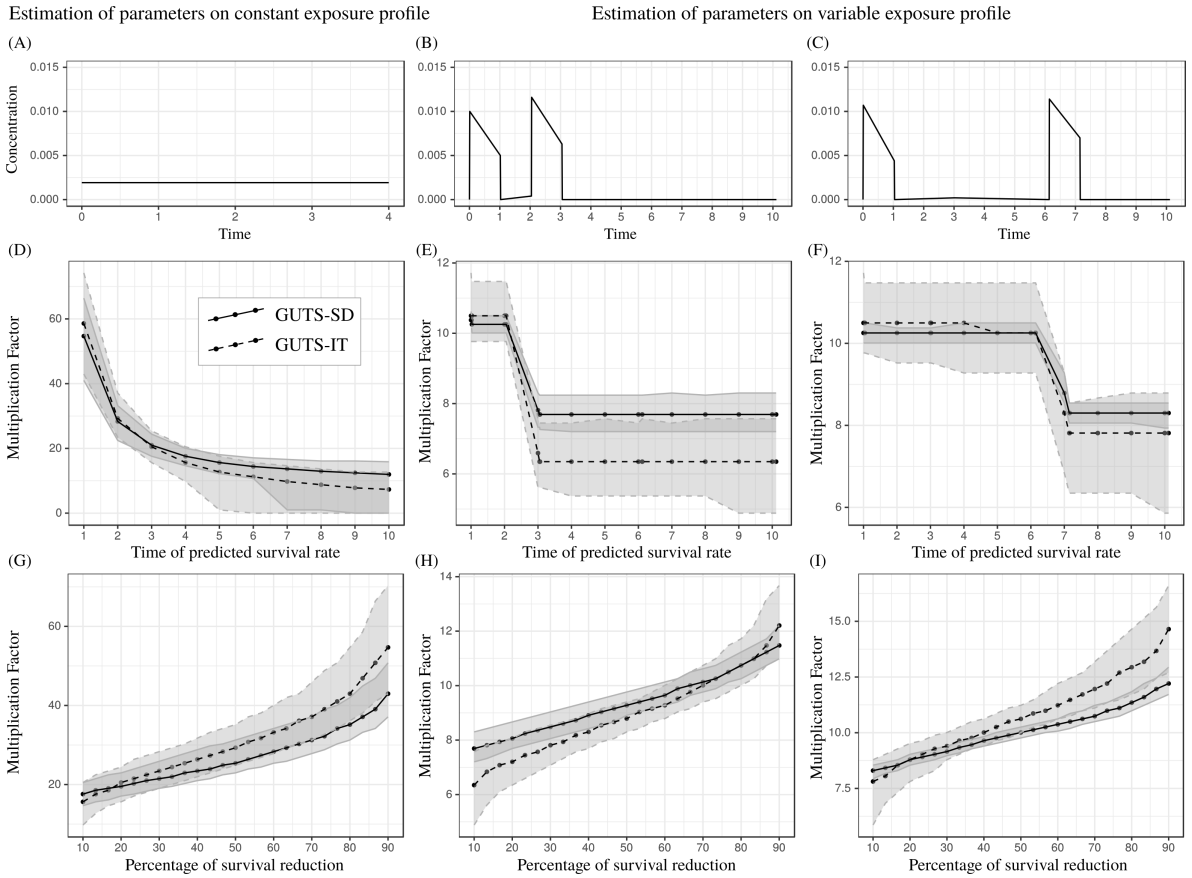

Figure 2: Comparison of $MF(x,t)$ for GUTS-RED-SD, solid lines, and GUTS-RED-IT, dashed lines, for Cypermethrin (see Supplementary Material for other compounds). Parameters are estimated on data under constant (A, D, G) and variable (B, C, E, F, H, I) concentration profiles. (A-C) Exposure profiles, (D-F) Multiplication factors estimated for a 10% reduction of survival (i.e. $MF(x = 10,t)$) along time. (G-I) Multiplication factors estimated at the end of experiments (time 4 for (G) and 10 for (H, I)) along the percentage of survival reduction.

based on real exposure profiles, it is important to pay close attention to the amplitudes and frequencies of pulses, as well as to the times at which multiplication factors are computed. As for the $LC(x,t)$, taking advantage of TKTD capabilities to predict at any time is of real interest to described the survival response under pulsed exposure.

3.3.2. $MF(x,t)$ as a function of percentage of survival reduction x

Logically, Figures 2-(G-I) show that the multiplication factor increases with the increase of the percentage of reduction of the survival rate. An interesting results is the non-linearity of this increase. As for the $LC(x,t)$, the uncertainty is greater at low and high percentages compared to what happens in the middle around 50% of survival reduction. As a consequence, it would be relevant to fix 50% as a standard for ERA.

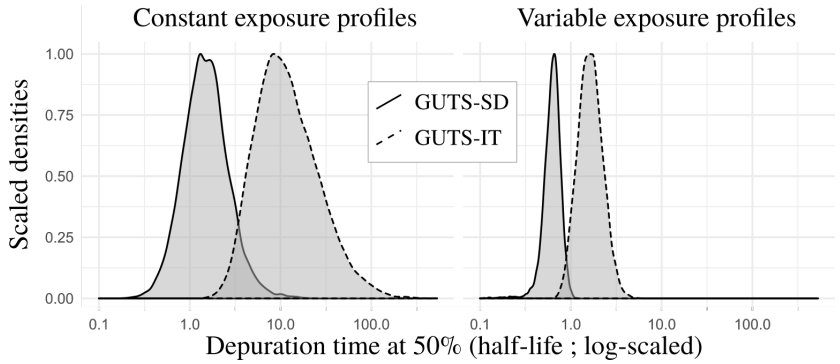

Figure 3: Distribution of estimated half-life for cypermethrin in GUTS-RED-SD and GUTS-RED-IT models for data sets under constant (left) and variables (right) exposure profiles. Median and 95% Credible interval of 50% depuration time are under constant exposure profiles 1.48 $[0.502, 5.00]$ for GUTS-RED-SD and 10.8 $[3.21, 68.5]$ for GUTS-RED-IT, and under variable exposure profiles 0.633 $[0.386, 0.890]$ for GUTS-RED-SD and 1.62 $[0.917, 3.06]$ for GUTS-RED-IT.

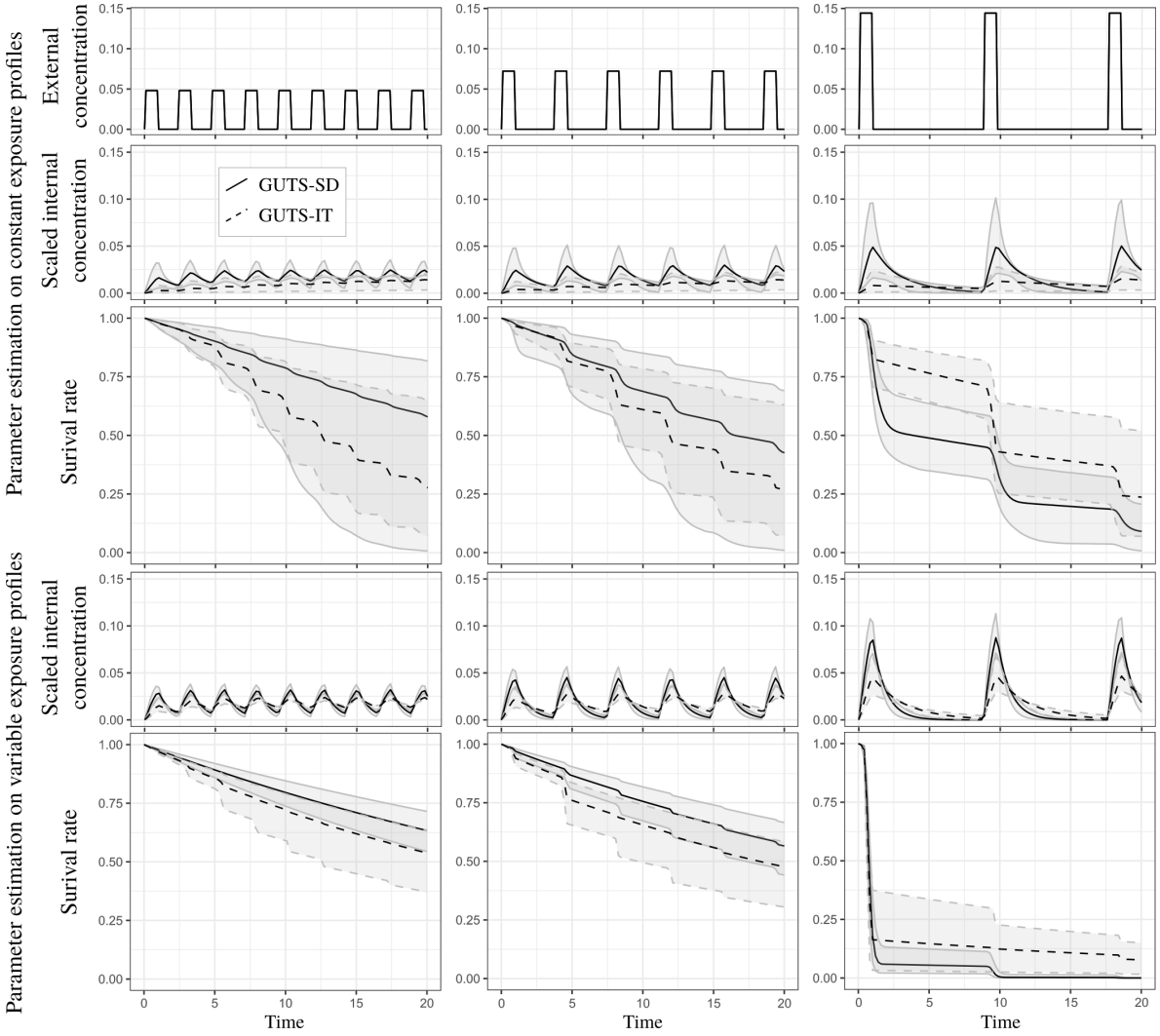

Figure 4: Survival rate over time with GUTS-RED-SD and GUTS-RED-IT models (respectivelly solid and dashed lines) under different exposure profiles with the same area under the curve (differences are in the duration time after pulses and in the maximal concentration of pulses). Parameters were estimated from the Cypermethrin data set, either under constant (upper panel of the figure) or time-variable (lower panel of the figure) exposure.

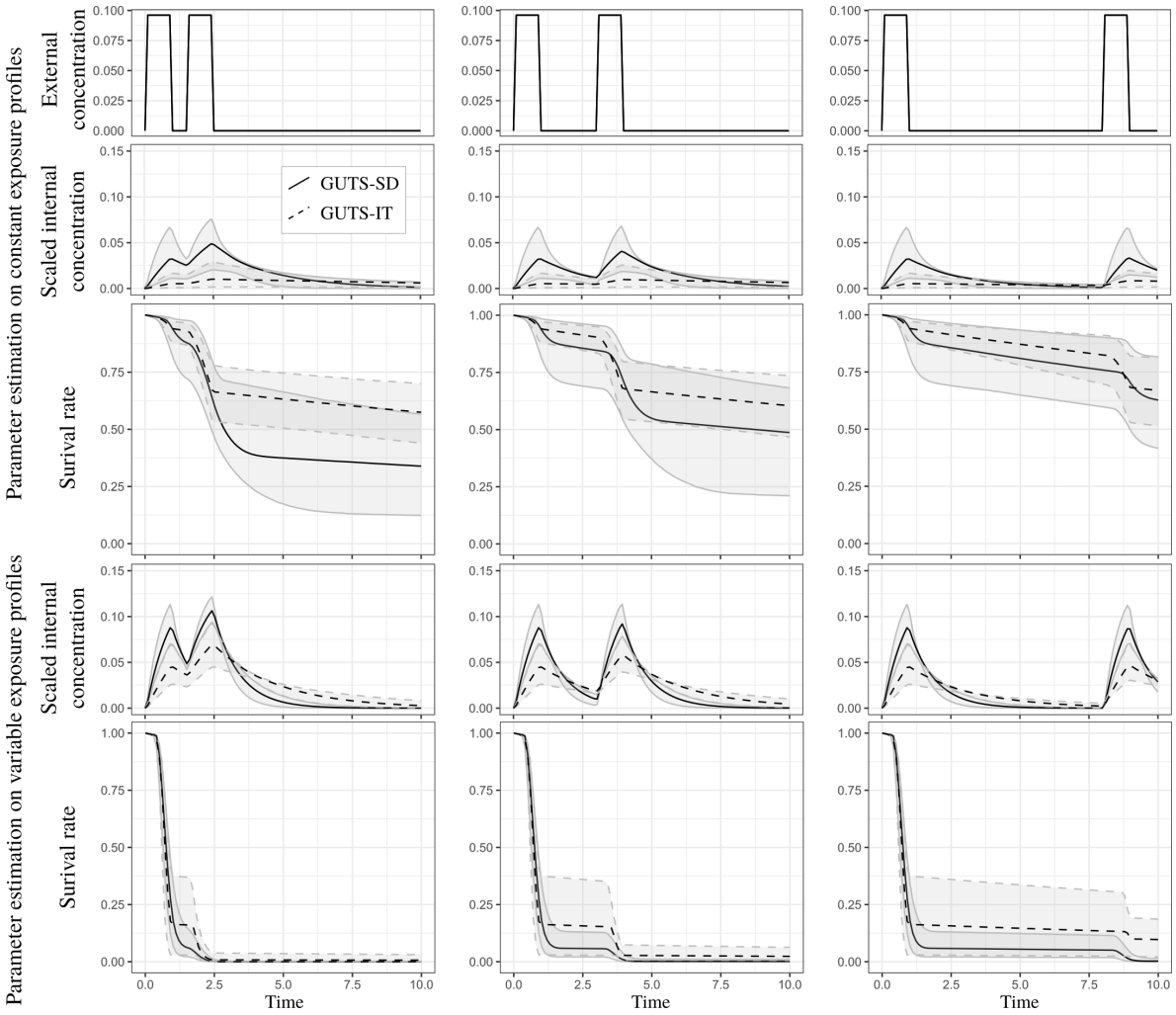

Figure 5: Survival rate over time with GUTS-RED-SD and GUTS-RED-IT models (respectivelly solid and dashed lines) under two pulsed exposure profile with the same area under the curve (differences are in the duration time between the two pulses). Parameters were estimated from Cypermethrin data set, either under constant (upper panel of the figure) or time-variable (lower panel of the figure) exposure.

3.4. Effect of the depuration time on the predicted survival rate

3.4.1. Patterns of internal scaled concentration

The dominant rate constant, k_d, regulating the kinetics of the toxicant is always greater for GUTS-RED-SD compared to GUTS-RED-IT, so that the depuration time for GUTS-RED-SD is always smaller than for GUTS-RED-IT (see Figure 3 and Supplementary Material). As a consequence, under time-variable exposure concentration, the internal scaled concentration with GUTS-RED-SD has greater amplitude than with GUTS-RED-IT (Figures 4, 5 and Supplementary Material). In other words, toxicokinetics with GUTS-RED-IT is more smooth than with GUTS-RED-SD. The compensation of the difference in k_d, and therefore in the scaled internal concentration comes from the other paramaters: threshold z and killing rate k_k for GUTS-RED-SD, and median threshold m_w and shape β for GUTS-RED-IT. However, the calibration of models being based on the same observed number

of survivors, threshold parameter z for GUTS-RED-SD and the median of threshold m_w for GUTS-RED-IT are shifted .

3.4.2. Variation in the number of pulses in exposure profiles

A first step has been to explore the effect of the number of pulses (9, 6 and 3 pulses of one day each) along a period of 20 days with the same cumulative amount of contaminant in external concentration after the 20 days (Figure 4 and Supplementary Material). From a conservative approach for ERA, whatever the model, GUTS-RED-SD or GUTS-RED-IT, it seems better to have few pulses of high amplitude than frequent pulses of low amplitude. Indeed, the survival rate over time with only 3 high pulses is lower than the survival rate under frequent lower exposure. This is confirmed in Supplementary Material for Malathion and Propiconazole data sets. With GUTS mechanistic models, the higher is the pulse, the higher is the scaled internal concentration and so is the damage. Thus, from these simulations, we do not see the effect of the depuration time which could help individual to recover when reducing the frequency of peaks.

The comparison between constant and time-variable exposure profiles (Figure 4 and Supplementary Material) suggests that uncertainty is smaller when calibration has been done on data under time-variable exposure profile. The result is counterintuitive, especially since the number of time series was higher with constant exposure profiles what would reduce uncertainties of parameter estimates. If this result is confirmed, then it would be better to predict variable exposure profiles with parameters calibrated from time-variable exposure data sets.

3.4.3. Variation in the period between two pulses

In order to explore the effect of the depuration time, we simulated exposure profiles under two pulses with different time period between them (i.e., 1/2, 2 and 7 days). The cumulative amount of contaminant remains the same for the three simulations. Figure 5 shows that increasing the period between two pulses may increase the survival rate of individuals, whatever the model, GUTS-RED-SD or GUTS-RED-IT. This is a typical result of the depuration period which reduce the level of scaled internal concentration, and therefore reduces the damage. We can easily see that the highest scaled internal concentration is reached when the pulse interval is the smallest. In this situation, we clearly observe the addition of damages from the two pulses. Again, depuration time being different with the two GUTS models, results are also different. For ERA, having two close pulses being the most conservative, we recommend to perform such an experiment. However, the depuration time being the differentiating parameter of GUTS-RED-SD and GUTS-RED-IT, it is also relevant to add an experiment with two pulses separated by a long enough period in order to decorrelate their effect. Thus, having both correlated and uncorrelated experiments, we can better assess the influence of GUTS-RED-SD and GUTS-RED-IT hypothesis on the simulation outputs.

4. Discussion

4.1. Tracking uncertainties for environmental quality standards

Whatever the scientific field, risk assessment is by definition linked to the notion of probability, holding different uncertainties such as the variability between organisms and noises in observations. In this sense, tracking how the uncertainty propagates into models, from collected data to model calculations of toxicity endpoints that are finally used for EQS derivation is of fundamental interest for ERA. For ERA, having good fits over experimental data is not enough. Indeed, the key objective is the application of these fits to predict adverse effects under real environmental exposure profiles, and to derive robust EQS (Laskowski, 1995; Jager, 2011; Gray and Cohen, 2012; EFSA PPR, 2013; EFSA, 2018b). In this context, as we show in this paper, TKTD models calibrated under a Bayesian framework combines two great advantages: on one hand, TKTD models, such as the GUTS models, allow predictions of regulatory toxicity endpoints under any type of exposure profiles ; on the other hand, the Bayesian approach provides the marginal distribution of each parameter, and in this way, allows to track the uncertainty of any prediction of interest.

Previous studies investigating goodness-of-fit did not find typical differences between GUTS-RED-SD and GUTS-RED-IT models (Ashauer et al., 2013; Baudrot et al., 2018c). Here again, from the uncertainties in regulatory toxicity endpoints, we do not show evidence to choose GUTS-RED-SD compared to GUTS-RED-IT. A simple recommendation is therefore to use both and then to take the most conservative scenario in terms of ERA. With the 10 data sets we used and the 20 fittings we did, the four measures of goodness-of-fit show similar outputs for both GUTS-RED-SD and GUTS-RED-IT, under both constant and variables exposure profiles. The percentage of observed data lying in the 95 % predicted credible interval, denoted $\%PPC$ has the advantage of being linked to visual graphics, PPC plots, and is therefore easier to interpret for risk assessors and stackholders (Beck et al., 2016). However, it may hide a very large uncertainty due to its limitation to 100 % of covering. WAIC and LOO-CV criteria are more robust probability measures (Gelman et al., 2013). Since NRMSE is easy to calculate whatever the inference methods (e.g., Maximum Likelihood Estimation), it could be a relevant measure to check the goodness-of-fit of models.

4.2. What about the use and abuse of the lethal concentration?

After checking the quality of model parameter calibration, the next question is about the uncertainty in toxicity endpoints to derive EQS. Lethal concentrations are nowadays a standard for hazard characterization at levels of 10, 20 and 50 % effect on the population. Many criticisms have been addressed to the lethal and effective concentrations for $x\%$ of the population and other related measures (Jager, 2011). For instance, the classical way to compute the lethal concentration, at the final point, removes information provided by the observations made all along the experiment, and hides the time

dependency. For the lethal effect, a classical approach to limit the variability of time duration, is to consider a long enough exposure duration in order to obtain the incipient $LC(x, t \to +\infty)$ (Jager et al., 2006), that is when the $LC(x, t)$ reaches its asymptote and does not change with increasing duration of exposure as observed on Figure 1. We provide mathematical expression of $LC(x, t)$ convergence, and explicit results when $x = 50\%$ for both GUTS models. We can therefore use the joint posterior parameter distribution provided by the Bayesian inference to compute the distribution the incipient LC.

A consequence of the exponential decay of $LC(x, t)$ with increasing time t, is that the uncertainty is greater at early time where a small change in time t induces a great change in the $LC(x, t)$ whatever x. For this reason, classical measures of LC are done at the latest time of experiments. Hence, to compare $LC(x, t)$ of different compounds or species that may require different duration of experiments, using TKTD to extrapolate at other time points is of great advantage. Also, in order to reduce the uncertainty, extrapolation to greater time would be a preferable choice.

We show in this study that the uncertainty of $LC(x, t)$ is different according to percentage x under consideration (Figure 1). It appears this uncertainty is limited around 50%, while not specifically at 50%, what is in favor of the classical approach to return the LC_{50}. However, it is still of real importance to report the uncertainty of the toxicity endpoints since we show it can drastically change depending on the experimental design, the combination product-species.

Among the criticisms of the $LC(x, t)$, one is that it is meaningful only under a set of constant environmental conditions including a constant exposure profile (Jager et al., 2006; Jager, 2011). When the concentration of chemical compounds is highly variable over time in the environment, the use of toxicity endpoints based on toxicity data for constant exposure profiles may hide some processes, such as the response to pulses of exposure. This is the reason underlying the interest of multiplication factor for ERA (Ashauer et al., 2013).

4.3. What does it mean to take a margin of safety?

The deduction of a margin of safety from a multiplication factor, $MF(x, t)$, quantifies how far the exposure profile is below toxic concentrations (Ashauer et al., 2013). Then, a key question for risk assessors is to target the safest exposure duration, t, and percentage of effect on survival, x. Our study show a lower uncertainty around an x value of 50 %. Thus, to reduce the uncertainty of the $MF(x, t)$ estimation we recommend to select 50%, at least for comparison between studies. We also show that under constant exposure profiles, there is an asymptotic shape of the $MF(x, t)$ as it is for $LC(x, t)$. There is an incipient value of the multiplication factor for any x when t goes to a long time. Therefore, under constant profiles, we could recommend to use the latest time of the exposure profile for toxicity endpoints in order to reduce the sensitivity of the $MF(x, t)$ estimation. However, the $MF(x, t)$ is meaningful when applied on realistic exposure profiles which are rarely constant, and

our study shows that there is no asymptotic shape in such situations. In addition, we observed a great sensitivity of the multiplication factor around peaks in the exposure profiles, that is a high variation of the $MF(x,t)$ with a little change in time t. As a consequence, the assessors has to be very careful about the characteristics of pulses in the exposure profiles in order to understand how they drive changes in the multiplication factor. To do so, we recommend to compute the multiplication factor all along the period of the exposure profile, rather than choosing a single distribution at a specific time.

4.3.1. Depuration time

The survival response to pulses depends on the depuration time driven by the toxicokinetics part of the TKTD model. The kinetics of assimilation and elimination of compounds integrated within the toxicokinetic module is a fundamental part of ecotoxicological models (Wang and Fisher, 1999). In reduced GUTS models, GUTS-RED-SD and GUTS-RED-IT, we assume no measure of the internal concentration, which is therefore calibrated at the same time as other parameters included in the toxicodynamics part. The resulting "scaled internal concentration" is linked to a level of damage defined by the toxicodynamic which has two different hypotheses on the death mechanism for GUTS-RED-SD and GUTS-RED-IT. The mechanistic construction of the model, reflecting biological processes, may be misleading since the toxicokinetic is defined independently of the toxicodynamic part which is chosen afterwards. What is true in the mechanism is not in the inference process where the model parameters, from TK and TD parts, are calibrated all together. As a consequence, as illustrated with our results, the scaled internal concentration does not have the same biological meaning in GUTS-RED-SD and GUTS-RED-IT, and therefore cannot be directly compared between both models.

In both models of course, from the underlying mechanism, we know that damage is positively correlated with pulse amplitude: lower amplitude, lower damage, as we observed from Figure 4. A result is that, with the same cumulative amount of contaminant along an experiment, using fewer pulses reduces final survival rates. So, the most conservative experimental design is the one with fewer pulses of relatively high amplitudes.

Furthermore, from Figure 5, we bring to light the effect of the depuration time. When pulses are close, the organisms do not have time to depurate and therefore there is an addition of the damage and finally a cumulative effect on survival. As a consequence, on a long enough experiment, when pulses become less correlated in terms of cumulative damage, then the final survival rate increases. Because of this, we recommend an experimental design with two close pulses. However, to have a better calibration of the toxicokinetic parameter, it is important to also have two uncorrelated pulses in the experimental design.

Finally, our study reveals that the uncertainty for prediction under time-variable exposure profiles seems to be smaller when calibration was performed with data sets under time-variable rather than under constant exposure profiles. While this observation makes theoretical sense since predictions are

made on the same type of profile than calibration of parameters, further empirical studies have to be performed to confirm this point.

The environmental dynamics of chemical compounds can be highly variable depending on the whole environmental context (e.g., anthropogenics activities, geochemical kinetics, ecosystem processes) but also on the chemical and bio-transformation of the compound under study. Therefore, as a general recommendation, we would like to point out the relevancy of experimenting several type of exposure profiles. Basically, the control and both constant and time-variable exposure profiles including toxicologically dependent and independent pulses seem to be the minimum requirement.

4.4. Practical use of GUTS models

4.4.1. Optimization of experimental design

The complexity of environmental systems combined with the thousand of compounds produced by human activities implies to assess environmental risk for a huge set of species-compounds combination (Ashauer and Jager, 2018). As a direct consequence, optimizing experimental design in order to maximize the gain of high-quality information from experiments is a challenging requisite, where mechanism-based models combined with the Bayesian approach offer several tools (Albert et al., 2012). A next step of the present study is to use the joint posterior distribution of parameter, and the distribution of toxicity endpoints in order to quantify the gain of knowledge of several potential experiments in order to select the most informative. The next objective is thus to develop a framework that could help the construction of new experimental designs in order to minimize their complexity and their number while maximizing the robustness of toxicity endpoint estimates.

4.4.2. Implementation

Although their many advantages, TKTD models, and therefore GUTS models, still remain little used. This is due to their mathematical complexity based on differential equations that need to be numerically integrated when fitted to data (Albert et al., 2016). Associated to their promotion within regulatory documents associated to ERA, the use of GUTS models could be further extended when available within a software environment allowing their handling without immersing into technicalities. Nowadays, several software allow to overpass these difficulties (Jager and Ashauer, 2018; Albert and Vogel, 2017; Baudrot et al., 2018a), and a web-platform is also proposed (Baudrot et al., 2018d).

4.4.3. Limitations

Survival is the most often observed response of a chemical toxic effect in the environment, but sub-lethal effects may be more relevant to manage for ERA, to prevent community collapse (Baudrot et al., 2018b). While the lethal concentration decreases when time increases, other sub-lethal effects (e.g., reproduction, growth) does not always follow this pattern (Álvarez et al., 2006; Jager, 2011). The levels of concentration in acute toxicity tests are higher than those classically observed in the

environment. Therefore, under real environmental condition, sub-lethal effects may have more direct impacts on the population dynamics than effects on survival. Thus, it would be of real interest to encompass different effects in a global TKTD approach, in order to better predict when scaling-up at population and community levels (Jager, 2011).

Another well-known limitation is the derivation of EQS from specific species-compound combination. In order to extrapolate ecotoxicological information from a set of single species tests to a community, ERA uses Species Sensitivity (Weighted) Distribution, SS(W)D, which can be used to derive EQS covering a set of taxonomically different species (Duboudin et al., 2004). This calculation is classically applied on $LC(x,t)$ and could be easily done with $MF(x,t)$ with the benefit to be applied on time-variable exposure profiles (EFSA, 2018b).

4.5. Conclusion

As recently written by EFSA experts: "*uncertainty analysis is the process of identifying limitations in scientific knowledge and evaluating their implications for scientific conclusions*" (EFSA, 2018a). Description of uncertainties increases transparency and trust in scientific outputs and is therefore a key for an applied science such as ecotoxicology (Beck et al., 2016). Here, we evaluated the combination of mechanism-based models with the Bayesian inference framework to track uncertainties on toxicity endpoints used in regulatory risk assessment from one compound-one species survival bioassay. A lot of other kind of uncertainties emerge all along the decision chain, from the hazard identification to the characterization of risk. Focusing on uncertainty should be of concern at every steps, and above all, for any information returned by mathematical-computational models.

Acknowledgement

The authors are very grateful for inputs from Theo Brock on an earlier version of the manuscript. The authors also thank the French National Agency for Water and Aquatic Environments (ONEMA, now the French Agency for Biodiversity) for its financial support. The authors declare no competing interests.

References

Albert, C., Ashauer, R., Künsch, H., Reichert, P., 2012. Bayesian experimental design for a toxicokinetic–toxicodynamic model. Journal of statistical planning and inference 142 (1), 263–275.

Albert, C., Vogel, S., 2017. GUTS: Fast Calculation of the Likelihood of a Stochastic Survival Model. R package version 1.0.4.
URL `https://CRAN.R-project.org/package=GUTS`

Albert, C., Vogel, S., Ashauer, R., 2016. Computationally efficient implementation of a novel algorithm for the general unified threshold model of survival (guts). PLoS Comput Biol 12 (6), e1004978.

Álvarez, O. A., Jager, T., Redondo, E. M., Kammenga, J. E., 2006. Physiological modes of action of toxic chemicals in the nematode acrobeloides nanus. Environmental Toxicology and Chemistry 25 (12), 3230–3237.

Ashauer, R., Albert, C., Augustine, S., Cedergreen, N., Charles, S., Ducrot, V., Focks, A., Gabsi, F., Gergs, A., Goussen, B., et al., 2016. Modelling survival: exposure pattern, species sensitivity and uncertainty. Scientific Reports 6.

Ashauer, R., Hintermeister, A., Potthoff, E., Escher, B. I., 2011. Acute toxicity of organic chemicals to gammarus pulex correlates with sensitivity of daphnia magna across most modes of action. Aquatic toxicology 103 (1), 38–45.

Ashauer, R., Jager, T., 2018. Physiological modes of action across species and toxicants: the key to predictive ecotoxicology. Environmental Science: Processes & Impacts.

Ashauer, R., Thorbek, P., Warinton, J. S., Wheeler, J. R., Maund, S., 2013. A method to predict and understand fish survival under dynamic chemical stress using standard ecotoxicity data. Environmental toxicology and chemistry 32 (4), 954–965.

Baudrot, V., Charles, S., Delignette-Muller, M. L., Duchemin, W., Kon-Kam-king, G., Lopes, C., Ruiz, P., Veber, P., 2018a. morse: MOdelling Tools for Reproduction and Survival Data in Ecotoxicology. R package version 3.0.0.
URL `https://cran.r-project.org/web/packages/morse/index.html`

Baudrot, V., Fritsch, C., Perasso, A., Banerjee, M., Raoul, F., 2018b. Effects of contaminants and trophic cascade regulation on food chain stability: Application to cadmium soil pollution on small mammals–raptor systems. Ecological Modelling 382, 33–42.

Baudrot, V., Preux, S., Ducrot, V., Pavé, A., Charles, S., 2018c. New insights to compare and choose tktd models for survival based on an inter-laboratory study for *Lymnaea stagnalis* exposed to cd. Environmental science & technology 52, 1582–1590.

Baudrot, V., Veber, P., Gence, G., Charles, S., 2018d. Fit guts reduced models online: from theory to practice. Integrated environmental assessment and management.

Beck, N. B., Becker, R. A., Erraguntla, N., Farland, W. H., Grant, R. L., Gray, G., Kirman, C., LaKind, J. S., Lewis, R. J., Nance, P., et al., 2016. Approaches for describing and communicating overall uncertainty in toxicity characterizations: Us environmental protection agency's integrated risk information system (iris) as a case study. Environment international 89, 110–128.

Brock, T. C., 2009. Linking aquatic exposure and effects: risk assessment of pesticides. CRC Press.

Delignette-Muller, M. L., Ruiz, P., Veber, P., 2017. Robust fit of toxicokinetic–toxicodynamic models using prior knowledge contained in the design of survival toxicity tests. Environmental Science & Technology 51 (7), 4038–4045.

Duboudin, C., Ciffroy, P., Magaud, H., 2004. Effects of data manipulation and statistical methods on species sensitivity distributions. Environmental Toxicology and Chemistry 23 (2), 489–499.

ECHA, 2017. Guidance on information requirements and chemical safety assessment.

EFSA, 2018a. Guidance on uncertainty analysis in scientific assessments. EFSA Journal 16 (1).

EFSA, 2018b. Scientific Opinion on the state of the art of Toxicokinetic/Toxicodynamic (TKTD) effect models for regulatory risk assessment of pesticides for aquatic organisms. EFSA Journal (to appear).

EFSA PPR, 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA Journal 11 (7), 3290.

Ferson, S., 2005. Bayesian methods in risk assessment. Unpublished Report Prepared for the Bureau de Recherches Geologiques et Minieres (BRGM). New York.

Gabry, J., Mahr, T., 2017. bayesplot: Plotting for Bayesian Models. R package version 1.4.0. URL `https://CRAN.R-project.org/package=bayesplot`

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B., 2013. Bayesian data analysis. Chapman and Hall/CRC.

Gray, G. M., Cohen, J. T., 2012. Policy: rethink chemical risk assessments. Nature 489 (7414), 27.

Grimm, V., Berger, U., 2016. Robustness analysis: Deconstructing computational models for ecological theory and applications. Ecological modelling 326, 162–167.

Hommen, U., Forbes, V., Grimm, V., Preuss, T. G., Thorbek, P., Ducrot, V., 2016. How to use mechanistic effect models in environmental risk assessment of pesticides: case studies and recommendations from the setac workshop modelink. Integrated environmental assessment and management 12 (1), 21–31.

569 Isigonis, P., Ciffroy, P., Zabeo, A., Semenzin, E., Critto, A., Giove, S., Marcomini, A., 2015. A multi-
570 criteria decision analysis based methodology for quantitatively scoring the reliability and relevance
571 of ecotoxicological data. Science of the Total Environment 538, 102–116.

572 Jager, T., 2011. Some good reasons to ban ec x and related concepts in ecotoxicology.

573 Jager, T., Albert, C., Preuss, T. G., Ashauer, R., 2011. General unified threshold model of survival
574 - a toxicokinetic-toxicodynamic framework for ecotoxicology. Environmental Science & Technology
575 45 (7), 2529–2540.

576 Jager, T., Ashauer, R., 2018. Modelling survival under chemical stress. A comprehensive guide to the
577 GUTS framework. Version 1.0. Leanpud.

578 Jager, T., Heugens, E. H., Kooijman, S. A., 2006. Making sense of ecotoxicological test results: towards
579 application of process-based models. Ecotoxicology 15 (3), 305–314.

580 Laskowski, R., 1995. Some good reasons to ban the use of noec, loec and related concepts in ecotoxi-
581 cology. Oikos, 140–144.

582 Nyman, A.-M., Schirmer, K., Ashauer, R., 2012. Toxicokinetic-toxicodynamic modelling of survival of
583 gammarus pulex in multiple pulse exposures to propiconazole: model assumptions, calibration data
584 requirements and predictive power. Ecotoxicology 21 (7), 1828–1840.

585 Plummer, M., 2016. rjags: Bayesian Graphical Models using MCMC. R package version 4-6.
586 URL `https://CRAN.R-project.org/package=rjags`

587 Reinert, K. H., Giddings, J. M., Judd, L., 2002. Effects analysis of time-varying or repeated exposures
588 in aquatic ecological risk assessment of agrochemicals. Environmental Toxicology and Chemistry
589 21 (9), 1977–1992.

590 Syberg, K., Hansen, S. F., 2016. Environmental risk assessment of chemicals and nanomaterials—the
591 best foundation for regulatory decision-making? Science of the Total Environment 541, 784–794.

592 Wang, W.-X., Fisher, N. S., 1999. Assimilation efficiencies of chemical contaminants in aquatic inver-
593 tebrates: a synthesis. Environmental toxicology and chemistry 18 (9), 2034–2045.