

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

The manuscript presents analyses of TKTD models for survival, i.e. the GUTS models, which investigate properties of these models in the light of regulatory risk assessment. Based on an earlier published and analysed example data set, the authors derive analytical mathematical solutions of the reduced GUTS models under constant external exposure, and use these solutions in combination with the Bayesian inference method for model calibration and analyses of the propagation of parameter uncertainty to model predictions. Based on the derived mathematical framework, the authors analyse the uncertainty in model predictions for some scenarios, e.g. constant vs time-variable exposure, as well as properties of LC estimates over time, or as a function of the effect percentage. Further on, they test the effects of different time-variable exposure patterns on the model predictions.

The manuscript shows a very good mathematical quality, and interesting analyses, which are relevant for the application of TKTD models of the GUTS type for the evaluation of toxicity tests. The presented work is highly relevant and can be clearly recommended for publication in PCI ecology. Alone, the language is not fit for publication, and there are some formulations in the text which need to be rephrased in order to clarify the sentences or to avoid misunderstandings. In my review, I suggest some reformulations, and I corrected also some language issues, but I strongly recommend to let the manuscript be edited by a native speaker/expert.

Comments.

L7: As far as I'm aware of, EQS are used mainly in context of the European water framework directive, and much less in pesticide risk assessment. Therefore I would suggest to remove the term EQS from the text and to replace it by a more appropriate, e.g. just saying 'risk estimate'.

L10 "The classical toxicity endpoints are ~~the~~-x% effect/lethal concentrations at a specific time (i.e., EC/LC(x,t)), or ~~the~~-multiplication factors applied to environmental exposure profiles leading to x% of effect reduction at a specific time (i.e., MF(x,t))."

L10: the multiplication factors are for sure not a classical toxicity endpoint. Please re-formulate the sentence.

L12-14: What is likely to differ between species? Please formulate more clear.

L14 ff: Also, real-world-exposure profiles are hardly ever-constant over time, what makes impossible the use of classical dose-response models impossible 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 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-consequently also those of EQS. The innovative approach investigated in our paper is the use of the Bayesian framework to deal with analyse uncertainties raising in the calibration process and propagate d-them all along to the successive model predictions steps, including 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 EQSrisk/effect estimates.

L32: All used acronyms are undefined here! Please either explain or leave out.

Commented [AF1]: This is a very hard word, I suggest to formulate it a bit softer, e.g. use 'difficult' or 'imprecise'.

Commented [AF2]: This does not really make sense here, or?

L46: "the use **of** mechanistic models"

L52: this is very technical jargon, x% LC(x,t) etc. Is it possible to formulate this sentence (and also the abstract) in a more general, understandable way?

L59: "**The** GUTS-RED-SD model assumes..."

L71: "Quantifying uncertainties or **levels** of confidence..."

L96: "...data sets, as described in Ashauer et al. (2011) and Nyman t al., (2012).... testing in total five chemical compounds..."

Caption of Table 1: "and "Time points per profile" is the number of observation time points for each time series (each constant profiles consisted in-of 5 time-points)."

L125/126: It could be useful to point to the fact that estimated DRT were different for the SD and IT models, despite the TK model equations are identical.

L163: '...where **the** Normalized RMSE..."

L167-169: Please give more precise definition of the PPC, including equations.

L171-175: Also for WAIC and LOO-CV is could be interesting to see the equation. In addition, could you state what is the range of typical or usual values for these two measures?

L188: I tried to test equation 10, by filling in values to derive the LC, and ended up without results, because I needed to calculate tz, which requires a water concentration Cw. I felt like this was circular equations, since the LC value should be the concentration leading to x% effect, but I need to fill in another water concentration?!?! I do not exclude that I was just not smart enough to apply the formula, but please point out better and in more detail how to use the equation.

L189: "As mentioned in the Supplementary Material, under time-variable exposure, tz is also variable with time, while in the case of constant exposure, tz is exactly $-1/kd \ln(1 - z/Cw)$. When For increasing time increase, the LCSD (x, t) curve becomes a vertical line at point z, and when-for infinite time tends to infinity, the convergence is:"

L201-202: What do you mean by the last sentence, I do not understand. Please reformulate or delete.

Eq 17: Same issue with calculation of tz as described above.

L224 ff: "For all compounds, Table 2 shows that fitting on observed survival from testing under constant exposure profiles provides better fits than for using data from testing time-variable exposure profiles "

L232-234: I do not understand the meaning of this sentence, please formulate more clear or delete.

L238-242: "While NRMSE and %PPC tend to indicate better qualified matching, GUTS-RED-IT model predictions, the uncertainty is penalized with the 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."

Commented [AF3]: I do not understand.

L248: Rarely analysed?

Commented [AF4]: I do not understand, please formulate more clear or delete.

L250/251: Here and in many other places in the discussion, it is unclear whether you speak about uncertainty or variability. When I assume that uncertainty is the credible area around the median prediction, this sentence is not true. When you want to say that in the beginning of the tests the LCX values change quickly for increasing experimentation time, this is correct, but needs to be formulated more precise!

L254-256 Here you argue against L250/251. Please streamline this section and check the use of the words uncertainty and variability!

L262-264: "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." It sounds a bit funny that you recommend something that is already done. A possible rephrasing could be "Our analyses confirm that the evaluation of experimental data at the last day of an experiment is supported by theoretical considerations."

L272-274: It is not clear what the recommendation is here. Calculate LCX values for different X values for each experiment? This is not useful. Please comment on that in the text.

L278-279: I do not agree that the later the lower. There is a shift down in 2E and 2F, but then the MF is constant again. Please be more precise in the formulation.

L282: Replace 'surrounding' by 'at the time of exposure'

L283 ff: "Therefore, a recommendation would be to ~~wait for some times (e.g., several days) after a peak before computing multiplication factors only after some time (e.g., several days) after a peak.~~"

L285-287: Please check the language/formulations in this sentence.

L290: "describe"

Figure 3: replace 'half-life' by dominant rate constant. Correct inlay: GUTS-SD is GUTS-RED-SD and the same for IT. The same also for figure 4.

Figure 5: caption: "respectively"

L318: "for the Malathion..."

L318-321. I do not understand what you mean here. Please formulate more clearly!

L336: "Again, because of the different depuration times of the two GUTS models..."

L346: "noise"

L347: "tracking how uncertainty propagates"

L358: "Our study confirms that under specific consideration of uncertainties in regulatory toxicity endpoints no evidence to choose one of the GUTS-SD or GUTS-IT models over the other is given."

L360/361: Please clarify that the more conservative model should only be used if successfully validated.

L365: I'm not sure whether the PPC can be easier interpreted as e.g. the NRMSE, since also this indicator can be shown visually.

L369: Please add that the NRMSE was just recommended to be used in the recent EFSA scientific opinion on TKTD models.

L384: replace 'decay' by 'decrease'

L384 ff: Please check here also the use of the term 'uncertainty'; I think you speak about variability of the value, not about the uncertainty as expressed in credible intervals.

L394: Please write a correct end of the sentence.

L400: Please add the recent EFSA scientific opinion on TKTD models to the citation of Ashauer 2013. Ashauer used the term margin of safety, and the SO defined the multiplication factors.

L402-411: Please check whether you replace $MF(x,t)$ and $LC(x,t)$ by verbal descriptions, otherwise the section is hard to read. E.g. time-and level specific multiplication factor or lethal concentration.

L408: 't goes to a long time'. Please replace by 'for long times' or similar.

L412-417: I do not understand this section. Isn't it clear that the MF is changing abruptly in response to a peak? Still, the MF at the end of a tested exposure profile is most relevant, and since the computation of the MF all over time as you suggest is computational intensive, I do not see the advantage of this, but only confusion and increased effort. Please consider to change this section accordingly.

L424: 'scaled internal concentration' is not the updated terminology (Jager and Ashauer 2018), please check!

L424-431: Please check language and formulate more clearly what you want to say., I do not understand the meaning of this section.

L444: 'uncertainty of predictions under time-..."

L457: 'thousands'

L479: 'do not always follow'

L491: Since the content of the manuscript uses a lot from the recently published EFSA SO on TKTD models (e.g. multiplication factors, NRMSE), it would be only fair to acknowledge this by citing it in the conclusion paragraph.