

rbioacc: An R-package to analyze toxicokinetic data

Aude Ratier, Virgile Baudrot, Miléna Kaag, Aurélie Siberchicot, Christelle Lopes, Sandrine Charles

▶ **To cite this version:**

Aude Ratier, Virgile Baudrot, Miléna Kaag, Aurélie Siberchicot, Christelle Lopes, et al.. rbioacc: An R-package to analyze toxicokinetic data. Ecotoxicology and Environmental Safety, 2022, 242, pp.113875. 10.1016/j.ecoenv.2022.113875 . ineris-03846876

HAL Id: ineris-03846876

https://ineris.hal.science/ineris-03846876

Submitted on 21 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ecotoxicology and Environmental Safety 242 (2022) 113875

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

journal homepage: www.elsevier.com/locate/ecoenv

rbioacc: An R-package to analyze toxicokinetic data

Aude Ratier [a,b,1,2], Virgile Baudrot [a,c,1,3], Miléna Kaag [a,d], Aurélie Siberchicot [a,4], Christelle Lopes [a,5], Sandrine Charles [a,*,6]

[a] University of Lyon, University Lyon 1, UMR CNRS 5558, 43 boulevard du 11 novembre 1918, Villeurbanne 69100, France
[b] INERIS, Experimental Toxicology and Modelling Unit, UMR-I 02 SEBIO, parc Alata, Verneuil-en-Halatte 65550, France
[c] Qonfluens, Montpellier 34000, France
[d] INSA Lyon, Biosciences Department, 20 avenue Albert Einstein, Villeurbanne 69100, France

ARTICLE INFO

Edited by Dr. Hao Zhu

Keywords:
Bayesian inference
Bioaccumulation metrics
Ecotoxicology
Environmental risk assessment
One-compartment modelling
R-Shiny web interface
Toxicokinetic models
Uncertainty

ABSTRACT

The R-package rbioacc allows to analyse experimental data from bioaccumulation tests where organisms are exposed to a chemical (exposure) then put into clean media (depuration). Internal concentrations are measured over time during the experiment. rbioacc provides turnkey functions to visualise and analyse such data. Under a Bayesian framework, rbioacc fits a generic one-compartment toxicokinetic model built from the data. It provides TK parameter estimates (uptake and elimination rates) and standard bioaccumulation metrics. All parameter estimates, bioaccumulation metrics and predictions of internal concentrations are delivered with their uncertainty. Bioaccumulation metrics are provided in support of environmental risk assessment, in full compliance with regulatory requirements required to approve market release of chemical substances. This paper provides worked examples of the use of rbioacc from data collected through standard bioaccumulation tests, publicly available within the scientific literature. These examples constitute step-by-step user-guides to analyse any new data set, uploaded in the right format.

1. Introduction

Bioaccumulation is the organism's ability to absorb and concentrate chemicals in all or part of its body, firstly establishing a relationship between the external environmental concentration of this substance and its quantity in the organism (Mackay and Fraser, 2000). Assessing bioaccumulation is a component of international efforts to identify and control chemicals of environmental concern. Toxicokinetic (TK) modelling is currently used to describe, formalise, and predict the chemical bioaccumulation dynamics. In this paper, the chemical can be bioaccumulated in the organism by four routes of exposure: water, food, sediment, and/or pore water. At least in part, Environmental Risk Assessment (ERA) is based on OECD regulatory guidelines to perform bioaccumulation tests

This workflow can be performed with rbioacc, or directly on-line with the MOSAIC web platform at https://mosaic.univ-lyon1.fr/bioacc, making it easy for stakeholders to achieve this workflow. rbioacc provides ready-to-use functions to visualise and fully analyze bioaccumulation test data. Such data are internal concentration regularly measured over time all along the experiment during which organisms are exposed to a chemical (exposure phase with accumulation) and then put into a clean media (depuration phase). rbioacc automatically builds a generic one-compartment TK model according to the input data, and provides TK parameter estimates (appropriate uptake and elimination rates) and bioaccumulation metrics (e.g., the Bio-Concentration Factor, or BCF) by fitting the TK model on data under a Bayesian framework, which is particularly suitable to capture model and parameter uncertainty relevant to inform decision making. Hence, all

* Corresponding author.
E-mail addresses: Aude.RATIER@ineris.fr (A. Ratier), virgile.baudrot@qonfluens.com (V. Baudrot), milena.kaag@insa-lyon.fr (M. Kaag), aurelie.siberchicot@univ-lyon1.fr (A. Siberchicot), christelle.lopes@univ-lyon1.fr (C. Lopes), sandrine.charles@univ-lyon1.fr (S. Charles).
[1] These authors contributed equally to the work.
[2] ORCID: 0000-0001-8893-1587.
[3] ORCID: 0000-0003-1327-9728.
[4] ORCID: 0000-0002-7638-8318.
[5] ORCID: 0000-0001-7689-2234.
[6] ORCID: 0000-0003-4604-0166.

https://doi.org/10.1016/j.ecoenv.2022.113875
Received 11 February 2022; Received in revised form 16 June 2022; Accepted 8 July 2022
Available online 14 July 2022
0147-6513/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

parameters estimates and bioaccumulation metrics as well as predictions of internal concentrations in organisms are delivered with the quantification of their uncertainty. The overall uncertainty on TK parameter estimates is given as a joint posterior probability distribution, and marginally summarised for each parameter as a median and a 95 % credible interval (namely, the uncertainty range).

1.1. Statement of need

rbioacc is an R-package compatible with R version ≥ 4.1.0 (Core Team, 2021) and with all standard operating systems (MacOS, Linux, Windows). All rbioacc outputs have been compared with previously published results considering other TK model implementations under different software platforms. Giving very similar results, rbioacc was thus confirmed as fit-for-purpose to fit TK models on bioaccumulation test data (Charles et al., 2021b; Charles et al., 2021a; Ratier et al., 2021a; Ratier and Charles, 2022). All functions in rbioacc can be used without a deep knowledge neither of the underlying probabilistic modelling or of the inference methodology. Indeed, they were designed to behave as easily as possible, without requiring the user to provide prior values for input parameters. Meanwhile, models implemented in rbioacc can be used as a first step to explore new models which could appear as more appropriate for some situations, such as illustrated by Ratier et al. (2019) where several hypotheses of exposure routes were tested and then compared to select the most appropriate model. This is not yet possible with the current version of rbioacc, so that the user has to do it outside of the package. Note that rbioacc benefits from a user-friendly and freely available web interface, MOSAIC$_{bioacc}$, from which the same analyzes can be reproduced directly on-line without having to implement them in R programming. MOSAIC$_{bioacc}$ is directly accessible from the MOSAIC platform at https://mosaic.univ-lyon1.fr (Charles et al., 2018) or directly from https://mosaic.univ-lyon1.fr/bioacc (Ratier et al., 2021b).

1.2. Availability

rbioacc is available as an official R-package directly downloadable from CRAN at https://CRAN.R-project.org/package=rbioacc. Other package dependencies and system requirements are documented to help users with their installation issues. Besides, rbioacc is available on the GitHub repository https://github.com/aursiber/rbioacc/, especially for advanced users encountering troubles or having suggestions regarding the features or the hidden content of the rbioacc functions.

1.3. Installation

rbioacc is linked to Stan, a Hamiltonian MCMC sampler used to perform Bayesian inference with already implemented models, methods, and algorithms (Carpenter et al., 2017). This sampler is particularly efficient when parameters are highly correlated, as it is the case for TK models between uptake and elimination rates. Stan is fully integrated to R via the rstan (Stan Development Team, 2021) and rstantools (Gabry et al., 2020) packages. In addition, to use rbioacc, you need to install all the other R-packages dependencies: ggplot2, Rcpp (≥ 0.12.0), RcppParallel (≥ 5.0.1), testthat, ggmcmc, GGally, loo, stringr, zoo (≥ 1.8-9), BH (≥ 1.66.0), RcppEigen (≥ 0.3.3.3.0), StanHeaders (≥ 2.18.0), knitr, and rmarkdown. For this purpose, you can use the classical R command that is provided in the R-script (Supplementary Information, SI). rbioacc is also linked to the programming language C++ that speeds up calculations and runs simulations delivering the TK predictions. You should not have issues with C++ requirements since it is very well integrated in R.

1.4. Main features

rbioacc allows to fit more complex TK models than those classically used today, offering the possibility to account for multiple exposure routes (up to four, among water (Zhu et al., 2020), sediment (Svobodová et al., 2018), pore water and/or food (Lo et al., 2015; Lu et al., 2004; Wang et al., 2021)) as well as the possibility of biotransformation of the parent chemical into several phase I metabolites and the potential dilution by growth (Ashauer et al., 2012) of organisms (when growth measurements are available). The main functions in rbioacc are modelData() and modelData_ode() to format and visualise raw data as well as to build the corresponding TK model. The fitTK() function allows to fit a generic one-compartment TK model on the raw data; it provides estimates of the TK model parameters and the subsequent bioaccumulation metrics. Fitting outputs can be either displayed with plot() or summarised with quantile_table(). For the calculation of bioaccumulation metrics, both the bioacc_metric() and plot() functions may be run delivering the plot of the posterior probability distribution of the bioaccumulation metric density. The quantile() function gives a summary of this distribution with the median (50 % quantile) and the uncertainty range (2.5 % and 97.5 % quantiles). The time at which 95 % of the chemical is eliminated is provided by the t95() function. Underlying equations of the fitted TK model, automatically built by the package from the input data, can be displayed and stored with the equations() function, once the fitting process is complete. Other functions are available to check the goodness-of-fit (GoF) based on several criteria that we chose for their relevance and wide use in the Bayesian community: the Posterior Predictive Check (PPC, with the ppc() function); the comparison between marginal prior and posterior probability distributions of the TK parameters (with the plot_PriorPost() function); the correlation matrix between parameter estimates obtained from the joint posterior probability distribution (with the corrMatrix() function); the correlation plot between parameter estimates displayed from the join posterior probability distribution (with the corrPlot() function); the Potential Scale Reduction Factors of the parameters (PSRF, with the psrf() function, expected to be less than 1.01); the Widely Applicable Information Criterion (WAIC, with the waic() function) for model comparisons on a same data set; and the Monte Carlo Markov Chain (MCMC) chains ($n = 4$ by default) with the mcmcTraces() function. Additionally, the predict() and predict_manual() functions allow to perform predictions of internal concentrations within organisms, with or without previously observed data, thanks to previous estimated parameters values.

rbioacc currently handles constant and time-variable exposure concentration data according to a generic workflow:

1. Upload and format a data set;
2. Plot the data set;
3. Fit the appropriate TK model on the data;
4. Get model equations, parameter estimates and bioaccumulation metrics (chosen according to the exposure as automatically detected from the input data);
5. Get model predictions and their uncertainty superimposed to the data (Visual Posterior Check or VPC);
6. Check GoF criteria, prioritising the PPC and the associated percentage of the observed data that are comprised within the 95 % uncertainty range of their predictions (expected to be close to 95 %);
7. Perform validation of the model comparing simulations under a new exposure concentration and externally collected data.

Those steps are described in full details in a Tutorial available at https://doi.org/10.5281/zenodo.5092316, including the description on how to use all rbioacc features. More information on the model and the inference process used in rbioacc are given in an on-line User Guide available at http://lbbe-shiny.univ-lyon1.fr/mosaic-bioacc/data/user_guide.pdf but also in our companion paper of the MOSAIC$_{bioacc}$ web interface (Ratier et al., 2021b). Please refer to this documentation for further introduction to the use of rbioacc.

The data, however, must be properly formatted beforehand, in order to be uploaded in the R software (Core Team, 2021) as a data frame from

a .txt or a .csv file (that is comma, semicolon, or tabular separator). Each line of the data frame must correspond to a unique time point for a given replicate and a given exposure concentration of the contaminant. The data frame must contain at least four columns with predefined headings (Table 1): time, the time point of the measurement; expw, exps, expf, exppw the exposure concentration, from water, sediment, food, and/or pore water, respectively; replicate, a unique integer or string for each replicate (e.g., replicate a, b etc., replicate 1, 2, etc.); conc, the measured internal concentrations. Further columns can exist: concmℓ, for internal concentration measurements of metabolite ℓ (namely, concm1, concm2, …). Please note that only metabolites of phase I (deriving directly from the parent compound) in the metabolization process are considered; growth, for growth measurements (e.g., weight or size) of the organisms. Finally units must be fixed for time (in hours, minutes, days or weeks), for exposure concentrations (in $\mu g\ mL^{-1}$ or $\mu g\ g^{-1}$), for the measured concentrations (in $\mu g\ g^{-1}$) and for growth measurements (in g, mg, cm, mm or other). Replicate marking is dimensionless.

2. Brief worked examples

2.1. The ERA workflow

The ERA workflow proposed by the EFSA in the case of toxicokinetic-toxicodynamic (TKTD) models for the regulatory risk assessment of pesticides for aquatic organisms (Ockleford et al., 2018) consists of three steps: calibration, validation, and prediction. This workflow also applies to any TK model. The calibration step consists in fitting a TK model on bioaccumulation test data to get TK parameter estimates and deduce the bioaccumulation metrics: the BCF, the bio-sediment accumulation factor (BSAF) or the bio-magnification factor (BMF) depending on whether the exposure is by water, sediment, or food, respectively. All GoF criteria must be checked before proceeding to the validation step. The validation step consists in simulating the internal concentration over time as induced by the exposure to a chemical and in comparing the predictions to independent bioaccumulation data. Results are validated from a VPC and three additional quantitative criteria (Ockleford et al., 2018). Finally, the prediction step consists in making simulations under more realistic exposure scenarios to assess the risk in closer relationship with the real world. For this last step which completes the entire workflow, it is of the utmost importance for the TK model to have been thoroughly calibrated and validated beforehand for the chemical-species combination under consideration.

To be in full compliance with regulatory guidelines and the above-mentioned workflow, the modus operandi with rbioacc is illustrated below to be followed step-by-step with any other data set.

2.2. Calibration step

To first illustrate the calibration step, we built two basic examples using standard bioaccumulation data sets available within rbioacc thanks to the data() function. The first data set was collected from a laboratory bioaccumulation test with the rainbow trout (*Oncorhynchus mykiss*) exposed to 2 different concentrations of a highly hydrophobic compound in spiked water for 49 days. There was 1 replicate per concentration. The duration of the depuration phase was 97 days. Internal concentrations were monitored at regular time points (Crookes and Brooke, 2011). The second example concerned the species *Chironomus tentans*, a freshwater invertebrate, exposed to benzo[*a*]pyrene in spiked sediment for 3 days. Only 1 exposure concentration was tested with 2 replicates. The duration of the depuration phase was 3 days. The internal concentrations for both the parent compound and its phase I metabolite were monitored at regular time points (Schuler et al., 2003).

Table 1
Example of a data set ready to be uploaded in rbioacc.

time	conc	expw	replicate
0	0.000	0.0044	A
3	0.225	0.0044	A
7	0.355	0.0044	A
14	0.553	0.0044	A
21	0.658	0.0044	A
28	0.785	0.0044	A

2.2.1. Upload data and fit the TK model accordingly

For the first example, because 2 different exposure concentrations were tested, one of them must be chosen before performing the analysis with rbioacc. For the second example, only 1 concentration has been tested; thus, the TK analysis can directly be launched with rbioacc. See the R-code given in SI to be copied and pasted, to identically reproduce all the results.

2.2.2. Results

The major results provided by rbioacc are the predictions of the internal concentrations from the fitted model, superimposed to the observed data over time with the plot() function. The posterior probability distribution of the bioaccumulation metric(s) is (are) given by the bioacc_metric() function; this function applies whatever the exposure route(s) and the elimination process(es). If bioacc_metric() is used with its default option, it provides the kinetic bioaccumulation metric(s). The steady-state bioaccumulation metric(s) is also available if the data show having more or less reached the steady-state at the end of the accumulation phase. A summary of the TK model parameter probability distributions is given by the quantile_table() function. In addition, rbioacc provides the equations of the underlying model, automatically built from the input data with the equations() function. Fig. 1 below shows the internal concentrations predicted from the fitted model and superimposed to the observed data for example 1: *Oncorhynchus mykiss* exposed to a hydrophobic chemical.

Fig. 2 below shows the internal concentrations predicted from the fitted model and superimposed to the observed data for example 2: *Chironomus tentans* exposed to benzo[*a*]pyrene.

2.2.3. Other goodness-of-fit criteria

Once the predictions of the internal concentrations are visually checked (VPC, Figs. 1 and 2) several other GoF criteria require to be checked to evaluate how relevant are the fitting outputs, including the parameter estimates especially. For example, the PPC plot allows to compare each observed value to its prediction from the fitted TK model associated with its 95 % uncertainty range. If the fitting process is correct, it is expected to have 95 % of the observed values inside their 95 % uncertainty range. With rbioacc, the PPC can be obtained with the ppc() function as illustrated in Fig. 3 below for example 2: we got 94.6 % of data included within their uncertainty range. Additional GoF can also be checked, such as the comparison of prior and posterior probability distributions with the plot_PriorPost() function, the correlation matrix of the model parameters (with the corrMatrix() and/or corrPlot() functions), the PSRF for each parameter with the psrf() function, and the traces of the MCMC chains for each parameter with the mcmcTraces() function (see SI). If several models have been fitted to the same data set then the WAIC can be asked for their comparison with the waic() function.

2.3. Validation step

To illustrate the validation step with rbioacc, we chose a data set concerning *Spirostomum ambiguum* exposed to the pharmaceutical product fluoxetine at 0.025 $\mu g\ mL^{-1}$ under spiked water for 6 days (Nalecz-Jaweski et al., 2020). These data were first used to calibrate the appropriate TK model and to get the joint posterior probability distribution of the model parameters (results not shown). Once calibrated, the model was used to predict what may happen at 0.1 $\mu g\ mL^{-1}$ of exposure

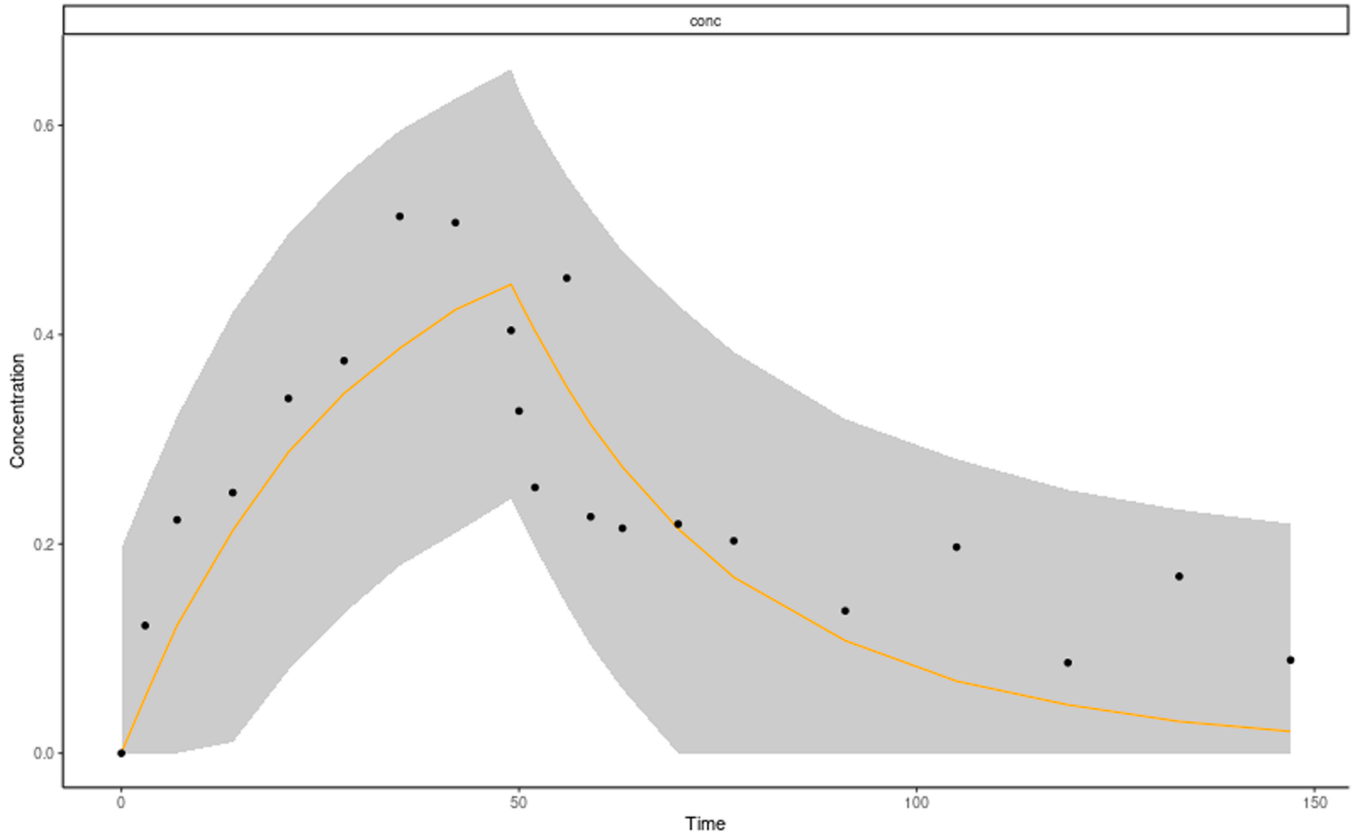


Fig. 1. Fitting plot provided by rbioacc for the example 1 with *Oncorhynchus mykiss*. Black dots are the observed data. The predicted median is symbolised by the orange plain line and the 95 % uncertainty range by the grey area.

for which data were also available. Then, predictions were superimposed to these independent bioaccumulation data (Fig. 4).

2.4. Prediction step

If the previous validation step provides relevant outputs, then the prediction step can be undertaken for further simulations at untested exposure concentrations or for a different accumulation time. In this perspective, both the predict() or predict_manual() functions of rbioacc can be used. From the previous validation step with *Spirostomum ambiguum*, we performed simulation at 0.05 $\mu g\ mL^{-1}$ (see the supplementary Figure S1). This prediction was obtained by propagating the uncertainty on TK parameter estimates from the pre-exitent calibration step with *Spirostomum ambiguum*.

Another situation may require the use of the prediction step. Indeed, it may happen that only mean or median values for each parameter are available. These values may come from the scientific literature or from previous experiments for which the associated raw TK data have not been archived or lost over time. In such situations, neither calibration nor validation can be performed. Thus, only the predict_manual() function can be used as illustrated below (Fig. 5) from median values of TK parameters provided for *Oncorhynchus mykiss* by the authors themselves (Crookes and Brooke, 2011) at 0.01 $\mu g\ mL^{-1}$. The exposure concentration 0.01 $\mu g\ mL^{-1}$ was used in Fig. 5, but the user can choose any other value, for example to design a new experiment for the same species-chemical combination. The counterpart of this approach is that no uncertainty can be propagated to the predictions (Fig. 5).

3. Bioaccumulation under time-variable exposure profiles

To mimic realistic environmental conditions, most often fluctuating than constant (*e.g.*, when collecting field data), it may need to predict the internal concentration under time-variable exposure profiles (Ashauer and Brown, 2013; Baudrot and Charles, 2019). Such situations can easily be explored with rbioacc. First, it requires to upload two data files (one is expected when the exposure concentration is constant): one text file with the time-variable exposure concentrations over time (at least two columns entitled time and Cwater; please notice that, despite this column heading, any exposure media can be considered, such as sediment, food or pore water) and another one with the corresponding internal concentration values over time (at least two columns entitled time and Cinternal); these internal concentrations may have been previously predicted with the fitted corresponding TK model by rbioacc. Secondly, it requires to call the modelData_ode() function which automatically manages the numerical integration of the ordinary differential equations that make the model behind and run simulations. Then the other functions can be used as illustrated above to get results and GOF criteria. The `R`-code in SI gives an example with a data set concerning *Sialis lutaria* exposed to a time-variable exposure profile of chlorpyrifos spiked water for 2 days (Rubach et al., 2010).

4. MOSAIC$_{bioacc}$: a web service interfacing `rbioacc`

4.1. Brief presentation

The MOSAIC$_{bioacc}$ web service has been designed to allow user-friendly TK modelling of complex situations in the field of ecotoxicology (Ratier et al., 2021b). It is developed under the R-shiny environment (Chang et al., 2021), providing an interactive interface to rbioacc. MOSAIC$_{bioacc}$ also exploits the generic one-compartment TK model (Charles et al., 2021a), meaning that organisms are considered as whole and unique compartments. Nevertheless, MOSAIC$_{bioacc}$ proposes a more complex TK model than what is classically used today, offering the possibility to account for multiple exposure routes (up to four, among

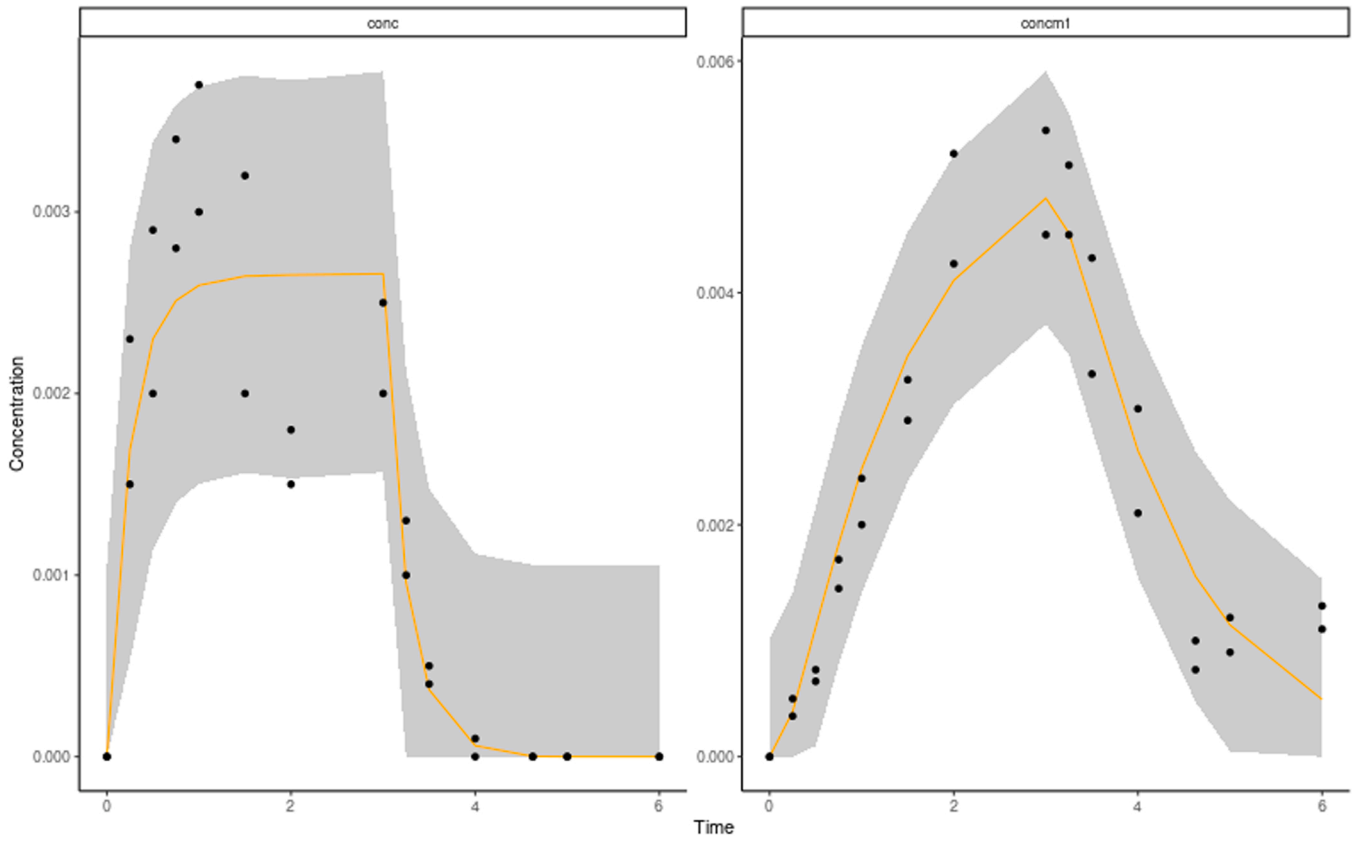


Fig. 2. Fitting plots provided by rbioacc for the example 2 with *Chironomus tentans*. Both fitting plots are displayed, for the parent compound (left) and its phase I metabolite (right). Black dots are the observed data. The predicted median curves are represented by the orange plain lines and the associated 95 % uncertainty range by the grey areas.

water, sediment, pore water and/or food) as well as the possibility of biotransformation of the parent chemical into several phase I metabolites; and the potential dilution by growth of organisms (growth measurements are additionally required). MOSAIC$_{bioacc}$ automatically builds the TK model appropriate to the uploaded data, benefiting from the rbioacc functions running behind on a dedicated server. MOSAIC$_{bioacc}$ delivers all the GoF criteria for the user to easily assess the quality of the fitting results. It is a tool really thought for the regulatory domain, but that can also be helpful for the academic research or education.

4.2. Last updates

Based on the clear observation that it is difficult to find enough raw data sets to validate generic modelling approaches, especially for TK models, we have recently developed a database with a collection of more than 200 accumulation-depuration data sets. They were extracted from the scientific literature and are now freely available on-line at http://lbbe-shiny.univ-lyon1.fr/mosaic-bioacc/data/database/TK_database.html. These data sets are associated with several outputs: the text file with raw data under the expected rbioacc input format; the full report with the fitting results and the GoF criteria as provided either by MOSAIC$_{bioacc}$ or rbioacc; and a link to the original paper (Ratier and Charles, 2022). MOSAIC$_{bioacc}$ recently shifted from an all-embarked web application to be a user interface of the newly developed rbioacc package. It allows all users to bring MOSAIC$_{bioacc}$ and its self-explanatory functions to their own device for an easy integration of its methodology to every workflow. Furthermore, new features were added: a validation and prediction tools with the recent functions implemented in rbioacc. The present section showcases these very last updates of MOSAIC$_{bioacc}$ still improving its intuitive use for any user without any need to invest in the underlying technicalities.

4.2.1. Aesthetic

The previous version of MOSAIC$_{bioacc}$ was based on a scroll of only one web page. The new version works with several tabs unlocked once the user performed the actions required from the previous tab, except for the prediction and validation tabs that are always available. Globally, the same visual elements were kept, but they got reorganised in a tab structure with 5 different levels (Fig. 6): Data upload, Model and parameters, Results, Downloads, Prediction tool. The results of the fitting process were also compacted in a tab and column structure to better suit the new no-scroll policy. This tremendously reduced the length of the previously displayed unique web page. This revamp in a tab based interface, as well as other minor changes, were partly decided upon by analysing the behaviour of the potential users of MOSAIC$_{bioacc}$.

4.2.2. Architecture

The previous version of MOSAIC$_{bioacc}$ suffered from a mix of classical and reactive R coding, the biggest obstacle being the use of global variables. Global variables are generally not meant to be used in a reactive context, but can be forced if unavoidable for some very punctual cases. Using global variables allowed for a more classical way of coding, but it generated a series of scope related bugs, notably for the downloads, that needed to be fixed with lengthy patch ups. Furthermore, using global variables promotes excessive code re-evaluation, making the application less effective. Due to the long and iterative development process by different staff members, other issues appeared such as code duplication, repetitive operations on variables, creation of unused variables and non explicit, temporary variable names that ended up being kept. All in all, the 'server' file (server.R, required by a R-Shiny development) of the previous version of MOSAIC$_{bioacc}$ was approximately 4.5k lines long. For all these reasons, this file was entirely rewritten, which allows an easier code to maintain, to understand and to upgrade, by sectioning the code

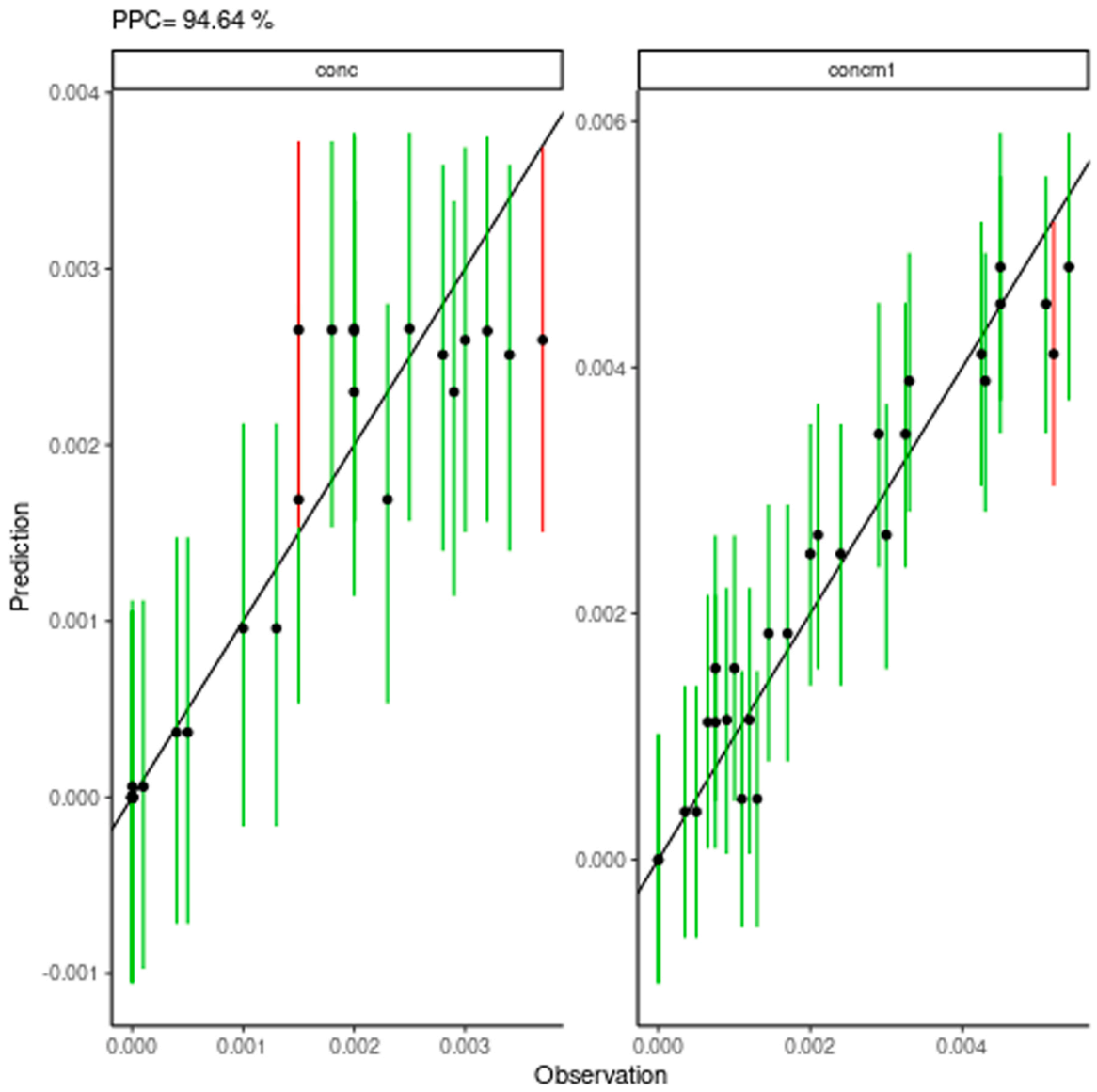


Fig. 3. Posterior Predictive Check (PPC) as given by rbioacc to check the goodness-of-fit for the example 2 with *Chironomus tentans*: PPC for the parent compound (left) and PPC for its phase I metabolite (right). Predicted against observed concentrations are symbolised by black dots (medians) while predicted 95 % credible intervals are displayed as vertical green (intersecting the first bisector) or red (not intersecting the first bisector) segments.

in multiple files; by introducing documentation with the roxygen2 package (Wickham et al., 2021) and more extensive comments through the code; and by strictly managing variable dependencies in an explicit way. With both the refactoring of the code and the integration symbolised, the server.R file became approximately 1k lines long, 1.5k lines long by taking into account the other files containing code used by the server file itself. The major benefit to this restructuring was to include rbioacc in the downloadable R-code from the application, gaining clarity and lines long, easier for a user to reuse the R code.

4.2.3. New features

A validation and prediction tools were added to MOSAIC$_{bioacc}$, for simulations of a chosen TK model with parameter values coming from either a previous calibration step, taken into account the joint probability distribution of the estimated TK parameters, or manually entered by the user as point parameter values. So, the validation and the prediction tools allow the user to either propagate the uncertainties on parameters or not, in order to perform a simulation for a new exposure concentration. The validation tool takes things a step further by allowing the visual comparison of the predicted accumulation-depuration curve to newly observed validation data; in this case, the uncertainty on parameter estimates is required so that a previous calibration step needs to have been performed earlier.

5. Conclusion

rbioacc was designed to help stakeholders and researchers to analyze TK data collected from bioaccumulation tests. This turnkey package automatically builds and fits the appropriate TK model, providing TK parameter estimates, GoF criteria and bioaccumulation metrics (namely, BCF/BSAF/BMF). In particular, it is designed to fulfil all requirements of regulators when examining market release for chemical approval. In this perspective, the associated MOSAIC$_{bioacc}$ web interface may reveals of particular help if supported by supervisory bodies.

Fundings

The authors are thankful to ANSES for providing the financial support for the development of the MOSAIC$_{bioacc}$ web tool (CNRS contract number 208483). This work was also made with the financial support of

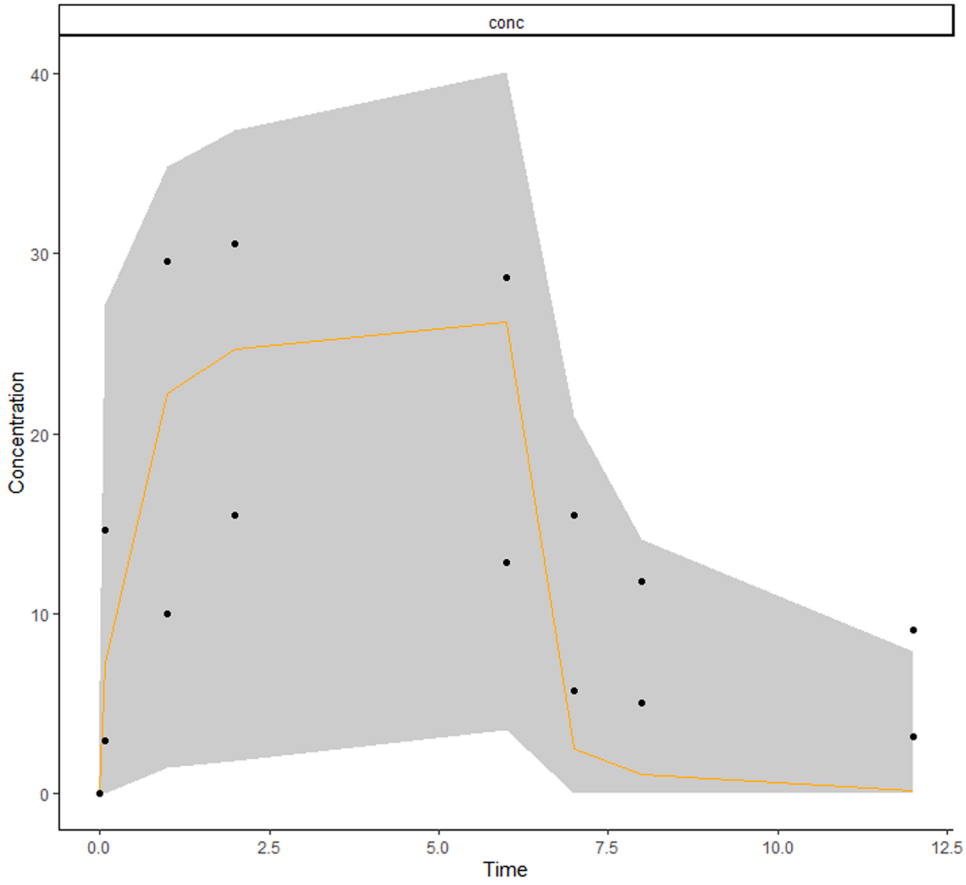


Fig. 4. Example of a validation step for the example with *Spirostomum ambiguum* exposed to fluoxetine at 0.1 $\mu g\ mL^{-1}$. The orange curve is the predicted median of the internal concentration from the TK parameters estimated at 0.025 $\mu g\ mL^{-1}$ during a previous calibration step. The grey area is the 95 % uncertainty range of the predictions, while black dots are the new observed data at 0.1 $\mu g\ mL^{-1}$.

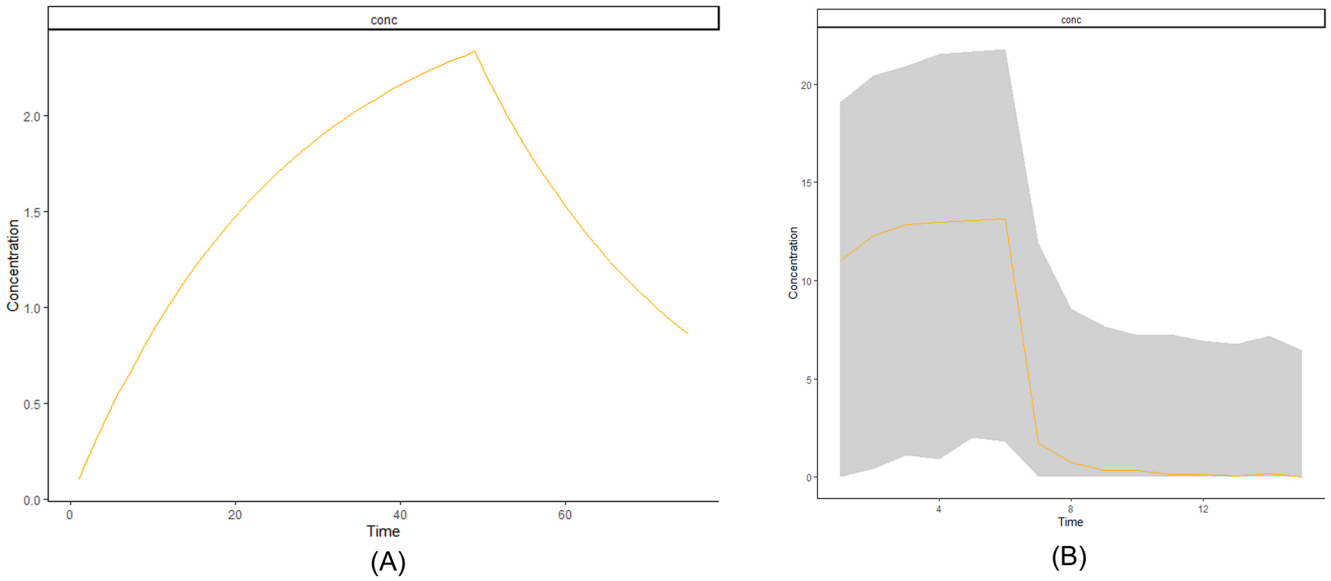


Fig. 5. Example of a prediction step for (A) fish *Oncorhynchus mykiss* exposed to a hydrophobic chemical and (B) *Spirostomum ambiguum* exposed to fluoxetine. The orange curve is the predicted median curve of the internal concentration predicted from the median TK parameter values provided in the original paper at 0.01 $\mu g\ mL^{-1}$ (Crookes and Brooke, 2011) for (A) and at 0.025 $\mu g\ mL^{-1}$ for (B) (Nalecz-Jaweski et al., 2020). The grey area is the associated 95 % uncertainty range.

the Graduate School H2O'Lyon (ANR-17-EURE-0018) and "Université de Lyon" (UdL), as part of the program "Investissements d'Avenir" run by "Agence Nationale de la Recherche" (ANR).

CRediT authorship contribution statement

Aude Ratier: Supervision, formal analysis, Data curation, Writing

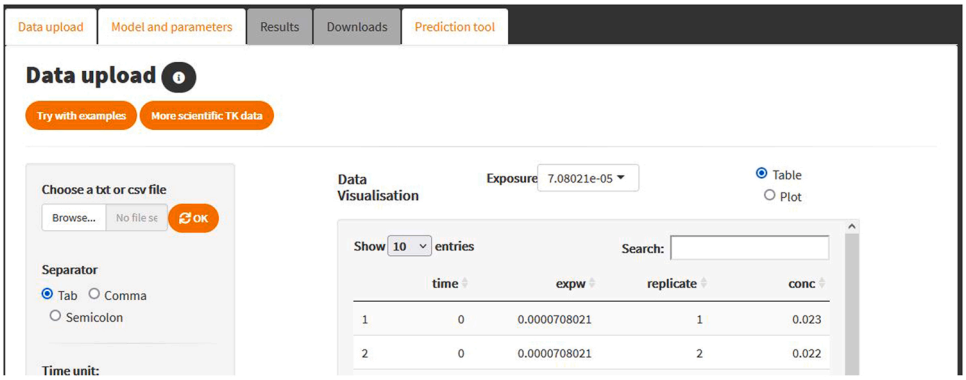


Fig. 6. Visual homepage of the MOSAIC$_{bioacc}$ first tab (Data upload); once a data set has been uploaded, the second tab (Model and parameters) is unlocked from which the TK model run can be asked thus unlocking Results and Downloads tabs in the following.

manuscript. **Virgile Baudrot:** (main developer symbolised) Conceptualisation, Methodology, Data curation, Visualisation, Writing manuscript. **Miléna Kaag:** Data curation, Conceptualisation, Methodology, Visualisation, Writing manuscript. **Christelle Lopes:** Supervision, Formal analysis, Data curation, Reviewing manuscript. **Aurélie Siberchicot:** Conceptualisation, Software maintenance, Formal analysis, Data curation, Reviewing manuscript. **Sandrine Charles:** Supervision, Funding acquisition, Project administration, Formal analysis, Data curation, reviewing manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is part of the ANR project APPROve (ANR-18-CE34-0013) for an integrated approach to propose proteomics for biomonitoring: accumulation, fate and multi-markers (https://anr.fr/Projet-ANR-18-CE34-0013). This work benefited from the French GDR “Aquatic Ecotoxicology" framework which aims at fostering stimulating scientific discussions and collaborations for more integrative approaches. Authors fully thank all contributors to rbioacc: Théo Ciccia, Ophélia Gestin, Gauthier Multari and Alain Pavé. This work was performed using the computing facilities of the CC LBBE/PRABI.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ecoenv.2022.113875.

References

Ashauer, R., Brown, C.D., 2013. Highly time-variable exposure to chemicals–toward an assessment strategy. Integr. Environ. Assess. Manag. 9, e27–e33. https://doi.org/10.1002/ieam.1421.

Ashauer, R., Hintermeister, A., O'Connor, I., Elumelu, M., Hollender, J., Escher, B.I., 2012. Significance of xenobiotic metabolism for bioaccumulation kinetics of organic chemicals in Gammarus pulex. Environ. Sci. Technol. 46, 3498–3508. https://doi.org/10.1021/es204611h.

Baudrot, V., Charles, S., 2019. Recommendations to address uncertainties in environmental risk assessment using toxicokinetics-toxicodynamics models. Sci. Rep. 9, 11432. https://doi.org/10.1101/356469.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: a probabilistic programming language. J. Stat. Softw. 76, 1–32. https://doi.org/10.18637/jss.v076.i01.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2021.shiny: Web application framework for R.

Charles, S., Ratier, A., Baudrot, V., Multari, G., Siberchicot, A., Wu, D., Lopes, C., 2021a. Taking full advantage of modelling to better assess environmental risk due to xenobiotics–the all-in-one facility mosaic. Environ. Sci. Pollut. Res. 29, 29244–29257. https://doi.org/10.1007/s11356-021-15042-7.

Charles, S., Ratier, A., Lopes, C., 2021b. Generic solving of one-compartment toxicokinetic models. J. Explor. Res. Pharmacol. Submitt. 6, 158–167. https://doi.org/10.1101/2021.05.06.442956v1.

Charles, S., Veber, P., Delignette-Muller, M.L., 2018. Mosaic: a web-interface for statistical analyses in ecotoxicology. Environ. Sci. Pollut. Res. 25, 11295–11302. https://doi.org/10.1007/s11356-017-9809-4.

Crookes, M.J., Brooke, D.N., 2011.Estimation of fish bioconcentration factor (BCF) from depuration data.Bristol, UK.

Gabry, J., Goodrich, B., Lysy, M., 2020.rstantools: Tools for developing R packages interfacing with Stan.

Lo, J.C., Campbell, D.A., Kennedy, C.J., Gobas, F.A., 2015. Somatic and gastrointestinal in vivo biotransformation rates of hydrophobic chemicals in fish. Environ. Toxicol. Chem. 34, 2282–2294. https://doi.org/10.1002/etc.3050. ⟨https://setac.onlinelibrary.wiley.com/doi/abs/10.1002/etc.3050⟩.

Lu, X., Reible, D.D., Fleeger, J.W., 2004. Relative importance of ingested sediment versus pore water as uptake routes for pahs to the deposit-feeding oligochaete Ilyodrilus templetoni. Arch. Environ. Contam. Toxicol. 47, 207–214. https://doi.org/10.1007/s00244-004-3053-x.

Mackay, D., Fraser, A., 2000. Bioaccumulation of persistent organic chemicals: mechanisms and models. Environ. Pollut. 110, 375–391. https://doi.org/10.1016/S0269-7491(00)00162-7. ⟨https://www.sciencedirect.com/science/article/pii/S0269749100001627⟩.

Nalecz-Jaweski, G., Wawryniuk, M., Giebultowicz, J., Olkowski, A., Drobniewska, A., 2020. Influence of selected antidepressants on the ciliated protozoan Spirostomum ambiguum: Toxicity, bioaccumulation, and biotransformation products. Molecules 25, 1476. https://doi.org/10.3390/molecules25071476.

Ockleford, C., Adriaanse, P., Berny, P., Brock, T., Duquesne, S., Grilli, S., Hernandez-Jerez, A.F., Bennekou, S.H., Klein, M., Kuhl, T., Laskowski, R., Machera, K., Pelkonen, O., Pieper, S., Smith, R.H., Stemmer, M., Sundh, I., Tiktak, A., Topping, C. J., Wolterink, G., Cedergreen, N., Charles, S., Focks, A., Reed, M., Arena, M., Ippolito, A., Byers, H., Teodorovic, I., Panel, E.P., 2018. Scientific opinion on the state of the art of toxicokinetic/toxicodynamic (tktd) effect models for regulatory risk assessment of pesticides for aquatic organisms. EFSA J. 16, 5377. https://doi.org/10.2903/j.efsa.2018.5377.

Core Team, R., 2021. R: A Lang. Environ. Stat. Comput.

Ratier, A., Charles, S., 2022. Accumulation-depuration data collection in support of toxicokinetic modelling. Sci. Data 9, 130. https://doi.org/10.1101/2021.04.15.439942.

Ratier, A., Lopes, C., Geffard, O., Babut, M., 2021a. The added value of bayesian inference for estimating biotransformation rates of organic contaminants in aquatic invertebrates. Aquat. Toxicol., 105811 https://doi.org/10.1016/j.aquatox.2021.105811.

Ratier, A., Lopes, C., Labadie, P., Budzinski, H., Delorme, N., Quéau, H., Peluhet, L., Geffard, O., Babut, M., 2019. A bayesian framework for estimating parameters of a generic toxicokinetic model for the bioaccumulation of organic chemicals by benthic invertebrates: proof of concept with pcb153 and two freshwater species. Ecotoxicol. Environ. Saf. 180, 33–42. https://doi.org/10.1016/j.ecoenv.2019.04.080.

Ratier, A., Lopes, C., Multari, G., Mazerolles, V., Carpentier, P., Charles, S., 2021b. New perspectives on the calculation of bioaccumulation metrics for active substances in living organisms. Integr. Environ. Assess. Manag. on-line 18, 10–18. https://doi.org/10.1101/2020.07.07.185835.

Rubach, M.N., Ashauer, R., Maund, S.J., Baird, D.J., Van Den Brink, P.J., den Brink, P.J., 2010. Toxicokinetic variation in 15 freshwater arthropod species exposed to the insecticide chlorpyrifos. Environ. Toxicol. Chem. 29, 2225–2234. https://doi.org/10.1002/etc.273.

Schuler, L.J., Wheeler, M., Bailer, A.J., Lydy, M.J., 2003. Toxicokinetics of sediment-sorbed benzo[a]pyrene and hexachlorobiphenyl using the freshwater invertebrates Hyalella azteca, Chironomus tentans, and Lumbriculus variegatus. Environ. Toxicol. Chem. 22, 439–449, 10.1897/1551-5028(2003)022<0439:TOSSBA>2.0.CO;2.

Stan Development Team, 2021. rstan: the R interface to Stan.

Svobodová, M., Šmídová, K., Hvězdová, M., Hofman, J., 2018. Uptake kinetics of pesticides chlorpyrifos and tebuconazole in the earthworm eisenia andrei in two different soils. Environ. Pollut. 236, 257–264. https://doi.org/10.1016/j.envpol.2018.01.082. ⟨https://www.sciencedirect.com/science/article/pii/S0269749117344755⟩.

Wang, H., Xia, X., Wang, Z., Liu, R., Muir, D.C., Wang, W.X., 2021. Contribution of dietary uptake to pah bioaccumulation in a simplified pelagic food chain: Modeling the influences of continuous vs intermittent feeding in zooplankton and fish. Environ. Sci. Technol. 55, 1930–1940. https://doi.org/10.1021/acs.est.0c06970 pMID: 33448220.

Wickham, H., Danenberg, P., Csardi, G., Eugster, M., 2021.roxygen2: In-line documentation for R.

Zhu, M., Wang, Z., Chen, J., Xie, H., Zhao, H., Yuan, X., 2020. Bioaccumulation, biotransformation, and multicompartmental toxicokinetic model of antibiotics in sea cucumber (Apostichopus japonicus). Environ. Sci. Technol. 54, 13175–13185. https://doi.org/10.1021/acs.est.0c04421.

Further reading

OECD, 2008. Test No 315: Bioaccumulation in Sediment-dwelling Benthic Oligochaetes, 315. OECD Publishing. https://doi.org/10.1787/2074577x.

OECD, 2012. Test No. 305: Bioaccumulation in Fish: Aqueous and Dietary Exposure, 305. OECD Publishing. https://doi.org/10.1787/9789264185296-en.