



HAL
open science

New perspectives on the calculation of bioaccumulation metrics for active substances in living organisms

Aude Ratier, Christelle Lopes, Gauthier Multari, Vanessa Mazerolles, Patrice Carpentier, Sandrine Charles

► To cite this version:

Aude Ratier, Christelle Lopes, Gauthier Multari, Vanessa Mazerolles, Patrice Carpentier, et al.. New perspectives on the calculation of bioaccumulation metrics for active substances in living organisms. Integrated Environmental Assessment and Management, 2022, 18, pp.10-18. hal-03343427

HAL Id: hal-03343427

<https://hal.science/hal-03343427v1>

Submitted on 20 Feb 2024

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Revised Manuscript [2021-04-19]

NEW PERSPECTIVES ON THE CALCULATION OF BIOACCUMULATION METRICS FOR ACTIVE SUBSTANCES IN LIVING ORGANISMS

Aude Ratier^a, Christelle Lopes^a, Gauthier Multari^a, Vanessa Mazerolles^b, Patrice Carpentier^b,

Sandrine Charles^{1,a}

^aUniversité de Lyon, Université Lyon 1, CNRS UMR5558, Laboratoire de Biométrie et Biologie Evolutive,
69100 Villeurbanne, France.

^bAnses, 14 rue Pierre et Marie Curie, 94701 Maisons-Alfort Cedex, France.

Acknowledgement

The authors are thankful to ANSES for providing the financial support. The MOSAIC_{bioacc} web tool is hosted at the Rhône-Alpes Bioinformatics Center PRABI (PRABI, 2020). This work benefited from the French GDR “Aquatic Ecotoxicology” framework which aims at fostering stimulating scientific discussions and collaborations for more integrative approaches. 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 was also made under the umbrella of 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). The authors are truly grateful to the anonymous colleagues who participated in testing MOSAIC_{bioacc} and for giving their feedback. The authors also are grateful for Benoît BRET for creating the logo for MOSAIC_{bioacc}. The authors declare no competing interests.

Data availability

Data are accessible directly within MOSAIC_{bioacc} at <https://mosaic.univ-lyon1.fr/bioacc>.

¹ Corresponding Author

Email address: sandrine.charles@univ-lyon1.fr (Sandrine Charles)

Revised Manuscript [2021-04-19]

1 **NEW PERSPECTIVES ON THE CALCULATION OF** 2 **BIOACCUMULATION METRICS FOR ACTIVE** 3 **SUBSTANCES IN LIVING ORGANISMS**

4 **ABSTRACT**

5 Today, only few ready-to-use and convenient decision-making tools are available in
6 ecotoxicology concerning accumulation and effects of chemical substances on organisms,
7 accounting for exposure situations that are known to be complex (routes of exposure,
8 metabolism, mixtures, etc.). This paper presents new perspectives on the generic calculation of
9 bioaccumulation metrics via the innovative web tool MOSAIC_{bioacc} ([http://mosaic.univ-](http://mosaic.univ-lyon1.fr/bioacc)
10 [lyon1.fr/bioacc](http://mosaic.univ-lyon1.fr/bioacc)). MOSAIC_{bioacc} provides all kind of bioaccumulation metrics associated with
11 their uncertainty whatever the species-compound combination. MOSAIC_{bioacc} expects
12 accumulation-depuration data as inputs, even with complex exposure and clearance patterns, to
13 quickly perform their relevant analysis. MOSAIC_{bioacc} intends to facilitate the daily work of
14 regulators, or any ecotoxicologist, who will freely benefit from a user-friendly on-line interface
15 that automatically fits toxicokinetic models without needs for users to invest in the technical
16 aspects to get bioaccumulation metrics estimates. MOSAIC_{bioacc} also provides all results in a
17 fully transparent way to ensure reproducibility.

18 **KEYWORDS**

19 Environmental Risk Assessment - Toxicokinetic models - Bayesian inference - Uncertainties –
20 User-friendly web platform

Revised Manuscript [2021-04-19]

21 INTRODUCTION

22 Faced with the current environmental challenges, linked in particular to environmental
23 pollution, ecotoxicology must today provide relevant and effective decision-making tools
24 regarding bioaccumulation and effects of chemical substances on living organisms. Such tools
25 must account for various exposure situations, environmentally realistic but complex (*e.g.*,
26 several routes of exposure, metabolism of substances, mixtures, etc.). Among available
27 methods, toxicokinetic-toxicodynamic (TKTD) models are now strongly recommended to
28 describe the link between exposure concentrations and effects on individual life-history traits
29 over time from experimental data collected through toxicity tests, even standard ones (EFSA
30 PPR Panel 2018). More specifically, the TK part of these models is used to relate the exposure
31 concentration to the time course of the internal concentration within organisms, considering
32 various processes such as accumulation, depuration, metabolization and excretion (known as
33 ADME processes). As some recent regulations, the EU regulation No 283/2013 for plant
34 protection products in marketing authorisation applications requires for example a
35 bioaccumulation test on fish according to OECD Test guideline 305 (OECD 2012), which
36 consists in an accumulation phase followed by a depuration phase. During the accumulation
37 phase, fish are exposed to a substance of interest at a range of concentrations, chosen according
38 to the assumed mode of action of the substance. After a certain time period fixed by the
39 experimenter, organisms are transferred to a clean medium for a depuration phase. The
40 concentration of the substance (and of its potential metabolites) is followed within fish at
41 regular time points during both phases leading *in fine* to the estimation of bioaccumulation
42 metrics. In this paper, for the sake of generality, we chose the generic expression
43 “bioaccumulation metrics” to denote either bioconcentration factors (BCF) used when exposure
44 is via water, biota-sediment accumulation factors (BSAF) when exposure is via sediment or
45 biomagnification factors (BMF) when exposure is via food. Bioaccumulation metrics appeared
46 to us as the best compromise regarding the wide diversity of terms used in the scientific

Revised Manuscript [2021-04-19]

47 literature (for example, USEPA (1994), Gobas et al. (2009) and Burkhard et al. (2012)).
48 Nevertheless, preference is often given to experimentally derived BCF estimates to be used for
49 secondary poisoning assessment under Biocidal Products Regulation (European Commission
50 2012). Bioaccumulation tests are of course not only limited to fish, even if a test according to
51 OECD Test guideline 305 (OECD 2012) is preferred when experimental information on
52 bioaccumulation is needed for PBT/vPvB assessment under REACH regulation (ECHA 2017;
53 European Commission 2006). Consequently, any other species can be used, such as benthic
54 invertebrates, terrestrial oligochaetes or birds, depending on the substance under consideration.
55 From a regulatory point of view, bioaccumulation metrics are key decision criteria used to
56 evaluate concentrations of active substances in food items of vertebrates (especially piscivorous
57 birds and mammals), making the estimation of these metrics with the most precision as possible
58 a highly crucial methodological challenge.

59 All bioaccumulation metrics rely on estimates of kinetic parameters as involved in
60 toxicokinetic (TK) models. In the past decades, many types of methods have been proposed to
61 get these estimates from simple TK models, most of them providing BCF estimates separately
62 considering the kinetics for both accumulation and depuration phases as observed in dedicated
63 experiments (OECD 2012). Nevertheless, TK model parameters are known to be highly
64 correlated, so that separating their estimation prevents to account for a mutual influence on their
65 uncertainty. Moreover, kinetic parameter estimates are only rarely provided with their
66 uncertainty, although this is now expected by the regulatory bodies (EFSA Scientific
67 Committee 2018). Consequently, concomitantly to the above-mentioned challenge,
68 environmental risk assessment could be improved if complete tools allowing for a
69 simultaneously estimation of all TK model parameters associated with their uncertainty would
70 be available in support of stakeholders that need to fulfil regulatory expectations. These tools
71 must also be easy to use to overcome the scepticism of regulators who are faced with multiple
72 TK models and implementation methods, while thinking about their standardization at the same

Revised Manuscript [2021-04-19]

73 time (Tan et al. 2020). An R-code was first proposed in 2016 (Aldenberg 2019) that allowed to
74 analyse data collected only from the OCED test guideline 305. In the same line of thought, a
75 spreadsheet was recently proposed by Gobas et al. (2020). Our paper goes beyond by
76 considering all type of species-compound accumulation-depuration data (not only those of the
77 OECD test guideline 305) which analysis leads to one or several bioaccumulation metrics of
78 interest.

79 Ratier et al. (2019) recently proposed a full revisit of the TK modelling approach based
80 on a unified inference method to estimate parameters of TK models for both accumulation and
81 depuration phases, simultaneously, automatically associating the uncertainties. This innovative
82 framework has been thought to make it possible to further incorporate the TK part into complete
83 TKTD models. Benefiting from this innovation, we present new perspectives for a facilitated
84 calculation of any type of bioaccumulation metrics (such as BCF/BSAF/BMF) thanks to the
85 new ready-to-use statistical web tool MOSAIC_{bioacc} (<http://mosaic.univ-lyon1.fr/bioacc>).
86 MOSAIC_{bioacc} runs one-compartment TK models that are automatically designed according to
87 the input data. MOSAIC_{bioacc} leads to bioaccumulation metrics associated with their uncertainty
88 propagated from the kinetic parameter estimates, without the need for users to invest underlying
89 technical aspects. MOSAIC_{bioacc} is free of use, fully integrated within the all-in-one facility
90 MOSAIC itself (<http://mosaic.univ-lyon1.fr>). MOSAIC_{bioacc} is regularly updated to always
91 offers the very latest conceptual advances related to TK models. Today, MOSAIC_{bioacc} allows
92 accounting for several exposure routes (water, pore water, sediment and food), for metabolism
93 of chemicals (if the input experimental data include measurements for both the parent chemical
94 and its metabolites), and for potential growth of organisms (if growth measurements are
95 included within the data set). The use of MOSAIC_{bioacc} only requires users to upload their
96 experimental data, collected via standard protocols or from home-made experimental designs.
97 Priority was first given to the calculation of BCF/BSAF/BMF because they are the widely used
98 bioaccumulation metrics in the current regulatory guidelines. MOSAIC_{bioacc} get their estimate

Revised Manuscript [2021-04-19]

99 as probability distributions that are summarized for users by the median (50th centile of the
100 distribution) and the 95% credible interval (delimited by the 2.5th and 97.5th centiles)
101 quantifying the uncertainty. In addition, the fitting plots and all model parameter estimates are
102 provided, followed by a collection of goodness-of-fit criteria allowing users to check the
103 relevance of the results. All outputs can be downloaded under different formats for further
104 inclusion into any home-made document (in particular the open-source programming code),
105 and a full report can also be directly downloaded gathering everything that is displayed to users
106 on the web page. These two latter features of MOSAIC_{bioacc} guarantee both reproducibility and
107 transparency of underlying calculations.

108 **MATERIALS AND METHODS**

109 MOSAIC_{bioacc} is part of the web platform MOSAIC (<https://mosaic.univ-lyon1.fr/>,
110 Charles et al. 2018). It was developed as a Shiny environment (Chang et al. 2020), available at
111 <https://mosaic.univ-lyon1.fr/bioacc> and hosted at the Rhône-Alpes Bioinformatics Center
112 PRABI (PRABI 2020). A user guide and an explanatory video are immediately available in the
113 introductory section of the application to fully assist users step by step in appropriating the tool
114 and its features. Details on underlying ordinary differential equations and their solving are also
115 provided within a detailed user guide (see supplemental data, **Annex 1**), thus ensuring all
116 required transparency as recommended by EFSA for a good modelling practice (EFSA PPR
117 Panel 2014).

118 ***Data uploading***

119 When using MOSAIC_{bioacc}, the first step is to upload input data (**Fig. 1-a**). MOSAIC_{bioacc}
120 expects to receive experimental exposure time-course data, including at least an accumulation
121 phase, as a .txt file or a .csv file (comma, semicolon or tabular separator) with a specific format.
122 Each line of the table corresponds to one time point for a given replicate and a given exposure
123 concentration of the contaminant. The data table must contain at least four columns, with the

Revised Manuscript [2021-04-19]

124 exact following headers, the order of columns being not important (**Table 1**): ‘time’ (the time
125 point of the measurement at the exposure concentration, in hours, minutes, days or weeks);
126 ‘expw’, ‘exppw’, ‘exps’ or ‘expf’: the exposure concentration of the contaminant in the
127 medium, that is water, pore water, sediment or food, respectively, all expressed in $\mu\text{g.mL}^{-1}$ or
128 in $\mu\text{g.g}^{-1}$); ‘replicate’ (a number or a character that is unique for each replicate, dimensionless);
129 and ‘conc’ (the concentration of the contaminant, and of its potential metabolites, measured
130 within organisms, in $\mu\text{g.g}^{-1}$). According to the experimental design, further columns can be
131 added in the data file: ‘expw’, ‘exppw’, ‘exps’ and/or ‘expf’ if several exposure routes are
132 considered together, ‘concm ℓ ’ (the concentrations of metabolite ℓ derived from the parent
133 compound within the organisms, *e.g.*, concm1, concm2, ...), and ‘growth’ (if growth
134 measurements of the organisms are available).

135 As shown on **Fig. 1-a**, MOSAIC_{bioacc} users can either upload their own data set with a click on
136 ‘Browse’ (by taking care about the expected format specification) or try MOSAIC_{bioacc} with
137 example data sets (in total six data sets are proposed, each with different characteristics). When
138 the upload is complete, users must manually select the appropriate separator, the time unit and
139 the duration of the accumulation phase; please note that, when using example data sets, these
140 fields are automatically filled in.

141 MOSAIC_{bioacc} first provides a table with the raw data allowing users to check if the data
142 were correctly entered. Users can also visualize a plot with the raw data (**Fig. 1-b**). In case data
143 were collected for several exposure concentrations and if the users have uploaded all, one must
144 be chosen for the MOSAIC_{bioacc} analysis. Note that only one file at a time can be analysed. Also,
145 when another exposure concentration from the same data file is chosen, the duration of the
146 accumulation phase is reset and framed in orange to invite users for update before to launch
147 new calculations. Example files provided in MOSAIC_{bioacc} are dedicated to assist users in
148 formatting their own data and to appropriate the different MOSAIC_{bioacc} features from the very
149 first step. This paper illustrates this step-by-step process based on a typical data set of a

Revised Manuscript [2021-04-19]

150 toxicokinetic bioassay where internal concentrations were collected in fathead minnows
151 (*Pimephales promelas*) exposed via contaminated water to a highly hydrophobic chemical
152 ($\log K_{ow} = 9.06$) at an exposure concentration of $0.0044 \mu\text{g}\cdot\text{mL}^{-1}$ over 49 days, with one
153 replicate at each time point. After 49 days, minnows were transferred in a clean medium for 98
154 days more (Crookes and Brooke 2011). The data set (CSV format) and the report with all results
155 (HTML format) can be downloaded directly from MOSAIC_{bioacc}.

156 ***Model and parameters***

157 All TK models considered in MOSAIC_{bioacc} describe organisms as single compartments
158 for which a first-order kinetic bioaccumulation model accounting for several exposure routes
159 and elimination processes can be expressed in a generic way as follows (Eqs. (1) to (4)):

$$160 \quad \begin{cases} \frac{dC_p(t)}{dt} = U - (E + M)C_p(t) & (1) \\ \frac{dC_{m_\ell}(t)}{dt} = k_{m_\ell}C_p(t) - k_{e_{m_\ell}}C_{m_\ell}(t) & (2) \end{cases} \quad \text{for } 0 \leq t \leq t_c$$

$$161 \quad \begin{cases} \frac{dC_p(t)}{dt} = -(E + M)C_p(t) & (3) \\ \frac{dC_{m_\ell}(t)}{dt} = k_{m_\ell}C_p(t) - k_{e_{m_\ell}}C_{m_\ell}(t) & (4) \end{cases} \quad \text{for } t > t_c$$

162 where $C_p(t)$ is the internal concentration of the parent compound at time t ($\mu\text{g}\cdot\text{g}^{-1}$),
163 $C_{m_\ell}(t)$ the internal concentration ($\mu\text{g}\cdot\text{g}^{-1}$) of metabolites ($\forall \ell = 1 \dots L$ with L the total number
164 of metabolites) at time t , U the sum of all uptake terms, E the sum of all elimination terms for
165 the parent compound, M the sum of all metabolization terms, k_{m_ℓ} the metabolization rate of
166 metabolite ℓ (time^{-1}) and $k_{e_{m_\ell}}$ the elimination rate of metabolite ℓ (time^{-1}). **Table 2** gives an
167 overview of all parameter and variable meaning. The dynamical system in equations (1) to (4),
168 corresponding to the deterministic part of the model, can explicitly be solved when the exposure
169 concentration is assumed to be constant over time (**Annex 1**). A Gaussian probability
170 distribution was assumed as the stochastic part of the final model, based on the quantitative

Revised Manuscript [2021-04-19]

171 continuous nature of the concentration variables for both the parent compound and its
172 metabolites within the organisms (Eqs. (5) and (6)):

$$173 \quad C_{obs,p}(t) \sim \mathcal{N}\left(C_p(t), \sigma_{C_p}^2\right) \quad (5)$$

$$174 \quad C_{obs,m_\ell}(t) \sim \mathcal{N}\left(C_{m_\ell}(t), \sigma_{m_\ell}^2\right) \quad (6)$$

175 Today MOSAIC_{bioacc} proposes data analyses by including until four exposure routes via
176 water, pore water, sediment, and/or food), until three processes of elimination, which are
177 excretion, growth dilution and biotransformation, with a maximum of 15 metabolites directly
178 deriving from the parent compound (*i.e.*, phase I metabolism). For example, if three metabolites
179 are considered ($L = 3$), the number of parameters involved in the most complete TK model
180 equals 19. In total, users have 112 possible models, automatically designed according to their
181 data (see supplemental data, **Annex 2**). For a given data set, the most complete model is built
182 up by default and first proposed to users, from which they can perform the MOSAIC_{bioacc}
183 analysis (**Fig. 1-c**). Users can also deselect some of the parameters (based on biological
184 hypotheses related to the most probable exposure route or by neglecting one elimination
185 process, for example). These choices lead to the automatic building of a nested TK sub-model
186 to fit again on the data.

187 For clarity reasons, this paper illustrates MOSAIC_{bioacc} features from a simple data set
188 (Pimephales_two.csv file) considering only the water exposure route (parameter k_{uw} for the
189 uptake rate) and the excretion process (parameter k_{ee} for the elimination rate), that is one of the
190 simplest TK models among the 112 possibilities. The corresponding equations for the
191 deterministic part of this TK model are given below (Eqs. (7) and (8)):

$$192 \quad \left\{ \begin{array}{l} \frac{dC_p(t)}{dt} = k_{uw} \times c_w - k_{ee} \times C_p(t) \quad \text{for } 0 \leq t \leq t_c \quad (7) \\ \frac{dC_p(t)}{dt} = -k_{ee} \times C_p(t) \quad \text{for } t > t_c \quad (8) \end{array} \right.$$

193 Finally, there are only three parameters to estimate: k_{uw} , k_{ee} and parameter σ_{C_p} of the
194 Gaussian distribution (Eq. (5)). An illustration of a more complex data set with

Revised Manuscript [2021-04-19]

195 biotransformation and growth processes for a benthic invertebrate is given as supplemental
196 material (**Annex 3**), as well as examples for exposure route by sediment (**Annex 4**) or food
197 (**Annex 5**).

198 Once the Pimephales_two.csv example file has been uploaded, all required fields related
199 to the experimental design are automatically filled in and raw data can be visualized, either as
200 a table (default) or a plot, and the button to launch calculations ('Calculate and Display') is
201 unlocked (**Fig. 1-c**). Once this button clicked, calculations start running with a progress bar
202 informing users about progress. When calculations are finished, results are displayed, as plots
203 or tables first for the bioaccumulation metrics, then for the fitting plots and some relevant
204 goodness-of-fit criteria.

205 ***Calculations***

206 *Bayesian inference*

207 Computations underlying MOSAIC_{bioacc} results are performed with JAGS (Plummer
208 2019) and the R software (R Core Team 2020, version 4.0.2) via the `rjags` and `jagsUI`
209 packages (Plummer 2019; Kellner 2019). Models are fitted to bioaccumulation data using
210 Bayesian inference via Monte Carlo Markov Chain (MCMC) sampling. For each model,
211 calculation running starts with a short sampling on three MCMC chains (5,000 iterations after
212 a burn-in phase of 10,000 iterations) using the Raftery and Lewis method (Raftery and Lewis
213 1992) to set the necessary thinning and the appropriate number of iterations in order to reach a
214 precise and accurate estimation of each model parameter. Thanks to `rjags`, model parameters
215 are retrieved as a joint posterior distribution from the likelihood of the observed data combined
216 with prior information on parameters. All details on this approach can be found in the original
217 research paper (Ratier et al. 2019) but also in many other papers in the field of ecotoxicology
218 (Billoir et al. 2011; EFSA PPR Panel 2018).

Revised Manuscript [2021-04-19]

219 *Choice of prior distributions*

220 For simplicity reasons, we hid the choice of priors to MOSAIC_{bioacc} users; hence, they
221 cannot be changed, except by downloading the open source programming code and handling it
222 directly within the R software. To ensure genericity of priors we chose non-informative (-5, 5)
223 log10-uniform distributions for all uptake and elimination rate constants, and non-informative
224 (0, A) uniform distributions for all standard deviations with a large A, here defined as five times
225 the maximum internal measured concentration, which is then removed from the data set, as
226 usually proceeded (Gelman 2006).

227 *Bioaccumulation metrics*

228 Bioaccumulation metrics are the first outputs delivered by MOSAIC_{bioacc}. From the
229 example chosen for this paper, the data analysis led to both the kinetic bioconcentration factor
230 (BCF_k) and the steady state bioconcentration factor (BCF_{ss}), with the following exact
231 mathematical expressions (Eqs. (9) and (10)):

232
$$BCF_k = \frac{k_{uw}}{k_{ee}} \quad (9)$$

233
$$BCF_{ss} = \frac{C_p(t_c)}{c_w} \quad (10)$$

234 where $C_p(t_c)$ is the internal parent compound concentration (in $\mu\text{g}\cdot\text{g}^{-1}$) at the end of the
235 accumulation phase (that is at $t = t_c$, in time) and c_w is the exposure contaminant concentration
236 in water ($\mu\text{g}\cdot\text{mL}^{-1}$). More details about calculations of bioaccumulation metrics at steady state
237 are provided in **Annex 6**.

238 **RESULTS AND DISCUSSION**

239 *Bioaccumulation metrics*

240 When users click on the ‘Calculate and Display’ button, some results are provided by
241 default. First, the BCF_k is given as a probability distribution (**Fig. 2-a**) and summarized with its
242 median and its 95% uncertainty limits, that is 95% credible interval delimited by the 2.5th and

Revised Manuscript [2021-04-19]

243 the 97.5th centiles of the posterior probability distribution (**Table 3-a**). If users ask for the BCF_{ss} ,
244 its probability distribution is also delivered (**Fig. 2-b**) also summarized with the median and the
245 95% uncertainty limits (**Table 3-a**). Credible intervals are crucial information to quantify the
246 uncertainty on parameter estimates. If data are available for several exposure routes (according
247 to the experimental design) and uploaded within MOSAIC_{bioacc}, the BCF/BSAF/BMF metrics
248 are displayed in separate tabs (see **Annex 4** for an example).

249 ***Predictions***

250 Following bioaccumulation metrics calculations, the fitted curve and its uncertainty
251 band superimposed to the observations is provided (**Fig. 2-c**). From the joint posterior
252 distribution of model parameters, MOSAIC_{bioacc} then provides the marginal posterior
253 distributions for each parameter, which are also summarized with quantiles in a table (**Table 3-**
254 **b**): medians (for point estimates) and 2.5th and 97.5th centiles (for 95% credible intervals).

255 ***Goodness-of-fit criteria***

256 After fitting plots, several goodness-of-fit criteria follow in a prioritized order chosen
257 based on their relevance and their ease of interpretation. The fitting quality of the model can be
258 first checked using the Posterior Predictive Check (PPC) plot: the idea is to compare each
259 observed value to its prediction from the fitted model at the corresponding exposure
260 concentration associated with its 95% credible interval. If the fit is correct, we expect to get
261 95% of the observed values falling within the 95% credible intervals of their predictions. As
262 shown on **Fig. 3-a**, x -axis locates the observed values, while the y -axis reports their median
263 predictions (black dots) with their 95% credible intervals (vertical segments).

264 The relevance of the inference process can also be checked using the comparison of
265 prior and posterior distributions for each model parameter. The overall expectation is to get a
266 narrower posterior distribution compared to the prior one for each parameter, reflecting that
267 data contributed enough to precisely estimate parameters (**Fig.3-b**). Users have the possibility
268 to select plots for deterministic (e.g., k_{uw} , k_{ee}) or stochastic (e.g., σ_{Cp}) parameters.

Revised Manuscript [2021-04-19]

269 Then, MOSAIC_{bioacc} provides a coloured matrix in order to see at a glance the most
270 correlated or anti-correlated parameters, in order to quickly diagnose potential problems of
271 precision due to highly correlated parameters. Moreover, MOSAIC_{bioacc} provides plots to
272 visualize correlations between parameters (**Fig. S1-a**). Such a plot is obtained by projecting the
273 joint posterior distribution as a matrix in planes of parameter pairs where contours have shapes
274 reflecting both the sign and the strength of the correlations (sub-diagonal). The correlation plot
275 also gives marginal posterior distribution of each model parameter (diagonal) and Pearson
276 correlation coefficients (upper diagonal). Correlations between parameters are important to
277 consider in particular when they are high (namely, greater than 0.75) what would mean that one
278 parameter estimate could considerably influence the other, and reciprocally. Users can display
279 the correlation plot for deterministic parameters only or for all parameters.

280 The convergence of MCMC chains can be checked with the Gelman-Rubin diagnostic
281 (Gelman and Rubin 1992) expressed via the potential scale reduction factor (PSRF) which is
282 expected to be close to 1.00 (**Fig. S1-b**). It can be also visually verified from the MCMC trace
283 plots, which show the time series of the sampling process leading to the posterior distribution
284 for each parameter; it is expected to get overlap of all MCMC chains (**Fig. S1-c**). Users can
285 visualize the MCMC trace plots for deterministic or stochastic parameters.

286 Finally, the Deviance Information Criterion (DIC) is provided. It is a penalized deviance
287 statistics accounting for the number of parameters that is only useful to compare several models
288 fitted to a same data set. Models with lower DIC values will be preferred. So, the DIC is only
289 useful when several sub-models are compared based on different choices of parameters from
290 all the possible combinations that the users can choose from the beginning according to the
291 uploaded data.

292 ***Downloads***

293 At the bottom of the result web page, all outputs, either separately or as a full report,
294 can be downloaded. Users can also download the entire R code corresponding to all calculations

Revised Manuscript [2021-04-19]

295 and graphs from the uploaded data set. This ensures transparency and reproducibility of all
296 MOSAIC_{bioacc} results. This R code can be used as a steppingstone to change default options or
297 to perform further analyses directly in the R software. For example, users can modify figures at
298 their convenience or make several analyses on several data sets at the same time.

299 **CONCLUSION**

300 Offering MOSAIC_{bioacc} as a new on-line service, free, user-friendly and ready-to-use,
301 raises important methodological issues: (1) automation of the inference process, in particular
302 for Bayesian inference; (2) options and choices in a transparent and facilitated way for users;
303 (3) default outputs whose order has been chosen to facilitate their step-by-step interpretation;
304 (4) easy accessibility to figures, tables and R code to be downloaded under different convenient
305 formats; and (5) a final full report of all MOSAIC_{bioacc} analyses. Besides its user-friendliness,
306 MOSAIC_{bioacc} is free of use while ensuring privacy of the uploaded data as well as transparency
307 and reproducibility of results, together with a short response time. MOSAIC_{bioacc} is particularly
308 useful to estimate parameters of TK models leading to predictions of chemical concentrations
309 bioaccumulated within living organism (whatever the species, aquatic, aerial or terrestrial) from
310 accumulation-depuration data, even standard ones. MOSAIC_{bioacc} could thus be of particular
311 interest for risk assessors and decision makers in their daily work of evaluating dossiers, *e.g.*,
312 for market authorisation of active substances (EFSA PPR Panel 2018). Indeed, all results
313 provided by MOSAIC_{bioacc} account for uncertainty and correlations between parameters,
314 making possible to reproduce any previous analysis that would need to be confirmed.
315 MOSAIC_{bioacc} can also be used for more exploratory research purposes by any environmental
316 scientists or ecotoxicologists when accumulation-depuration data are collected and need to be
317 analysed. MOSAIC_{bioacc} allows analyses for any species-compound combinations under
318 consideration even with biotransformation processes, allowing users to easily perform TK
319 analyses accounting for several exposure routes and phase I metabolites. MOSAIC_{bioacc} is

Revised Manuscript [2021-04-19]

320 available as a new statistical service dedicated to TK modelling approaches embedded within
321 the MOSAIC platform (Charles et al. 2018). This makes MOSAIC an all-in-one facility for
322 many applications. MOSAIC_{bioacc} will be shortly extended with a prediction tool, in order to
323 help in designing new TK experiments, for example, when a new species-compound
324 combination requires attention or when additional data are needed to get a better precision on
325 bioaccumulation metrics. To gain again in generality, an R-package (rbioacc) is already in
326 progress to include all the functionalities encompassed in MOSAIC_{bioacc}, which will also make
327 it possible to extend the use of TK models when the exposure concentrations vary over time,
328 and when it is necessary to consider non-first order kinetics.

329 **SUPPLEMENTAL DATA**

330 Supplemental data are available on-line.

Revised Manuscript [2021-04-19]

331 LIST OF FIGURES

332 **Figure 1.** Data uploading and user information as required from the MOSAIC_{bioacc} homepage.

333 **Figure 2.** Probability distributions of BCF_k (a) and BCF_{ss} (b). The middle-dotted line represents
334 the median value, while left and right dotted lines delimit the 2.5th and 97.5th centiles. (c):
335 Observed (black dots) and predicted contaminant concentrations in the organisms ($\mu\text{g}\cdot\text{g}^{-1}$),
336 where the median curve is displayed as the solid orange line and the uncertainty band as the
337 grey zone, delimited by the 2.5th and 97.5th centiles in dotted orange lines.

338 **Figure 3.** Some goodness-of-fit criteria as provided by MOSAIC_{bioacc}: (a) Posterior Predictive
339 Check (PPC) where observed values are read on the x -axis, while the y -axis reports median
340 predictions (black dots) and their 95% credible intervals (vertical segments, coloured in green
341 if they contain the observed value, in red otherwise); (b) prior (grey) and posterior (orange)
342 marginal distributions of parameters of the chosen TK model (here the simplest one).

343 LIST OF TABLES

344 **Table 1.** Example of a data set ready to be used in MOSAIC_{bioacc}. The data set must contain
345 four columns whatever their order: 'time' (time points of the measurements at the exposure
346 concentration, in time unit: here in days), 'conc' (contaminant concentrations measured within
347 organisms, that must be expressed in $\mu\text{g}\cdot\text{g}^{-1}$), 'expw' (the contaminant concentration in the
348 exposure medium, here water, that must be expressed in $\mu\text{g}\cdot\text{mL}^{-1}$), and 'replicate' (a number or
349 a character that is unique for each replicate, dimensionless).

350 **Table 2.** Meaning of parameters and variables of the TK model used in MOSAIC_{bioacc}.

351 **Table 3.** Example of BCF (a) and parameter (b) estimates expressed as medians (50th centile)
352 and 95% credibility intervals (2.5th - 97.5th centiles). Hyphens stand for dimensionless
353 parameters.

Revised Manuscript [2021-04-19]

354 REFERENCES

355 Aldenberg T. 2019. bmcfR: Tools for Modeling Bioaccumulation Potential in Fish.
356 Version 0.4-18 (bcmfR_0.4-18.tar.gz). [https://www.oecd.org/chemicalsafety/testing/section-3-](https://www.oecd.org/chemicalsafety/testing/section-3-environmental-fate-behaviour-software-tg-305.htm)
357 [environmental-fate-behaviour-software-tg-305.htm](https://www.oecd.org/chemicalsafety/testing/section-3-environmental-fate-behaviour-software-tg-305.htm)

358 Billoir E., Delhaye H., Clément B., Delignette-Muller M.L., Charles S. 2011. Bayesian
359 modelling of daphnid responses to time-varying cadmium exposure in laboratory aquatic
360 microcosms. *Ecotoxicol. Environ. Saf.* 74(4):693-702. doi: 10.1016/j.ecoenv.2010.10.023

361 Burkhard L.P., Arnot J.A., Embry M.R., Farley K.J., Hoke R.A., Kitano M., Leslie H.A.,
362 Lotufo G.R., Parkerton T.F., Sappington K.G., Tomy G.T., Woodburn K.B. 2012. Comparing
363 laboratory and field measured bioaccumulation endpoints. *Integr Environ Assess Manag.*
364 8(1):17-31. doi: 10.1002/ieam.260

365 Chang S., Cheng W., Allaire J. J., Xie Y., McPherson, J. 2020. Shiny: Web Application
366 Framework for R. Version 1.5.0. <http://shiny.rstudio.com/>

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

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

372 European Commission. 2006. COMMISSION REGULATION (EU) No 1907/2006 of
373 the European Parliament and of the Council of 18 December 2006 concerning the Registration,
374 Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European
375 Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC)
376 No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive
377 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and
378 2000/21/EC. Official Journal of the European Union 1–849.

Revised Manuscript [2021-04-19]

379 European Commission. 2012. REGULATION (EU) No 528/2012 of the European
380 Parliament and of the Council of 22 May 2012 concerning the making available on the market
381 and use of biocidal products. Official Journal of the European Union.

382 European Commission. 2013. COMMISSION REGULATION (EU) No 283/2013 of 1
383 March 2013 setting out the data requirements for active substances, in accordance with
384 Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the
385 placing of plant protection product. Official Journal of the European Union.

386 ECHA. [ECHA] European Chemicals Agency. 2017. Guidance on Information
387 Requirements and Chemical Safety Assessment. Chapter R.11: PBT/vPvB Assessment.
388 doi:10.2823/128621

389 EFSA PPR Panel. [EFSA] European Food Safety Authority. 2014. Scientific Opinion
390 on good modelling practice in the context of mechanistic effect models for risk assessment of
391 plant protection products. EFSA Journal. 12:3589. doi: 10.2903/j.efsa.2014.3589

392 EFSA Scientific Committee. [EFSA] European Food Safety Authority. 2018. Guidance
393 on Uncertainty Analysis in Scientific Assessments. EFSA Journal. 16:1–39. doi:
394 10.2903/j.efsa.2018.5123

395 EFSA PPR Panel (Ockleford C., Adriaanse P., Berny P., Brock T., Duquesne S., Grilli
396 S., Hernandez-Jerez A.F., Bennekou S.H., Klein M., Kuhl T., Laskowski R., Machera K.,
397 Pelkonen O., Pieper S., Smith R.H., Stemmer M., Sundh I., Tiktak A., Topping C.J., Wolterink
398 G., Cedergreen N., Charles S., Focks A., Reed M., Arena M., Ippolito A., Byers H., Teodorovic
399 I.). [EFSA] European Food Safety Authority. 2018. Scientific Opinion on the state of the art of
400 Toxicokinetic/Toxicodynamic (TKTD) effect models for regulatory risk assessment of
401 pesticides for aquatic organisms. EFSA Journal. 16(8):5377. doi: 10.2903/j.efsa.2018.5377

402 Gelman A., Rubin D.B. 1992. Inference from Iterative Simulation Using Multiple
403 Sequences. Stat. Sci. 7:457–472.

Revised Manuscript [2021-04-19]

- 404 Gelman, A. 2006. Prior Distribution for Variance Parameters in Hierarchical Models.
405 Bayesian Analysis. 3:5901–5906.
- 406 Gobas, F.A.P.C., de Wolf, W., Burkhard, L.P., Verbruggen, E., Plotzke, K. 2009,
407 Revisiting Bioaccumulation Criteria for POPs and PBT Assessments. Integr Environ Assess
408 Manag. 5: 624-637.
- 409 Gobas F.A.P.C., Lee Y.S., Lo J.C., Parkerton T.F., Letinski D.J. 2020. A Toxicokinetic
410 Framework and Analysis Tool for Interpreting Organisation for Economic Co-operation and
411 Development Guideline 305 Dietary Bioaccumulation Tests. Environ Toxicol Chem.
412 39(1):171-188. doi: 10.1002/etc.4599.
- 413 Kellner, K. 2019. jagsUI: A Wrapper Around 'rjags' to Streamline 'JAGS' Analyses.
414 Version 1.5.1. <https://cran.r-project.org/web/packages/jagsUI/index.html>
- 415 OECD. [OECD] Organisation for Economic Co-operation and Development. 2012. No.
416 305. Bioaccumulation in Fish: Aqueous and Dietary Exposure. Paris (FR): OECD. OECD.
417 doi: 10.1787/9789264185296-en.
- 418 Plummer M. 2019. rjags: Bayesian Graphical Models using MCMC. Version 4.10.
- 419 PRABI. 2020. Date Accessed: 2020-08-28. <http://www.prabi.fr>
- 420 Raftery A.E., Lewis S.M. 1992. [Practical Markov chain Monte Carlo]: Comment: one
421 long run with diagnostics: implementation strategies for Markov chain Monte Carlo. Stat. Sci.
422 7:493–497.
- 423 Ratier A., Lopes C., Labadie P., Budzinski H., Delorme N., Quéau H., Peluhet L.,
424 Geffard O., Babut M. 2019. A Bayesian framework for estimating parameters of a generic
425 toxicokinetic model for the bioaccumulation of organic chemicals by benthic invertebrates:
426 Proof of concept with PCB153 and two freshwater species. Ecotoxicol. Environ. Saf. 180:33-
427 42. doi:10.1016/j.ecoenv.2019.04.080
- 428 R Core Team, 2020. R: A Language and Environment for Statistical Computing.
429 Version 4.0.2.

Revised Manuscript [2021-04-19]

430 Tan Y.M., Chan M., Chukwudebe A., Domoradzki J., Fisher J., Hack C.E., Hinderliter
431 P., Hirasawa K., Leonard J., Lumen A., Paini A., Qian H., Ruiz P., Wambaugh J., Zhang F.,
432 Embry M. 2020. PBPK model reporting template for chemical risk assessment applications.
433 Regul. Toxicol. Pharmacol. 115. doi:10.1016/j.yrtph.2020.104691
434 USEPA. [USEPA] U.S. Environmental Protection Agency. 1994. Great Lake water
435 quality initiative technical support document for the procedure to determine bioaccumulation
436 factors 822-R-94-002. Office of water. Off. of Sci. and Tech., Washington, DC, USA.

Revised Manuscript [2021-04-19]

(a)

Choose a txt or csv file

Browse... No file selected

Separator

Tab

Comma

Semicolon

Time unit:

Days

Duration of accumulation phase:

49

Data Visualization Exposure: 4.1e-04 Table Plot

Parent

(b)

Model and parameters

All parameters are expressed in $days^{-1}$

(c)

Accumulation: (Select one or several exposure media)

Water (k_{uw}) Sediment (k_{us}) Food (k_{uf}) Pore water (k_{upw})

Depuration: (Select involved processes)

Elimination (k_{ee}) Dilution by growth (k_{eg}) Biotransformation (k_m)

Click [here](#) for more information about parameters meaning.

With:

$C_p(t)$ internal concentration of the parent compound at time t (in $\mu g.g^{-1}$)

c_i Exposure concentration of route i : w , s , f and pw , respectively for water, sediment, food and pore water exposure (in $\mu g.mL^{-1}$ or $\mu g.g^{-1}$). Consider that constant over time.

$$\frac{dC_p(t)}{dt} = k_{in} \times c_i - (k_{out}) \times C_p(t) \quad \text{for } 0 \leq t \leq t_e$$

$$\frac{dC_p(t)}{dt} = - (k_{out}) \times C_p(t) \quad \text{for } t > t_e$$

Click [here](#) for more information about the model and [here](#) for solving equations.

Click [here](#) for more information about the type of fitting algorithm used, the number of iterations required...

Calculations will be performed for the following exposure: $4.1.10^{-04} \mu g.mL^{-1}$

Calculations can take several minutes, please be patient.

Figure 1. Data uploading and user information as required from the MOSAICbioacc homepage.

Revised Manuscript [2021-04-19]

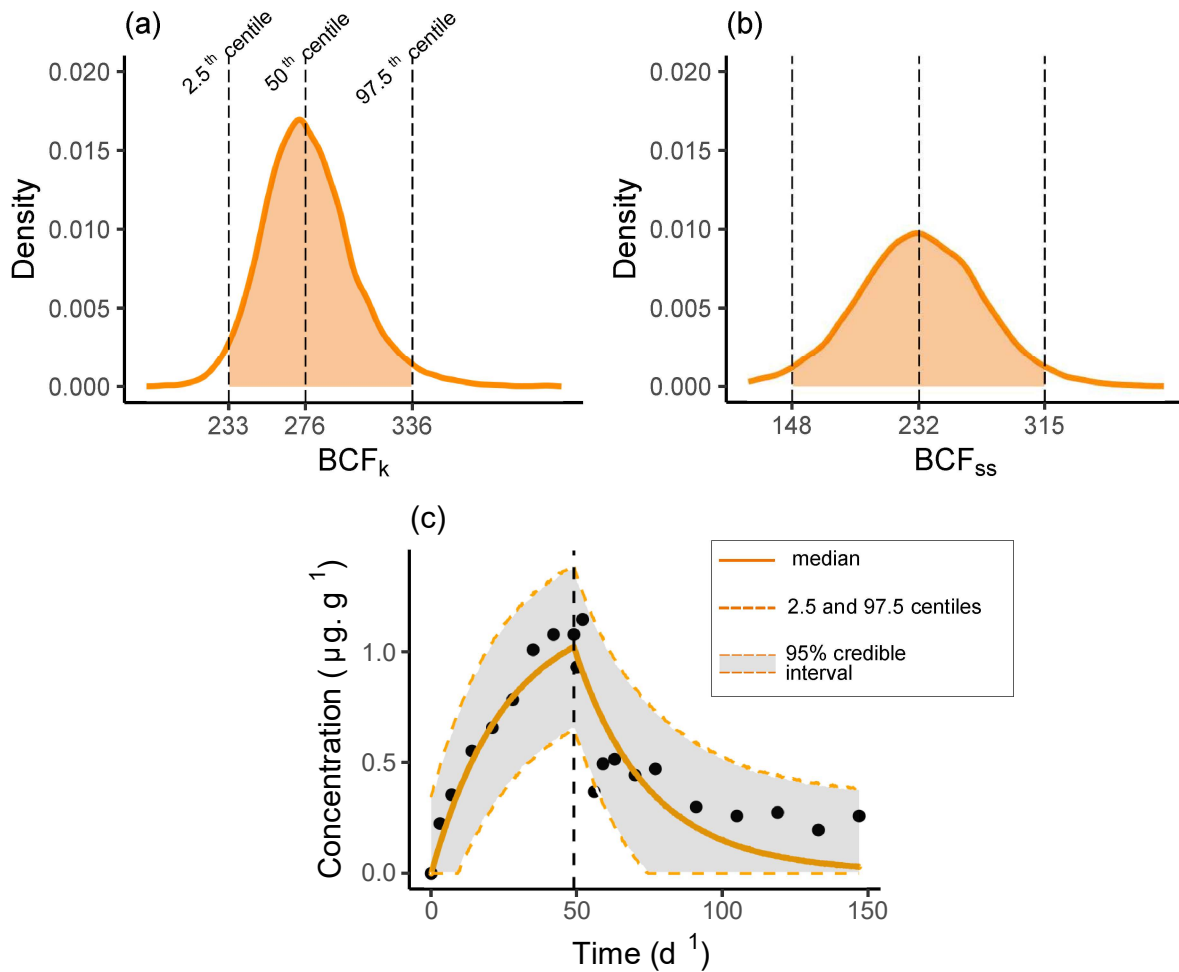


Figure 2. Probability distributions of BCF_k (a) and BCF_{ss} (b). The middle-dotted line represents the median value, while left and right dotted lines delimit the 2.5th and 97.5th centiles. (c): Observed (black dots) and predicted contaminant concentrations in the organisms ($\mu\text{g}\cdot\text{g}^{-1}$), where the median curve is displayed as the solid orange line and the uncertainty band as the grey zone, delimited by the 2.5th and 97.5th centiles in dotted orange lines.

Revised Manuscript [2021-04-19]

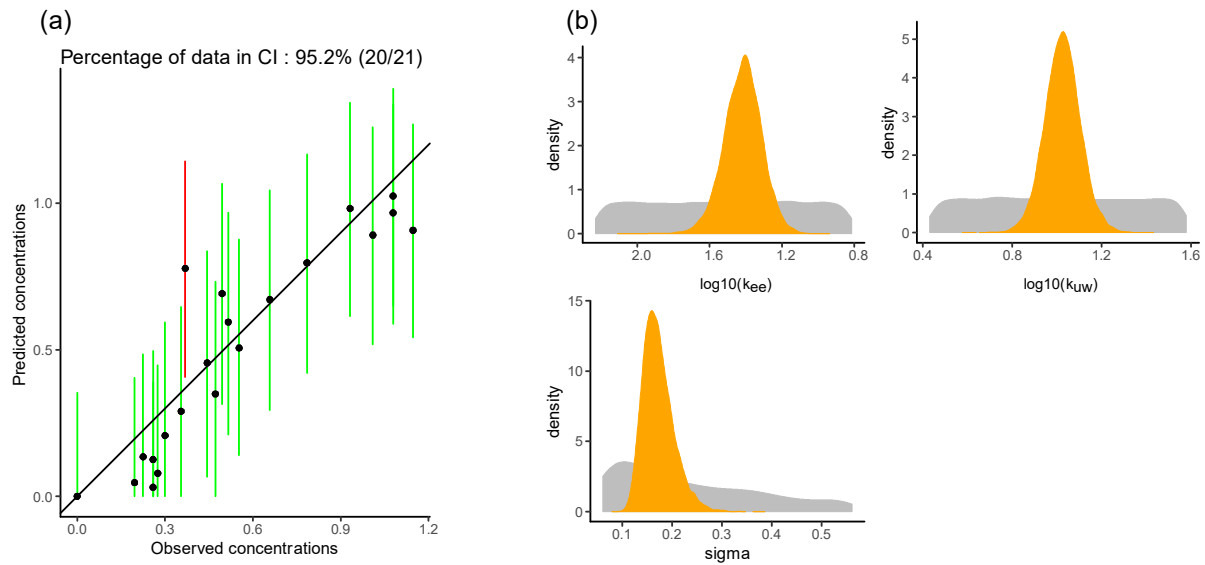


Figure 3. Some goodness-of-fit criteria as provided by MOSAIC_{bioacc}: (a) Posterior Predictive Check (PPC) where observed values are read on the x-axis, while the y-axis reports median predictions (black dots) and their 95% credible intervals (vertical segments, coloured in green if they contain the observed value, in red otherwise); (b) prior (grey) and posterior (orange) marginal distributions of parameters of the chosen TK model (here the simplest one).

Revised Manuscript [2021-04-19]

Table 1. Example of a data set ready to be used in MOSAIC_{bioacc}. The data set must contain four columns whatever their order: ‘time’ (time points of the measurements at the exposure concentration, in time unit: here in days), ‘conc’ (contaminant concentrations measured within organisms, that must be expressed in $\mu\text{g}\cdot\text{g}^{-1}$), ‘expw’ (the contaminant concentration in the exposure medium, here water, that must be expressed in $\mu\text{g}\cdot\text{mL}^{-1}$), and ‘replicate’ (a number or a character that is unique for each replicate, dimensionless).

time	conc	expw	replicate
0	0.000	0.0044	2
3	0.225	0.0044	2
7	0.355	0.0044	2
14	0.553	0.0044	2
21	0.658	0.0044	2
28	0.785	0.0044	2

Revised Manuscript [2021-04-19]

Table 2. Meaning of parameters and variables of the TK model used in MOSAIC_{bioacc}.

Symbol	Meaning
i	index of exposure routes, $i = 1 \dots I$
j	index of elimination processes, $j = 1 \dots J$
ℓ	index of metabolites, $\ell = 1 \dots L$
I	total number of exposure routes
J	total number of elimination processes
L	total number of metabolites
t	time (in time unit)
c_i	exposure concentration of route i (in $\mu\text{g} \cdot \text{mL}^{-1}$)
$C_p(t)$	internal concentration of the parent compound at time t (in $\mu\text{g} \cdot \text{g}^{-1}$)
$C_{m_\ell}(t)$	internal concentration of metabolite ℓ (in $\mu\text{g} \cdot \text{g}^{-1}$)
k_{u_i}	uptake rate of exposure route i (expressed in time^{-1})
k_{e_j}	elimination rate of elimination process j (expressed in time^{-1})
k_{em_ℓ}	elimination rate of metabolite ℓ (expressed in time^{-1})
k_{m_ℓ}	biotransformation rate of metabolite ℓ (expressed in time^{-1})
t_c	duration of the accumulation phase (in time unit)
\mathcal{N}	Gaussian probability distribution
$C_{obs,p}(t)$	internal measured concentration of the parent compound at time t (in $\mu\text{g} \cdot \text{g}^{-1}$)
$C_{obs,m_\ell}(t)$	internal measured concentration of metabolite ℓ (in $\mu\text{g} \cdot \text{g}^{-1}$)
σ_{C_p}	standard deviation for internal concentration of the parent compound (in $\mu\text{g} \cdot \text{g}^{-1}$)
σ_{met_ℓ}	standard deviation for the internal concentration of metabolite ℓ (in $\mu\text{g} \cdot \text{g}^{-1}$)
$U = \sum_{i=1}^I k_{u_i} c_i$	sum of all uptake terms
$E = \sum_{j=1}^J k_{e_j}$	sum of all elimination terms for the parent compound
$M = \sum_{\ell=1}^L k_{m_\ell}$	sum of all elimination terms for metabolite ℓ

Revised Manuscript [2021-04-19]

Table 3. Example of BCF (a) and parameter (b) estimates expressed as medians (50th centile) and 95% credibility intervals (2.5th - 97.5th centiles). Hyphens stand for dimensionless parameters.

	2.5 th	50 th	97.5 th	Units
(a)				
BCF_k	233	276	338	-
BCF_{ss}	147	233	317	-
(b)				
k_{uw}	7.388	10.59	15.45	d ⁻¹
k_{ee}	0.02307	0.03851	0.06091	d ⁻¹
σ_{Cp}	0.1249	0.1684	0.2445	μg.g ⁻¹