Tutorial MOSAIC$_{\text{growth}}$

A. Ratier, D. Wu, G. Multari, C. Lopes, S. Charles

September 25, 2020

Contents

1 Introduction

1.1 Objectives

The purpose of this document is to introduce you how to use the MOSAIC$_{\text{growth}}$ application (https://mosaic.univ-lyon1.fr/growth). This application is based on the R software[1] and especially the `rjags` library (version 4.10)[2], to provide a dose-response (DR) analysis of growth toxicity under a Bayesian framework. MOSAIC$_{\text{growth}}$ is developed as an R-Shiny interface (version 1.5.0)[3].

1.2 Context

The MOSAIC$_{\text{growth}}$ application is a turn-key web tool providing a dose-response (DR) analysis of growth toxicity test data under a Bayesian framework, including an estimation of the x% effective toxicity value, that can be an x% effective rate (ER$_x$), an x% effective concentration (EC$_x$) or any other expression of your choice. For clarity reasons, we will use the abbreviation ER$_x$ in the application. Growth measurement might be any quantitative continuous variable describing the growth of organisms (*e.g.*, shoot length and dry weight for plants). This tool makes it possible to analyse one single or multiple data set(s) and to get various ouputs, such as a summary table of ER$_x$ estimates. This summary table of ER$_x$ estimates includes not only medians and 95% credible intervals, but also censored ER$_x$ values accounting for their uncertainty compared to the range of tested concentrations. These censored ER$_x$ values can be used for future SSD analyses in the MOSAIC$_{\text{SSD}}$ application https://mosaic.univ-lyon1.fr/ssd. More details about the underlying modelling and the process of censoring can be found in the **vignette**.

The MOSAIC$_{\text{growth}}$ application also provides a prediction tool for growth data. This tool first allows the users to simulate a DR curve for given point values of parameters and chosen concentrations according to a planned experiment, but also to propagate the parameter uncertainty from a previous DR analysis in order to predict a DR curve for growth for a new range of concentrations. Such a tool can be helpful in designing future experiments for a given species/compound combination.

1.3 Installation

If you are using the web interface (https://mosaic.univ-lyon1.fr/growth), you don't need to install anything.

However, if you want to run the R script (downloadable from the application) by yourself, you need to install:

- the R software[1]. Refer to https://cran.r-project.org/ to proceed.
- the JAGS software[2]. Refer to http://sourceforge.net/projects/mcmc-jags/ to proceed.
- the `rjags` package[2]. You can install it directly from the R software > Tools > Install Packages > rjags or from the CRAN website http://cran.r-project.org/web/packages/rjags/index.html.
- Others R packages necessary to run the application: `tidyverse`, `gridExtra`, `ggmcmc`, `GGally`.

Here is an example of the R code to install the required packages:

```
if(is.element('rjags', installed.packages()[,1]) == FALSE)
  {install.packages('rjags')}

if(is.element('tidyverse', installed.packages()[,1]) == FALSE)
  {install.packages('tidyverse')}

if(is.element('gridExtra', installed.packages()[,1]) == FALSE)
  {install.packages('gridExtra')}

if(is.element('ggmcmc', installed.packages()[,1]) == FALSE)
  {install.packages('ggmcmc')}

if(is.element('GGally', installed.packages()[,1]) == FALSE)
  {install.packages('GGally')}
```

2 Step 1: Data uploading

When using MOSAIC$_{\text{growth}}$, the first step is to upload input data (**Fig. 1**):

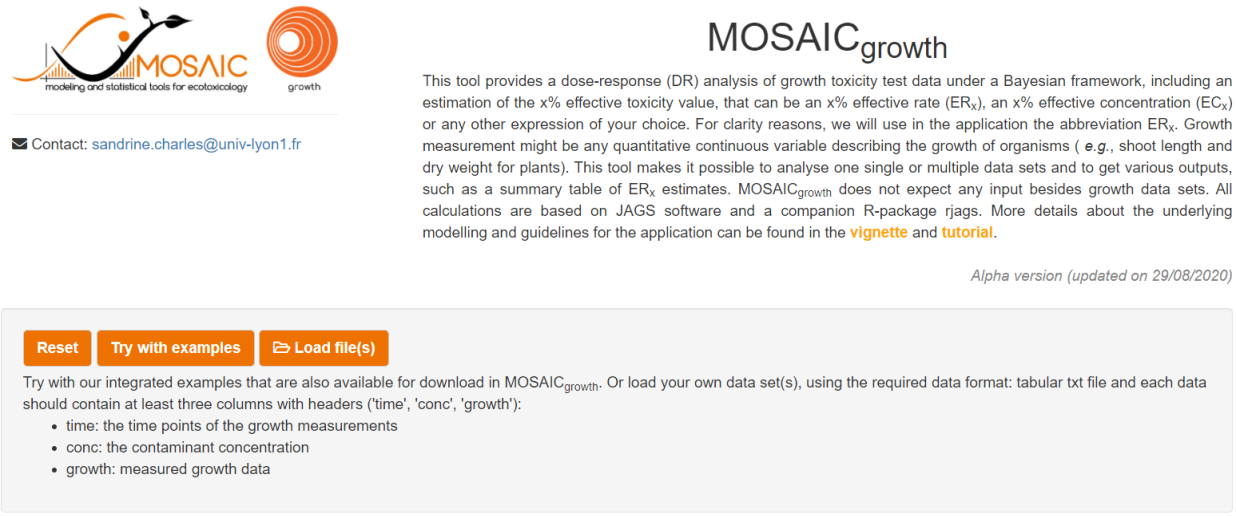

Figure 1. *Data uploading and user information to enter.*

2.1 Format

You can upload your own data (click on 'Load file(s)') by taking care about the format specification of your file. MOSAIC$_{\text{growth}}$ expects to receive data as a tabular .txt file. Each line of the table corresponds to an exposure concentration of the contaminant for a given time point and a growth measurement. The table must contain the three following columns, with exact header names (the order of column does not matter):

- 'time': the time point of the measurement;
- 'growth': the dose of the contaminant;
- 'conc': the measured growth data of the organism.

If required, you can add a column which contains the replicates, as follows:

- 'replicate': a number or a string that is unique for each replicate.

Here is an example:

Table 1. *Example of first lines of a data set ready to be uploaded.*

replicate	time	conc	growth
1	21	0	2.88
2	21	0	3.96
3	21	0	3.25
4	21	0	3.29
5	21	0	3.90
6	21	0	3.15

In MOSAIC$_{\text{growth}}$, you can upload one data set or several ones to simultaneously perform DR analyse for a set of species. This may be the preferred option in the perspective of an SSD analysis.

2.2 Example data

Some example files are provided in order to test the application and better appropriate the functioning[4-7]:
- **chlordan-daphnia.txt**[4]: exposure of *Daphnia magna* to chlordane (6 concentrations including the control, expressed in $\mu g/L$) at day 21. Length data is collected (expressed in mm).
- **cadmium-daphnia.txt**[5]: exposure of *Daphnia magna* to cadmium (5 concentrations including the control, expressed in $\mu g/L$) during 21 days. 10 time-points and 4 replicates of 10 animals. Length data is collected (expressed in mm).
- **copper-daphnia.txt**[5]: exposure of *Daphnia magna* to copper (5 concentrations including the control, expressed in $\mu g/L$) during 21 days. 16 time-points and 3 replicates of 20 animals. Length data is collected (expressed in mm).
- **zinc-daphnia.txt**[5]: exposure of *Daphnia magna* to zinc (4 concentrations including the control, expressed in $\mu g/L$) during 21 days. 15 time-points and 3 replicates of 20 animals. Length data is collected (expressed in mm).
- **subst01-lymnaea.txt**[6]: exposure of snails to a given substance (6 concentrations including the control, expressed in $\mu g/L$) at day 56. Length of shell is collected (expressed in mm).
- **plant01.txt to plant10.txt**[7]: plant species 1 to 10 exposed to a given product (the same one for other plants of the available examples in the application) during 21 days for the vegetative vigour test.

You can try MOSAIC$_{\text{growth}}$ with these example data sets (**Fig. 2**):

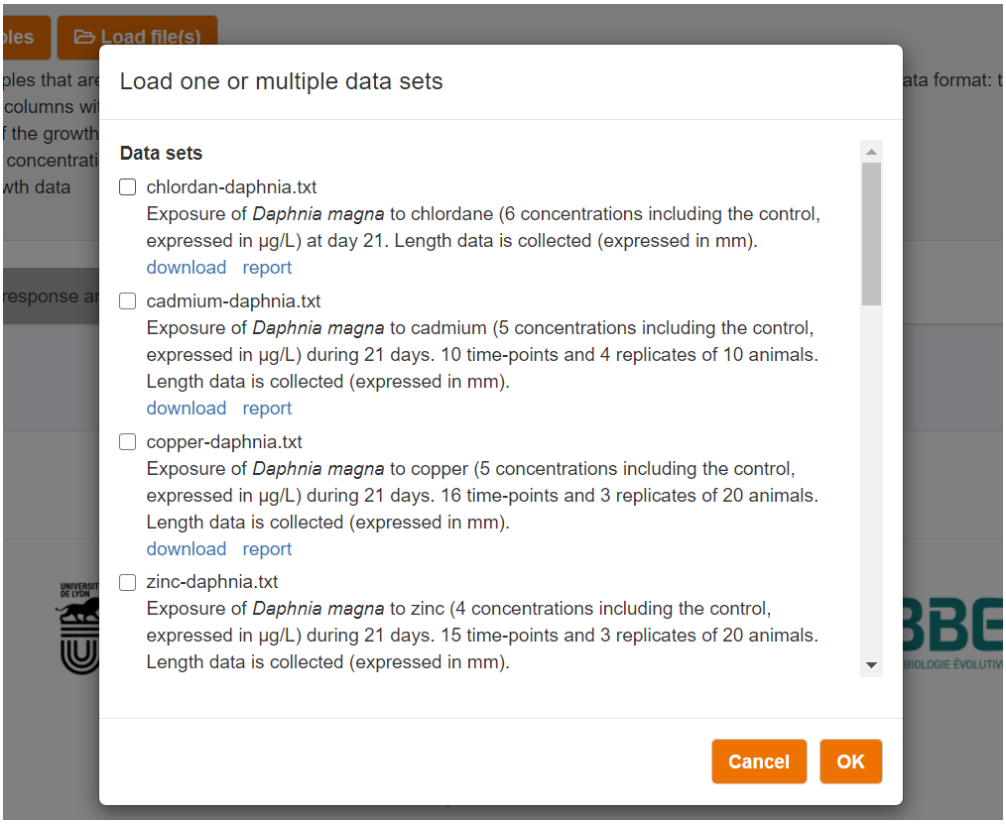

Figure 2. *Examples available in MOSAIC$_{\text{growth}}$.*

For each example data set, you can download data and the report.

2.3 Data visualization

In case you upload several data sets, select the one for which you want to visualize the data. Its name is reminded at the top of the data visualization section (**Fig. 3**).

When the upload is complete, you have the possibility to change the x- and y-axis labels. Then you have to manually select the exposure, growth and time units. These inputs will be used for plotting results.
There are two types of visualizations on the right in this section: plot (default) and table (**Fig. 3** and **4**), which allow you to check if the file is correctly uploaded. For the table visualization, you can select to show 10, 25, 50 or 100 entries per page.

If the inputs are correct, you can move on 'Dose-response analysis' section.

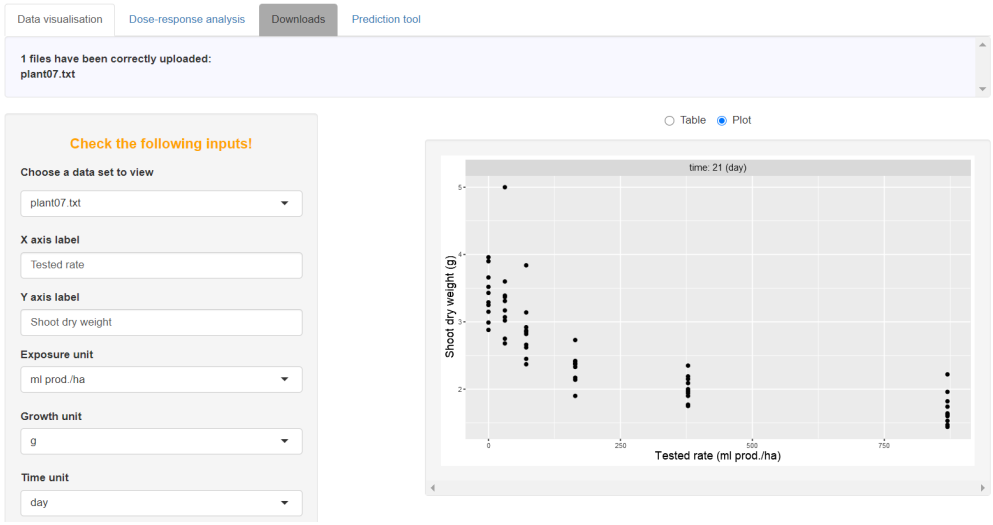

Figure 3. *Plot of the uploaded data for the selected file.*

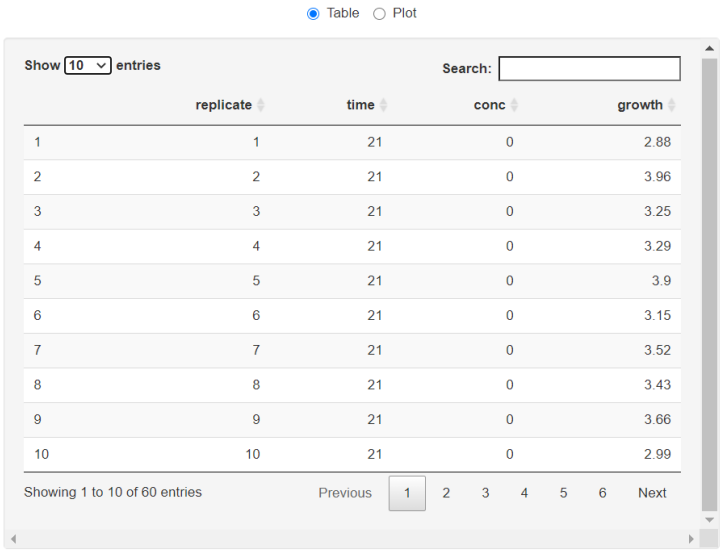

Figure 4. *Table of the uploaded data for the selected file.*

3 Step 3: Results

You will find at the end of this user-guide an appendix (**section 8**) which gathers other types of results which can be obtained with other data sets and how to interpret them.
To illustrate the result section, we will use the example file 'plant07.txt', plant species 7 exposed to an herbicide during 21 days of a vegetative vigour test. Shoot dry weight data (expressed in mm) was collected.

At first, select the data set(s) on which you want to perform DR analysis and ER_x calculation (**Fig. 5**). Then, select the target time. If multiple data sets are selected, be aware that a common target time is required to perform a similar DR analysis for each of them.

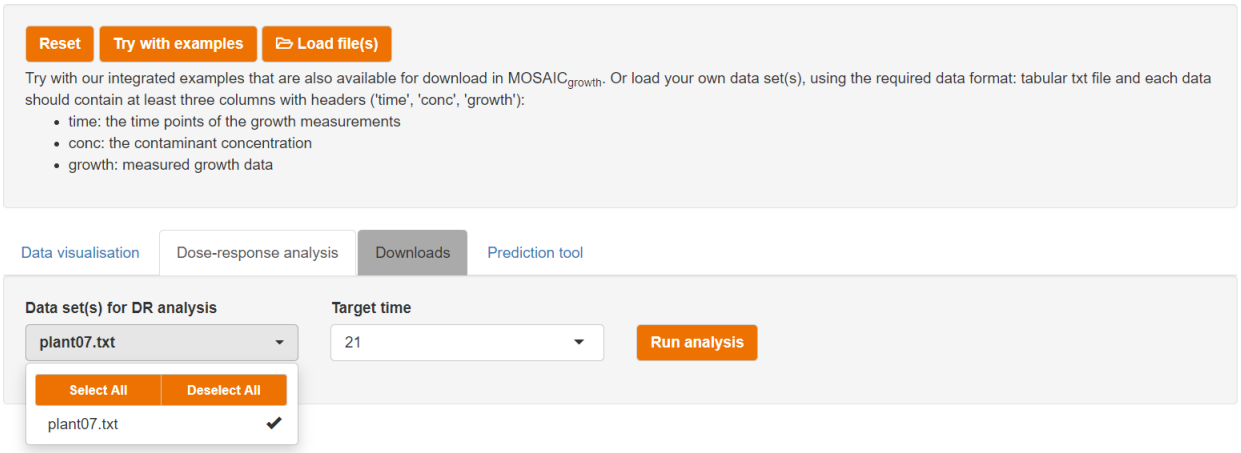

Figure 5. *Selection of the data set(s) before ER_x and DR analysis.*

When the desired data sets are selected, click on 'Run analysis'. Calculations can take a while to perform.

3.1 ER_x estimate

As a first result, we provide the ER_x value for the selected x by the user. After the selection of the x value(s) (choice within 5, 10, 25, 50, 75, 90), you can select one or more data sets to calculate the corresponding ER_x. Then click on 'calculate ER_x' and 'Display all ER_x' (**Fig. 6**). In the example on **Fig. 6**, the rate that gives 50% effect (growth inhibition, ER_{50}) is 683.21 [509.36; 950.76] ml prod./ha.

ER_x estimates are summarized in a table of ER_x estimates for all performed dose-response analyses (**Fig. 6**). This output provides median values and the 95% credible interval for the selected x value(s) and file(s). 'Censored Value' in the table stands for censored ER_x according to the criterion based on the ratio of probabilities and a decision threshold T equal to 0.5. For more details about censored ER_x, please consult the **vignette**).

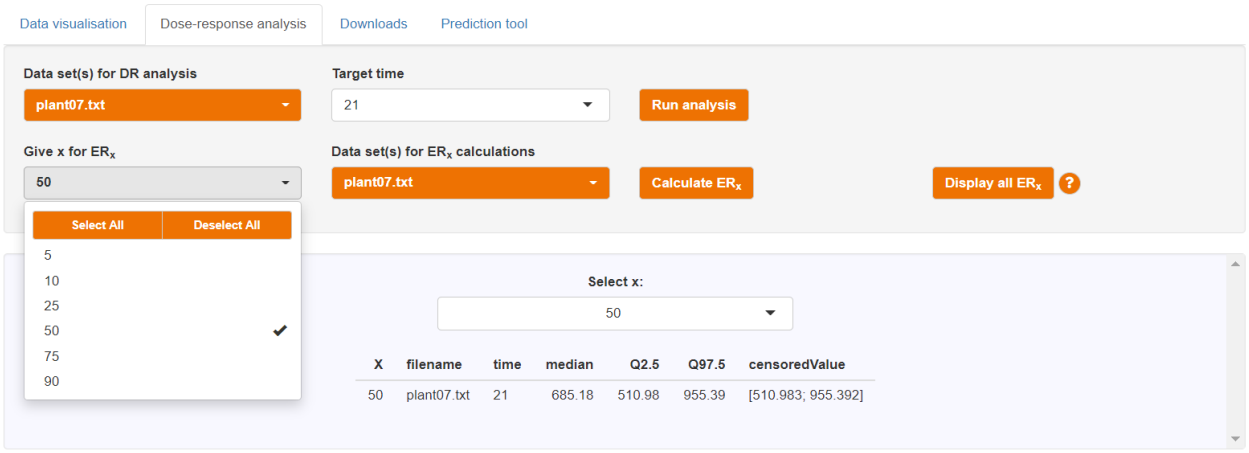

Figure 6. *Selection of inputs to calculate and display the* ER_x.

3.2 Dose-response analysis

The results of the DR analysis for the example data set are displayed on the left of **Fig. 7**, reminding the target time. Three outputs are given: DR curve, parameter estimates and ER_x estimates.

3.2.1 DR curve

We first provide the fitted dose-response curve superimposed to the observations (**Fig. 7**, black dots): the orange plain line is the median curve, the gray zone the uncertainty band delimited by 2.5% and 97.5% quantiles in orange dotted lines. You have the possibility to plot the results according a logarithm scale for the x-axis.

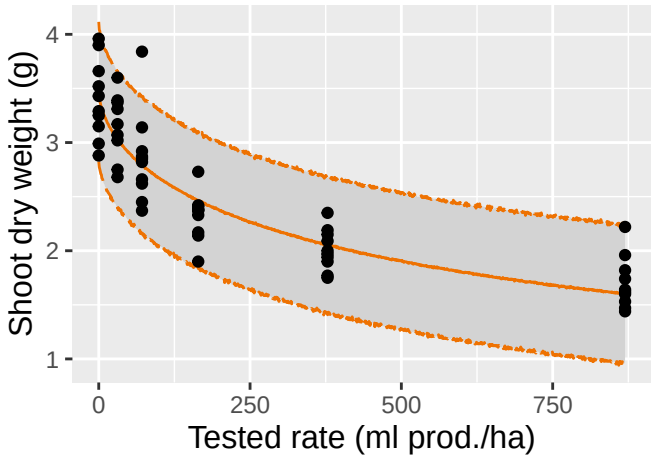

Figure 7. *Measured (black dots) and predicted growth data according to the contaminant rate. Median predictions are symbolized by the orange plain line and the uncertainty bands by the gray zone which is delimited by the 2.5% and 97.5% quantiles in orange dotted lines.*

3.2.2 Parameter estimates

We also provide parameter estimates. From the joint posterior distribution, we can obtain the marginal posterior distribution for each parameter, which can be summarized by the median and the 95% credible

intervals (**Table 2**).

Table 2. *Example of parameters medians (50% quantile) with their 95% credible intervals (2.5% - 97.5% quantiles from the joint posterior distribution).*

	median	Q2.5	Q97.5
b	0.6317317	0.5006001	0.7925764
d	3.4633903	3.2855732	3.6490226
e	682.3091823	509.4969732	957.3108039
sigma	0.3106841	0.2607767	0.3810995

Where:

- b is the shape parameter (the "slope" of the dose-response curve);
- d corresponds to growth in control data (*i.e.* in absence of contaminant);
- e corresponds to the ER_{50} and sigma is the standard deviation of growth data.

If you want more information on the meaning of these parameters, we invited you to read the **vignette**.

3.2.3 ER_x

If an x value is chosen and the ER_x calculated, the summary and the density of the ER_x probability distribution are displayed (**Fig. 8**).

'CensoredValue' in the provided table (**Fig. 8**) stands for censored ER_x according to the criterion based on the ratio of probabilities and a decision threshold T equal to 0.5. For more details about the censored ER_x, please consult the **vignette**.
The black curve represents the posterior probability distribution of the ER_x. The two solid vertical lines delimit the 95% credible interval of the ER_x estimate and the dashed vertical line corresponds to the value of the highest tested rate (max_rate). The orange surface represents the probability for the ER_x to lie between Q2.5 and max_rate. The ratio of probabilities mentioned above is the ratio of the probability that the ER_x lies within Q2.5 and max_rate over the probability that the ER_x lies within Q2.5 and Q97.5. It equals to the orange surface divided by 0.95.

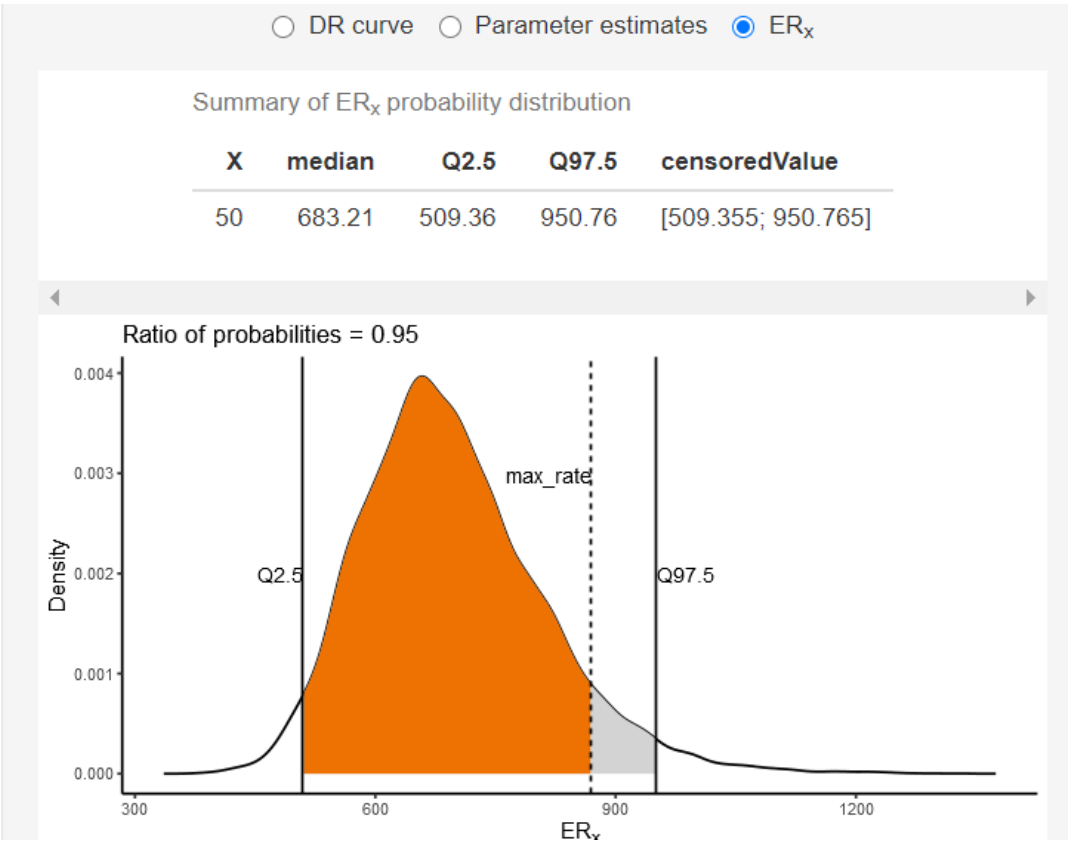

Figure 8. *Selection of inputs to calculate and display the* ER_x.

3.3 Goodness-of-fit criteria

Goodness-of-fit criteria are given below in our prioritised order; the PPC and the prior-posterior comparison are the most important to check; if they do not correspond to the expectation, you must consider your results with an even more particular attention. As an indication, if at least two criteria are fulfilled, the results obtained can be considered as good enough.
We suggest that you refer to the appendix at the end of the document for more details (**section 8**), in particular to know to deal with results far from the expected ones.

3.3.1 Posterior Predictive Check (PPC)

The PPC shows the observed values against their corresponding estimated predictions (black dots), along with their 95% credible interval (vertical segments). If the fit is correct, we expect to see 95% of the data within the intervals. Ideally, observations and predictions should coincide, so we would expect to see black dots along the first bisector $y = x$ (plain black line). The 95% credible intervals are coloured in green if they overlap this line, in red otherwise. In the following example (**Fig. 9**), 98.3% of the measured data ($n = 58/59$) are in the 95% credible intervals of their predictions.

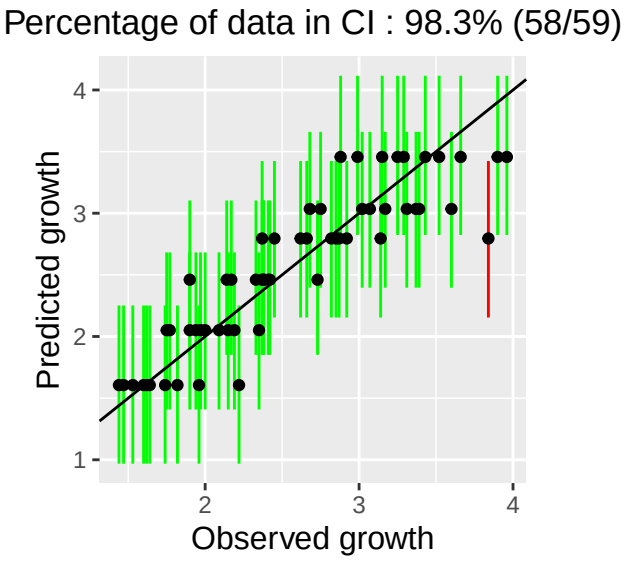

Figure 9. *Example of a PPC: predicted against measured concentrations (black dots) and predicted 95% credible intervals (vertical green and red segments).*

3.3.2 Prior and posterior distributions

The prior and posterior distributions are illustrated in **Fig. 10**. The prior distribution is represented by the gray area and the posterior distribution by the orange area. The accuracy of the model parameter estimation can be visualized by comparing prior and posterior distributions: the overall expectation is to get a narrower posterior distribution compared to the prior one, what reflects that data contributed enough to precisely estimate parameters.

In the given example (**Fig. 10**), marginal posterior distributions for d, b, e and σ are narrower (orange area) than their respective prior distributions (grey area).

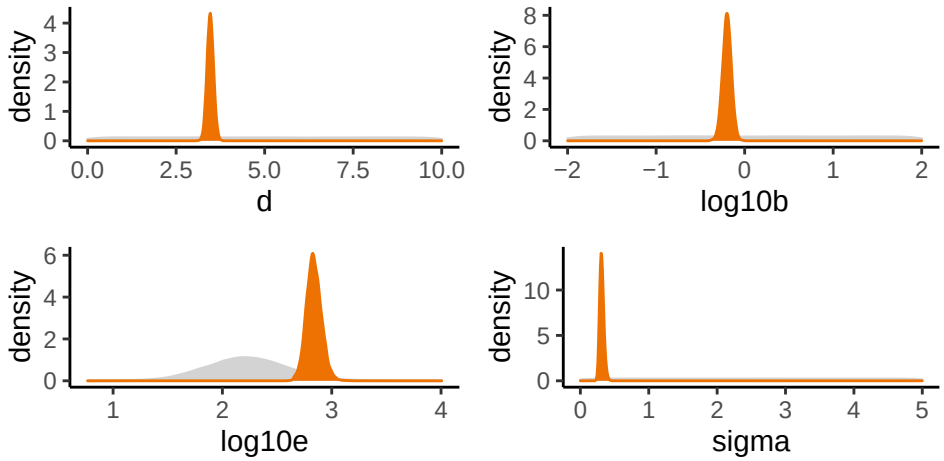

Figure 10. *Example of prior (gray) and posterior (orange) probability distributions for each parameter.*

3.3.3 Correlations between parameters

It is also recommended to check for correlations between parameters (**Fig. 11**).

Correlations between parameters are visualized by projecting the joint posterior distribution in a plot matrix with planes of parameter pairs (**Fig. 11**, lower triangular elements), marginal posterior distribution of each model parameter (**Fig. 11**, diagonal), and Pearson correlation coefficients (**Fig. 11**, upper triangular elements). Correlations are expected to be low (reflected by "potatoid" shapes of density lines in orange, *e.g.*, $log_{10}b$ and $log_{10}e$ in **Fig. 11**); a leaning elliptical shape translates high correlations (positive if leaning to the right, negative if leaning to the left).

If two parameters are highly correlated, this means that the estimate obtained for one of these two parameters will strongly influence the estimate of the other parameter. This high correlation is often due to the model structure itself, but can also come from data if they are not in accordance with the requirements of the model fitting process.

3.3.4 Potential Scale Reduction Factors (PSRF)

Convergence of the Monte Carlo Markov Chain (MCMC) can be checked with the Gelman-Rubin diagnostic expressed with the potential scale reduction factor (PSRF). Approximate convergence is diagnosed when the PSRF is close to 1.00 (**Fig. 12**)[8].

In the example on **Fig. 12**, the PSRF is equal to 1 for each model parameter, thus the convergence of the MCMC was correctly achieved for the used number of iterations when fitting the model.

3.3.5 Deviance Information Criterion (DIC)

This criterion, denoted DIC, is a penalized deviance statistics accounting for the number of parameters for use in model comparison fitted on a same data set. Models with lower DIC values will be preferred[9]. DIC value can be negative. However, DIC value itself is not important, what matters is the difference between two DICs what will determine which model is the most appropriate to choose.

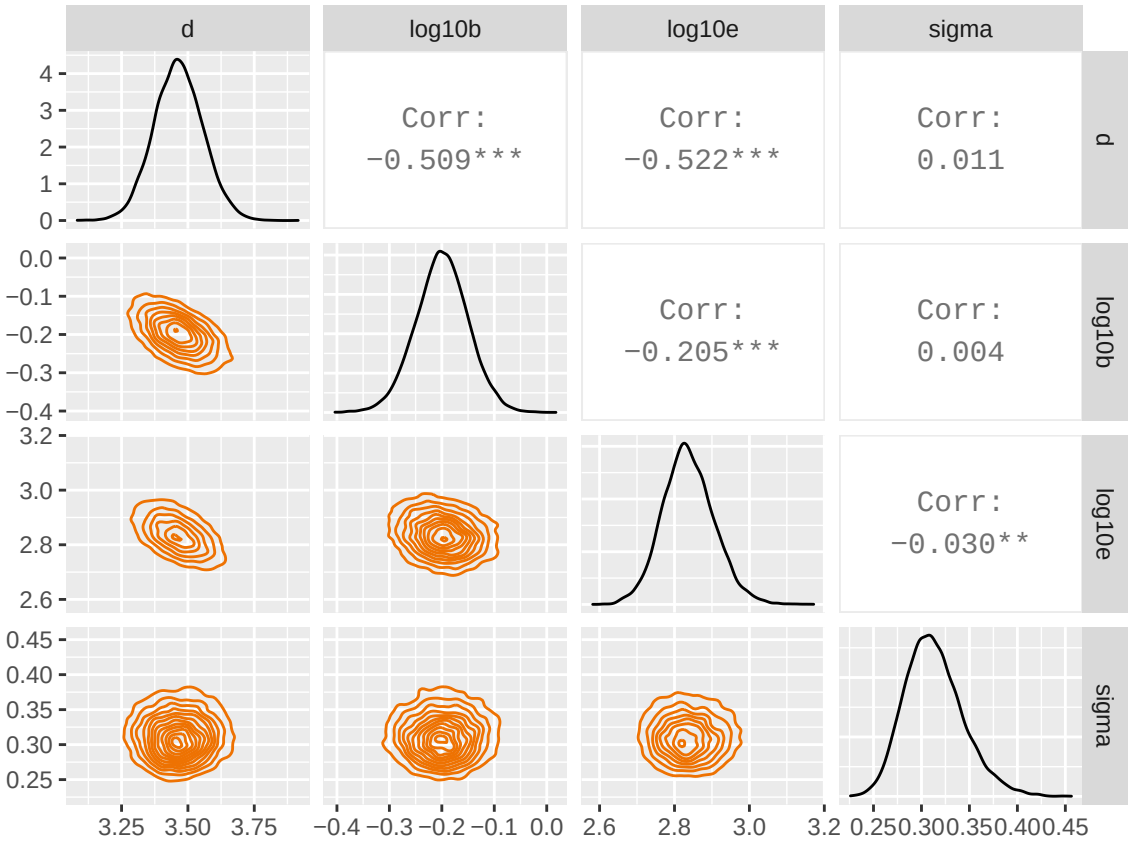

Figure 11. *Example of parameter correlations.*

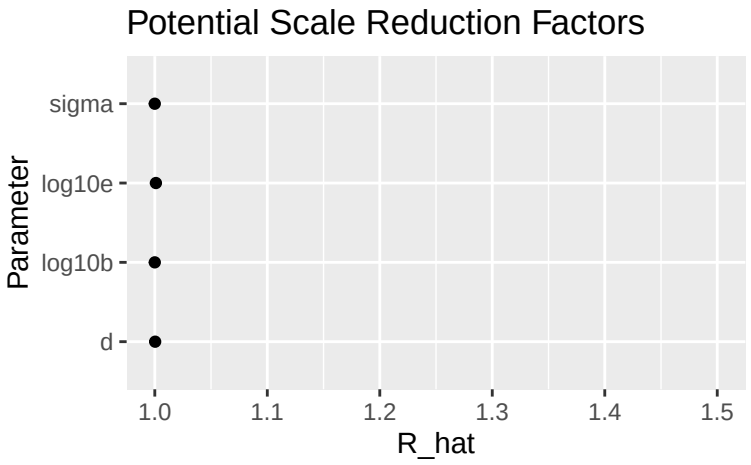

Figure 12. *Examples of PSRF.*

3.3.6 Traces of MCMC iterations

A traceplot is also an essential plot for assessing convergence and diagnosing of MCMC. It shows the time series of the sampling process leading to the posterior distribution. Different colours are used for each of the chains (here three) to assess within-chain variability. The user must check whether all MCMC converge towards the same distribution limit (overlapping of the chains). This can be verified visually by observing the simulated values for each node of interest as a function of the number of iterations (**Fig. 13**).

In the following example, the three MCMC overlap and converge towards the same distribution limit for each parameter. Thus, the algorithm has suitably converged.

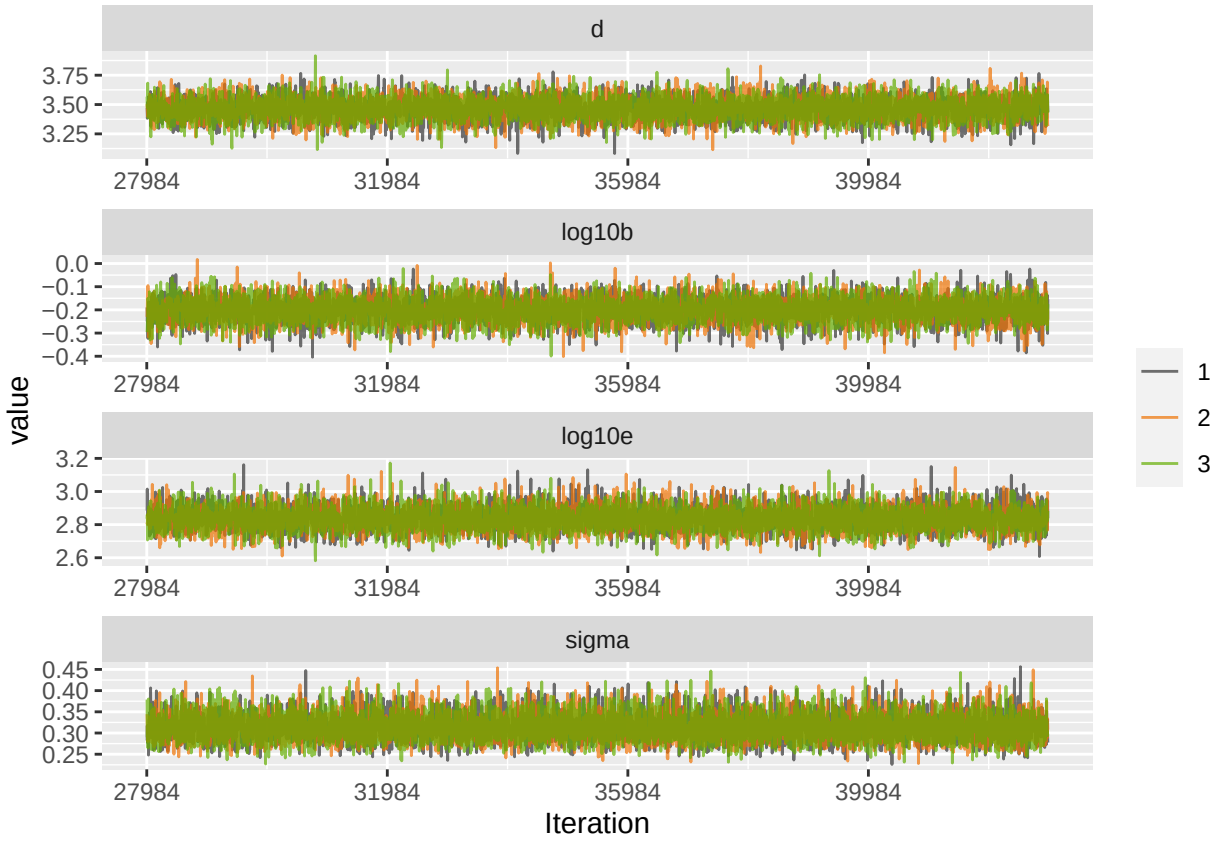

Figure 13. *Example of overlapping MCMC.*

4 Step 4: Downloads

In MOSAIC$_{\text{growth}}$, you can download the following documents from the 'downloads' section:

- **Full report** (.pdf, .html and .doc): a report providing results on all performed dose-response analyses.
- **Joint posterior distribution** (.txt and .csv): a file containing a joint posterior distribution of parameters (a sample of parameter estimates) for a single data set.
- **Output ER$_x$** (.txt and .csv): a file containing a summary table of ER$_x$ estimates for all data sets.
- **Output ER$_x$ for MOSAIC$_{\text{SSD}}$** (.txt): a file containing a set of censored ER$_x$ values expressed in two columns. The first column is for the lower bounds of the censored ER$_x$ values and the second column for higher bounds. You can upload this .txt file in the MOSAIC$_{\text{SSD}}$ application to fit a species sensitivity distribution (SSD) and get an HC$_p$ estimate.
- **Predicted data**(.txt and .csv): a file containing predicted growth data as a function of contaminant rate. The file allows to plot a dose-response curve with your favourite software, as you can see in our application.
- **R script**(.txt): an R script allowing to reproduce previous analyses.
- **Single report** (.pdf, .html and .doc): a report providing results on a single dose-response analysis.

5 Step 5: Prediction tool

The MOSAIC$_{\text{growth}}$ application also provides a prediction tool for growth data ('Prediction tool' section).

- This tool allows interactive simulations of a dose-response model based on a three-parameters log-logistic function to describe the dose-response relationship between a contaminant and a growth measurement (see vignette for more details).

- This tool also allows the users to propagate the parameter uncertainty from a previous DR analysis into the prediction of a DR curve for growth with a new range of concentrations.

Such a tool can be helpful in designing future experiments for a given species/compound combination.

The first step is to enter concentration values separated by a semi-colon. You can try with this example (**Fig. 14**): 40; 80; 160; 320; 640. Then, you have to choose if parameters are distributed or not:

- If not, you have to enter a single value of your choice for each parameter (d, $log_{10}b$ and $log_{10}e$, **Fig. 14**).

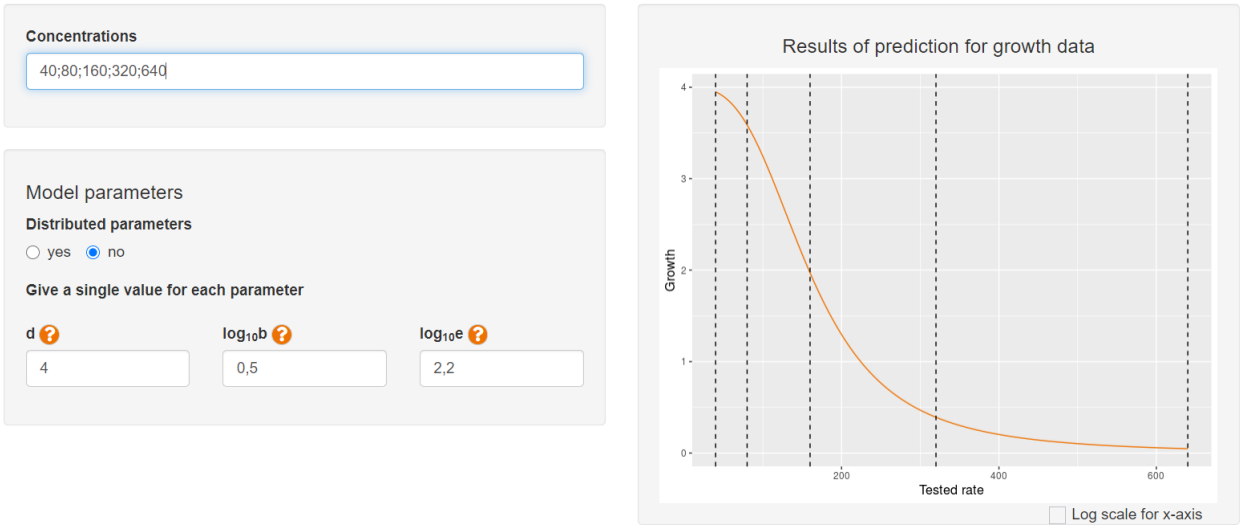

Figure 14. *Selection of inputs to perform predictions and corresponding results.*

- If parameters are distributed, you can choose a joint posterior distribution which comes directly from a previous dose-response analysis performed in MOSAIC$_{\text{growth}}$ (**Fig. 15**) or from a .txt file (**Fig. 16**). The file should be a tabular .txt file and contain four columns with headers ("d", "log10b", "log10e", "sigma"). Such a file may have been saved after running MOSAIC$_{\text{growth}}$ application.

Results of the prediction tool for growth data are visualized on the right. You have the possibility to ask for a log scale for x-axis at the bottom right of the plot.

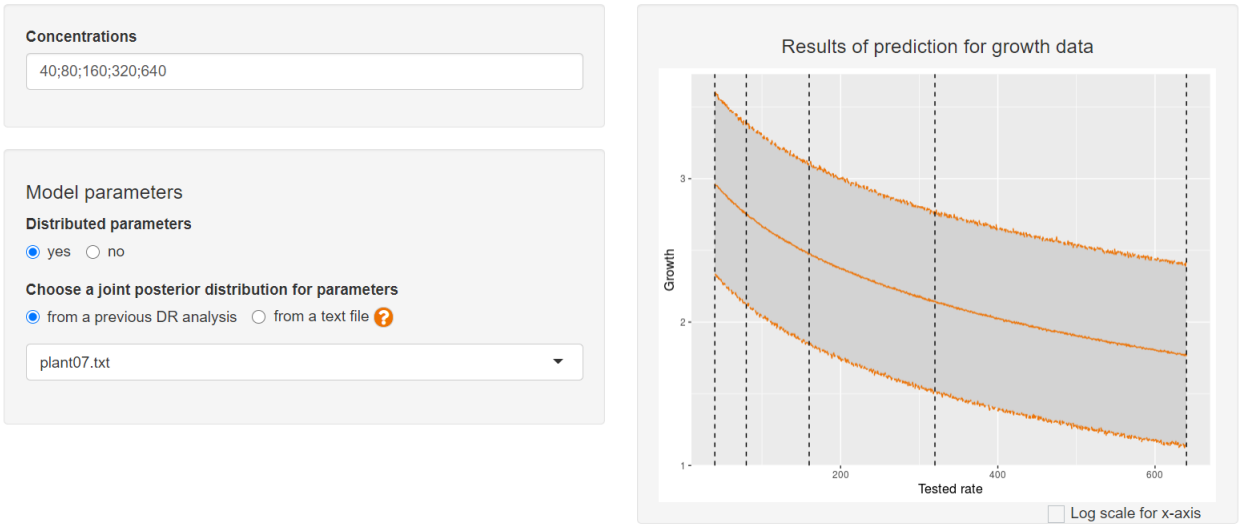

Figure 15. *Selection of inputs when parameters are distributed and obtained from a previous DR analysis.*

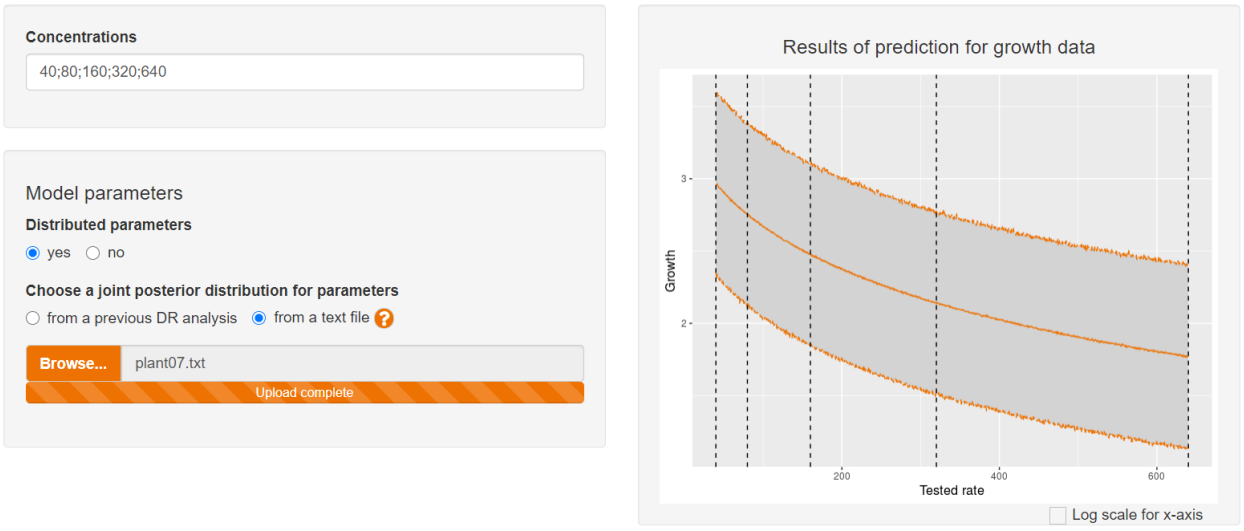

Figure 16. *Selection of inputs when parameters are distributed and uploaded from a .txt file.*

6 References

[1] Plummer, M. 2019. rjags: Bayesian Graphical Models using MCMC. R package version 4-10. https://CRAN.R-project.org/package=rjags.

[2] R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[3] Chang, W., Cheng, J., Allaire,J.J., Xie, Y. and McPherson, J. 2020. shiny: Web Application Framework for R. R package version 1.4.0.2. https://CRAN.R-project.org/package=shiny.

[4] Manar, R., Bessi, H., Vasseur, P. 2009. Reproductive effects and bioaccumulation of chlordane in *Daphnia magna*. Environnemental Toxicology and Chemistry **28**:2150–2159. https://doi.org/10.1897/08-564.1.

[5] Billoir, E., Delignette-Muller, M.L., Péry, A.R.R., Charles, S. 2008. A Bayesian Approach to Analyzing Ecotoxicological Data. Environnemental Science and Technology **42**:8978–84. https://doi.org/10.1021/es801418x.

[6] Ducrot, V., Askem, C., Azam, D., Brettschneider, D., Brown, R., Charles, S., Coke, M., Collinet, M., Delignette-Muller, M.L., Forfait-Dubuc, C., Holbech, H., Hutchinson, T., Jach, A., Kinnberg, K.L., Lacoste, C., Le Page, G., Matthiessen, P., Oehlmann, J., Rice, L., Roberts, E., Ruppert, K., Davis, JE., Veauvy, C., Weltje, L., Wortham, R., Lagadic, L. 2014. Development and validation of an OECD reproductive toxicity test guideline with the pond snail *Lymnaea stagnalis* (Mollusca, Gastropoda). Regulatory Toxicology and Pharmacology **70**:605–614. https://doi.org/10.1016/j.yrtph.2014.09.004.

[7] Charles, S., Wu, D., Ducrot, V. 2020. How to account for the uncertainty from standard toxicity tests in species sensitivity distributions: an example in non-target plants. preprint in bioRxiv. https://doi.org/10.1101/2020.07.02.183863.

[8] Gelman, A. and Rubin, D. 1992., Inference from iterative simulation using multiple sequences. Statistical Science **7**:457-511. https://www.jstor.org/stable/2246093.

[9] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Linde, A.V.D. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series **64**:583–639. https://doi.org/10.1111/1467-9868.00353.

7 Glossary

Dose-Response (DR): describes the magnitude of the response of an organism, as a function of exposure (or doses) to a stimulus or stressor (usually a chemical) after a certain exposure time.

Effective Concentration ($\mathbf{EC}_x$): the concentration that gives x % of effect (EC_x).

Effective Rate ($\mathbf{ER}_x$): the rate that gives x % of effect (ER_x).

Hazardous Concentration ($\mathbf{HC}_p$): hazardous concentration for p % of the species.

Monte Carlo Markov Chain (MCMC): A method which comprise a class of algorithms for sampling from a probability distribution. By constructing a Markov chain that has the desired distribution as its equilibrium distribution, one can obtain a sample of the desired distribution by recording states from the chain.

Potential Scale Reduction Factor (PSRF): Gelman-Rubin diagnostic to check the convergence of the MCMC.

Species Sensitivity Distribution (SSD) analysis: an approach to define safe levels for toxic compounds in an ecosystem. It is based on the assumption that species sensitivity to a given contaminant can be described by a probability distribution estimated from toxicity experiments.

8 Appendix: a guide in the interpretation of results

In practice, you may encounter situations where the results are not ideal, like with the example file 'plant01.txt'. This appendix will allow you to better interpret the results in such cases. For the goodness-of-fit criteria, we suggest you to have at least two good criteria to consider analysis given by $\text{MOSAIC}_{\text{growth}}$ as receivable.

8.1 ER_x

As illustrated for the ER_{50} on **Fig. 17**; the density of the ER_x probability distribution can be bimodal rather to be unimodal as in ideal situations. If at least two of the other criteria are validated, this can be disregarded. If not, your experimental data could be not sufficient to performed ER_x calculations.

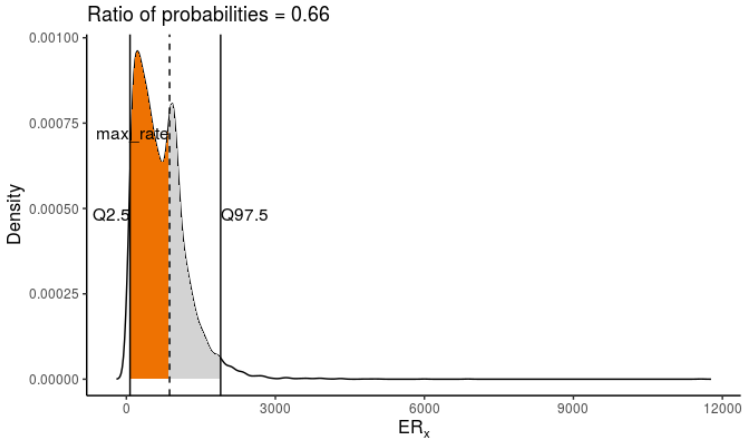

Figure 17. *Example of a bimodal probability distribution of the ER_x.*

8.2 Dose-response curve

As illustrated in **Fig. 18**, almost no effect was observed at 21 days, resulting in a large 95% credible band around growth predictions when plotted against the contaminant rate. It is an information to consider when you will interpret results because this large uncertainty will propagate to all the other model outputs, including the ER_x (**Fig. 17**).

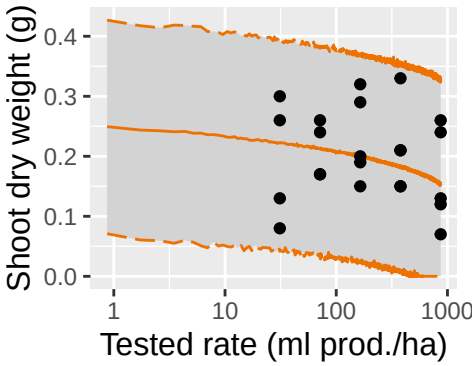

Figure 18. *Measured (black dots) and predicted growth data of the organism versus contaminant rate. Median predictions are symbolized by the orange plain line and the uncertainty bands by the gray zone which is delimited by the 2.5% and 97.5% quantiles in orange dotted lines.*

8.3 Parameter estimates

Large 95% credible intervals can sometimes be obtained for some parameters, especially for parameter b (**Table 3**). Such a situation leads to non precise estimate of the corresponding parameters, what can compromise their use for predictions. This can be explained by the few experimental data or because no effect is observed in your experiment. Thus, it is an information to consider when you will interpret the results.

Table 3. *Example of parameter medians (50% quantile) with large 95% credible intervals (2.5% - 97.5% quantiles).*

	median	Q2.5	Q97.5
b	0.63	0.01	63.21
d	0.27	0.19	0.38
e	878.93	74.25	2077.29
sigma	0.08	0.06	0.11

8.4 PPC

If the fit is correct, it is expected to get 95% of the data within the 95% credible intervals of their predictions. So, if the range of the percentage of data within the credible intervals is between 92 and 96%, calculations and predictions can be considered as good enough. If the percentage is under 92%, calculations and predictions are considered as underestimated. If the percentage is upper 96%, calculations and predictions are considered as overestimated.

On **Fig. 19**, you can see a counter-example with large uncertainties of the model predictions leading to 100% of the data within there credible intervals.

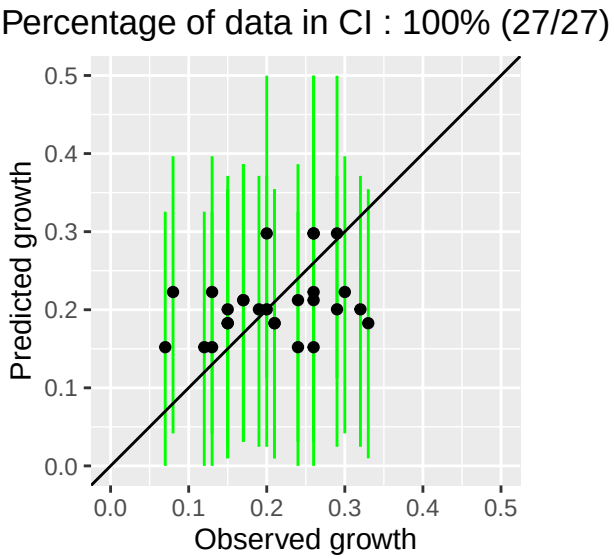

Figure 19. *Example of a PPC where there is an overestimation of the model predictions.*

8.5 Prior and posterior distributions

We remind you that prior distributions are defined by default to be the most generic as possible. However, it can happen that your data would require other prior distributions (*e.g.*, inspired by literature or by a previous study leading to parameters estimations outside of the default value as used in $\text{MOSAIC}_{\text{growth}}$).

The accuracy of the model parameter estimation can be visualized by comparing prior and posterior distributions: the overall expectation is to get a narrower posterior distribution compared to the prior one, what reflects that data contributed enough to precisely estimate parameters.
If one of the posterior distribution for a model parameter has bounds close to the lower or the upper bound of the priors distributions (*e.g.* $log_{10}b \simeq -2$ or $log_{10}b \simeq 2$), then the prior distribution may be not well defined. If a bimodal distribution is observed for one parameter, as illustrated for d on the top left of **Fig. 20**, then the inference process needs to be questioned. Conversely, the fit can be considered as correct if you obtain prior and posterior distributions as illustrated on the bottom of **Fig. 20**.

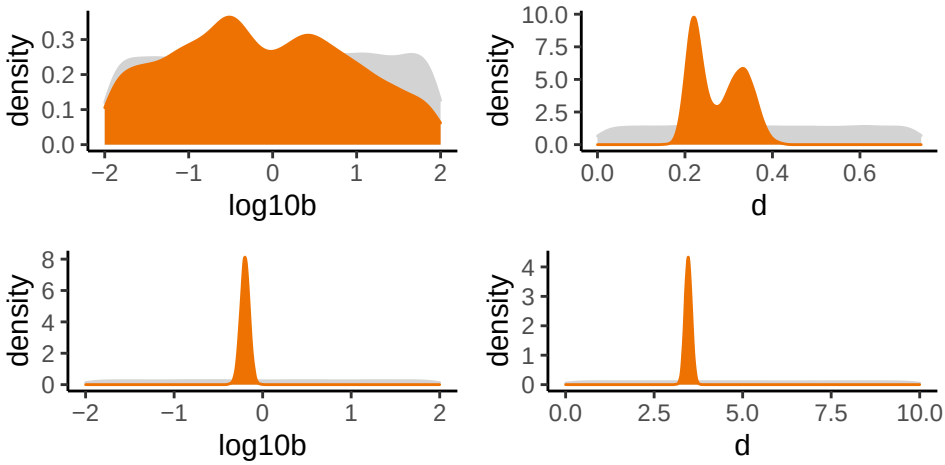

Figure 20. *Questionable posterior distribution on the top panel, and a posterior distribution as expected at the bottom panel.*

In $\text{MOSAIC}_{\text{growth}}$, it is not possible to change the prior distributions of parameters directly in the application. To do this, we suggest you to download the R code and to change the prior distributions directly in the R software. We remind you that to define the prior distributions you should not have a look at your data, but only on previous experiments, literature data, or expert knowledge.

8.6 Correlations between parameters

If a high correlation is obtained between two parameters (*e.g.*, more than 0.7 or less than -0.7 for the Pearson correlation coefficient), it is an information to consider, not necessarily a bad result. It means that the estimate obtained for one of these two parameters will strongly influence the estimate of the other one (for example, d and $log_{10}b$ are anti-correlated in **Fig. 21**). Such an high correlation may be due to the model structure itself so that it cannot be avoided.

Sometimes, you may get a bimodal posterior distribution for one or several parameters what translates into a double maximum on density plots (parameter d in **Fig. 21**). This may be due to not enough large priors. Indeed, to make the application as global as possible, we defined priors for each parameter the most global as possible. However, depending on the experimental conditions, the parameters may not be really included within the chosen range of values. In such a case, we recommend you to contact us sandrine.charles@univ-lyon1.fr if you are not experimented with Bayesian inference and R software.

8.7 PSRF

This criterion must be as close as possible to 1 for each model parameter to ensure that the between-chain variability is small compared to the within-chain variability. Based on our experience, from a value of 1.03,

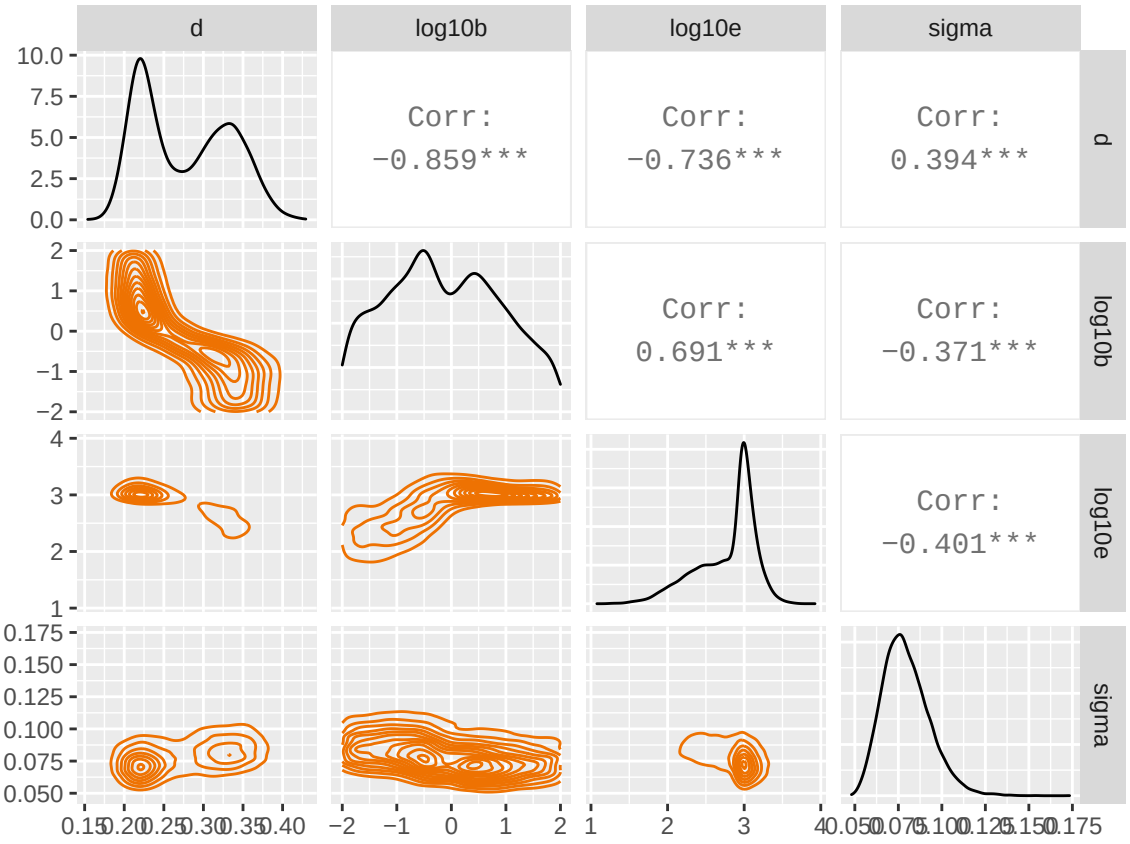

Figure 21. *Example of parameter correlations.*

the results should be questioned. Most often, such a case appears when priors are not well defined or when the data do not contain enough information. One of the solution may be to increase the number of iterations in the MCMC by using the R script directly.

8.8 DIC

The DIC is not criterion to consider to check the goodness-of-fit itself. However, it is crucial to consider when two models or more are compared after having been fitted on a same data set.

8.9 Traces of MCMC

You must check whether the MCMC converge towards the same distribution limit (overlapping of the chains).