Practice with with `morse` and `MOSAIC`

Elements of correction

Sandrine CHARLES & Christelle LOPES

Contact:sandrine.charles@univ-lyon1.fr

Master EPET

```
require(morse)
require(rjags)
require(lattice)
require(GGally)
```

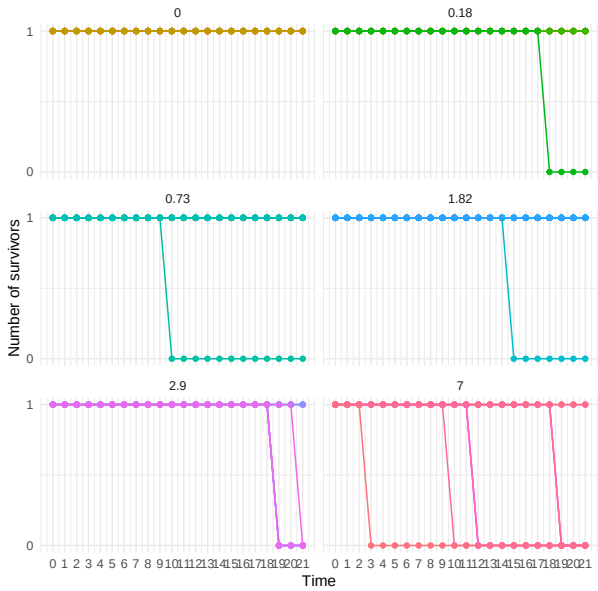
1 *D. magna* exposed to chlordan

Step 1.1

```
data(chlordan)
dataset <- survData(chlordan)
summary(dataset)
```

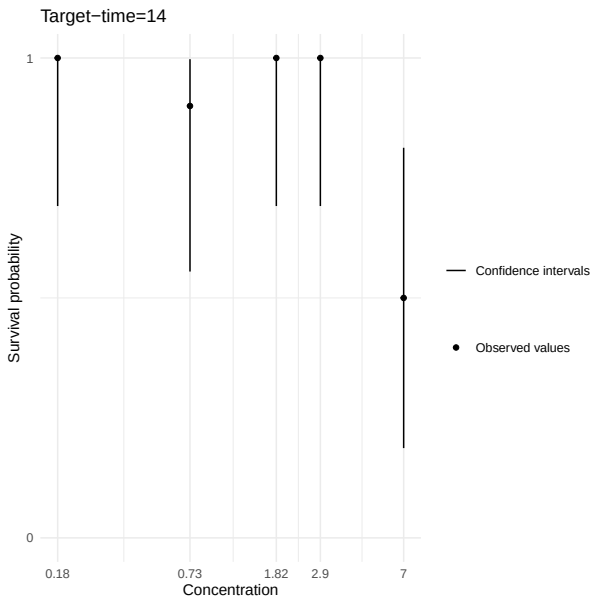
```
Number of replicates per time and concentration:
      time
conc    0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21
  0    10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
  0.18 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
  0.73 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
  1.82 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
  2.9  10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
  7    10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10

Number of survivors (sum of replicates) per time and concentration:
      0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21
0    10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
0.18 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10  9  9  9  9
0.73 10 10 10 10 10 10 10 10 10 10  9  9  9  9  9  9  9  9  9  9  9  9
1.82 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10  9  9  9  9  9  9  9
2.9  10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10  4  4  3
7    10 10 10  9  9  9  9  9  9  9  8  8  5  5  5  5  5  5  5  2  2  2
```

```
plot(dataset)
```



Step 1.2

```
plotDoseResponse(dataset, target.time=14,
                 addlegend=TRUE, log.scale=TRUE, main="Target-time=14")
```



Step 1.3

```
fit <- survFitTT(dataset, target.time=14, quiet=TRUE,
                 lcx=c(5, 15, 25, 45, 75))
summary(fit, quiet=TRUE)$Qpost
```

```
        50%      2.5%     97.5%
b 1.843e+00 8.747e-01 3.474e+00
e 7.443e+00 4.693e+00 1.625e+01
```

```
summary(fit, quiet=TRUE)$QLCx
```

```
            50%      2.5%     97.5%
LC5  1.507e+00 3.357e-01 3.110e+00
LC15 2.910e+00 1.219e+00 4.941e+00
LC25 4.116e+00 2.223e+00 6.883e+00
LC45 6.671e+00 4.234e+00 1.355e+01
LC75 1.355e+01 7.457e+00 4.817e+01
```

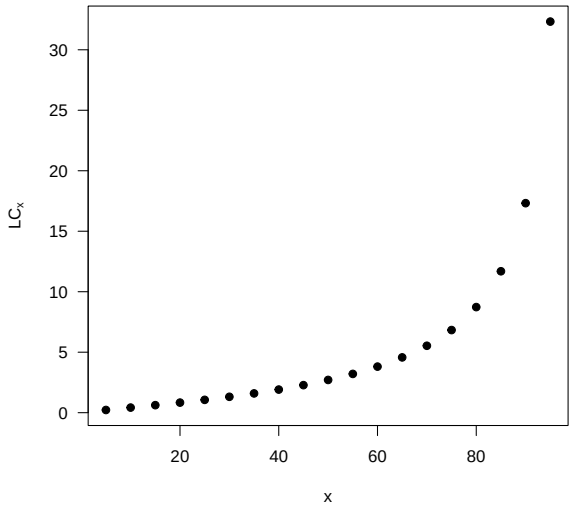
When reducing the target time, a warning message is delivered:

```
The LC50 estimation (model parameter e) lies outside the range of tested concentration
and may be unreliable as the prior distribution on this parameter is defined from this
range !
```

Parameter e is not well estimated, so that LC_x calculations are suspicious.

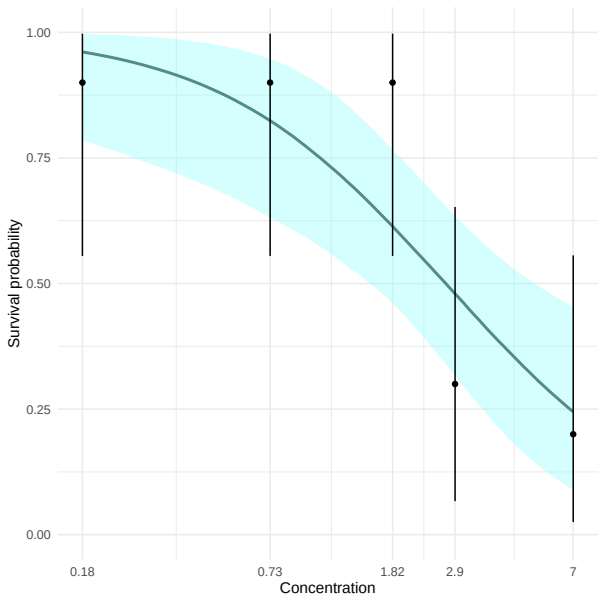
We decide to keep a target time at 21 days.

```
fit <- survFitTT(dataset, target.time=21, quiet=TRUE,
                 lcx=seq(5,95,5))
lethal.conc <- summary(fit, quiet=TRUE)$QLCx
plot(seq(5,95,5), lethal.conc$`50%`, pch=19, las=1,
     xlab="x", ylab=expression(LC[x]))
```



Step 1.4

```
plot(fit, adddata=TRUE, log.scale=TRUE,
     fitcol="darkslategray4", cicol=NA, ribcol="darkslategray1")
```



2 *D. magna* exposed to cadmium

Step 2.1

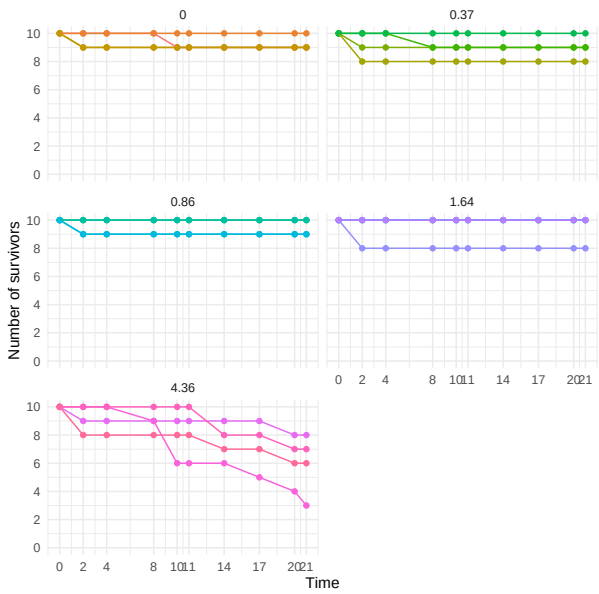
```
data(cadmium1)
dataset <- survData(cadmium1)
summary(dataset)
```

```
Number of replicates per time and concentration:
      time
conc   0 2 4 8 10 11 14 17 20 21
  0    4 4 4 4  4  4  4  4  4  4
  0.37 4 4 4 4  4  4  4  4  4  4
  0.86 4 4 4 4  4  4  4  4  4  4
  1.64 4 4 4 4  4  4  4  4  4  4
  4.36 4 4 4 4  4  4  4  4  4  4

Number of survivors (sum of replicates) per time and concentration:
      0  2  4  8 10 11 14 17 20 21
0    40 38 38 38 37 37 37 37 37 37
0.37 40 37 37 36 36 36 36 36 36 36
0.86 40 38 38 38 38 38 38 38 38 38
1.64 40 38 38 38 38 38 38 38 38 38
4.36 40 37 37 36 33 33 30 29 25 24
```

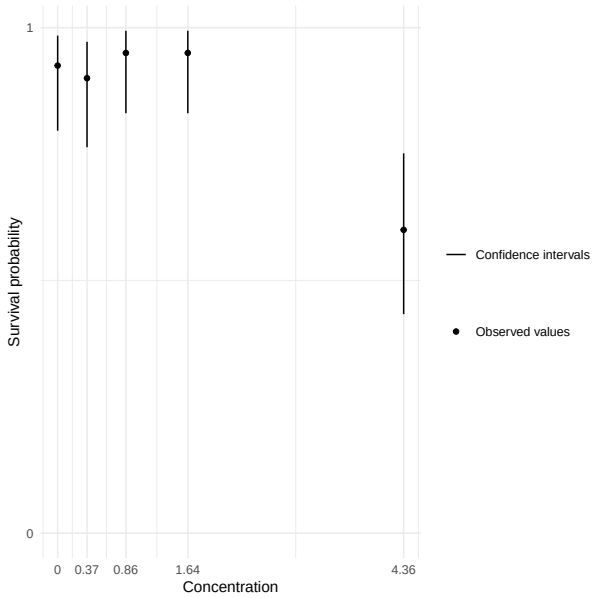
We notice that we have 4 replicates of 10 individuals for each tested concentrations, instead of 10 replicates of 1 individual. In this experiment, daphnids have not been followed daily.

Step 2.2

```
plot(dataset)
```



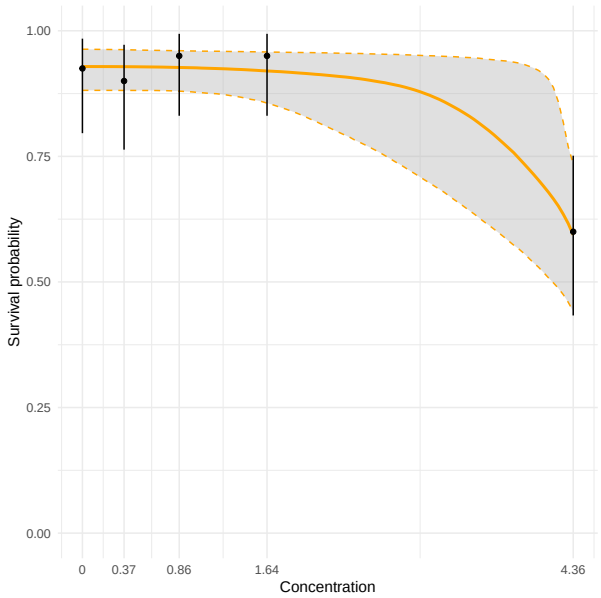
Step 2.3 `plotDoseResponse(dataset)`



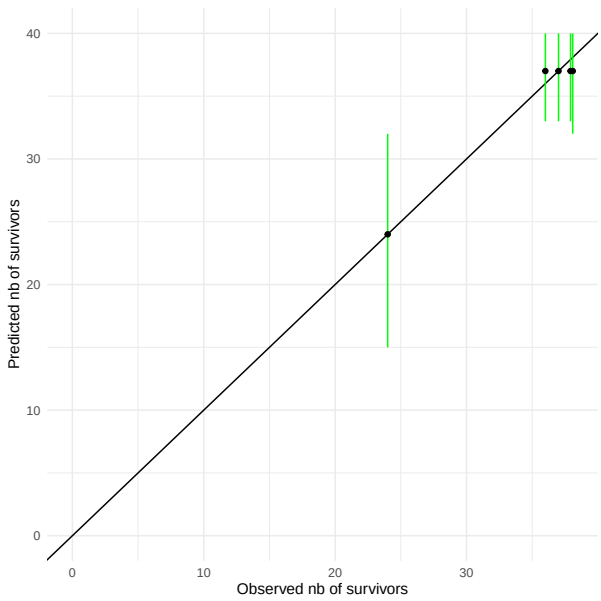
Step 2.4
```
fit <- survFitTT(dataset, quiet=TRUE, lcx=c(10, 20, 50))
summary(fit)
```

```
Summary:

The  loglogisticbinom_3  model with a binomial stochastic part was used !

Priors on parameters (quantiles):

       50%      2.5%     97.5%
b 1.00e+00 1.259e-02 7.943e+01
d 5.00e-01 2.500e-02 9.750e-01
e 1.27e+00 3.792e-01 4.254e+00

Posteriors of the parameters (quantiles):

        50%      2.5%     97.5%
b 6.430e+00 2.057e+00 6.211e+01
d 9.288e-01 8.814e-01 9.632e-01
e 4.692e+00 4.301e+00 6.371e+00

Posteriors of the LCx (quantiles):

           50%      2.5%     97.5%
LC10 3.427e+00 1.908e+00 4.261e+00
LC20 3.905e+00 2.724e+00 4.356e+00
LC50 4.692e+00 4.301e+00 6.371e+00
```

Step 2.5 `plot(fit, adddata=TRUE)`



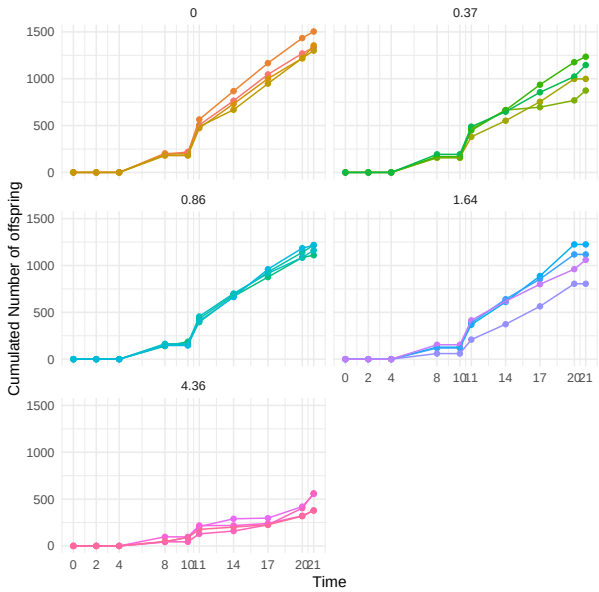
`ppc(fit)`



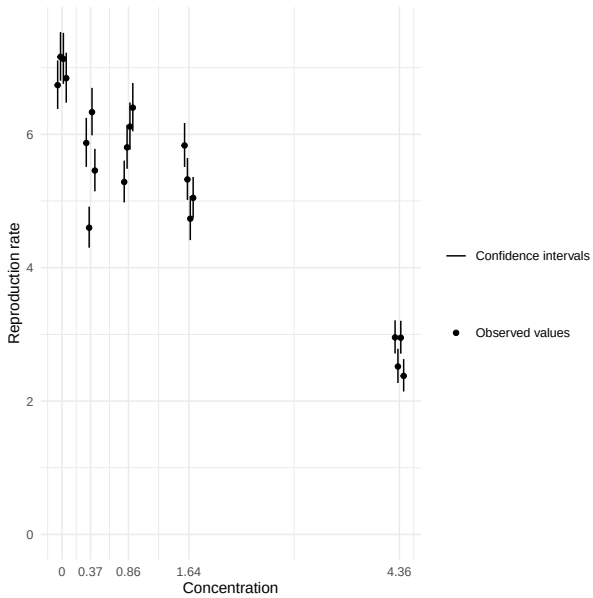
Fit results seem good regarding the data and despite they are sparse. But, because data are sparse, we got a large credibility band after the fourth tested concentrations. This means that we have to be cautious, for example, with LC_x estimates when x is high: for $x = 50\%$, the upper uncertainty bound is greater than the highest tested concentration, making the LC_{50} estimate not precise.

Step 2.6 Reproduction analysis

```
dataset <- reproData(cadmium1)
plot(dataset)
```




```
plotDoseResponse(dataset, addlegend=TRUE)
```

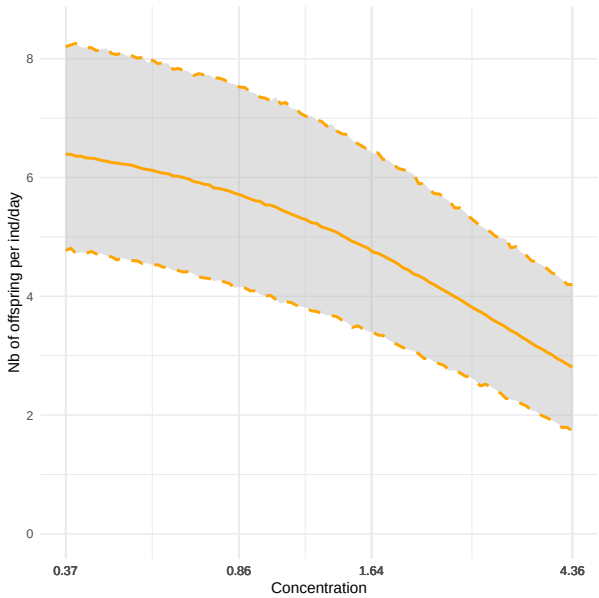



```
fit <- reproFitTT(dataset, quiet=TRUE, ecx=c(10, 20, 50))
summary(fit)
```

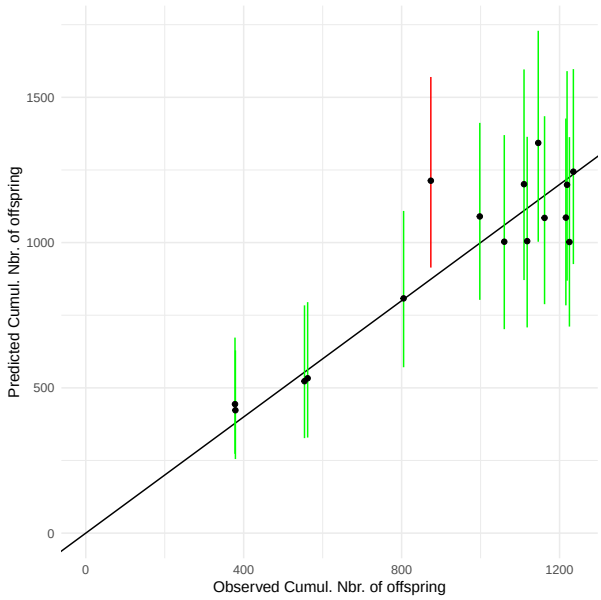
```
Summary:

The  loglogistic  model with a Gamma Poisson stochastic part was used !

Priors on parameters (quantiles):

            50%      2.5%     97.5%
b     1.000e+00 1.259e-02 7.943e+01
d     6.968e+00 6.762e+00 7.175e+00
```

```
e     1.936e+00 8.741e-01 4.290e+00
omega 1.000e+00 1.585e-04 6.310e+03

Posteriors of the parameters (quantiles):

            50%      2.5%     97.5%
b     1.197e+00 7.768e-01 1.786e+00
d     6.942e+00 6.731e+00 7.147e+00
e     3.220e+00 2.605e+00 4.082e+00
omega 9.802e-02 4.741e-02 2.383e-01

Posteriors of the ECx (quantiles):

           50%      2.5%     97.5%
EC10 5.108e-01 1.943e-01 9.649e-01
EC20 1.004e+00 5.472e-01 1.548e+00
EC50 3.220e+00 2.605e+00 4.082e+00
```

```
plot(fit, log.scale=TRUE)
```



```
ppc(fit)
```



As already said before, because data are sparse, the credibility band is large (we also have large credible intervals around EC_x). Despite this fact, only one of the prediction segments is red, so that fit results are finally not so bad with this dataset.

Comparing the sensitivity of *D. magna* to both chlordan and cadmium, means look at toxicity indices on both survival and reproduction for both substances. As biologically expected, EC_x estimates are lower than LC_x ones.

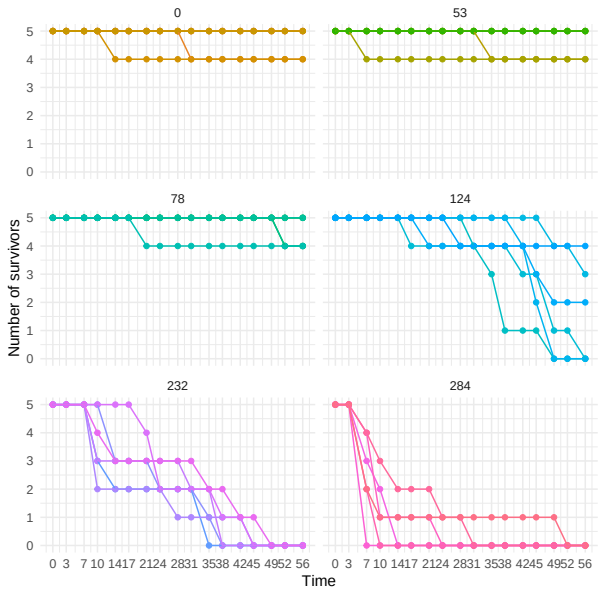
Because of the large uncertainty on parameter estimates with the dataset 'cadmium1', it is more appropriate to make a decision based on the EC_{20}, that is a toxicity index estimated with a better precision (even if it is still quite large).

Then, because the EC_{20} for cadmium ($1.0 \in [0.55; 1.6]$) is greater than the one for chlordan ($0.39 \in [0.22; 0.64]$), we can say that *D. magna* is more sensitive to chlordan than to cadmium.

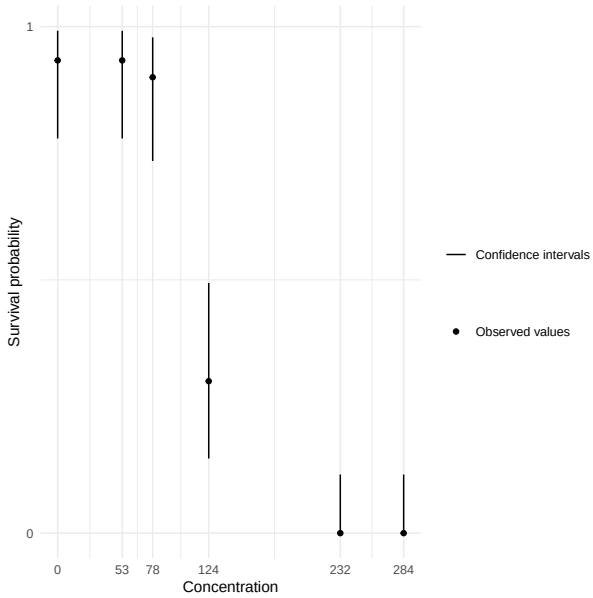
3 *L. stagnalis* exposed to cadmium

Step 3.1 Survival analysis

```
data(cadmium2)
dataset <- survData(cadmium2)
plot(dataset)
```


0
53
78
124
232
284
Number of survivors
Time
0 3 7 10 1417 2124 2831 3538 4245 4952 56

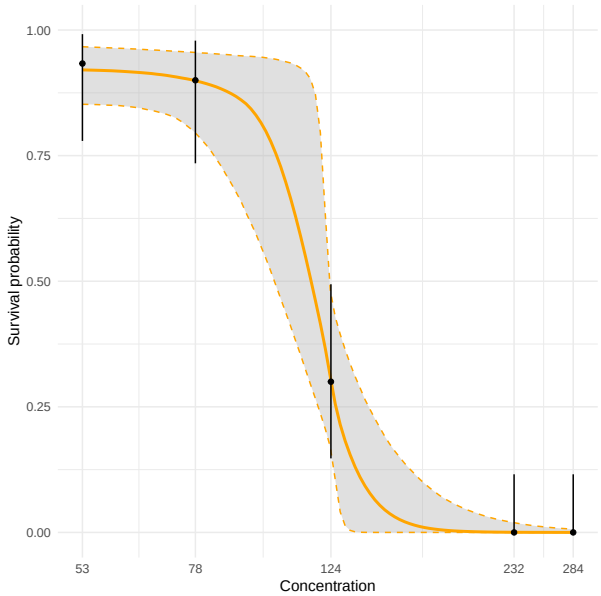
```
plotDoseResponse(dataset, addlegend=TRUE)
```



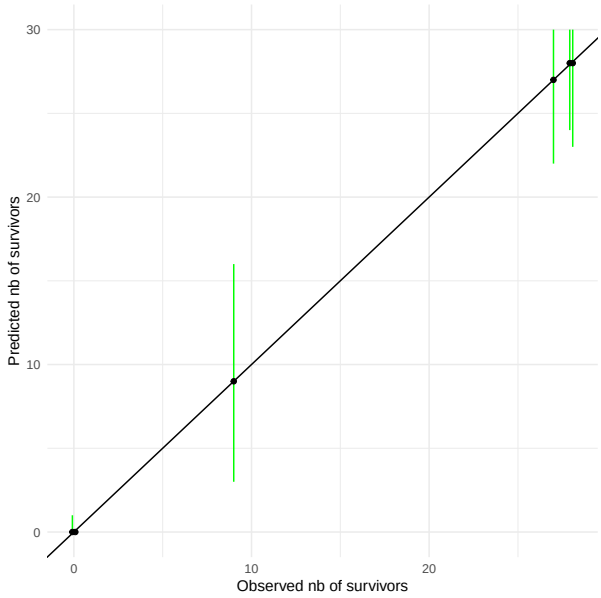

```
fit <- survFitTT(dataset, quiet=TRUE, lcx=c(10, 20, 50))
summary(fit)
```

```
Summary:

The  loglogisticbinom_3  model with a binomial stochastic part was used !

Priors on parameters (quantiles):

        50%      2.5%     97.5%
b 1.000e+00 1.259e-02 7.943e+01
d 5.000e-01 2.500e-02 9.750e-01
e 1.227e+02 5.390e+01 2.793e+02

Posteriors of the parameters (quantiles):

        50%      2.5%     97.5%
b 1.165e+01 5.394e+00 7.086e+01
d 9.227e-01 8.532e-01 9.711e-01
e 1.177e+02 1.043e+02 1.246e+02

Posteriors of the LCx (quantiles):

           50%      2.5%     97.5%
LC10 9.678e+01 7.266e+01 1.191e+02
LC20 1.041e+02 8.349e+01 1.206e+02
LC50 1.177e+02 1.043e+02 1.246e+02
```

```
plot(fit, log.scale=TRUE, adddata=TRUE)
```



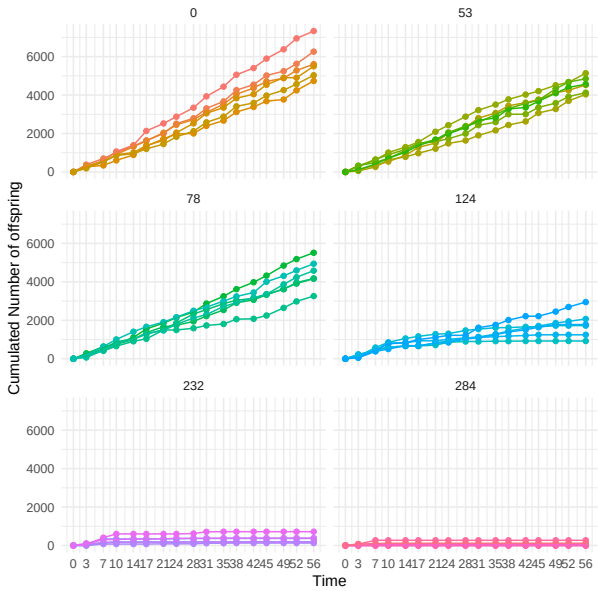

```
ppc(fit)
```



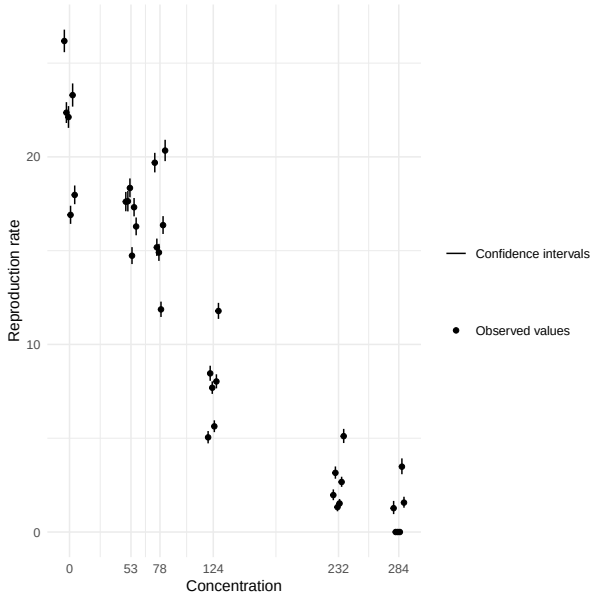
Very good fit results for survival.

Reproduction analysis

```
dataset <- reproData(cadmium2)
plot(dataset)
```




```
plotDoseResponse(dataset, addlegend=TRUE)
```

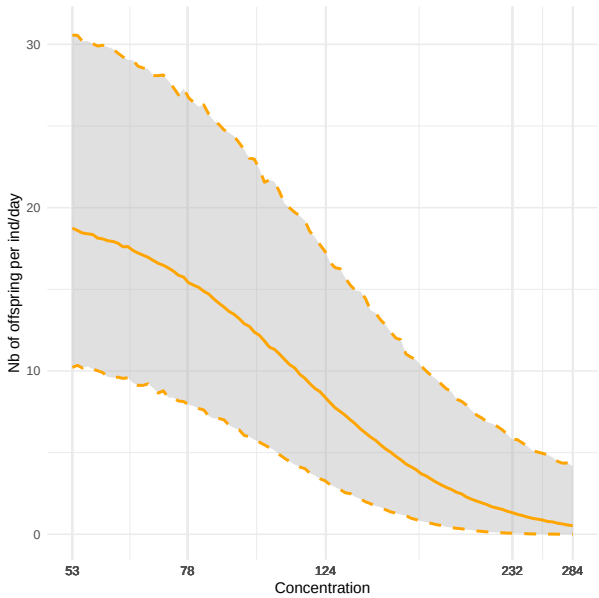



```
fit <- reproFitTT(dataset, quiet=TRUE, ecx=c(10, 20, 50))
summary(fit)
```

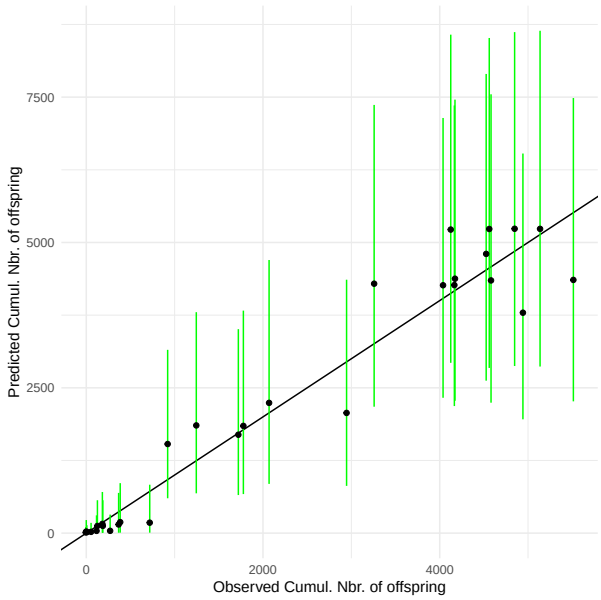
```
Summary:

The  loglogistic  model with a Gamma Poisson stochastic part was used !

Priors on parameters (quantiles):

            50%      2.5%     97.5%
b     1.000e+00 1.259e-02 7.943e+01
d     2.147e+01 1.871e+01 2.424e+01
e     1.488e+02 7.902e+01 2.804e+02
omega 1.000e+00 1.585e-04 6.310e+03

Posteriors of the parameters (quantiles):
```

```
            50%      2.5%     97.5%
b     3.367e+00 2.608e+00 4.330e+00
d     2.072e+01 1.838e+01 2.329e+01
e     1.130e+02 9.565e+01 1.335e+02
omega 1.262e+00 7.221e-01 2.461e+00

Posteriors of the ECx (quantiles):

           50%      2.5%     97.5%
EC10 5.879e+01 4.269e+01 7.786e+01
EC20 7.477e+01 5.777e+01 9.458e+01
EC50 1.130e+02 9.565e+01 1.335e+02
```

```
plot(fit, log.scale=TRUE, adddata=TRUE)
```



```
ppc(fit)
```



As for survival, we got very good fit results for reproduction, even if uncertainty limits are larger.

Step 3.2 Comparison of sensibilities to cadmium between *D. magna* and *L. stagnalis*

Based on both the LC_{50} and the EC_{50} estimates, the species *D. magna* is much more sensible than *L. stagnalis*:

	LC_{50}	EC_{50}
D. magna	4.7	3.2
	$[4.3; 6.5]$	$[2.6; 4.1]$
L. stagnalis	236	137
	$[210; 253]$	$[114; 177]$

Due to the large uncertainty on parameter estimates with data on *D. magna* exposed to cadmium, reasoning on the EC_{20} and the LC_{20} is more reasonable. Nevertheless, the conclusion will be the same.

4 Quantitative continuous data analysis

Step 4.1 Go back to `http://mosaic.univ-lyon1.fr/` and choose menu `Growth`. This is the way to use MOSAIC at your office. During this training course, please use the pedagogical server which ensures that all participants can perform exercises at the time:

`http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/`.

Step 4.2 If you wish, you can first visualize our demo video (duration: `08:39`), by simply clicking on the video link or directly here: `http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/MOSAICgrowth.mp4`

Step 4.3 Data visualisation

- Try first with example `plant07.txt`. Notice that either a plot and a table are provided to check data have been correctly uploaded. Note that the user may personalize X- and Y-axis labels and choose the appropriate units for exposure, measurements and time.

> For example data set that are provided within MOSAIC, X- and Y-axis labels as well as units for exposure, measurements and time are automatically filled in. There is no need to change anything here, nevertheless you can personalize fields as you want.

- An innovative feature in this module is the possibility to upload several files **at the same time**. Try to upload these additional examples `chlordan-daphnia.txt`, `cadmium-daphnia.txt`, `subst01-lymnaea.txt`, `plant03.txt` and `plant10.txt`. What do you notice? Which data set is displayed? Try to change file.

  The list of selected files appears within the light grey box, while a warning is displayed below standing that some data files have no common time with the others. This means that a similar dose-response analysis (that is for the same exposure duration, also called the target time) will be not possible for these files simultaneously.
  Not selecting `subst01-lymnaea.txt` solves the issue; indeed, for this file the exposure duration is 56 days, while for all the others it is 21 days.
  Notice that by default the last uploaded data file is displayed, so here `plant10.txt`. The scrolling menu allows you to choose the data you want to visualize.

Step 4.4 Dose-response analysis

- Keeping several files uploaded, go to sub-menu `Dose-response analysis`. Run the analysis for example `plant07.txt`. What is the estimate of the EC_{50} (denoted ER_{50} within the module). Provides the median and the 95% credible interval.

  We get $ER_{50} = 685.05\,[511.67; 949.79]$.

- From goodness-of-fit results, get the correlation plot. What do you notice? Don't hesitate to read the help text below the figure.

  Parameter σ is not correlated to the three others (potatoid shapes of contour lines), while a more or less strong correlation appears between b, d and e (left-leaning ellipses).

- Ask for the calculation of ER_{25}, ER_{50} and ER_{75}. What do you notice in terms of precision of the estimates?

  When x increases, the precision of the ER_x estimates decreases, underlined by an increasingly large credible interval.

- Run additional analyses for examples `plant03.txt` and `plant10.txt`. Look at the ER_{50} estimate from file `plant10.txt`. What do you notice? How could you explain such a result?

  The ER_{50} estimate from file `plant10.txt` is very high, far from the highest tested concentration equal to 2000), and the estimate is very imprecise; a way to quantify the precision is to calculate a kind of coefficient of variation as $CV = \frac{Q97.5-Q2.5}{median}$ that is expected to be less than 0.5: here it is equal to 1.3. This means that this ER_{50} estimate must be considered with caution, as it corresponds to an extrapolation far from the concentration range of the experiment.
  Such a disappointing result is due to the fact that the concentration range has not been chosen accordingly to the sensitivity of the plant species towards the substance to which it has been exposed. Indeed, it is a priori expected, when designing a toxicity test, to go from 0 to 100% of response or effect within the chosen concentration range.

- Is the PPC plot suitable enough to trust provided results for further ERA?

  This question stands for file `plant10.txt`, where 100% of segments are green what translates a not enough correct fit of the model on these data.

- Ask the calculation of ER_{50} for `plant03.txt` and `plant10.txt`. Ask for the table displaying all ER_{50} for files `plant03.txt`, `plant07.txt` and `plant10.txt`. Which plant species is the most sensitive?

Using the ER_{50} to compare the plant species is not really appropriate as the one for `plant10` is not precisely estimated. So, rather looking at the ER_{25} estimates from the three data files, `plant10` is undoubtedly the less sensitive, while `plant03` and `plant07` are equally sensitive with very close ER_{25} estimates.

Step 4.5 Downloads

- Go to sub-menu `Downloads`.
- For example `plant07.txt`, download the single report (choose the Word format).
- Download the text file of the joint posterior distribution.

A file entitled `outputPosterior_plant07.txt` is downloaded within the appropriate folder on your computer. This file is a text file (.txt) with four columns corresponding to the joint posterior distribution of the four model parameters: b, d and e for the deterministic part (the three-parameters log-logistic model), and σ for the stochastic part (here a normal distribution).

Step 4.6 Prediction tool

- Go to sub-menu `Prediction tool`.
- Enter a concentration range, for which parameter value of the three-parameters log-logistic model will be available. For example: `40; 80; 160; 320; 640`.

Note that fields `Target Time` and `Unit` ; the user should fill in them to keep in mind that the provided simulation will be made for a given set of parameter values that have previously been estimated at a given target time (namely the exposure duration, expressed in a given unit). Hence, the prediction is valid for a similar exposure duration.

- **As a first try**, choose non distributed parameters (default option).
- **Option (1)**: enter a single value for each parameter, based on your experience, based on expert knowledge or coming form the literature for example. Keep in mind that these values where obtained for a given exposure duration, chosen by the experimenter, usually equal to the duration of the experience (e.g., the standard duration of 21 days for most of the toxicity tests performed under an OECD guideline). Try option (1) with parameter values entered by default as they are; please note that parameter b and e are expected to be given in log_{10}. What do you notice?

With default parameter values, we can notice that the prediction (given as a median curve) goes from 0 to 100% of effect, meaning the the chosen concentration range will be suitable to test the toxicity of the species/substance combination of interest.

- Change the previous value of parameter e from $log_{10}(e) = 2.2$ to $log_{10}(e) = 3$. What do you notice?

Changing parameter e from $log_{10}(e) = 2.2$ to $log_{10}(e) = 3$ provides a prediction for which the chosen concentration range leads to less than 50% of effect at the highest concentration. Hence, if this concentration range is kept as it is, it will not be possible to precisely estimate the ER_{50} once data will have been collected. It is thus strongly recommended to choose another concentration range.

- **Option (2)**: get median parameter estimates from a DR analysis that has already been performed on a data set (either with $\text{MOSAIC}_{\text{growth}}$ or with any other software), under experimental conditions that are close enough to those for which you want to make the prediction; in particular pay attention to simulate for a similar exposure duration. Try option (2) with parameter estimates obtained from `plant07.txt` data set. What do you notice?

Parameter values should be around $b = 0.63$, $d = 3.46$ and $e = 685.05$; use a dot to separate decimals. Don't forget to convert b and e in $\log_{10}$-scale.
We obtain a predicted curve very similar to the median curve of the fitting plot obtained during the DR analysis from the `plant07.txt` data set; the difference is in the concentration range.

- Try again option (2), but with parameter estimates obtained from `plant10.txt` data set. Keep again the same concentration range. What do you notice?

Parameter values should be around $b = 0.73$, $d = 8.63$ and $e = 6797.48$; use a dot to separate decimals. Don't forget to convert b and e in $\log_{10}$-scale.
As previously, we obtain a predicted curve very similar to the median curve of the fitting plot obtained during the DR analysis from the `plant07.txt` data set; the difference is in the concentration range.

- Based on the previous simulation, find a concentration range that would be more appropriate.

A possible choice could be the following series: `800; 1600; 3200; 6400; 12800`, that allows to predict more than 50% of effect at the highest concentration.

- **As a second try**, choose **distributed parameters**. This necessary means that you previously downloaded a joint posterior distribution of the DR model parameters, what is today only possible with the MOSAIC$_{\text{growth}}$ module, or that you previously performed a DR analysis with MOSAIC$_{\text{growth}}$.
- **Option (1)**: coming back to the initially proposed concentration range (namely `40; 80; 160; 320; 640`), upload file `outputPosterior_plant07.txt` you may have already downloaded earlier. If not, go to option(2). Which difference(s) do you observe with previous predictions?

Since parameters are now distributed according to the joint posterior distribution, the prediction is associated with a 95% credible band, giving an idea about the precision around the prediction.

- **Option (2)**: choose option `from a previous DR analysis`, that is supposed to have been performed on data set `plant07.txt` earlier. What do you notice?

We get the same prediction as previously.

- Change the concentration range and see what happens.

For example, taking the same concentration range obtained earlier, namely `800; 1600; 3200; 6400; 12800`, we predict a 50% of effect at concentration 12800 associated here with the precision of the prediction (quite large).

For further details on MOSAIC$_{\text{growth}}$, you can download and read the Tutorial available on line from here: `http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/Tutorial.pdf`
A vignette also accompanies the Tutorial, with details on the modelling framework used in MOSAIC$_{\text{growth}}$: `http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/vignette.pdf`.
Its reading is optional for the use of the application.

5 Bioaccumulation factors calculation

Step 5.1 Go back to `http://mosaic.univ-lyon1.fr/` and choose menu `bioacc`. This is the way to use MOSAIC at your office. During this training course, please use the pedagogical server which ensures that all participants can perform exercises at the time:

`http://umr5558-shiny.univ-lyon1.fr/mosaic-bioacc/`.

For further details on MOSAIC$_{\text{bioacc}}$, you can download and read the user-guide available on line by clicking on the dedicated button.

Step 5.2 Data visualisation

- Try first with example `Male_Gammarus_Single.txt`. Notice that either a plot and a table are provided to check data have been correctly uploaded. Note that the user has nothing to do, everything is automatically filled in.
- Download example file `Male_Gammarus_Single.txt` and browse the file. Notice that the user has to enter the correct separator, the time unit and the duration of the accumulation phase. Based on the description of the example, fill in the correct inputs.

  Tab separator, days for time unit and 4 for the duration of the accumulation phase.

- Try with example file `Oncorhynchus_two.csv` and look at the data. Notice that the user has now the choice between two exposure concentrations.

  Two exposure concentrations are available: $4.1.10^{-4}$ and $4.4.10^{-3}$ $\mu g.mL^{-1}$

- Try with example file `Male_Gammarus_seanine.csv`. Notice that in plot visualisation, parent concentrations, metabolite concentrations and growth measurements are available.

Step 5.3 Model and parameters

- Upload again example file `Male_Gammarus_Single.txt` and go to sub-menu `Model and parameters`. The parameters of the model are automatically selected according to the uploaded data. In addition, the equations of the model (for the deterministic part only) are displayed with the corresponding variables. For this data set, how many parameters are there? And How many equations? Can you deduce the exposure routes (water, sediment, pore water and/or food?) and elimination processes (excretion, growth dilution and/or biotransformation?) that were considered in the experiment?

  There are 3 parameters: k_{uw}, k_{ee} and σ_{Cpred}. Two equations are given: one stands for the accumulation phase ($0 \leq t \leq t_c$), the other for the depuration phase ($t > t_c$). The exposure route is by water and the elimination process is excretion (elimination).

- Idem with example file `Chironomus_benzo-a-pyrene.csv`.

  There are 6 parameters: k_{us}, k_{ee}, k_{m1}, k_{em1}, σ_{Cpred} and $\sigma_{Cmetpred1}$. Three equations are given, the first one stands for the accumulation phase of the parent compound ($0 \leq t \leq t_c$), the second for the depuration phase of the parent compound ($t > t_c$) and the last one for the metabolite compound (both accumulation and depuration phases). The exposure route is by sediment and the elimination processes are excretion (elimination) and biotransformation.

- Idem with example file `Male_Gammarus_seanine.csv`.

  There are 16 parameters: k_{uw}, k_{ee}, k_{eg}, g_0, g_{max}, k_{m1}, k_{m2}, k_{m3}, k_{em1},k_{em2}, k_{em3}, σ_{Cpred}, $\sigma_{Cmetpred1}$, $\sigma_{Cmetpred2}$, $\sigma_{Cmetpred3}$ and σ_{Gpred}. Six equations are given, the first one stands for the accumulation phase of the parent compound ($0 \leq t \leq t_c$), the second for the depuration phase of the parent compound ($t > t_c$), the third, fourth and fifth ones for the metabolite compound (both accumulation and depuration phases, respectively for metabolite 1, 2 and 3), and the last one stands for growth. The exposure route is by water and the elimination processes are excretion (elimination), growth dilution and biotransformation.

- Upload again example file `Male_Gammarus_Single.txt` and run the analysis with a click on the "Calculate and display" button. Calculations can take several minutes depending on the data set.

Step 5.4 Results and goodness-of-fit criteria

- File `Male_Gammarus_Single.txt`, what is the estimate of the BCF_k? Provide the median and the 95% credible interval.

The bioaccumulation factor provided is the BCF (bioconcentration factor), because the exposure route is by water. Median BCF_k is 18103, and its 95% credible interval is: [14401-23473].

- Is it reasonable to ask for the BCF_{ss}? If yes, provide its median and its 95% credible interval.

The steady-state is not reached at the end of the accumulation phase. Thus, it is not relevant to ask the BCF_{ss} for this example.

- What are the estimates of parameters k_{uw} and k_{ee}? Provide their median and 95% credible intervals.

k_{uw}: 616.4 [548.5-685.4] d^{-1} k_{ee}: 0.03388 [0.02213-0.04614] d^{-1}

- From `Goodness-of-fit` sub-menu, get the Posterior Predictive Check (PPC). What is the percentage of data within the prediction intervals? Do not hesitate to read the help-text next to the figure.

The percentage of data within the prediction intervals is 100%. We expect to have 95%. Here the problem comes from very large uncertainties on predictions that are thus not precise enough to be confident within.

- Get the `Priors and Posteriors` plot. Are posterior distributions narrower than prior ones?

Yes, the posterior distributions (orange) are narrower than prior ones (grey) for each parameter.

- Go to the `Correlation plot`. What do you notice?

k_{uw} and k_{ee} are highly positively correlated (0.82), but this is mainly due to the model itself here.

- Go to the `Potential Scale Reduction Factors` (PSRF). Are the values acceptable for convergence for all parameters?

The values are acceptable for each parameter (equal to 1.00).

- Do the MCMC chains overlap well for each parameter?

Yes, the thee MCMC chains overlap well for each parameter.

- Idem with example files `Chironomus_benzo-a-pyrene.csv` and `Male_Gammarus_seanine.csv`. Note that calculations can take several minutes, especially for the second file. It would be quicker to directly use the R code in RStudio, but that would require to deeply go to the code itself (in case you would like to do such a way, please ask us).

`Chironomus_benzo-a-pyrene.csv` Go to the report, directly downloadable from MOSAIC$_{\text{bioacc}}$. The bioaccumulation factor provided is the BSAF (biota-sediment bioaccumulation factor), because the exposure route is by sediment. The median $BSAF_k$ is 1.1829 (equal to median), and its 95% credible interval is: [0.8066-3.0964]. The steady-state seems to be reached, thus: $BSAF_{ss}$ = 0.1949[0.1111-0.2771].
k_{us}: 0.8036 [0.4943-323.8] d^{-1}
k_{m1}: 1.709 [1.333-2.283] d^{-1}
k_{ee}: 2.401 [0.2594-1728] d^{-1}
k_{em1}: 0.8479 [0.6077-1.159] d^{-1}
PPC: 92.31% for the parent compound (acceptable) and 100% for the metabolite (large uncertainties on predictions). Prior and posterior plots: yes, the posterior (orange distribution) is narrower than the prior distribution (grey distribution) for each parameter. However, for k_{us}, we can observe large distribution queues. For the correlation plot: k_{us} with k_{ee} and k_{m1} with k_{em1} are highly correlated (> 0.8). PSRF: the values are correct for each parameter. Idem with the MCMC plot, the 3 chains overlap for each parameter. For k_{us} and k_{ee} we observe at the beginning that one chain does not overlap, but at the end, the 3 MCMC have converged.

`Male_Gammarus_seanine.csv` Go to the report, directly downloadable from MOSAIC$_{\text{bioacc}}$. The bioaccumulation factor provided is the BCF because the exposure route is by water. 50% BCF_k is 1969, and its 95% credible interval is: [62-233651]. This large interval is due to the large parameter estimates, especially the posterior distribution of k_{uw} and k_{ee} which have large distribution queues (see below, the parameter estimates and prior posterior plots). The steady-state seems to be reached, thus: BCF_{ss} = 71[18-121].
k_{uw}: 16740 [918.9 - 91770] d^{-1}
k_{ee}: 4.164 [1.91e-05 - 1076] d^{-1}
k_{eg}: 0.3092 [0.2261 - 0.3718] d^{-1}
k_{m1}: 73.27 [0.165 - 1191] d^{-1}
k_{m2}: 0.5166 [0.371 - 0.7093] d^{-1}
k_{m3}: 0.1957 [0.1153 - 1.116] d^{-1}
k_{em1}: 561 [0.6479 - 9304] d^{-1}
k_{em2}: 0.123 [1.35e-03 - 0.2172] d^{-1}
k_{em3}: 0.7808 [0.3172 - 8.234] d^{-1}
g_{max}: 0.507 [0.5003 - 0.549] g
g_0: 0.2127 [0.5003 - 0.2483] g
PPC: 95% for the parent compound and the second metabolites (acceptable), 100% for the first and third metabolites, as well as for growth (large uncertainties on predictions). Prior and posterior plots: globally, the posterior (orange distribution) is narrower than the prior distribution (grey distribution) for each parameter. However, for k_{eg}, k_{m2} and k_{em2}, we can observe large distribution queues (*e.g.*, 10^{-4} to 2 for k_{eg}). Correlation plot: according to the coloured matrix, we notice a positive correlation between k_{uw} and k_{ee}, k_{m1} and k_{em1}, k_{m2} and k_{em2}, and k_{m3} and k_{em3}. In addition, negative correlations are observed for g_{max} and g_0 with k_{eg}. PSRF: the values are correct for each parameter. Idem with the MCMC plot, the 3 chains overlap for each parameter, except for k_{uw}, k_{ee}, k_{m1} and k_{em1}. GOF are not really good with this data set, probably due to a lack of data regarding the number of parameters to estimate.

- What do you notice for the bioaccumulation factor with data set `Chironomus_benzo-a-pyrene.csv`?

The bioaccumulation factor provided is the BSAF (biota-sediment bioaccumulation factor), because the exposure route is by sediment. Besides, we can observe on the $BSAF_k$ density plot a very small distribution queue. In the table on the right, the median value and the 95% credible interval of $BSAF_k$ are reasonable.

Step 5.5 Downloads

- Go to sub-menu `Downloads`.
- For example `Male_Gammarus_Single.txt`, download the report (choose the Word format).

- Download the text file of the joint posterior distribution and the R code.