MOSAIC

Efficiency for Environmental Risk Assessment

https://mosaic.univ-lyon1.fr

Sandrine CHARLES (UCBL - LBBE)

sandrine.charles@univ-lyon1.fr

April the 6^{th}, 2021

Université Claude Bernard Lyon 1

Table of content

Introduction

Dose-response modelling

Parameter inference

Environnemental Risk Assessment - The MOSAIC platform

Table of content

Introduction

Dose-response modelling

Parameter inference

Environnemental Risk Assessment - The MOSAIC platform

Ecotoxicology

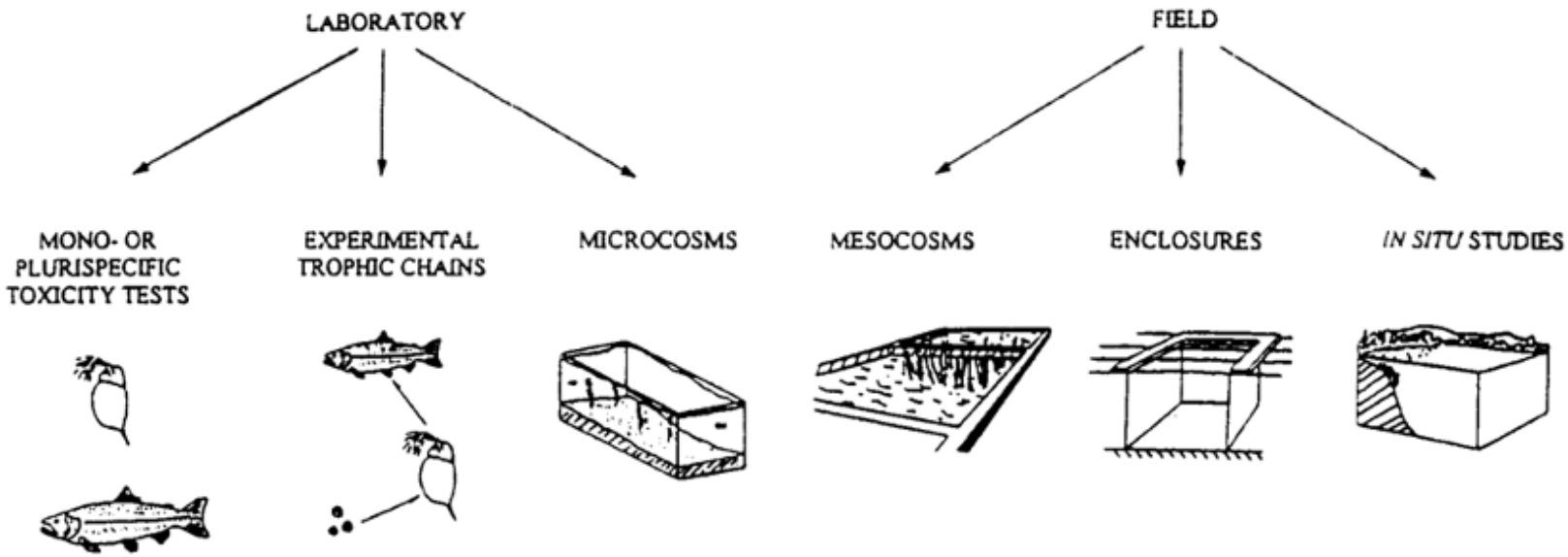
- A scientific field at the bridge of chemistry, toxicology and ecology

  "The branch of toxicology concerned with the study of toxic effects, caused by natural or synthetic pollutants, to the constituents of ecosystems, animals (including humans), vegetables and microorganisms, in an integrated context" [Truhaut, 1977]

  "Ecology in the presence of toxicants" [Chapman, 2002]

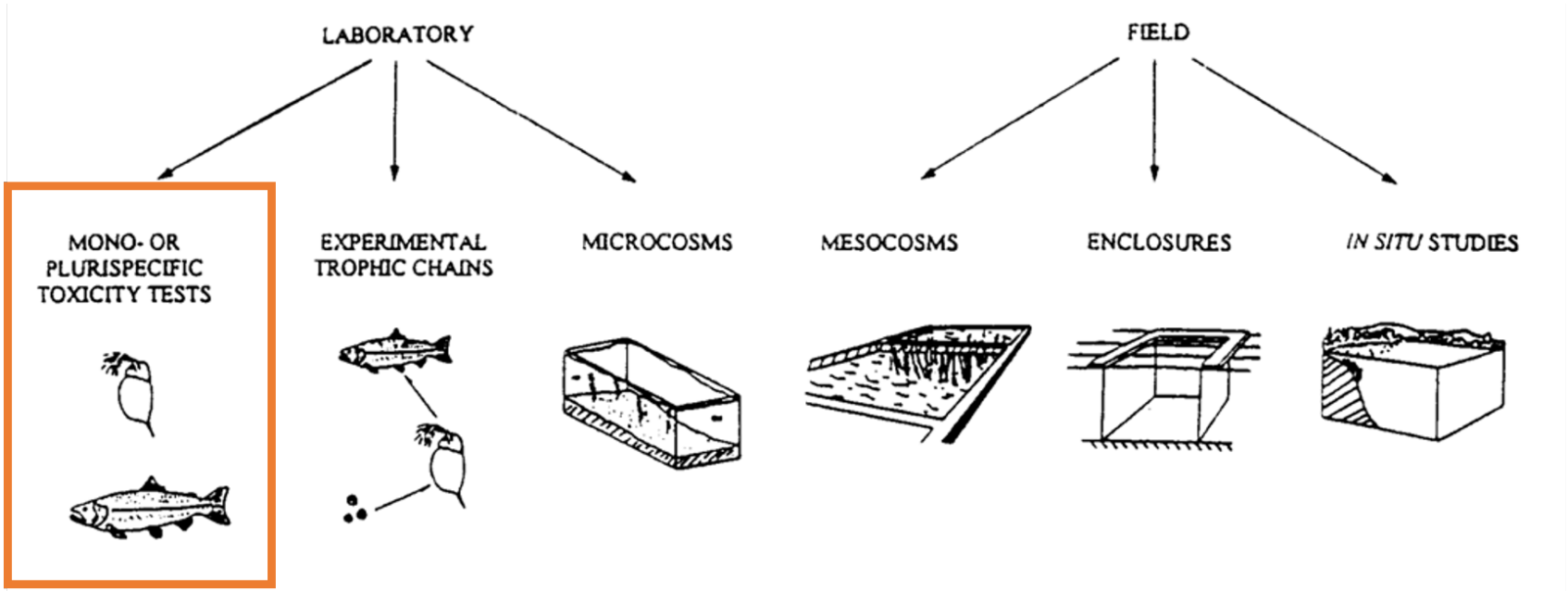
- In ecotoxicology, the answer of the ecosystem to environmental perturbations (physical, chemical and/or biological) is studied in all compartments of the biosphere (air, soil and water) and at all levels of biological organization [Walker et al., 2006]

A variety of experimental devices



[Caquet et al., 1996]

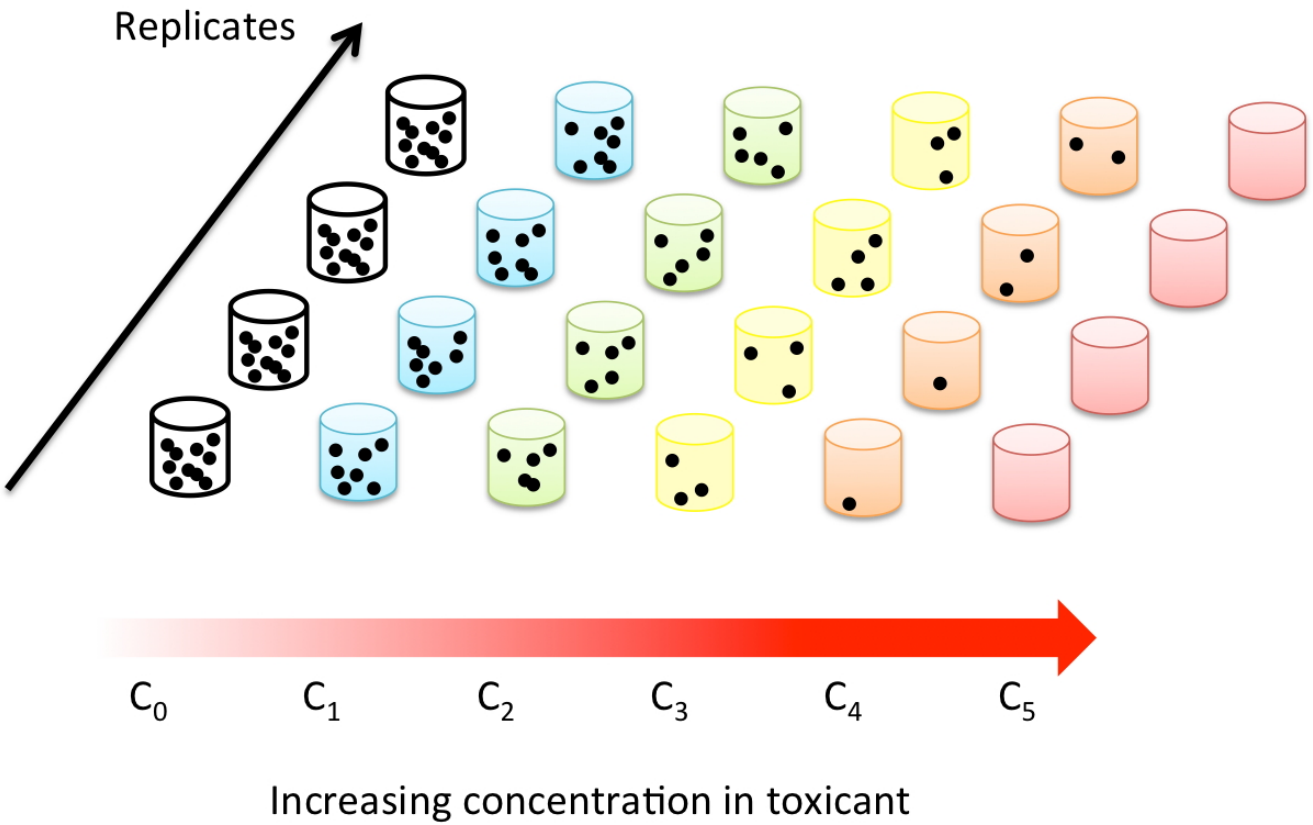
A variety of experimental devices



[Caquet et al., 1996]

Standard experimental design

Under standardized protocols, individuals are counted **over time**, that is at regular time points.
Endpoints can be survival, growth and/or reproduction for example.



Example of a toxicity test

Daphnia magna, acute immobilisation test (OECD 202, 1984) and chronic reproduction test (OECD 211, 2012)

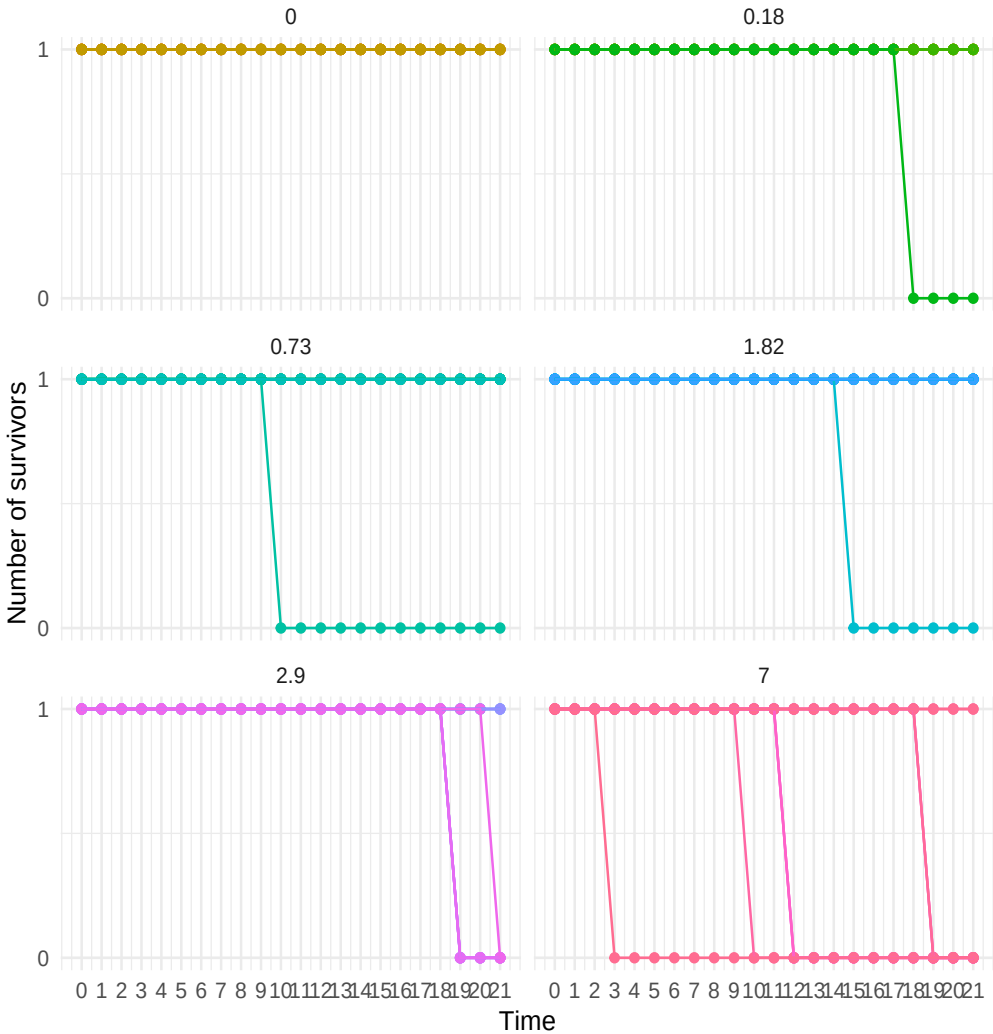
Daphnia magna

Acute test: the number of immobile daphnids is determined for each concentration at 24 and 48 hours.

Chronic test: offsprings are daily counted during 21 days.

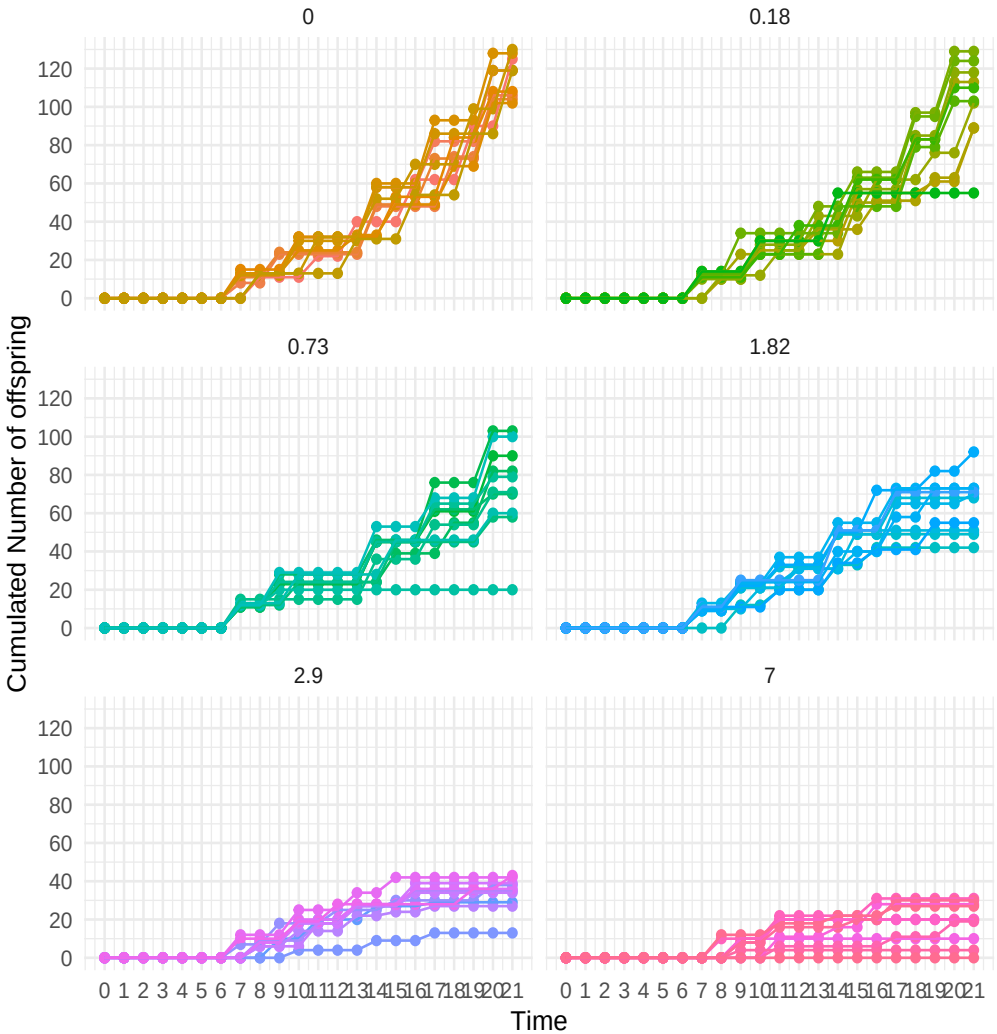
Example of survival data

Effect of chlordane on *D. magna* survival during 21 days (10 replicates of 1 individual):



Example of reproduction data

Effect of chlordane on *D. magna* reproduction during 21 days (10 replicates of 1 individual):



NOEC/LOEC: severely criticized for multiple reasons

Shortcomings

- necessarily one of the tested concentrations (hence strongly dependent on the experimental design);
- based on a wrong interpretation of the p-value (*absence of evidence is not evidence of absence*);
- strongly dependent on the sample size
  $\rightarrow$ unprotectrive with small sample sizes: the lower the sample size, the higher the NOEC;
- cannot always be determined (e.g., if the first concentration leads to a significant difference);
- no uncertainty limits are associated.

x% effective or lethal concentrations (EC_x/LC_x)

Alternative to the NOEC, EC_x/LC_x are now **strongly recommended**.
$\rightarrow$ obtained by fitting a dose-response model to toxicity test data at a chosen target time point, then deriving the dose which corresponds to a given effect level (usually x =10, 20 or 50%).

Advantages

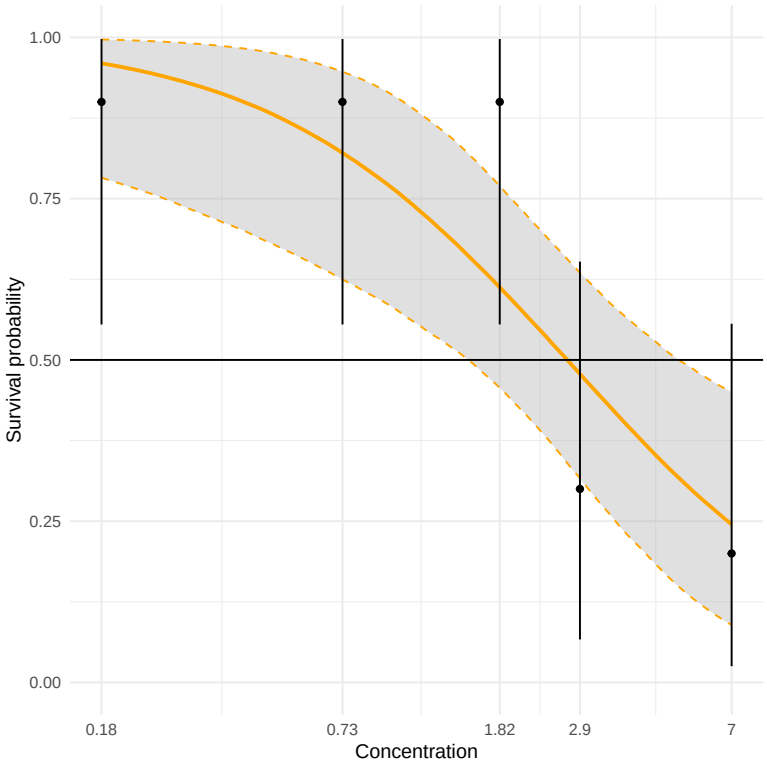
- capture and account for the whole dose-response curve;
- slightly dependent on the experimental design;
- may be associated to uncertainty limits.

Shortcomings

- sometimes technical difficulties when fitting;
- choice of a model;
- choice of an effect level x;
- choice of the exposure duration.

Example of LC_x estimation

Use of survival data at the end of the experiment (day 21)



	Median	2.5%	97.5%
LC_5	0.22	0.0074	0.71
LC_{10}	0.41	0.033	1.04
LC_{20}	0.82	0.16	1.6
LC_{50}	**2.67**	**1.50**	**5.3**

Table of content

Introduction

Dose-response modelling

Parameter inference

Environnemental Risk Assessment - The MOSAIC platform

(Dose or Concentration)-(response or effect) relationships?

A few vocabulary:

- **Dose** refers to the internal concentration, *i.e.*, the amount of toxicant **within** the body of organisms. But in ecotoxicology, only the **exposure concentration** is usually known.
  $\rightarrow$ We rather speak about **concentration**-response or effect relationships.
- Concentration-**response** relationships refer to the link between the exposure concentration and the proportion of individuals responding with an all-or-none effect.
- Concentration-**effect** relationships refer to the link between the exposure concentration and the magnitude of the induced biological change, measured in appropriate units.

What is a concentration-response/effect relationship?

A **concentration-response/effect relationship** is a simple X-Y graph relating increasing levels of exposure (X) to the response/effect (Y) **at a certain exposure time**.

Examples of **responses**:

- **Quantal** (or binary) data, expressed as proportion or probability (*e.g.*, mortality or immobilization).

Examples of **effects**:

- Ordered descriptive categories (*e.g.*, severity of a lesion);
- **Counts** (or discrete) (*e.g.*, reproduction products like eggs or clutches);
- **Continuous** measurements (*e.g.*, body size).

What is a regression model to be fitted on data?

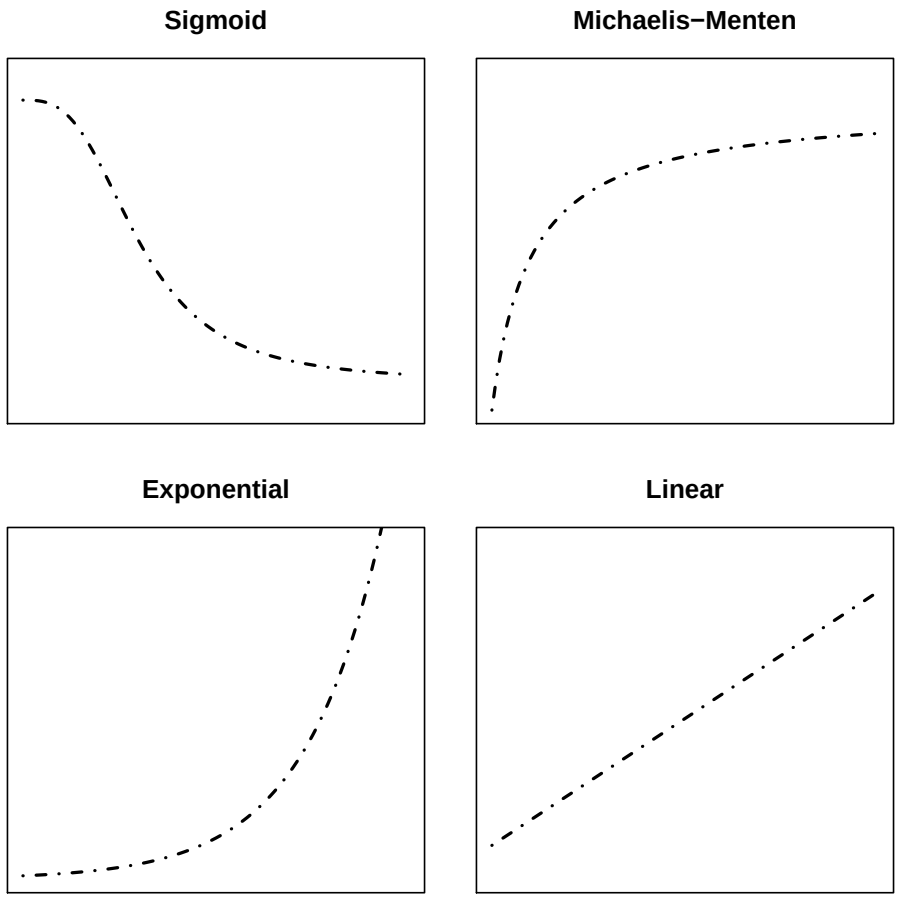
From concentration-response/effect experiments, if there is a reasonable number of concentrations (usually ≥ 5) of the toxicant and a reasonably well-behaved response/effect, it is straightforward to **fit a regression model**.

A regression model relating a dependent variable Y (the response or the effect) to an explanatory variable X (the concentration) is composed of two parts:

1. a **deterministic part**, which describes the mean value (or curve) (*e.g.*, a log-logistic model);
2. a **stochastic part**, which represents the distribution around the mean curve (*e.g.*, a normal distribution).

$\rightarrow$ Each part depends on the nature of data to analyze.

Four shapes to describe dynamic in life science

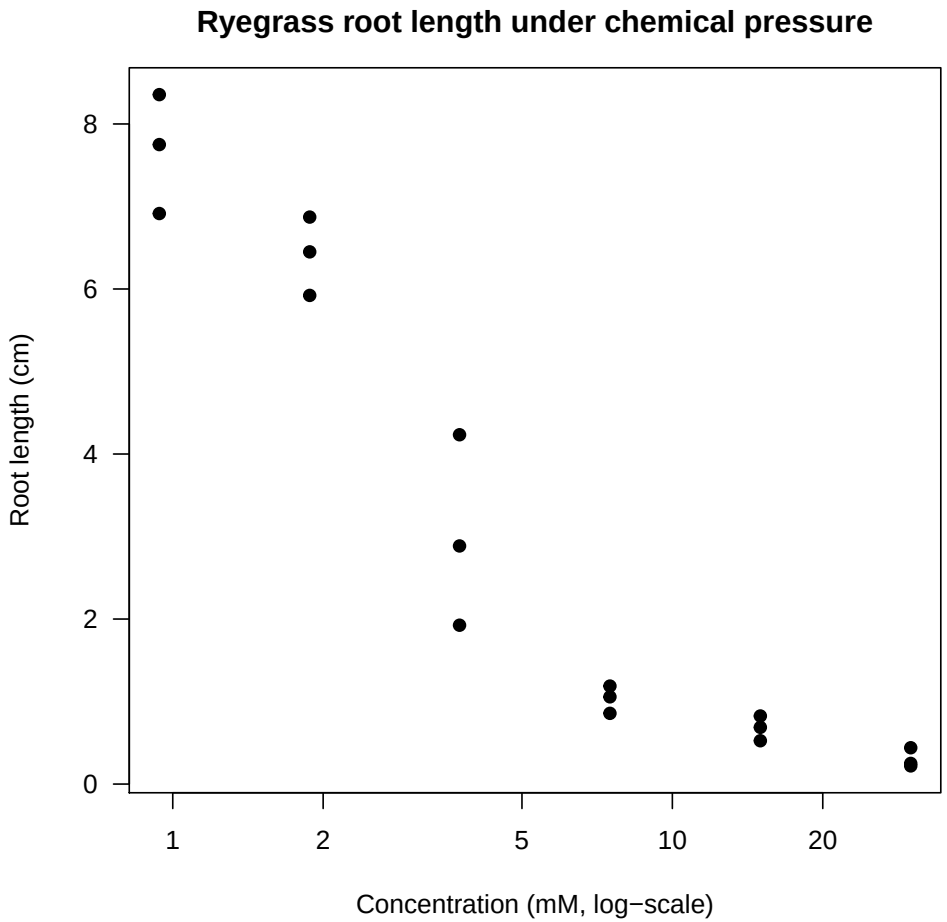


From http://bioassay.dk/bioassay/.

Example of **sigmoidal** relationship

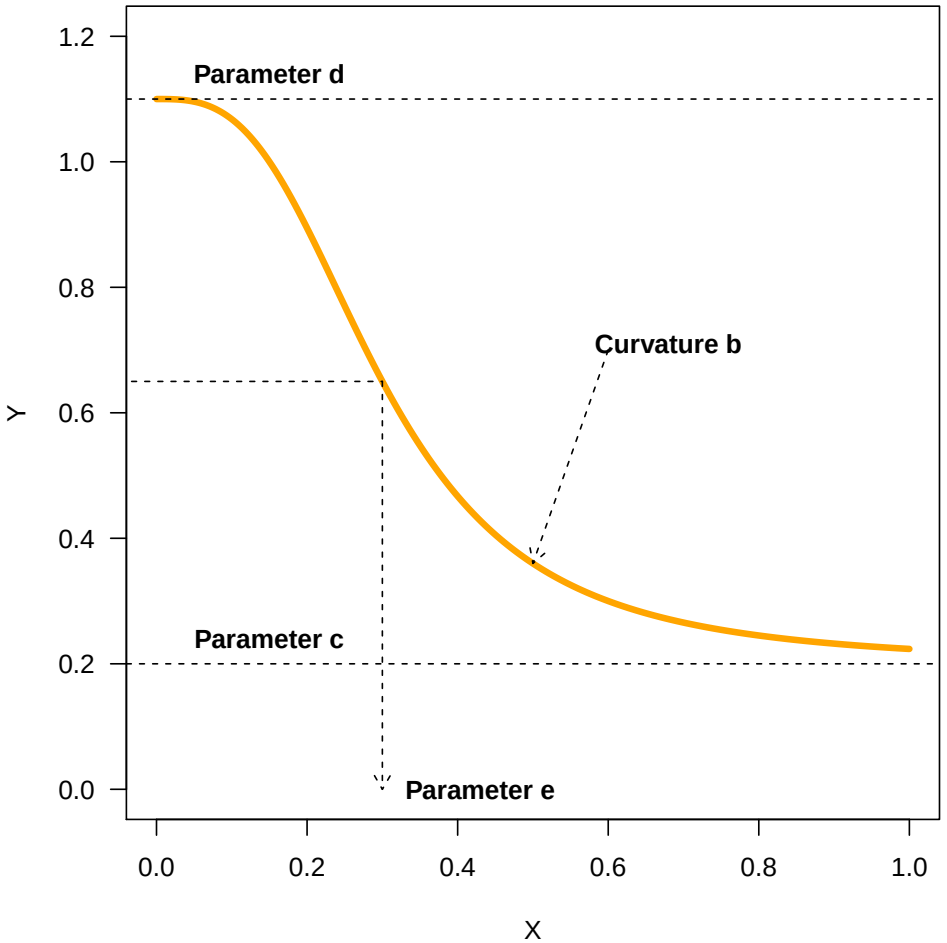
Data from a single dose-effect relationship between root lengths of perennial ryegrass (*Lolium perenne* L.) and concentration of ferulic acid.

$(n = 24)$



The log-logistic model - Graph

$b = 3, c = 0.2, d = 1.1$ and $e = 0.3$



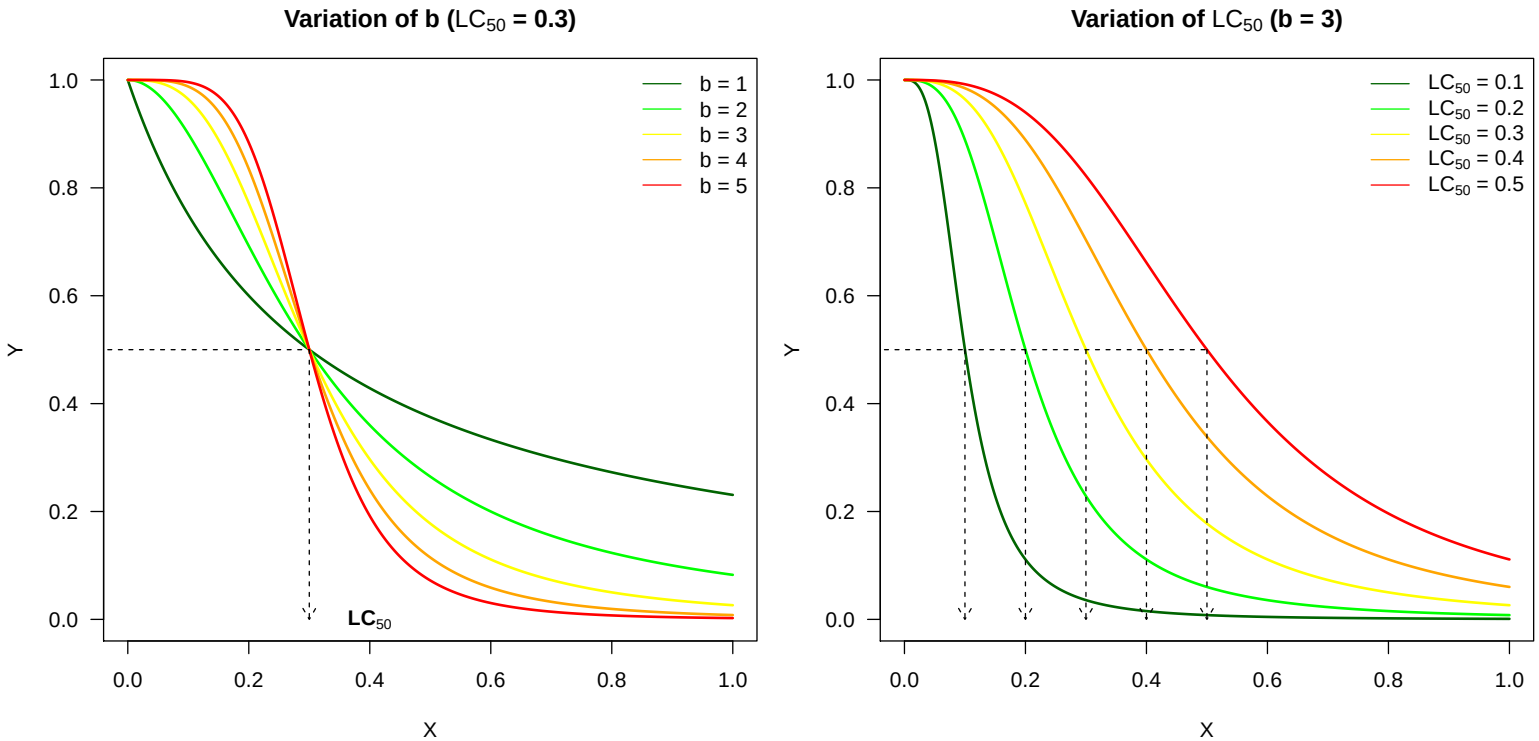
$$Y = c + \frac{d - c}{1 + (\frac{X}{e})^b}$$

$e = EC_{50}$ or LC_{50}
(here, in arbitrary unit)

For **survival**, d is the natural mortality (may be fixed to 1) and c is fixed to 0.

The log-logistic model - Morphology

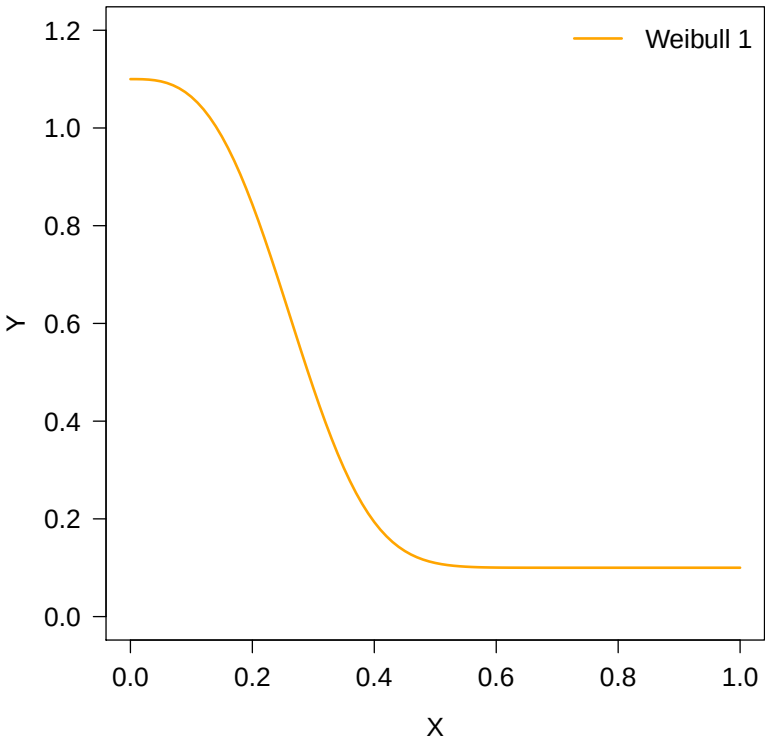
Case of survival, with $c = 0$ and $d = 1$: $e = LC_{50}$.

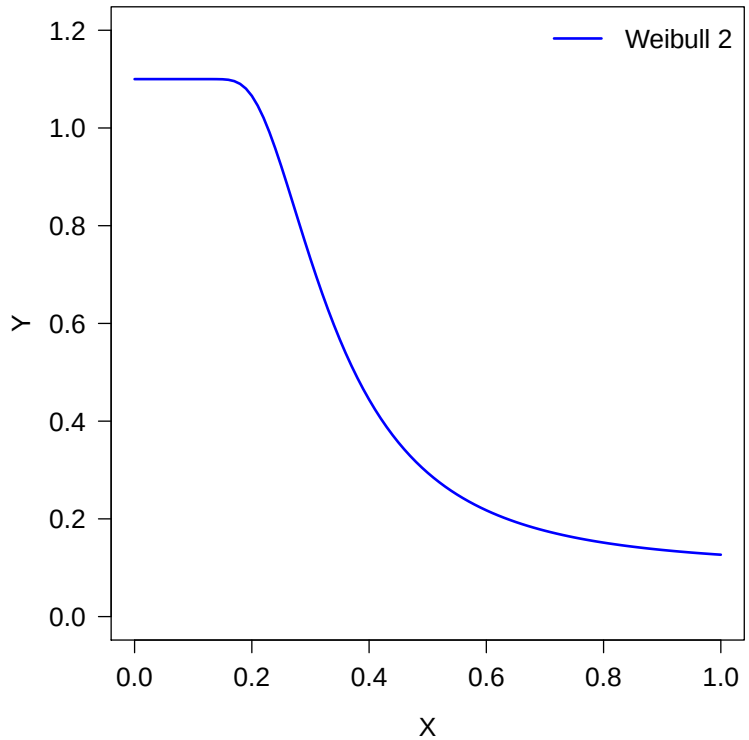


Other sigmoidal models

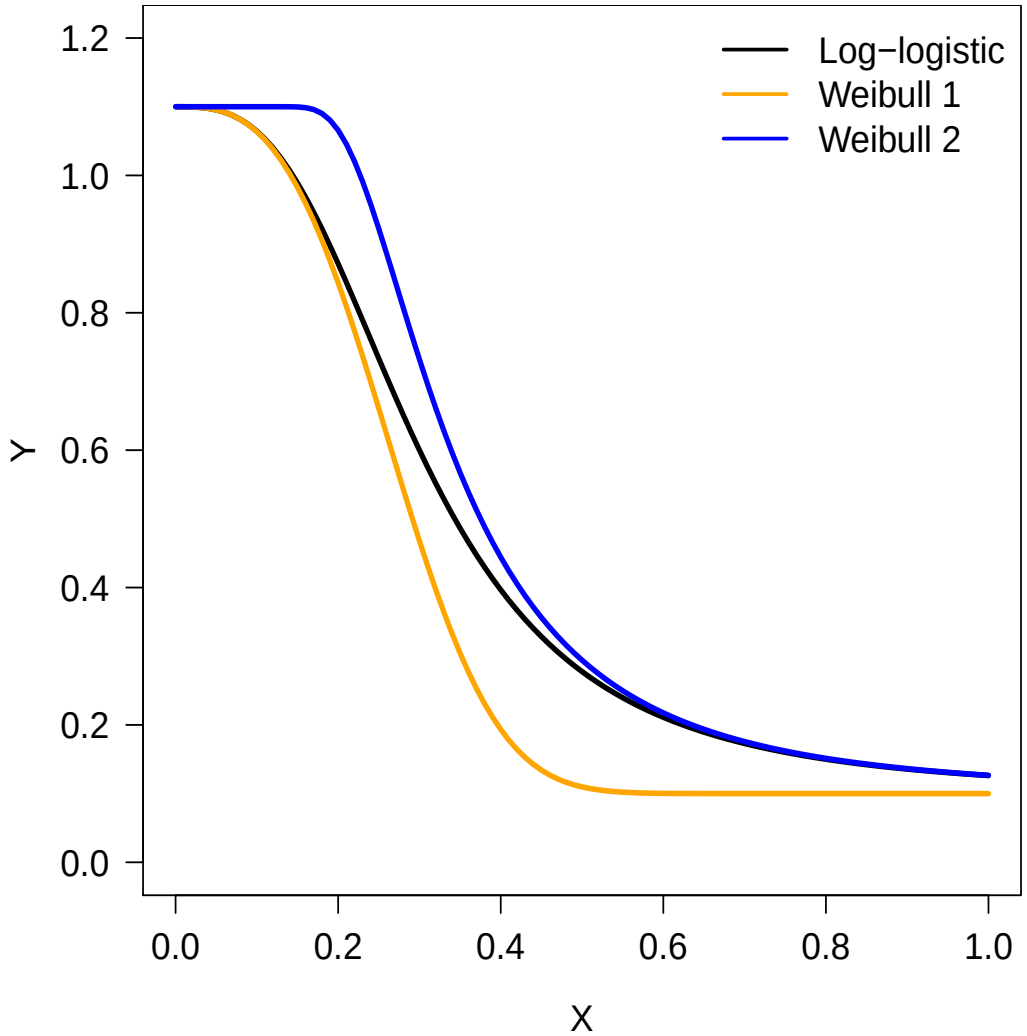
Many other models exist to describe sigmoidal shapes of dose-response curves as for example the Weibull's models:

$$Y = c + (d - c)e^{-(\frac{x}{e})^b} \quad \text{or} \quad Y = c + (d - c)(1 - e^{-(\frac{x}{e})^b})$$





Comparison of sigmoidal curves



About the stochastic part

Its role

Modelling of the variability **around** the mean tendency of the data
$\rightarrow$ requires the choice of the appropriate probability distribution

- **Quantal (or binary) data**: use of a binomial distribution.

$$Y \sim \mathcal{B}\left(p(X, \theta), n\right)$$

- **Count (or discrete) data**: use of a Poisson distribution

$$Y \sim \mathcal{P}(\lambda) \quad \text{with} \quad \lambda = f(X, \theta)$$

- **Continuous data**: use of a normal distribution

$$Y \sim \mathcal{N}(f(X, \theta), \sigma)$$

In brief: a wide variety of models

1. A **deterministic part**: linear or non-linear,
   and its associated parameters:
   for example, (α, β) or (b, c, d, e);

2. A **stochastic part**: Gaussian or not,
   and its associated parameters:
   for example, p (binomial), λ (Poisson) or σ (normal).

A battery of regression types

- **Gaussian linear regression**: simple linear regression, polynomial regression, multiple regression,...;
- **Generalized linear regression**: logistic regression, Poisson regression, multinomial logit regression, probit regression,...;
- **Gaussian non-linear regression**: least-square regression, simple or multiple;
- **Generalized non-linear regression**.

Table of content

Introduction

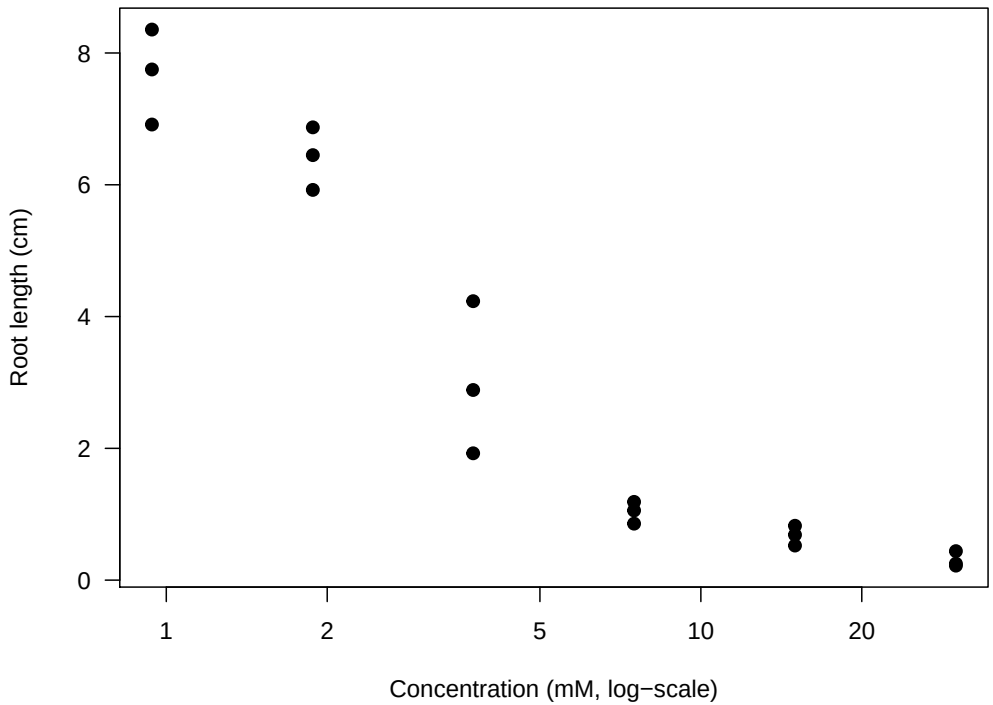
Dose-response modelling

Parameter inference

Environnemental Risk Assessment - The MOSAIC platform

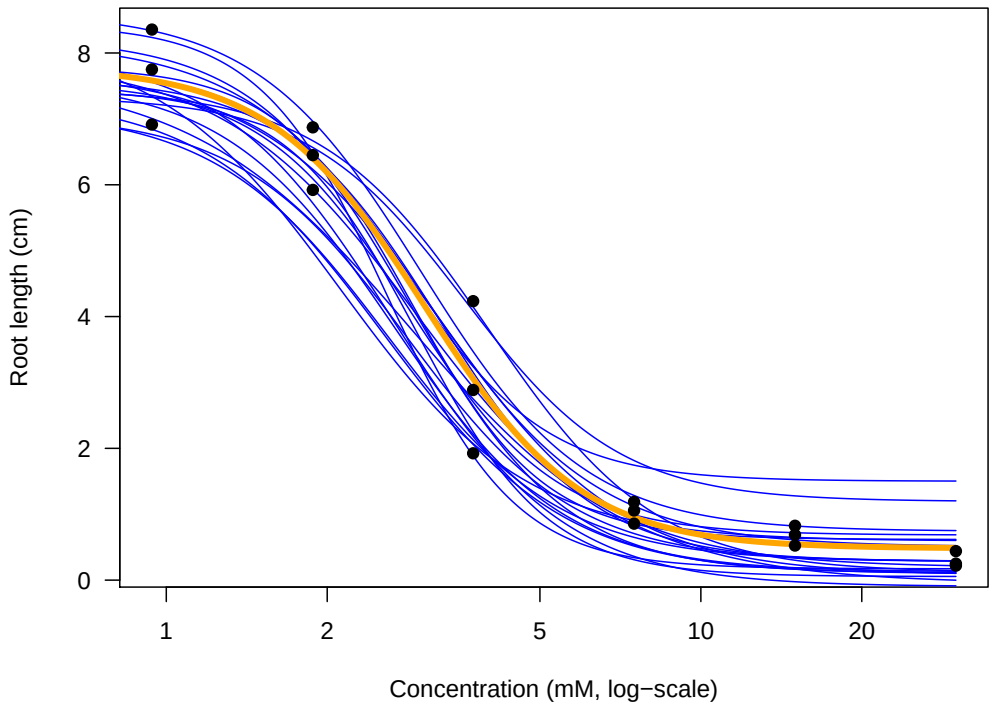
What is inference ?

Inference implies the use of both observed data and a model.



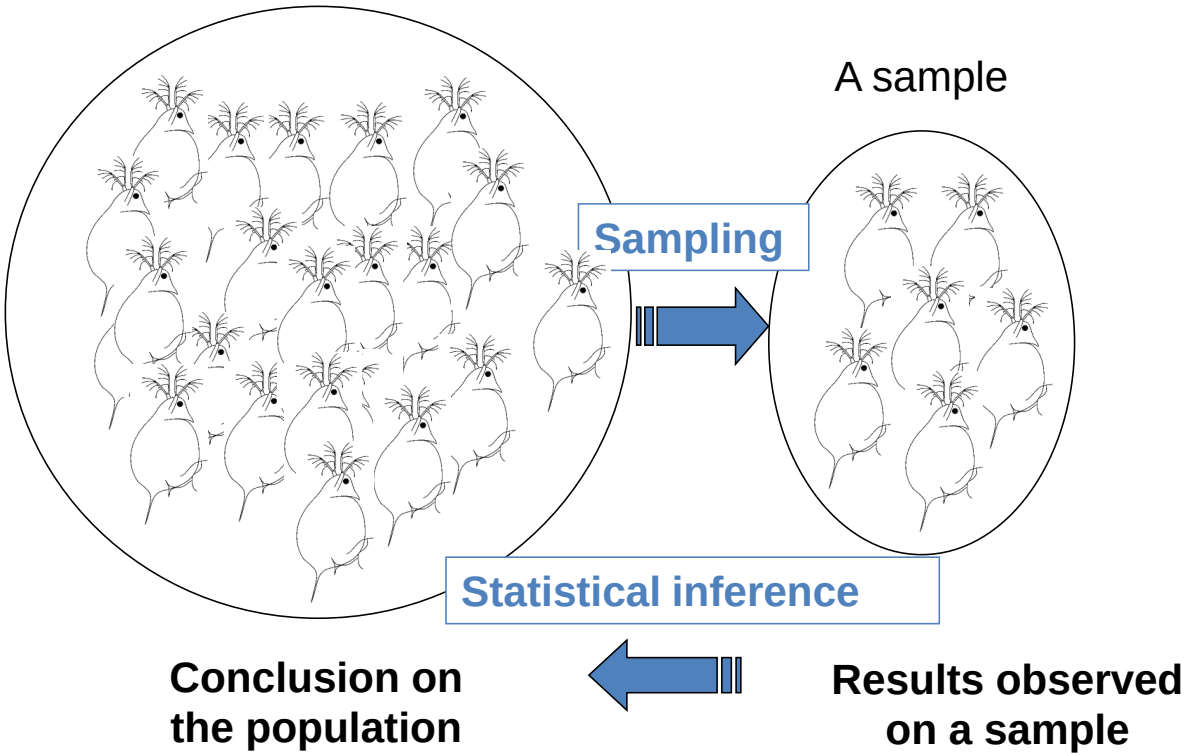
What is inference ≡ Get parameter estimates

Several criteria may provide the best fit parameter values.



Inference: generalization to population

Inference necessarily implies **generalization** from a sample to population, and the calculation of **uncertainty** in the estimated parameters, especially uncertainty due to the sampling error.



Inference: how to perform?

Two main ways of practicing:

The **frequentist** framework

Based on the principle of maximizing the probability of the data given the model, namely the **likelihood** $P(Y|\theta)$.

The **Bayesian** framework

Based on the principle of maximizing the probability of the model given the data, namely the **joint posterior distribution** $P(\theta|Y)$, combining both the likelihood and **prior information** available on parameters **in advance**.

Frequentist framework

- **Parameter θ is supposed fixed but unknown**;
- Parameter inference only uses observed data;
- Confidence intervals are based on repeated sampling from the model, the probability being associated to the relative occurrence frequency of an outcome.

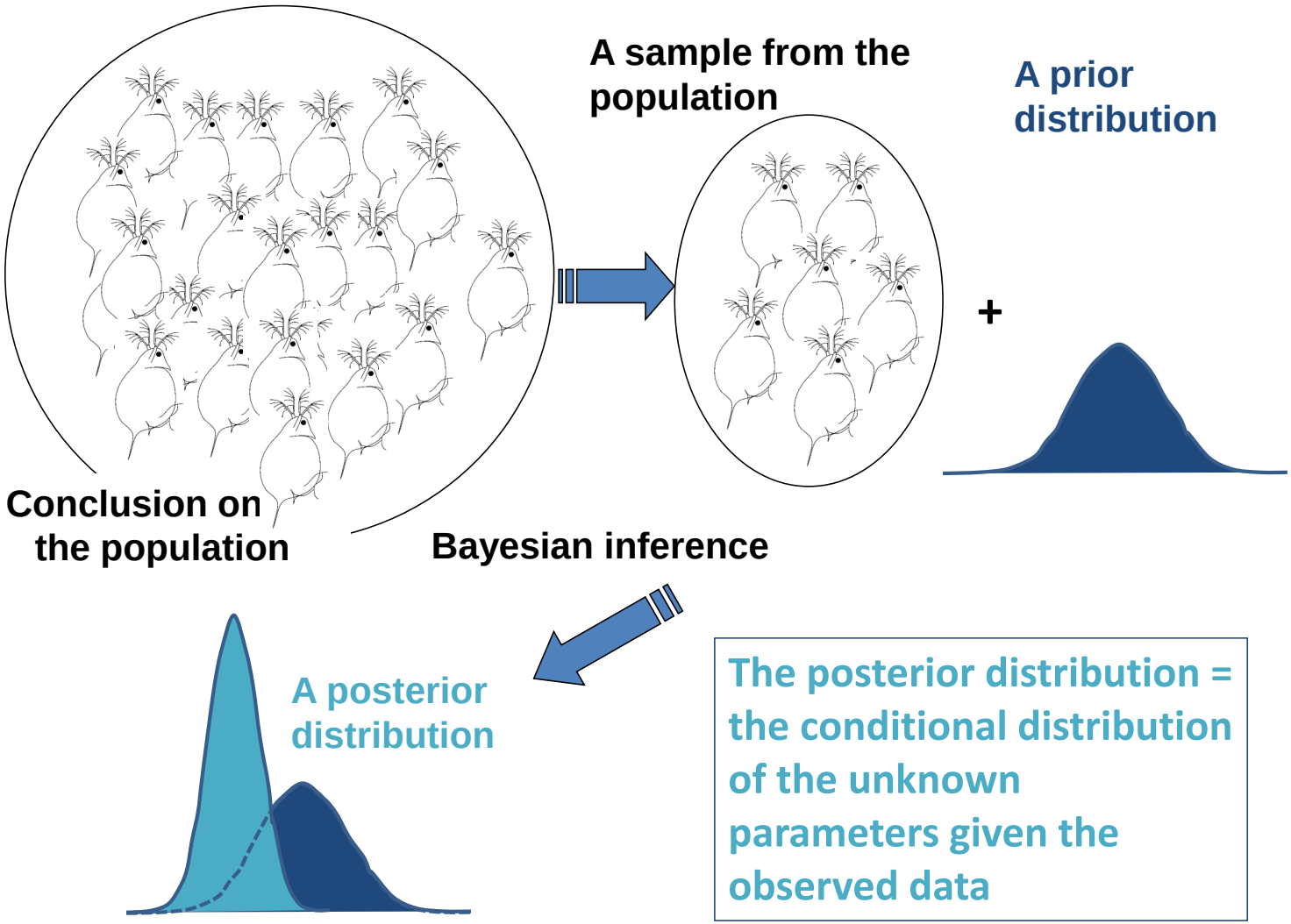
Frequentist framework

- **Parameter θ is supposed fixed but unknown**;
- Parameter inference only uses observed data;
- Confidence intervals are based on repeated sampling from the model, the probability being associated to the relative occurrence frequency of an outcome.

Bayesian framework

- **Parameter θ is considered as a random variable, associated to a probability distribution**;
- Parameter inference uses both observed data and prior information (prior distribution);
- Credible intervals are defined from the posterior distribution and can be easily interpreted: **95% is the probability that the true parameter value lies within its 95% credible interval**.

The Bayesian framework in pictures



Advantages of the Bayesian framework

Use of the posterior distribution for parameter estimation

- **Point estimate:**
  Mean, median or mode of the posterior distribution
- **Interval estimate:**
  Definition of a **credible interval** (or Bayesian confidence interval) from posterior distribution quantiles:
  $\rightarrow 2.5\%$ and 97.5% quantiles for a 95% credible interval.
  Easy interpretation: the probability that the parameter lies in a 95% credible interval is 95%.
- There is **no more need** of hypothesis tests, nor of p-value calculation: one can make decisions directly from the posterior distribution.

Inference: tools for practice

Under the **frequentist** framework

TOX Rat®

Excel macro : REGTOX

Under the **Bayesian** framework

Table of content

Introduction

Dose-response modelling

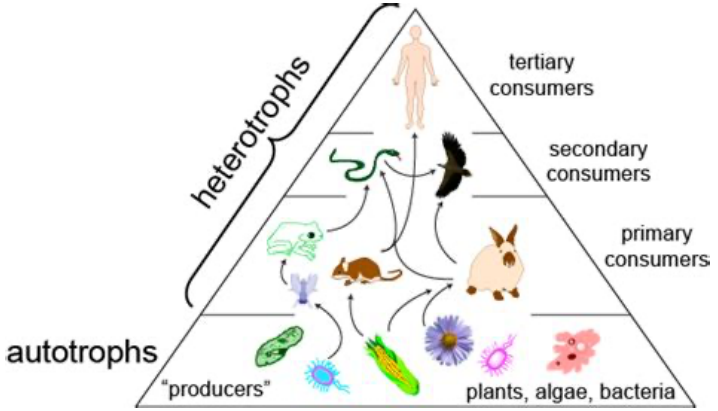
Parameter inference

Environnemental Risk Assessment - The MOSAIC platform

Environmental Risk Assessment (ERA)

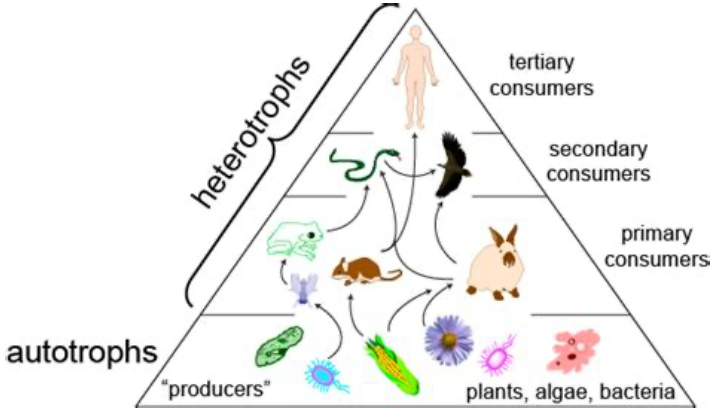
Environmental risk assesment

An evidence



Environmental risk assesment

An evidence

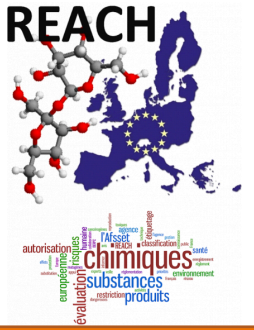


A priority

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
DE LA TRANSITION
ÉCOLOGIQUE
ET SOLIDAIRE

EUR-Lex



Environmental risk assessment

All concerned

One substance
=
One registration

ECHA
EUROPEAN CHEMICALS AGENCY

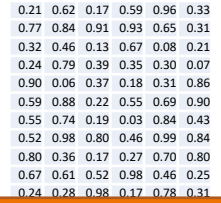
Environmental risk assessment

All concerned

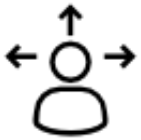
One substance
=
One registration

ECHA
EUROPEAN CHEMICALS AGENCY

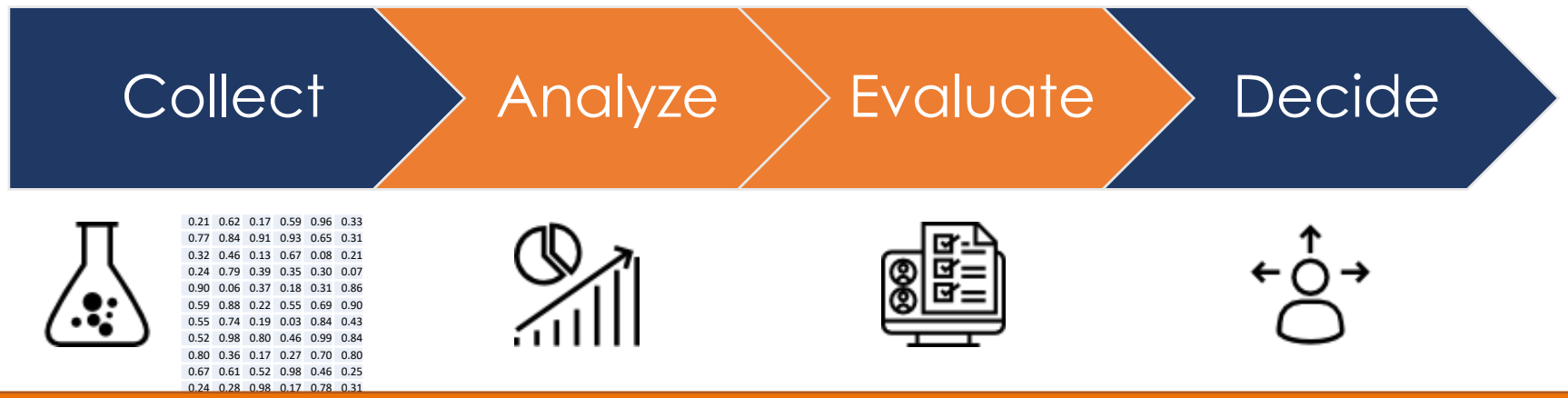
Collect → Analyze → Evaluate → Decide



0.21	0.62	0.17	0.59	0.96	0.33
0.77	0.84	0.91	0.93	0.65	0.31
0.32	0.46	0.13	0.67	0.08	0.21
0.24	0.79	0.39	0.35	0.30	0.07
0.90	0.06	0.37	0.18	0.31	0.86
0.59	0.88	0.22	0.55	0.69	0.90
0.55	0.74	0.19	0.03	0.84	0.43
0.52	0.98	0.80	0.46	0.99	0.84
0.80	0.36	0.17	0.27	0.70	0.80
0.67	0.61	0.52	0.98	0.46	0.25
0.24	0.28	0.98	0.17	0.78	0.31



Environmental risk assessment



MOSAIC, the solution

OECD
BETTER POLICIES FOR BETTER LIVES

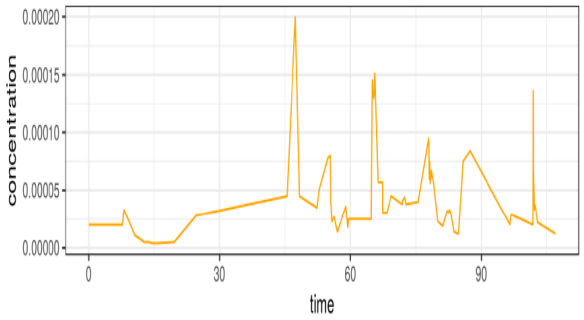
ISO

https://mosaic.univ-lyon1.fr

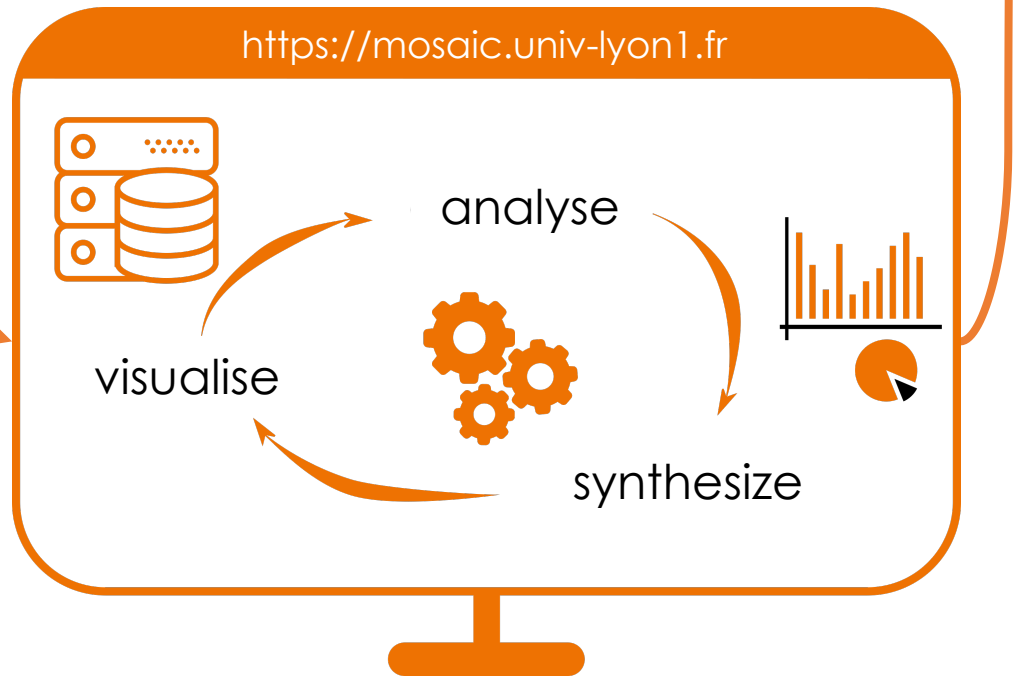
analyse

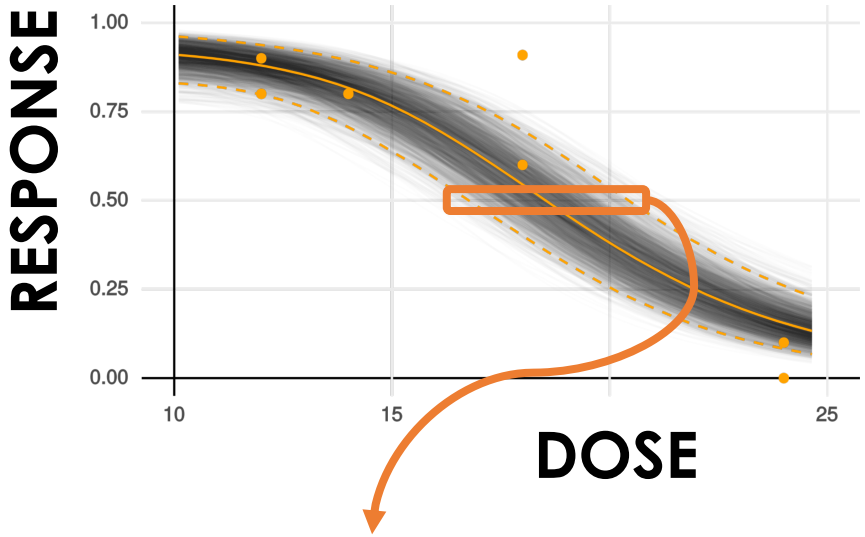
visualise

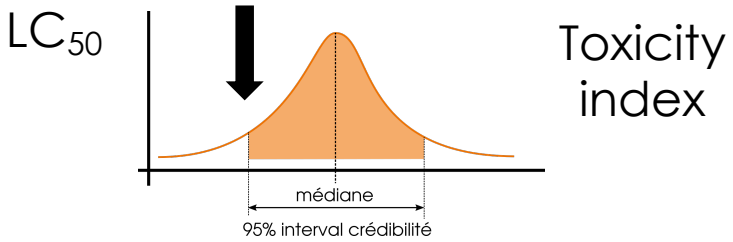
synthesize



MOSAIC, the solution







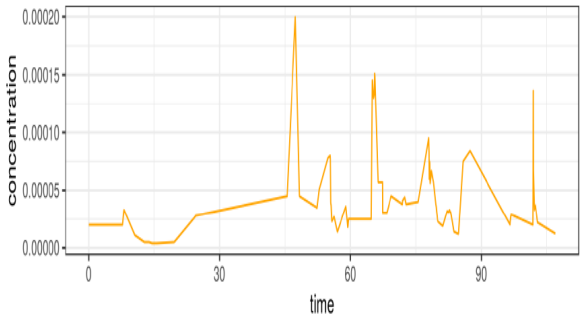
MOSAIC, the solution

https://mosaic.univ-lyon1.fr

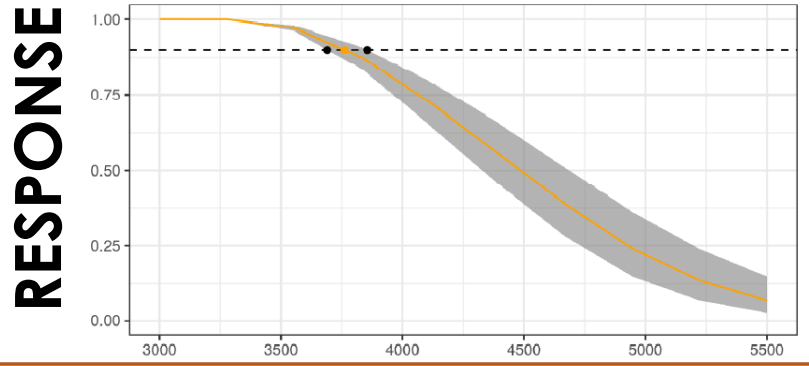
analyse

visualise

synthesize

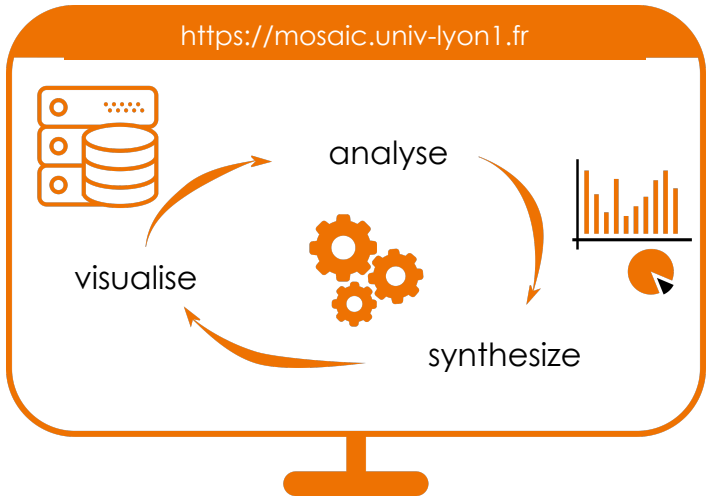


TKTD modeling



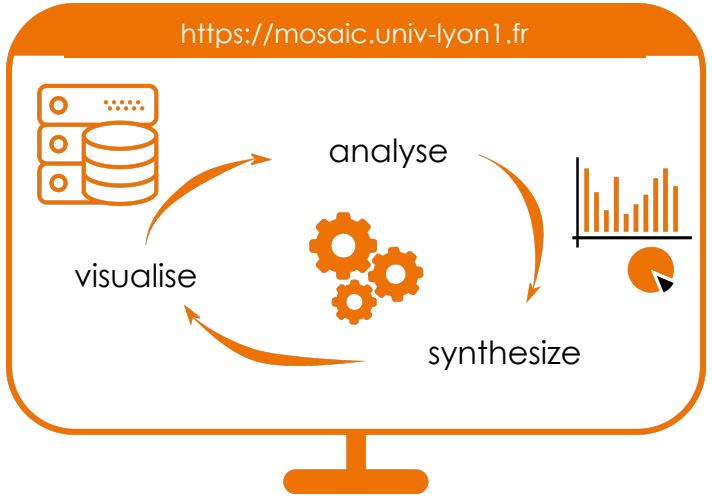
MOSAIC, its strengths

- Transparency and reproducibility



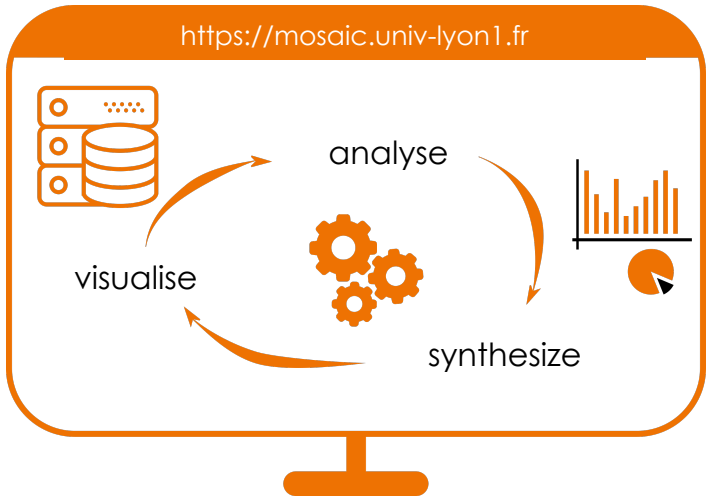
MOSAIC, its strengths

- Transparency and reproducibility



- Bayesian statistics

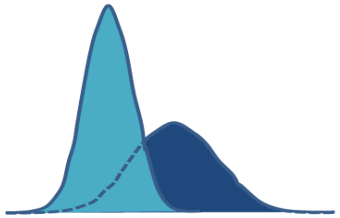
MOSAIC, its strengths



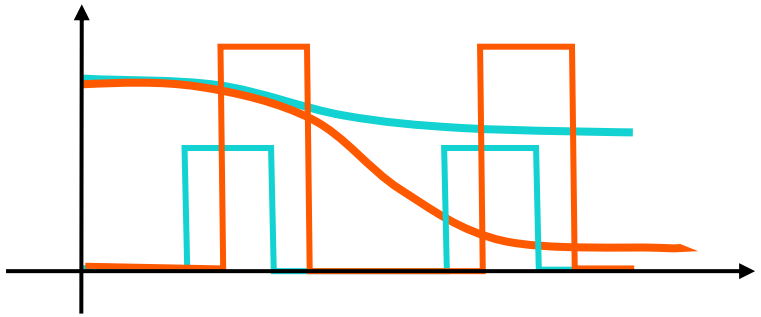
- Transparency and reproducibility

GLP

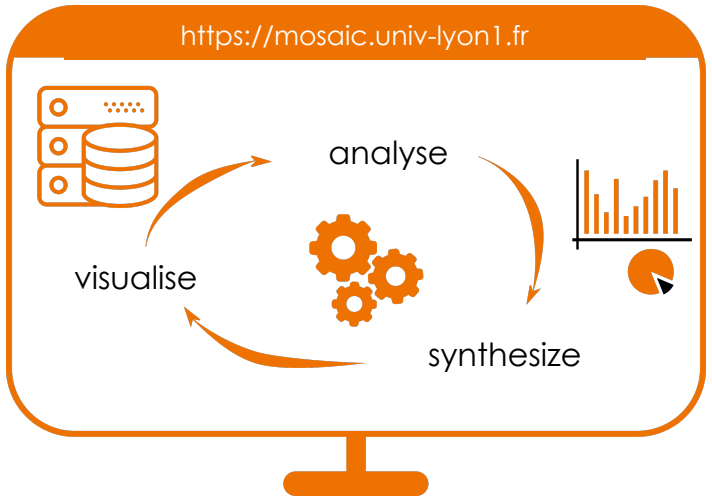
- Mechanistic modeling approach
- Bayesian statistics



- Predict to better prevent



MOSAIC, a first prescriber



OECD
BETTER POLICIES FOR BETTER LIVES

243
Adopted:
29 July 2016

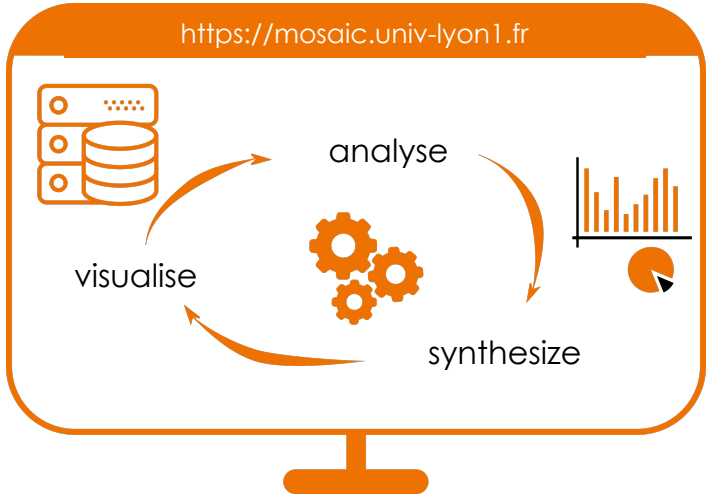
OECD GUIDELINE FOR THE TESTING OF CHEMICALS

EC_x

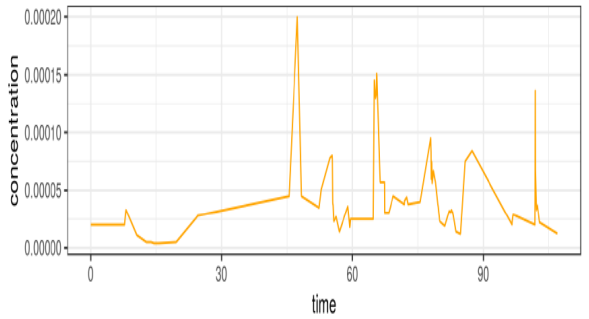
Lymnaea stagnalis Reproduction Test

52. EC_x values, including their associated lower and upper credible/confidence limits, are estimated using any appropriate statistical method based on a regression analysis of the number of clutches (or eggs) per individual-day. Even if any statistical software can be used for regression analysis (3), the user-friendly web-platform MOSAIC_repro, freely available at http://pbil.univ-lyon1.fr/software/mosaic/reproduction/, is recommended because the procedures implemented within this software were developed during the validation process of the *L. stagnalis* Reproduction Test (see details in ANNEX 7).

MOSAIC, in support of regulation



efsa
European Food Safety Authority



→ TKTD modeling

modeling and statistical tools for ecotoxicology

http://mosaic.univ-lyon1.fr

financial Supports institutions

MOSAIC

For who? For what?

MOSAIC is a turnkey decision-making tool for ecotoxicologists, regulators and industrials.

Without the need to immerse into extensive mathematical and statistical technicalities, users are given advanced and **innovative methods** for a valuable quantitative environmental risk assessment.

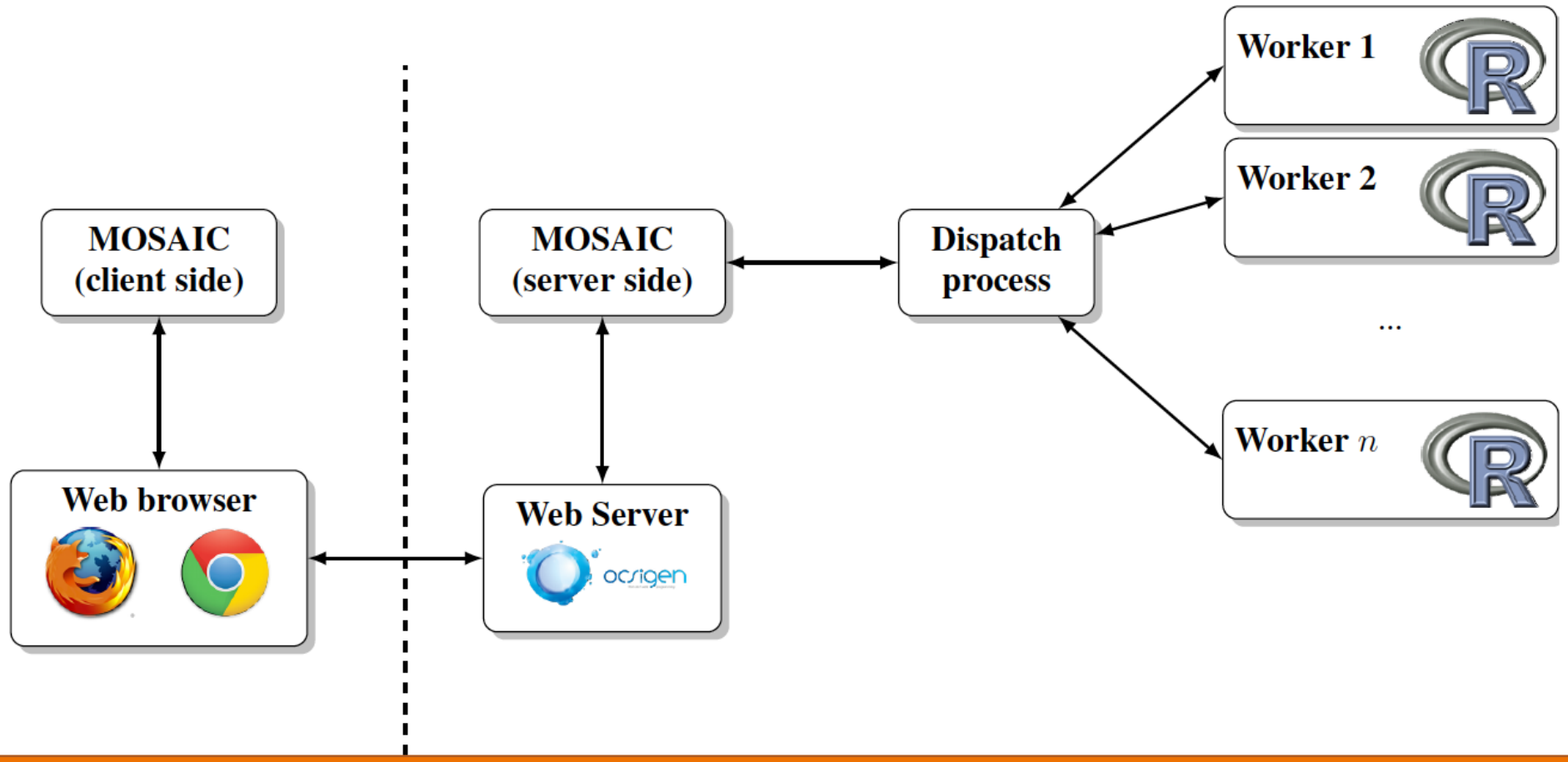
Contact

http://pbil.univ-lyon1.fr/software/mosaic/
mosaic@univ-lyon1.fr

http://goo.gl/Y2UOMp

MOSAIC

The dark side



Contact

http://pbil.univ-lyon1.fr/software/mosaic/
mosaic@univ-lyon1.fr

http://goo.gl/Y2UOMp

MOSAIC

Two modules for survival

MOSAIC surv

Classical dose-response analysis of bioassay **survival** data, with descriptive summaries of the data and **x% lethal concentrations** (LCx) estimates under a **Bayesian framework**.

MOSAIC GUTS

Toxicokinetic-toxicodynamic (TKTD) analysis of survival data, fitted with a General Unified Threshold model of Survival (GUTS) model to estimate **threshold concentrations and x% lethal concentrations** (LCx) under a **Bayesian framework**.

MOSAIC

Two modules
for reproduction and SSD

MOSAIC repro

Classical dose-response analysis of bioassay **reproduction** data, in addition with descriptive summaries of the data and estimates of **x% effective concentrations** (ECx) under a **Bayesian framework**.

MOSAIC SSD

Species Sensitivity Distribution fitted to estimate **hazardous concentration for p%** (HCp) of the species.

Parameters of the probability distribution are estimated from toxicity thresholds under a **frequentist framework**.

MOSAIC

Two new modules
2020

MOSAIC growth

Classical dose-response analysis of bioassay data of **growth-type**, in addition with descriptive summaries of the data and estimates of **x% effective concentrations** (ECx) under a **Bayesian framework**.

MOSAIC bioacc

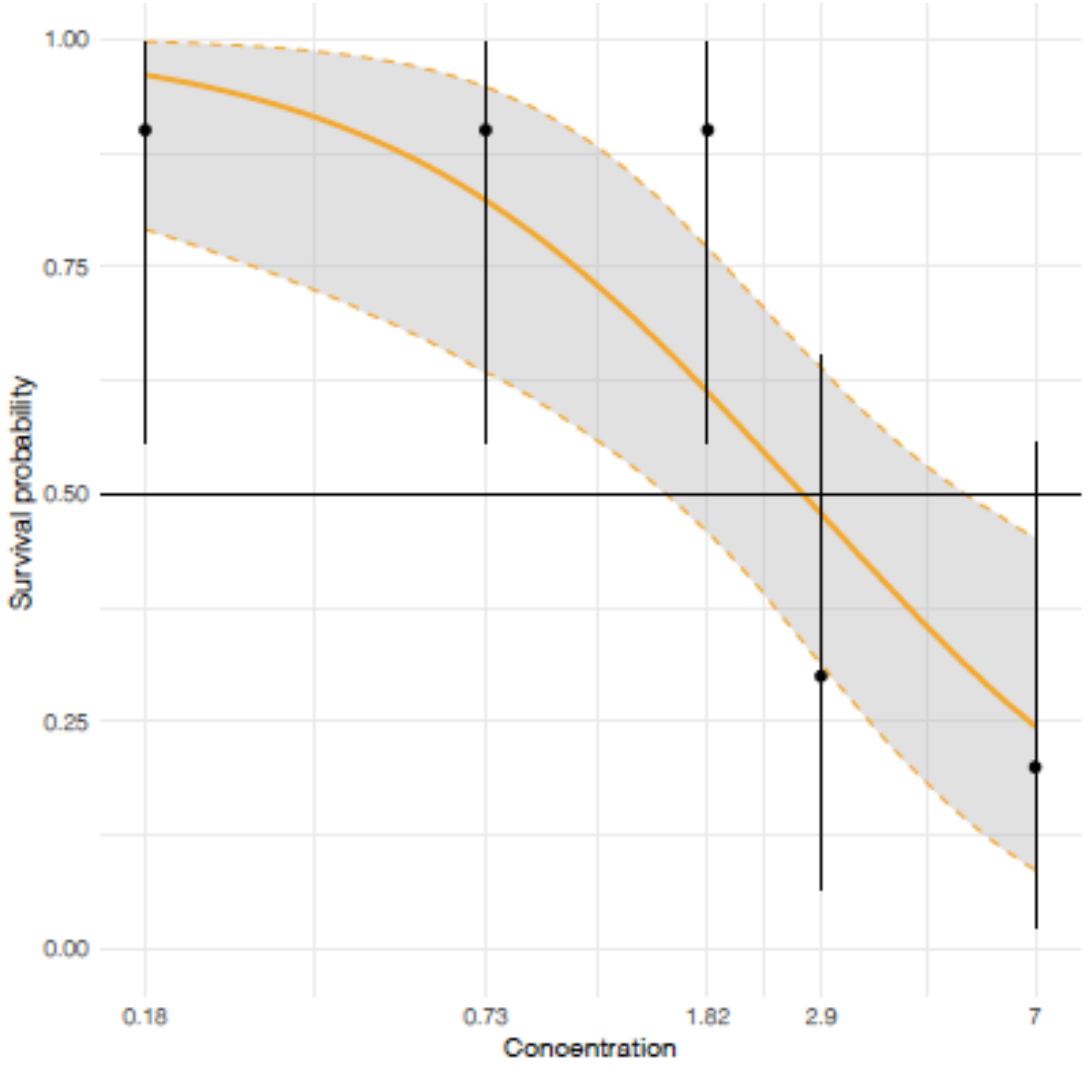
Provides **bioaccumulation factors** (BCF/BMF/BAF) from the fitting of a toxicokinetic (TK) model on accumulation-depuration data under a **Bayesian framework**. Fulfils all requirements of regulators when examining applications for market authorization of active substances.

MOSAIC$_{surv}$

- **Input**: data from standard bioassays where survival has been recorded through time at different concentrations. Data from different replicates of a same experimental condition are pooled.

- **Model**: a three parameters log-logistic model (concentration-exposure part) associated with a binomial stochastic part (binary data).

- **Output**: LCx values (x = 5, 10, 20, …, 80%) expressed as median estimates associated with their 95% credible intervals.



R script

The following script may be used to reproduce the above analysis, or as a starting point to produce custom graphs or large-scale analyses. It relies on a R package called MORSE which is released under the GPL license.

```
# This script was generated by MOSAIC, a web application dedicated to
# ecotoxicology. It is available at http://pbil.univ-lyon1.fr/software/mosaic/

# For any further question, please contact us at mosaic@univ-lyon1.fr

library(morse)

# Your input file 'chlordan' should be present in the current working directory
dat <- read.table(file='chlordan',sep='\t',header=T)

rdat <- reproData(dat)
plot(rdat, target.time = max(dat$time))
plot(as.survData(rdat), target.time = max(dat$time))
```

MOSAIC

Features

- Free usage
- User-friendly interface
- Data privacy
- Reproducible results
- Open methods

Contact

http://pbil.univ-lyon1.fr/software/mosaic/
mosaic@univ-lyon1.fr

http://goo.gl/Y2UOMp

MOSAIC

How to cite MOSAIC?

- Charles S, Veber P, Delignette-Muller ML. 2018. MOSAIC: a web-interface for statistical analyses in ecotoxicology. *Environ. Sci. Pollut. Res*. 25:11295–11302.

- Kon Kam King G, Veber P, Charles S, Delignette-Muller ML. 2014. MOSAIC_SSD: a new web tool for species sensitivity distribution to include censored data by maximum likelihood. *Environ. Toxicol. Chem.* 33:2133–9.

- Baudrot V, Veber P, Gence G, Charles S. 2018. Fit GUTS reduced models online: from theory to practice. *Integr. Environ. Assess. Manag.* 14:625–630.

Contact

http://pbil.univ-lyon1.fr/software/mosaic/

mosaic@univ-lyon1.fr

http://goo.gl/Y2UOMp

MOSAIC

Usage statistics (overview)

Showing 3351 entries

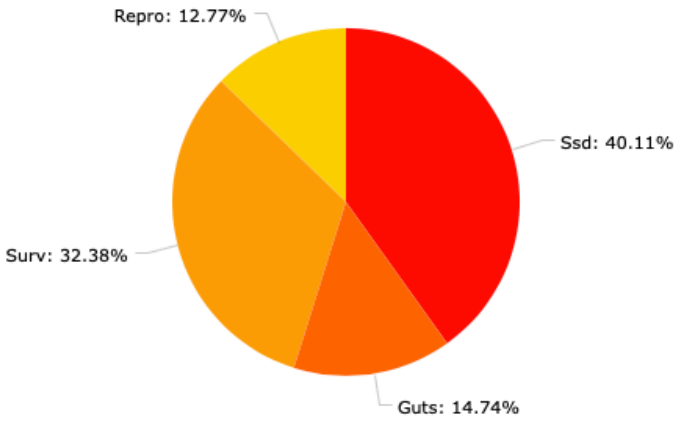
Unique users

553

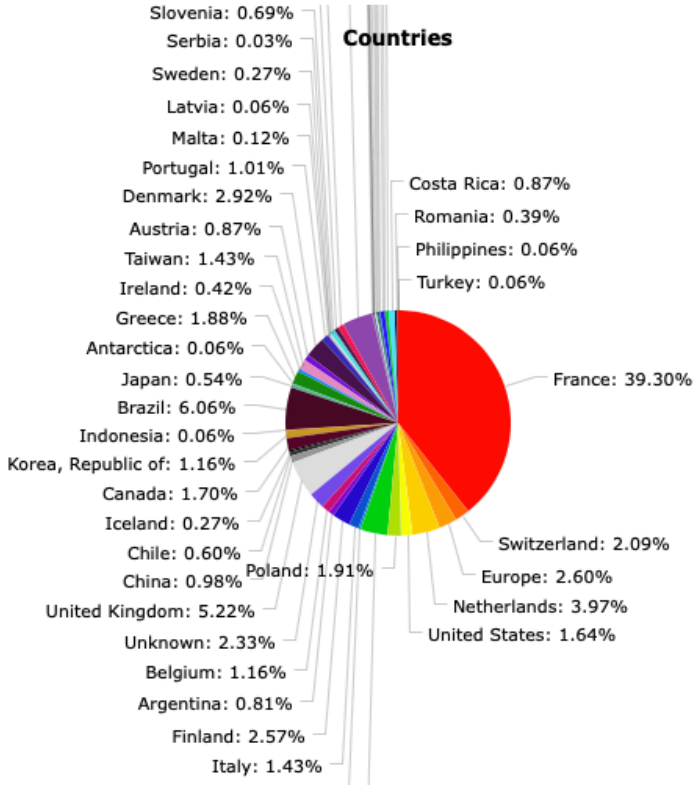
New users

553

Modules



Countries



MOSAIC

Usage statistics (last month)

Showing 271 entries

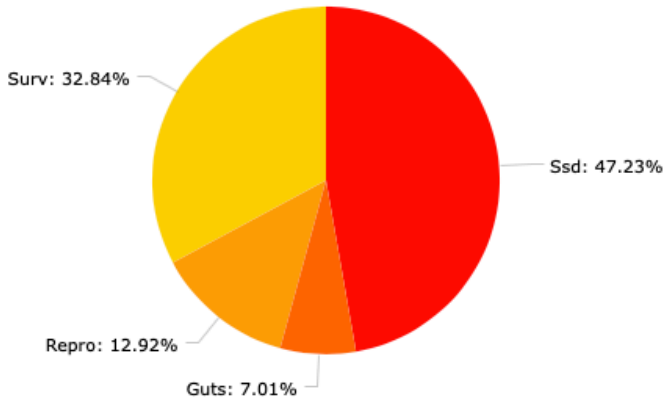
Unique users

78

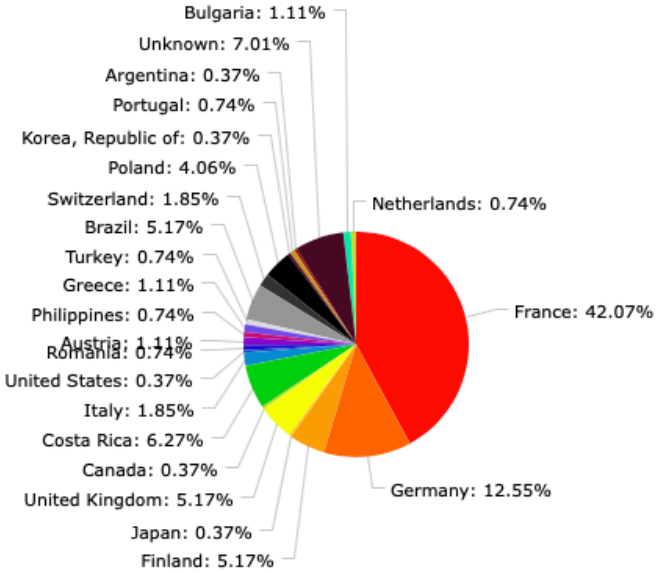
New users

78

Modules



Countries



http://mosaic.univ-lyon1.fr