

## MOSAIC : une plate-forme web pour l'analyse statistique des données d'écotoxicologie

Les effets des substances chimiques sur les organismes vivants se mesurent au cours du temps sur la survie, la reproduction et/ou la croissance individuelles, et le plus souvent selon des protocoles standardisés (normes ISO ou OCDE) pour garantir le contrôle des conditions de réalisation des bioessais ainsi que leur reproductibilité. Ces données expérimentales sont classiquement analysées, en fin d'expérience, avec un modèle concentration-réponse (CR) qui permet d'estimer une concentration létale ou effective à x% (LCx/ECx) avec son incertitude. Outre le fait que la dépendance au temps des données n'est pas pleinement exploitée, une approche CR suppose par ailleurs implicitement que la concentration d'exposition reste constante tout au long de l'expérience. Il est donc difficile d'extrapoler les résultats à des scénarios d'exposition réalistes, c'est-à-dire variables au cours du temps. C'est pourquoi il est maintenant recommandé d'utiliser des modèles toxicocinétiques-toxicodynamiques (TKTD) qui permettent d'intégrer la dynamique temporelle de l'exposition à l'estimation du risque [1].

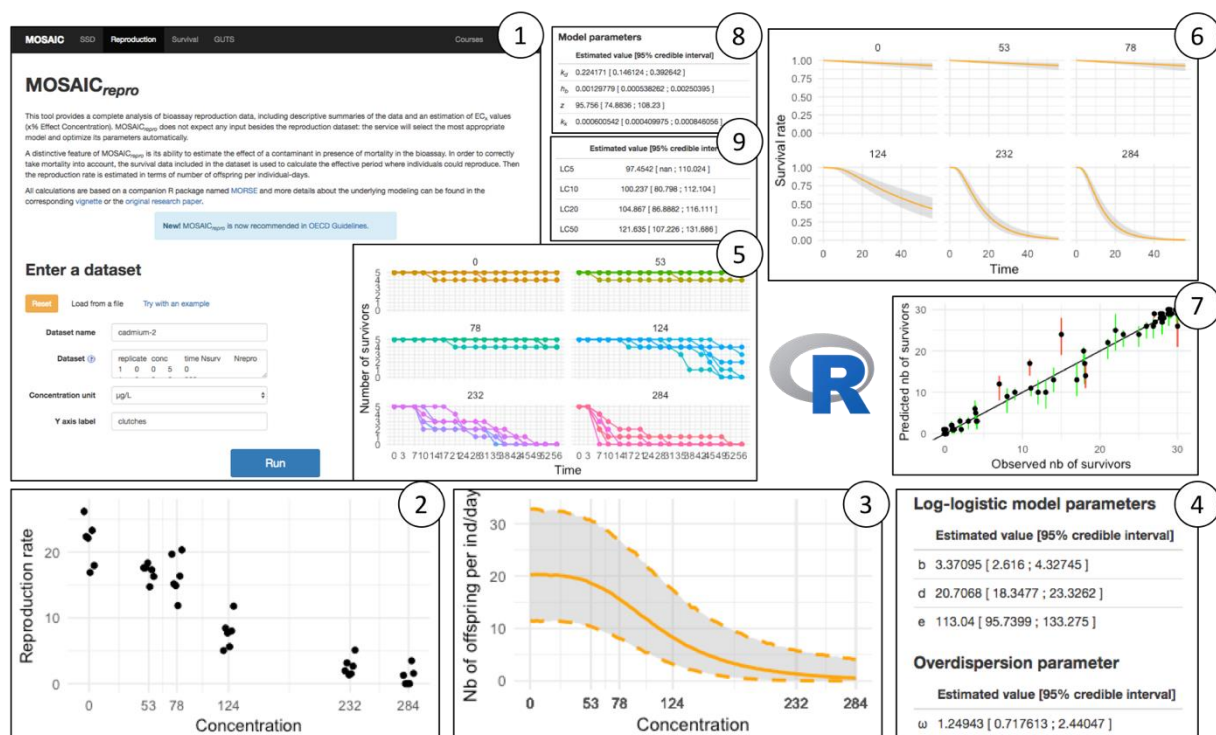
Que ce soit pour une analyse CR classique ou pour une analyse TKTD, la nature des données, qu'elles soient binaires (e.g. survie), de comptage (e.g., reproduction) ou continues (e.g. croissance), doit être prise en compte correctement d'un point de vue statistique. Ainsi, en plus du choix d'un modèle mathématique qui va décrire de manière appropriée la tendance moyenne des observations (e.g. modèle log-logistique pour une analyse CR, ou modèle GUTS pour une analyse TKTD de la survie), il convient également de choisir le bon modèle stochastique pour décrire la variabilité des données autour de la tendance moyenne (e.g., loi binomiale ou binomiale conditionnelle, loi de Poisson, loi normale). Enfin, estimer les paramètres du modèle choisi avec leur incertitude, et propager ces incertitudes à l'estimation des indicateurs de toxicité comme la LCx ou la ECx, nécessite de faire le choix d'une méthode d'inférence statistique suffisamment flexible et fiable.

Au quotidien, pour un écotoxicologue ou un évaluateur du risque, tous ces choix peuvent s'avérer ardu, de même que leur mise en œuvre logicielle. Beaucoup d'entre eux s'accordent d'ailleurs pour dire qu'il manque aujourd'hui cruellement d'un outil « clé en main » qui permette de réaliser ces différentes analyses (CR et TKTD) de manière conviviale et sans avoir à se préoccuper de la technicité statistique sous-jacente. C'est précisément ce qu'offre la plate-forme web MOSAIC, acronyme de *MOdeling and StAtistical tools for ecotoxiCology* : <http://pbil.univ-lyon1.fr/software/mosaic/> [2]. MOSAIC s'utilise librement à travers un navigateur Internet. Tous ses modules sont développés en langage OCaml (<https://ocaml.org/>) et sont hébergés sur un serveur Ocsigen (<http://ocsigen.org/>) localisé au Pôle Rhône-Alpin de Bio-Informatique (PRABI, <http://www.prabi.fr/>). MOSAIC intègre un interpréteur embarqué du logiciel de statistique R [3] ainsi qu'un système de distribution des calculs. La plate-forme MOSAIC tire profit des machines multi-processeurs et fournit ses résultats avec un temps de calcul raisonnable en cas d'utilisateurs multiples en parallèle. L'ensemble des calculs repose sur deux librairies R : la librairie 'morse' pour les analyses CR et TKTD [4], la librairie 'fitdistrplus' pour l'ajustement de distributions [5].

MOSAIC propose aujourd'hui trois types d'analyses qui sont détaillées ci-après : (i) des analyses CR classique pour la survie et la reproduction ; (ii) des analyses TKTD pour la survie (modèles GUTS) ; (iii) des analyses de la distribution de sensibilité des espèces (SSD).

## MOSAIC pour des analyses concentration-réponse

MOSAIC offre deux modules d'analyse de données d'écotoxicité selon une approche CR classique, l'un pour les données de survie [6], l'autre pour les données de reproduction [7]. Par défaut, ces analyses CR sont réalisées en fin d'expérience par ajustement d'un modèle log-logistique à trois paramètres sur lequel se greffe une partie stochastique appropriée à la nature des données, et ce de manière transparente pour l'utilisateur qui n'a besoin que de fournir ses données expérimentales en entrée (Figure 1.1). Les paramètres des modèles sont estimés par inférence Bayésienne. Les résultats sont rendus sous forme graphique pour les données brutes (Figures 1.2 et 1.5), les courbes ajustées aux données (Figures 1.2 et 1.6) et la comparaison des observations aux valeurs prédites (Figure 1.7). Les paramètres ainsi que les LCx ou les ECx sont rendus sous forme d'une valeur médiane et d'un intervalle d'incertitude à 95% (Figures 1.4, 1.8 et 1.9).



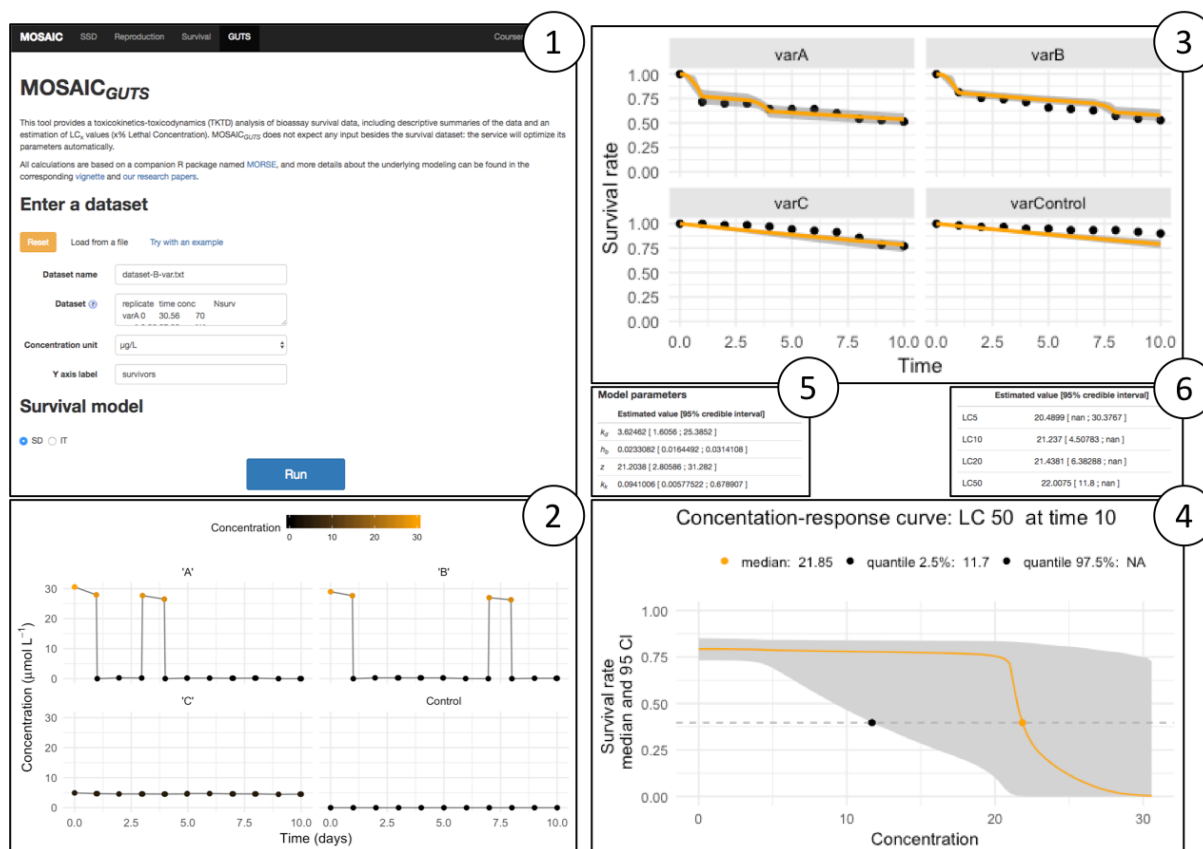
**Figure 1 :** Morceaux choisis de l'interface web MOSAIC lors d'analyses CR et/ou TKTD sur le jeu de données 'cadmium2' disponible en ligne (<http://pbil.univ-lyon1.fr/software/mosaic/data/cadmium2>) : (1) Page de saisie des données expérimentales (exemple du module 'repro') ; (2) Représentation graphique du taux de reproduction cumulé en fin d'expérience, en fonction de la concentration ; (3) Modèle concentration-réponse de la reproduction en fonction de la concentration : courbe médiane (ligne orange en trait plein) et son incertitude (zone grisée délimitée par les lignes pointillées orange) ; (4) Estimation des paramètres du modèle ajusté aux données de reproduction ; (5) Représentation graphique des données de survie (exemple 'cadmium2') en fonction du temps et pour les différentes concentrations testées ; (6) Ajustement du modèle GUTS-SD sur les données de survie ; (7) Graphique du posterior predictive check issu de l'ajustement du modèle GUTS-SD ; (8) Estimation des paramètres du modèle GUTS-SD ; (9) Estimation des concentrations létales à 5, 10, 20 et 50%, obtenues à partir des résultats de l'ajustement du modèle GUTS-SD.

## MOSAIC pour des analyses TKTD

Encore peu utilisés du fait de leur complexité de mise en œuvre, les modèles TKTD présentent de nombreux avantages pour analyser des données d'écotoxicologie. Ils permettent de décrire de manière mécaniste le mode d'action des contaminants, de dériver des paramètres indépendants de la durée d'exposition (comme la NEC, 'No-Effect-Concentration'), de modéliser des données recueillies sous exposition à des profils de concentrations qui varient au cours du temps, et de faire des prédictions dans des situations environnementales qui n'ont pas pu être testées.

expérimentalement. Un autre avantage majeur est que ces modèles TKTD permettent d'estimer une concentration létale ou effective pour n'importe quel x% et n'importe quelle durée d'exposition, avec son incertitude.

Le module 'GUTS' de MOSAIC permet de mettre en œuvre des modèles TKTD pour analyser des données de survie dans le cadre du *General Unified Threshold model of Survival* (GUTS) qui donne une estimation de la probabilité de survie liée à l'effet du contaminant lorsque sa concentration dépasse un certain seuil. Selon que ce seuil est le même pour tous les individus (modèle GUTS-SD) ou qu'il est distribué au sein de la population d'individus (modèle GUTS-IT), les prédictions sont différentes [8]. Le module 'GUTS' de MOSAIC offre la possibilité d'ajuster soit le modèle GUTS-SD soit le modèle GUTS-IT, que les données aient été recueillies à concentration constante (tests de toxicité standard) ou à concentration variable dans le temps (e.g., exposition à des pics de contaminant régulièrement espacés). Comme le montre la Figure 2, MOSAIC-GUTS fournit une représentation de données brutes (Figure 2.2), une visualisation de l'ajustement (Figure 2.3), la représentation graphique de la LCx calculée par défaut à la fin du temps d'exposition (ici 10 jours) et pour un x=50% choisi (Figure 2.4).



**Figure 2 :** Morceaux choisis du module 'GUTS' de MOSAIC lors d'une analyse TKTD sur un jeu de données à concentration variable (ici, quatre profils d'exposition différents) : (1) Page de saisie des données expérimentales (ici 'dataset-B-var.txt') ; (2) Représentation graphique des profils de concentrations variables ('A', 'B', 'C', 'Control') ; (3) Ajustement du modèle GUTS-SD avec la courbe médiane (ligne orange) et son incertitude (zone grisée) superposées aux données expérimentales ; (4) Représentation graphique de la LC<sub>50</sub> à 10 jours, avec sa courbe médiane et son incertitude ; (5) Estimation des paramètres fournis avec une médiane et les quantiles à 2.5 et 97.5% définissant ainsi un intervalle d'incertitude ; (6) Estimation des concentrations létales à 5, 10, 20 et 50%, obtenues à partir des résultats de l'ajustement du modèle GUTS-SD.

## MOSAIC pour des analyses SSD

Un dernier module complète la panoplie MOSAIC permettant d'analyser la distribution de sensibilité d'un pool d'espèces à un contaminant [9]. Les données en entrée sont des indicateurs de toxicité, comme des LCx ou des ECx telles qu'estimées dans les autres modules de MOSAIC, que l'on peut injecter dans l'analyse en tenant compte de leur incertitude. Le module 'SSD' de MOSAIC permet en effet de traiter ces données censurées par intervalles, c'est ce qui en fait son originalité au regard des outils concurrents.

## Contacts

Sandrine CHARLES, Virgile BAUDROT et Philippe VEBER

[mosaic@univ-lyon1.fr](mailto:mosaic@univ-lyon1.fr)

Laboratoire de Biométrie - Biologie Évolutive  
Université de Lyon; Université Lyon 1  
CNRS; UMR 5558  
43 boulevard du 11 novembre 1918  
F-69622 Villeurbanne Cedex, France



## Pour en savoir plus

<http://pbil.univ-lyon1.fr/software/mosaic/>  
<http://pbil.univ-lyon1.fr/software/mosaic/survival>  
<http://pbil.univ-lyon1.fr/software/mosaic/reproduction>  
<http://pbil.univ-lyon1.fr/software/mosaic/guts>  
<http://pbil.univ-lyon1.fr/software/mosaic/ssd>  
<https://cran.r-project.org/web/packages/morse/vignettes/modelling.pdf>  
<https://cran.r-project.org/web/packages/morse/vignettes/tutorial.html>

## Bibliographie citée

1. EFSA. 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. *EFSA J.* 11:1–268.
2. Charles S, Veber P, Delignette-Muller ML. 2017. MOSAIC: a web-interface for statistical analyses in ecotoxicology. *Environ. Sci. Pollut. Res.* doi:10.1007/s11356-017-9809-4.
3. R Core Team. 2015. R: A Language and Environment for Statistical Computing. Available from <https://www.r-project.org>.
4. Baudrot V, Charles S, Delignette-Muller ML, Duchemin W, Kon-Kam-King G, Lopes C, Ruiz P, Veber P. 2018. morse: Modelling Tools for Reproduction and Survival Data in Ecotoxicology. Available from <https://cran.r-project.org/package=morse>.
5. Delignette-Muller ML, Dutang C. 2015. fitdistrplus : An R Package for Fitting Distributions. *J. Stat. Softw.* 64:1–34.
6. Forfait-Dubuc C, Charles S, Billoir E, Delignette-Muller M. 2012. Survival data analyses in ecotoxicology: critical effect concentrations, methods and models. What should we use? *Ecotoxicology*. 12:1072–1083.
7. Delignette-Muller ML, Lopes C, Veber P, Charles S. 2014. Statistical handling of reproduction data for exposure-response modeling. *Environ. Sci. Technol.* 48.
8. Baudrot V, Preux S, Ducrot V, Pavé A, Charles S. 2018. New insights to compare and choose TKTD models for survival based on an inter-laboratory study for *Lymnaea stagnalis* exposed to Cd. *Environ. Sci. Technol.* doi:10.1021/acs.est.7b05464.
9. 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.