

R au secours des écotoxicologues

G. Kon Kam King^a, P. Veber^a, M.L. Delignette-Muller^{a,b} and S. Charles^{a,c}

^aUniversité de Lyon; Université Lyon 1
Laboratoire de Biométrie - Biologie Evolutive
CNRS; UMR 5558
43 boulevard du 11 novembre 1918
69622 Villeurbanne Cedex
guillaume.kon-kam-king@univ-lyon1.fr
philippe.veber@univ-lyon1.fr

^bUniversité de Lyon
VetAgro Sup Campus Vétérinaire de Lyon
1 avenue Bourgelat
69280 Marcy l'Etoile
marielaure.delignettemuller@vetagro-sup.fr

^cInstitut Universitaire de France
103 boulevard Saint-Michel
75005 Paris
sandrine.charles@univ-lyon1.fr

Mots clefs : interfaçage web, analyse de données de bioessais, distribution de sensibilité des espèces.

Le défi majeur que doit relever aujourd'hui l'écotoxicologie est de se munir d'outils de modélisation prédictifs, intégrés dans un cadre décisionnel standardisé, et dont les autorités de régulation et les décideurs puissent directement tirer profit pour une meilleure gestion des impacts potentiels des substances chimiques sur les populations et les communautés à protéger. Dans cette perspective, il apparaît nécessaire d'offrir aux écotoxicologues le moyen d'exploiter au mieux les données qu'ils acquièrent. Les données en écotoxicologie sont issues de bioessais, généralement standardisés selon des normes (OCDE, ISO), en toxicité aiguë ou chronique, au cours desquels sont mesurées la survie, la reproduction et/ou la croissance d'organismes modèles. Ce sont donc des données temporelles dépendantes d'une gamme croissante de concentrations testées.

Classiquement, l'analyse statistique de ce type de données se réduit à l'estimation de concentrations critiques d'effet à un temps donné (le plus souvent en fin d'essai), soit en utilisant des tests d'hypothèse (par exemple pour déterminer une NOEC, *No Observed Effect Concentration*), soit en ajustant des modèles paramétriques concentration-réponse (par exemple pour estimer des EC_x, *x% Effective Concentration*). Malgré l'existence d'un guide OCDE [1], la jungle des statistiques paraît généralement impénétrable à l'écotoxicologue : il gaspille une bonne partie de ses données à ne les considérer qu'en fin d'essai, il doit choisir parmi une multitude de tests/méthodes/modèles celui qui "convient le mieux", la nature de ses données ne lui permet pas toujours de vérifier les conditions d'applications de ces tests/méthodes/modèles, il ne dispose pas d'un outil "clé en main", fiable statistiquement, convivial et simple d'utilisation...

Face aux macros Excel faites maison ou aux logiciels boîte noire parfois hors de prix, R pourrait être l'alternative salvatrice. Pour des analyses statistiques simples, il existe déjà quelques packages dédiés comme par exemple `drc` ou `fitdistrplus`. Le package `drc` permet l'estimation des paramètres des modèles concentration-réponse par maximum de vraisemblance [2] ; le package `fitdistrplus` permet d'ajuster des distributions paramétriques par maximum de vraisemblance, le cas échéant en tenant compte de la présence de données censurées [3]. Par contre, pour des analyses plus complexes, par exemple avec des données de comptage qui nécessitent d'utiliser des modèles d'erreur non standard ou avec des données dépendantes du temps, il faut des programmes R plus sophistiqués. Et dans l'optique que ces programmes R soient utilisés par des écotoxicologues, que les aspects techniques et théoriques sous-jacents rebutent généralement, il est indispensable d'investir dans la conception d'interfaces utilisateur faciles d'utilisation. L'accessibilité d'un tel outil repose sur deux aspects : (1) l'interface graphique rend transparente l'utilisation plus experte de l'interpréteur R ; (2) elle propose un paramétrage par défaut ou automatique des procédures d'estimation.

Dans cet exposé, nous commencerons par illustrer la valeur ajoutée de R pour des analyses statistiques simples de données d'écotoxicologie. Nous dévoilerons ensuite notre nouvelle interface, `MOSAIC_SSD` (*Modelling and Statistical Tools for Ecotoxicology*, <http://pbil.univ-lyon1.fr/software/mosaic/>), qui permet de modéliser de façon simple et conviviale la distribution de sensibilité d'une communauté d'espèces à une substance chimique donnée, pour en extraire par exemple une concentration qui affecte $x\%$ (ou qui protège $1-x\%$) des espèces considérées (HCx , $x\%$ Hasard Concentration). Nous montrerons également comment cette interface pallie les manques actuels des logiciels existants, en permettant d'une part de calculer des intervalles de confiance bootstrap autour des HCx , et d'autre part de prendre en compte des données censurées. Enfin, nous discuterons des spécificités inhérentes à la conception de ce type d'interface. Nous évoquerons l'approche suivie pour embarquer un interpréteur R dans un serveur web, en particulier les aspects relatifs au parallélisme et à l'appel de fonctions R depuis le langage utilisé par le serveur.

Références

- [1] OCDE (2006) Current approaches in the statistical analysis of ecotoxicity data: a guidance to application. Technical report, Organisation for Economic Cooperation and Development.
- [2] Ritz, C., & Streibig, J. (2005). Bioassay analysis using R. *Journal of Statistical Software*, 12(5), 1–22.
- [3] Delignette-Muller ML, Pouillot R, Denis JB, Dutang C. (2009). `fitdistrplus`: Help to Fit of a Parametric Distribution to Non-censored or Censored Data. Available at <http://cran.at.r-project.org/web/packages/fitdistrplus/index.html> [accessed 7 May 2013].