

L2 Ecologie et Informatique

TD/TP Ecotoxicologie

Sandrine CHARLES (sandrine.charles@univ-lyon1.fr)
Christelle LOPES (christelle.lopes@univ-lyon1.fr)

Avril 2022

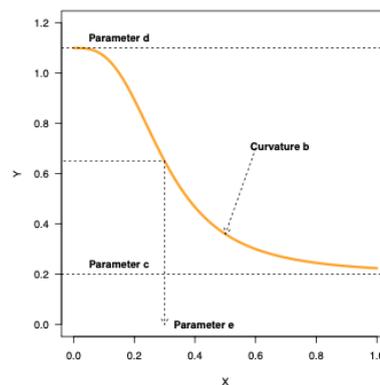
Introduction

Le but de ce document est de vous guider dans l'utilisation de la plateforme web MOSAIC pour ajuster des modèles dose-réponse (DR) sur des données issues de tests standards de toxicité, et les simuler. Tous les aspects théoriques utiles ont été abordés en CM.

Le modèle utilisé ici pour décrire l'effet d'un contaminant sur les traits d'histoire de vie d'une espèce après une durée d'exposition donnée est le modèle log-logistique, dont la forme générale à 4 paramètres est rappelée ici :

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

où $Y(X)$ est la valeur du trait mesurée à la fin de l'exposition à la concentration X , c est l'asymptote inférieure (à savoir la valeur du trait quand la concentration X tend vers l'infini), d est la valeur initiale (à savoir la valeur du trait en condition non contaminée, *i.e.*, pour $X = 0$), e est la concentration conduisant à 50% d'effet sur la réponse mesurée (encore appelée EC_{50} ou LC_{50} selon le trait), et b est un paramètre de courbure (sans unité).



A cette partie déterministe est associée une partie stochastique qui dépend de la nature de la donnée Y à expliquer.

1 Analyse de données quantales (ou binaires)

On s'intéresse dans cette partie aux effets de contaminants sur des traits d'histoire de vie dont la réponse se caractérise par des données dites binaires. Ceci est notamment le cas pour la survie des individus (un individu est soit mort, soit vivant), le taux de survie étant le trait d'histoire de vie le plus couramment utilisé et le plus facile à suivre expérimentalement quand on s'intéresse aux espèces animales. Les exemples que nous allons utiliser ici concernent donc la survie d'individus exposés, mais il existe d'autres traits aux données binaires, comme le taux d'émergence de larves de diptère, le taux d'enfouissement de vers oligochètes, le taux de germination des graines, le taux de floraison des plantes... Les modèles utilisés pour ces données binaires comportent donc une partie stochastique adaptée de type Binomiale.

Step 1.1 Allez à <http://mosaic.univ-lyon1.fr/> et choisissez le menu `surv`, puis sous-menu `Standard`.

Step 1.2 Essayez en premier lieu avec l'exemple `chlordan` (pensez à lire la description du jeu de données que vous utilisez pour savoir au moins sur quelle espèce il a été réalisé) : que remarquez-vous une fois que les données ont été automatiquement téléchargées ? Regardez le jeu de données entier en agrandissant la fenêtre ou en utilisant la barre de défilement.

Step 1.3 Cliquez sur `run` et passez en revue les éléments suivants :

- En attendant les résultats de l'ajustement, combien de concentrations ont été testées ?
- Combien de réplicats par concentration ont été réalisés ?
- Une fois les résultats de l'ajustement obtenus, que représentent les figures `Raw data plot`, `Dose-response plot` et `Fitted curve` ?
- Qu'observez-vous sur les figures si vous changez l'échelle des abscisses (`Change scale for x-axis`) ?
- A quoi correspond le paramètre e ? Que pensez-vous de son estimation ?
- A quoi correspond le paramètre b ? Que pensez-vous de son estimation ?
- Que remarquez-vous concernant les estimations de LC_x ?
- Au regard de la figure du PPC, est-ce que l'ajustement vous semble correct ou non ?
- Revenez en haut de la page. Vous pouvez imprimer le rapport fourni par MO-SAIC, contenant l'ensemble des résultats obtenus (format pdf).

Step 1.4 Essayez maintenant avec l'exemple `cadmium1`. Quelles différences observez-vous sur les résultats ?

Step 1.5 Ces deux jeux de données concernent l'espèce *Daphnia magna* exposée au chlordane et au cadmium. Au vu des résultats, à quel contaminant la survie de cette espèce est-elle la plus sensible ?

2 Analyse de données de comptage (ou discrètes)

On s'intéresse dans cette partie aux effets de contaminants sur des traits d'histoire de vie dont la réponse se caractérise par des données discrètes de comptage. Ceci concerne notamment tous les traits associés à la reproduction : nombre d'œufs, nombres de juvéniles, nombre de pontes... Les modèles utilisés pour décrire ces données de comptage comportent donc une partie stochastique adaptée, de type Poisson par exemple.

Step 2.1 Revenez à <http://mosaic.univ-lyon1.fr/> et choisissez le menu **Repro**.

Step 2.2 Essayez le jeu de données exemple `chlordan`. Vous remarquerez que vous avez en premier lieu une représentation de la survie, avant l'analyse des données de reproduction (la survie est implicite ici car pour se reproduire, un individu doit être vivant...). Concernant les résultats pour la reproduction, dans la partie **Reproduction data description** : à quoi correspondent les représentations dans **Raw data plots** ?

Step 2.3 Comparez vos résultats avec ceux obtenus sur la survie. A quoi correspond ici le paramètre e ? A-t-il la même valeur ? Que pouvez-vous dire de son estimation ?

Step 2.4 Essayez maintenant avec le jeu de données `cadmium1`. Comparez vos résultats à ceux obtenus précédemment sur la survie.

Step 2.5 Au vu des résultats, à quel contaminant la reproduction de cette espèce est-elle la plus sensible ?

3 Analyse de données quantitatives continues

Enfin, on s'intéresse dans cette dernière partie aux effets de contaminants sur des traits d'histoire de vie dont la réponse se caractérise par des données quantitatives continues. Ceci concerne notamment tous les traits associés à la croissance : taille, poids... Les modèles utilisés pour décrire ces données quantitatives continues comportent donc une partie stochastique Gaussienne (loi Normale).

Step 3.1 Retournez à <http://mosaic.univ-lyon1.fr/> et choisissez le menu **Growth**. Durant ce TD, merci d'utiliser le serveur pédagogique pour permettre à tous de réaliser les exercices en même temps : <http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/>.

Step 3.2 Si vous le souhaitez, vous pouvez visualiser notre vidéo de démonstration (durée: 08:39), en cliquant simplement sur le lien vidéo ou directement ici : <http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/MOSAICgrowth.mp4>

Step 3.3 **Visualisation des données : onglet Data visualisation**

- Essayez le jeu de données exemple `plant07.txt`. Notez que vous pouvez visualiser votre jeu de données soit sous forme de tableau, soit sous forme de graphique pour vous assurer qu'il a été correctement téléchargé. Notez que vous pouvez personnaliser les noms des axes et choisir une unité appropriée pour la concentration d'exposition, les mesures de croissance et le temps.

- Une nouvelle fonctionnalité est disponible dans ce module : la possibilité de télécharger plusieurs fichiers **en même temps**. Essayez de charger ces exemples supplémentaires : `chlordan-daphnia.txt`, `cadmium-daphnia.txt`, `subst01-lymnaea.txt`, `plant03.txt` et `plant10.txt`. Que remarquez-vous ? Quel jeu de données est affiché ? Essayez de changer de fichier.

Step 3.4 Analyse Dose-Réponse : onglet Dose-response analysis

- En gardant ces plusieurs fichiers chargés, allez au sous-menu **Dose-response analysis**. Exécutez l'analyse pour le jeu de données `plant07.txt`. Quelle est l'estimation de la EC_{50} (dénommée ER_{50} ici). Donnez la valeur médiane et l'intervalle de crédibilité à 95%.
- Dans la partie **Goodness-of-fit criteria**, regardez la figure de **Correlation**. Que remarquez-vous ? N'hésitez pas à lire les texte d'aide en-dessous des figures.
- Demandez le calcul des ER_{25} , ER_{50} et ER_{75} . Que remarquez-vous en terme de précision des estimations ?
- Exécutez les analyses pour les jeux de données `plant03.txt` et `plant10.txt`. Regardez en particulier l'estimation de la ER_{50} pour le jeu de données `plant10.txt`. Que remarquez-vous ? Comment pouvez-vous expliquer un tel résultat ?
- Pour tous ces jeux de données, les graphiques PPC sont-ils suffisamment adaptés pour faire confiance aux résultats fournis dans une perspective ultérieure d'évaluation du risque (ERA) ?
- Demandez le calcul de la ER_{50} pour `plant03.txt` et `plant10.txt`. Demandez à créer le tableau fournissant toutes les ER_{50} pour les fichiers `plant03.txt`, `plant07.txt` et `plant10.txt` grâce au bouton **Display all ER_x** . Quelle plante est la plus sensible ?

Step 3.5 Téléchargement : onglet Downloads

- Allez au sous-menu **Downloads**.
- Pour le jeu de données `plant07.txt`, téléchargez le **Single report** (choisissez le format Word).
- Téléchargez le fichier texte de la **joint posterior distribution**.

Step 3.6 Outil de prédiction

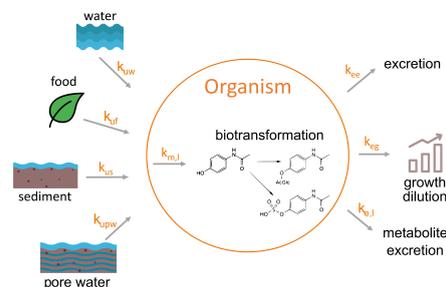
- Allez dans le sous-menu **Prediction tool**.
- Entrez une gamme de concentrations pour laquelle la valeur des paramètres du modèle log-logistique à 3 paramètres est disponible, par exemple 40; 80; 160; 320; 640.
- **Premièrement**, choisissez des paramètres non distribués (option par défaut).
- **Option (1)** : fixer une valeur unique pour chaque paramètre. Ces valeurs choisies peuvent provenir de connaissances d'expert, de la littérature..., obtenus pour une durée d'exposition donnée. Vous pouvez essayer avec les valeurs données par défaut. Notez que les paramètres b et e sont donnés en \log_{10} . Que remarquez-vous ?
- Changer la valeur du paramètre e de $\log_{10}(e) = 2.2$ à $\log_{10}(e) = 3$. Que remarquez-vous ?

- **Option (2)** : prendre les valeurs médianes des estimations obtenues à partir de l'analyse DR réalisée en amont. Essayez cette option (2) avec les valeurs estimées pour le jeu de données `plant07.txt`. Que remarquez-vous ?
- Ré-essayez l'option (2), mais avec les estimations obtenues pour le jeu de données `plant10.txt`, tout en gardant la même gamme de concentrations. Que remarquez-vous ?
- A partir de la simulation précédente, trouvez une gamme de concentrations plus appropriée.
- **Deuxièmement**, choisissez **distributed parameters**. Cela nécessite d'avoir préalablement téléchargé la distribution jointe *a posteriori* des paramètres du modèle DR ajusté précédemment, ce qui est aujourd'hui possible à partir du module `MOSAICgrowth`.
- **Option (1)** : revenez à la gamme de concentrations précédemment proposée, à savoir 40; 80; 160; 320; 640. Chargez le fichier `outputPosterior_plant07.txt` que vous avez dû télécharger précédemment. Quelle(s) différence(s) observez-vous avec les prédictions précédentes ?
- **Option (2)** : choisissez l'option `from a previous DR analysis`, qui est supposée avoir été préalablement réalisée sur le jeu de données `plant07.txt`. Que remarquez-vous ?
- Changez la gamme de concentrations et regardez ce qu'il advient.

Pour plus de détails sur `MOSAICgrowth`, vous pouvez télécharger le tutoriel disponible en ligne ici : <http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/Tutorial.pdf>

4 Métriques de Bioaccumulation

On s'intéresse ici à l'étape de bioaccumulation des contaminants, étape préalable à l'évaluation des effets. Les expérimentations visant à étudier la bioaccumulation consistent à exposer des organismes à une certaine concentration en contaminant pendant un temps donné (phase d'accumulation) puis à les laisser dans un milieu "propre" sans contaminant pendant un certain temps (phase de dépuración). Pendant ces deux phases, des individus sont échantillonnés de manière aléatoire à différents pas de temps et leur concentration interne en contaminant est mesurée. A ces données sont ajustés des modèles toxico-cinétiques (TK), permettant de prédire la concentration bioaccumulée dans les organismes en fonction de la concentration d'exposition, en considérant différents processus, appelés processus ADME (Absorption, Distribution, Métabolisation, Elimination):



Les modèles TK les plus simples pour décrire ces processus sont des modèles à un compartiment où l'organisme entier est considéré comme un seul et unique compartiment dans lequel rentre le contaminant et duquel il "sort".

Durant la phase d'accumulation ($0 < t < t_c$ avec t_c la durée de cette phase), le modèle TK à un compartiment peut s'écrire sous la forme générale suivante :

$$\frac{dC(t)}{dt} = \sum_{i=1}^n [k_i \times C_i(t)] - \sum_{j=1}^p [k_j \times C(t)]$$

où $C(t)$ est la concentration en contaminant à l'intérieur de l'organisme, k_i correspond au taux d'accumulation de la substance selon la voie d'exposition i à la concentration $C_i(t)$ dans l'environnement au temps t , et k_j correspond au taux d'élimination de la substance selon le processus j .

Durant la phase de dépuration ($t > t_c$), seuls les processus d'élimination agissent et le modèle s'écrit :

$$\frac{dC(t)}{dt} = - \sum_{j=1}^p [k_j \times C(t)]$$

A cette partie déterministe est associée une partie stochastique gaussienne.

Pour résumer les capacités de bioaccumulation d'une substance par une espèce donnée, les évaluateurs du risques utilisent des "métriques". On distingue trois métriques selon la voie d'exposition :

- **BCF (Bio-Concentration Factor)** si l'exposition se fait par l'eau.
- **BSAF (Biota-Sediment Accumulation Factor)** si l'exposition se fait par le sédiment.
- **BMF (Bio-Magnification Factor)** si l'exposition se fait par la nourriture (voie trophique).

Pour chacune des ces métriques, on en distingue deux types :

- les métriques **à l'équilibre** ("steady state": BCF_{ss} , $BSAF_{ss}$ et BMF_{ss}) si l'organisme est à l'équilibre avec son milieu à la fin de la phase d'accumulation (i.e. il a atteint un plateau constant pendant au moins 3 pas de temps). Dans ce cas, la métrique est calculée par le ratio entre la concentration interne à la fin de la phase d'accumulation et la concentration d'exposition.
- les métriques **cinétiques** (BCF_k , $BSAF_k$ et BMF_k) si l'équilibre n'est pas atteint. Dans ce cas, la métrique est calculée par le ratio entre les taux d'accumulation et les taux d'élimination estimés.

Nous allons maintenant illustrer l'ajustement de modèles TK et le calcul de ces différentes métriques sur des jeux de données exemple à l'aide de $MOSAIC_{bioacc}$.

Step 4.1 Revenez à <http://mosaic.univ-lyon1.fr/> et choisissez le menu bioacc. Pendant ce cours, merci d'utiliser le serveur pédagogique suivant:

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

Pour plus de détails sur MOSAIC_{bioacc}, vous pouvez télécharger et lire le guide d'utilisation fourni en ligne.

Step 4.2 Visualisation des données

- Essayez avec le jeu de données exemple `Male_Gammarus_Single.txt`. Notez que soit une figure, soit un tableau est fourni pour vérifier que les données ont été correctement téléchargées. Tout est automatisé, l'utilisateur n'a rien à faire.
- Téléchargez le fichier exemple `Male_Gammarus_Single.txt` et parcourez-le. Notez que l'utilisateur doit entrer le bon spérateur, l'unité de temps et la durée de la phase d'accumulation. A partir de la description de l'exemple, remplissez les champs correctement.
- Essayez avec l'exemple `Pimephales_two.csv` et regardez les données. Notez que l'utilisateur doit choisir entre 2 concentrations d'exposition.
- Essayez avec l'exemple `Male_Gammarus_seanine.csv`. Notez dans les graphiques que la concentration du composé parent, celle des métabolites ainsi que des mesures de croissance sont disponibles.

Step 4.3 Modèle et paramètres

- Téléchargez à nouveau le fichier `Male_Gammarus_Single.txt` et allez au sous-menu **Model and parameters**. Les paramètres sont automatiquement sélectionnés selon les données téléchargées. De plus, les équations du modèle (de la partie déterministe seulement) sont fournies avec les variables correspondantes. Pour ce jeu de données, combien y a-t-il de paramètres? Combien d'équations? Pouvez-vous déduire les voies d'exposition (eau, sédiment, eau interstitielle et/ou nourriture?) et les processus d'élimination (excretion, dilution par la croissance et/ou biotransformation?) qui sont considérés dans l'expérience?
- Idem avec l'exemple `Chironomus_benzo-a-pyrene.csv`.
- Idem avec l'exemple `Male_Gammarus_seanine.csv`.
- Téléchargez à nouveau `Male_Gammarus_Single.txt` et exécutez les analyses en cliquant sur le bouton **"Calculate and display"**. Les calculs peuvent prendre plusieurs minutes selon le jeu de données.

Step 4.4 Résultats and critères de "goodness-of-fit"

- A partir de l'analyse de l'exemple `Male_Gammarus_Single.txt`, quelle est l'estimation du BCF_k ? Donnez sa médiane et son intervalle de crédibilité à 95%.
- Est-il raisonnable de demander le calcul du BCF_{ss} ? Si oui, donnez sa médiane et son intervalle de crédibilité à 95%.
- Quelle est l'estimation des paramètres k_{uw} et k_{ee} ? Donnez leur médiane et leur intervalle de crédibilité à 95%.
- A partir du sous-menu **Goodness-of-fit**, interprétez le Posterior Predictive Check (PPC). Quel est le pourcentage de données dans les intervalles de prédiction? N'hésitez pas à lire le texte d'aide à côté de chaque figure.

- Interprétez la figure `Priors and Posteriors`.
- Regardez la figure `Correlation plot`. Que remarquez-vous ?
- Allez au `Potential Scale Reduction Factors (PSRF)`. Les valeurs sont-elles acceptable pour considérer la convergence pour tous les paramètres ?
- Idem avec les exemples `Chironomus_benzo-a-pyrene.csv` et `Male_Gammarus_seanine.csv`. Notez que les calculs peuvent prendre plusieurs minutes, notamment pour le second exemple.
- Que remarquez-vous concernant le facteur de bioaccumulation du jeu de données `Chironomus_benzo-a-pyrene.csv`?

Step 4.5 Téléchargements

- Allez au sous-menu `Downloads`.
- Pour l'exemple `Male_Gammarus_Single.txt`, vous pouvez télécharger le rapport.