

Modélisation et méthodes numériques pour les sciences du vivant.

Projets

Pierre en jaune; Paul en bleu et vert les deux.

L'**objectif** de ce projet est

- de vous permettre de vous familiariser avec le langage de programmation R qui vous servira également dans d'autres cours et potentiellement dans vos stages ou vie professionnelle,
- de vous plonger dans un exemple particulier de modèle en biologie, sur lequel vous **mettrez en pratique** les différentes méthodes que l'on verra tout au long du cours.
- de découvrir que différentes approches de modélisation existent pour un même domaine d'application et d'en appréhender les spécificités.

Un projet correspond à l'**étude de 2 articles scientifiques, notés M (article pour *modèle*) et C (article pour *commentaire*), présentant différentes approches de modélisation du même système** ou processus biologique (ou tout du moins d'un système proche). **L'un des modèles présent dans l'article M devra être codé sous R et analysé** à l'aide des méthodes vues en cours. **Les autres modèles**, issus de l'article C à trouver par vous-mêmes, serviront à la comparaison et à la **prise de recul**.

Le **contenu** du rapport dépend bien sûr de votre sujet, mais quelques points génériques doivent apparaître (le barème est donné à titre indicatif et des adaptations y sont apportées pour prendre en compte les difficultés spécifiques de chaque sujets) :

1. Présentation du problème **biologique** et du contexte (synthèse bibliographique, une page environ, citant vos sources) - 2 points

2. Présentation du **modèle choisi et de vos simulations** - 4 points.

— Rechercher quelques informations sur les auteurs de l'article M : sont-ils plutôt biologistes ou mathématiciens ? Quels sont leurs thèmes principaux de recherche ?

— Donner en quelques lignes l'objectif du travail de modélisation réalisé par les auteurs.

— Présenter votre propre démarche de modélisation, de manière résumée si elle est similaire à celle des auteurs et de manière plus détaillée pour les points que vous n'auriez pas pu reproduire (modèle trop complexe, valeurs manquantes de paramètres ou de variables initiales, etc). En tous les cas, essayez de lister clairement les hypothèses sous-jacentes au modèle, ainsi que les différentes variables et paramètres du modèle, par exemple sous forme d'un tableau du style :

Var/par ¹	Symbole	Nom (unité)	Valeur ²	Détermination
Var	M(t)	Masse du bébé (kg)	M(0)=3.5kg	Mesure expérimentale par les auteurs
Par	k	coeff. de croissance ($g.j^{-1}$)	25	Estimée
...				

(les items de ce tableau peuvent être adaptés selon votre sujet).

— Mettez en valeur quelques résultats choisis, en les comparant à ceux des auteurs (par exemple via une figure représentant côte à côte une copie du graphe des auteurs et vos propres sorties).

— Expliquez succinctement la structure de votre code et vos choix de programmation ainsi que les tests effectués pour vérifier l'absence d'erreurs dans votre code. **Les scripts R doivent être rendus.** Ils ne doivent pas être copiés dans le rapport, sauf sous forme d'extraits dans des figures si c'est utile à la compréhension de votre travail (ou de vos difficultés), ou bien en annexe.

3. Vos **analyses** du modèle :

— **Analyse d'incertitude/de sensibilité du modèle** - 3 points

— **étudier lorsque c'est possible l'identifiabilité structurelle et/ou pratique du modèle ; estimation paramétrique du modèle à partir de données expérimentales** - 3 points

- Proposer un autre modèle, qui peut être une modification ou une extension du modèle précédent, et appliquer les critères de sélection de modèle pour comparer (de manière quantitative si votre nouveau modèle est suffisamment simple ; de manière qualitative si vous vous êtes lancés dans une approche plus complexe)³ - 3 points
- Commentaires et interprétations
- 4. Chercher un article C traitant du même sujet ou d'un sujet proche, décrivant une approche de modélisation différente, de préférence plus globale ou plus complexe, ou une review, que vous présenterez et commenterez en donnant votre opinion - 3 points
- 5. un petit paragraphe de perspectives et conclusions ; il inclura votre évaluation de l'article M via une note sur 5 accompagnée de quelques mots de commentaires (erreurs éventuelles repérées, facilité à reproduire le travail...). - 1 point
- 6. Qualité de la rédaction et de la mise en forme du rapport sur 1 point.

À rendre

Rapport intermédiaire A mettre sur Edunao, section Projets, pour le **11 octobre**. C'est tout simplement l'embryon de votre rapport final. Indiquer le plan prévisionnel du rapport final. Non noté mais obligatoire ; inclure les scripts R en fichiers attachés, toujours commentés.

Rapport final : A rendre pour le **10 novembre à 23h** : environ une dizaine de pages, sans compter vos scripts R en fichiers attachés. A télécharger au format **pdf** sur Edunao.

Il s'agit d'un projet d'application **personnel**. Au maximum 4-5 personnes peuvent choisir un même sujet. Vous pouvez vous entraider si besoin, mais chacun doit avoir **son propre code et rapport** ; l'article commenté pour la prise de recul est spécifique à chacun. Les sujets n'ont pas tous la même difficulté ; il en sera tenu compte dans la notation.

Nouveauté 2022 : autorisation de rendre un rapport pour 2 à condition que :

- vous choisissiez un binôme d'un niveau pas trop différent du vôtre
- soit indiqué pour chaque paragraphe (voire pour chaque figure) le nom du rédacteur principal
- le paragraphe concernant la présentation du 2ème article doit être dédoublé puisque chacun doit en choisir un différent.

Annexes

Détails sur l'analyse de sensibilité

Il s'agit d'effectuer une analyse de sensibilité globale (de préférence) ou locale sur votre modèle. On peut par exemple explorer les méthodes du package 'sensitivity' (<http://rss.acs.unt.edu/Rdoc/library/sensitivity/html/sensitivity-package.html>), qui contient les méthodes les plus classiques (documentation : faire help("sensitivity") sous R).

1. Commencez par identifier une ou des variables de sortie d'intérêt de votre modèle, Y (peut être une moyenne pondérée de plusieurs sorties).
Remarque : il faudra réitérer l'analyse pour chaque variable indépendante de sortie.
2. Choisissez pour chaque paramètre d'entrée, $X_i, i = 1..n$, une gamme plausible de variation et une forme de densité de probabilité en expliquant votre choix (incertitudes de type A ou B, ...).
3. Préliminaire - analyse d'incertitudes : tracer un profil de la variabilité de Y sous une des différentes hypothèses choisies pour les variations des entrées. Extraire certaines statistiques descriptives (moyenne, écart-type, quantiles, ...)
4. Selon la démarche vue en cours, tester, si approprié, l'une des méthodes suivantes :
 - calcul des indices de régression (standardisés ou partiels).

3. Pour les élèves VSE, ce peut être l'occasion de mettre en oeuvre la partie sur les modèles stochastiques, si le sujet s'y prête

- screening par la méthode de Morris
- méthodes de décomposition de la variance (Sobol, FAST). Calcul des indices de sensibilité totaux S_{Ti} lorsque disponibles.

Inclure dans le rapport : quelques lignes d'explication, les résultats, la durée du calcul⁴, interprétation et commentaires des résultats. Comparer les méthodes.

Détails sur l'estimation paramétrique et la comparaison de modèles

Pour l'estimation paramétrique : selon les projets, pour accéder aux données expérimentales, il pourra être nécessaire de les extraire des graphes. Pour extraire les données d'un graphique, de nombreuses applications existent⁵. Si aucun jeu de données expérimentales n'est disponible, se créer un jeu de données virtuel à partir de simulations puis donner de nouvelles valeurs aux paramètres et voir si l'on retrouve bien les valeurs ayant servi à générer ces données virtuelles. Cette procédure est classiquement employée pour tester les méthodes d'estimation et est appliquée en préalable à l'utilisation de données expérimentales réelles.

On pourra tester, parmi d'autres et selon la nature du modèle, les méthodes suivantes : modèle linéaire (`lm` ou `glm`), modèle non linéaire (`nls(y~a*exp(bt), data, start)`, `gnls` du package `nlme`). Si vous avez besoin d'implémenter votre propre fonction objectif (par exemple pour une méthode du max de vraisemblance avec une erreur non gaussienne), on peut utiliser `optim`.

(*facultatif*) Une étude d'identifiabilité des paramètres, lorsque c'est possible⁶.

Enfin la comparaison de modèles pourra s'appuyer, là encore selon les modèles considérés et en particulier leurs nombres de paramètres respectifs, sur différents critères : dits de 'goodness-of-fit' (vraisemblance, moindres carrés, etc) directement issus de l'estimation, de validation croisée (utilisation de la fonction `predict` associée aux résultats renvoyés par `lm`, `glm`, `nls`), ou de type AIC/BIC (méthodes du même nom sous R).

4. `proc.time()` ou `system.time()`

5. par exemple, capturer l'image à l'aide Irfanview puis l'ouvrir avec ImageJ puis :

- choisir le 5ème symbole, 'straight line', et tracer un segment d'une longueur connue (pris sur l'axe des ordonnées). Faire Analyze/'Set Scale' et donner cette longueur connue.

- faire les mesures : pour chaque valeur, tracer un segment et faire Ctrl+M.

ou bien : GraphClick, PlotDigitizer (application web) <http://arohatgi.info/WebPlotDigitizer/app/>

6. si le modèle est suffisamment simple : identifiabilité structurelle puis par exemple package R `fme` pour l'identifiabilité pratique