

«MTK» : un package pour unifier les démarches d'exploration numérique de modèles

Juhui WANG ^a, Robert Faivre ^b, Hervé Richard ^c, Hervé Monod ^a

^a MaIAGE - INRA, Domain de Vilvert, 78352 Jouy en Josas, France
Juhui.Wang@jouy.inra.fr, Herve.Monod@jouy.inra.fr

^b MIAT, UR875 - INRA, Auzeville, 31326 Castanet-Tolosan, France
Robert.Faivre@toulouse.inra.fr

^c BioSp - INRA, Domaine Saint-Paul, Site Agroparc, 84914 Avignon, France
Hervé.Richard@avignon.inra.fr

Mots clefs : Expérience numérique, Analyse de sensibilité, Gestion de workflow, Classes S4.

Introduction : L'exploration numérique de modèles couvre un très large spectre de méthodes allant de la planification d'expérience à l'analyse de sensibilité en passant par la propagation d'incertitude, la calibration, l'optimisation, etc. Due à la complexité croissante des modèles, elle est aujourd'hui devenue un outil indispensable des modélisateurs. D'ailleurs, de nombreux packages sous R lui sont consacrés [1, 2] : *sensitivity*, *diceDesign*, *lhs*, *planor*, *multisensi*, *FME*, *puma*, etc. Cependant, face à une offre en méthodes et en logiciels de plus en plus riche et en constante évolution, il est difficile pour un non-spécialiste de s'y retrouver, même pour des méthodes relativement simples. C'est dans ce contexte que le projet « mtk » a été développé. Il a pour but de rassembler l'ensemble des méthodes d'exploration numérique de modèles (existantes et à venir) dans un environnement générique, et ce avec une démarche unifiée et une syntaxe standardisée. Il peut être soit intégré dans les plateformes de modélisation, soit utilisé directement sous R.

Méthode : L'approche que nous proposons repose sur une restructuration des démarches utilisées par la communauté scientifique de l'exploration numérique de modèles. Elle consiste à décomposer les workflows en des tâches génériques et élémentaires de sorte que chaque méthode peut être représentée comme une combinaison de ces tâches. Ceci nous permet de nous affranchir des difficultés liées à l'hétérogénéité des démarches utilisées, qui, conjuguée avec la diversité des syntaxes préconisées dans les logiciels, constitue une des principales difficultés pour l'exploitation efficace des méthodes disponibles. Par exemple, dans la littérature, l'analyse de sensibilité peut être entreprise de différentes façons. Certains séparent les tâches de génération de plans d'expérience de celles de la simulation de modèle ou du calcul de l'indice de sensibilité, et ne réalisent qu'une partie des trois tâches citées. D'autres, à l'inverse, imposent la réalisation de l'ensemble des tâches en une et une seule commande et rendent impossible l'analyse des données produites avec d'autres méthodes. Dans le cadre de ce travail, nous proposons une démarche unifiée qui repose sur une représentation homogène des facteurs d'entrée et sur une décomposition de l'expérimentation numérique en cinq tâches principales : 1) choix des facteurs d'entrée et de leurs distributions d'incertitude ; 2) plan d'expérience ou d'échantillonnage pour déterminer les combinaisons de niveaux des facteurs à simuler ; 3) simulation proprement dite pour obtenir les sorties du modèle ; 4) analyse des résultats de simulation ; 5) restitution des résultats. Ainsi, toutes les démarches d'analyse de sensibilité peuvent être considérées comme des enchaînements partiels ou complets de ces tâches principales. Des filtres peuvent être ajoutés pour prendre en compte des situations atypiques et complexes. Ils sont souvent utilisés pour transformer ou importer des données dans le but de connecter les tâches principales.

Résultats : Exploitant la technologie S4 offerte par R, le package « mtk » repose sur une architecture orientée-objet à trois composants : la gestion des facteurs, la gestion du workflow et la gestion des accès aux ressources externes. Chaque composant assure une partie de services et gère l'échange de données et services avec les autres composants via des

interfaces communes. Ainsi, un composant ne connaît ses partenaires qu'à travers ces interfaces. Ceci permet de favoriser un développement plus efficace du projet, et par la même occasion, facilitera la maintenance évolutive et corrective du package car l'évolution ou la modification du package impactent surtout l'interaction entre les composants sans pour autant remettre en cause leur structure fondamentale.

Dans la version actuelle, seules les parties liées à l'analyse de sensibilité sont implémentées, et un nombre important de méthodes sont déjà couvertes telles *Fast*, *Sobol*, *Morris*, *Basic Monte-Carlo*, *Regression*, *LHS*, *PLMM* (*Polynomial Linear MetaModel*) [3], etc. Le package « *mtk* » gère nativement les modèles implémentés en R et comme une application externe. Des classes de type plugin sont disponibles, et elles permettent l'intégration des modèles implémentés dans les plateformes de modélisation. Les développements en cours concernent l'embarquement non-intrusive des modèles (qui peuvent être programmés en différents langages) dans les plateformes de calcul à haute performance.

Des outils pour l'extension du package sont fournis. Ils permettent l'intégration de nouveaux modèles ou méthodes écrits en R sans les reprogrammer, et ce de façon transparente et non intrusive.

Une autre spécificité du package « *mtk* » est sa capacité à prendre en charge la notion de virtualisation du calcul : d'une part, certaines tâches réalisées au sein d'un processus peuvent être implémentées en dehors du package, et d'autre part, les données produites sont conservées avec leur contexte de production. Un calcul sera refait si et seulement si un des éléments du contexte a été modifié. Ceci permet de maximiser la réutilisation des données tout en garantissant leur consistance.

L'exemple ci-dessous montre la simplicité d'utilisation du package sur le cas du modèle *Ishigami* où la définition des facteurs a été importée via la commande `data(Ishigami.factors)` :

```
# Chargement du package 'mtk' :
  library(mtk)
# Création des facteurs :
  data(Ishigami.factors)
# Création des tâches : design, evaluate, analyze.
  d <- mtkFastDesigner(listParameters=list(n=1000))
  e <- mtkIshigamiEvaluator()
  a <- mtkFastAnalyser()
# Création et lancement du workflow :
  w <- mtkExpWorkflow(expFactors=Ishigami.factors,
                     processesVector=c(design=d,evaluate=e,analyze=a))
  run(w)
# Mise en rapport des résultats :
  summary(w)
```

Références

- [1] Gilles Pujol, Bertrand Iooss, Alexandre Janon with contributions from Paul Lemaitre, Laurent Gilquin, Loic Le Gratiet, Taieb Touati, Bernardo Ramos, Jana Fruth and Sebastien Da Veiga (2014). The package « *sensitivity* », <http://cran.r-project.org/web/packages/sensitivity>.
- [2] Hervé Monod, Annie Bouvier, André Kobilinsky (2014). The package « *planor* », <http://cran.r-project.org/web/packages/planor>.
- [3] Faivre, R. (2013). Exploration par construction de métamodèles. In: Robert Faivre, B. Iooss, S. Mahévas, D. Makowski, Hervé Monod, Analyse de sensibilité et exploration de modèles. Applications aux sciences de la nature et de l'environnement (p. 159-194). Collection Savoir-Faire. Versailles, FRA : Editions Quae, 2013.