



MAIAGE, INRA, UNIVERSITÉ PARIS-SACLAY, 78350,
JOUY-EN-JOSAS, FRANCE

MÉMOIRE DE STAGE
M1 MATHÉMATIQUES APPLIQUÉES

Analyse de sensibilité sur des modèles à sorties dynamiques



Stagiaire : Romain Narci

Encadrants : Marie-Luce Taupin, Hervé Monod, Caroline Bidot

Professeur référent : Christine Keribin

2 Mai 2016 — 31 Août 2016

Table des matières

Remerciements	3
Introduction	4
Contexte	4
Objectifs	4
1 Analyse de sensibilité	6
1.1 Préambule	6
1.1.1 Indices de sensibilité de premier ordre et totaux	6
1.1.2 Méthode d'estimation des indices de sensibilité : eFAST	6
1.2 Analyse de sensibilité dynamique	7
1.2.1 Principe de la méthode	7
1.2.2 Mise en oeuvre dans R	7
1.3 Analyse de sensibilité après réduction de dimension	9
1.3.1 Principe général	9
1.3.2 Méthodes de réduction de dimension	10
1.3.3 Indices de sensibilité généralisés	11
1.3.4 Mise en oeuvre dans R	11
2 Classification	13
2.1 Qualité d'une classification	13
2.1.1 Notations	13
2.1.2 Inerties intra-cluster, inter-cluster et totale	13
2.2 Méthode des <i>K</i> -means	14
2.2.1 Algorithme	14
2.2.2 Remarques	15
2.2.3 Mise en oeuvre dans R	15
2.3 Influence des facteurs d'entrée sur les clusters	15
2.3.1 Comparaison des distributions des facteurs d'entrée	15
2.3.2 Analyse de sensibilité sur les distances	16
3 Application sur le modèle Mhydas (plateforme OpenFluid)	17
3.1 Le modèle	17
3.2 Analyse de sensibilité dynamique	19
3.2.1 Application sur le cas pluie faible	19
3.2.2 Application sur le cas pluie forte	21
3.3 Analyse de sensibilité après réduction de dimension	22
3.3.1 Application sur le cas pluie faible	23
3.3.2 Application sur le cas pluie forte	26
3.4 Classification	32
3.4.1 Partition obtenue par la méthode des <i>K</i> -means	32
3.4.2 Influence des facteurs d'entrée sur les 4 clusters	34
3.5 Conclusion sur le modèle Mhydas	36
Conclusion	38
Références	40
Annexes	41

Table des figures

1	Indices de sensibilité normalisés et non-normalisés : cas du modèle ToyCrop.	8
2	Sortie graphique de l'analyse de sensibilité par ACP : cas du modèle Cantis.	12
3	Représentation spatialisée d'un bassin versant défini par des parcelles et un réseau hydrographique.	17
4	Evolution en fonction du temps du débit à l'exutoire pour deux années climatiques.	19
5	Sortie graphique de l'analyse de sensibilité dynamique : cas pluie faible.	20
6	Indices de sensibilité totaux et d'interactions normalisés : cas pluie faible.	21
7	Sortie graphique de l'analyse de sensibilité dynamique : cas pluie forte.	22
8	Représentation des composantes de la base : cas pluie faible	23
9	Sortie graphique de l'analyse de sensibilité par ACP : cas pluie faible.	24
10	Sortie graphique de l'analyse de sensibilité sur base polynomiale : cas pluie faible.	25
11	Indices de sensibilité généralisés : cas pluie faible.	26
12	Représentation des composantes de la base : cas pluie forte.	27
13	Sortie graphique de l'analyse de sensibilité par ACP : cas pluie forte.	28
14	Influence de MultiKs et de MultiTheta sur H_1	29
15	Influence de MeanCellRS et de MultiKs sur H_2	30
16	Sortie graphique de l'analyse de sensibilité sur base polynomiale : cas pluie forte.	31
17	Indices de sensibilité généralisés : cas pluie forte.	32
18	Evolution en fonction du temps du débit à l'exutoire dans le cas pluie forte pour des couleurs différentes selon l'appartenance à un cluster.	33
19	Résultats de l'analyse de sensibilité sur les distances entre les courbes et les centres des clusters.	35
20	Evolution en fonction du temps des sorties : cas du modèle Cantis.	42
21	Evolution en fonction du temps de la biomasse accumulée pour deux années climatiques.	45
22	Evolution en fonction du temps de la fraction d'eau transpirable pour deux années climatiques.	46
23	Représentation de la base des polynômes : cas du modèle Cantis.	47
24	Influence de kz sur les H_j	48
25	Influence de kmz sur les H_j	48

Remerciements

Je tiens à remercier tout d'abord mes encadrants : Marie-Luce Taupin, Hervé Monod et Caroline Bidot. Leurs compétences et leurs disponibilités ont permis à mon stage de se dérouler d'une façon exemplaire. Je les remercie aussi pour leur bienveillance et la régularité du soutien qu'ils m'ont apporté, tout au long de l'élaboration de mon rapport de stage et la préparation de la soutenance. Un grand merci à Caroline qui a tout le temps été présente durant mon stage, que ce soit pour m'aider sur des points techniques mais aussi logistiques (par exemple le séjour de trois jours à l'INRA d'Avignon).

La collaboration avec le CATI IUMA, et en particulier avec Samuel Buis, Sébastien Roux, François Lafolie, David Crevoisier et Ronan Trépos, a été passionnante et je tiens à les en remercier.

Je remercie mon professeur référent Christine Keribin pour son suivi et ses excellents conseils tout au long de l'année du Master 1.

Je remercie enfin les membres de l'unité MaIAGE dont la bonne humeur et la bonne entente m'ont permis de m'épanouir dans cet environnement de la Recherche.

A tous, je dois mon envie croissante de poursuivre plus tard dans ce domaine.

Introduction

Contexte

Ce stage a été réalisé dans le cadre du Pari Scientifique « Approches multi-plateformes pour la résolution de problèmes avancés en Analyse de Sensibilité » du CATI ¹ IUMA ² et du réseau MEXICO ³. J'ai travaillé au sein de l'unité MaIAGE ⁴ de l'INRA ⁵ de Jouy-en-Josas en collaboration avec des chercheurs et des ingénieurs des centres INRA d'Avignon, de Montpellier et de Toulouse. L'objectif de ce projet est la mise à disposition pour les utilisateurs des plateformes de méthodes d'analyse de sensibilité adaptées aux sorties dynamiques. L'idée est de privilégier des méthodes plutôt génériques, en travaillant sur l'aspect pédagogique. L'unité MaIAGE s'intéresse à ces méthodes depuis plusieurs années. H. Monod, C. Bidot et M. Lamboni ont notamment développé un package R dédié, **multisensi**. Mon stage est basé sur l'utilisation de ce package afin de tester ces méthodes dans des contextes variés auxquels sont confrontés les modélisateurs de plusieurs plateformes de modélisation des départements EA ⁶ et MIA ⁷ de l'INRA. L'un des objectifs de ces plateformes est de mettre à disposition de leurs utilisateurs un ensemble de logiciels et de données permettant de développer, utiliser et interpréter les résultats de modèles agronomiques. Au cours de ce stage, nous avons étudié des méthodes adaptées d'analyse de sensibilité à ce type de modèles.

Plus précisément, nous avons travaillé sur l'application de méthodes d'analyse de sensibilité sur trois modèles à sorties dynamiques provenant de trois plateformes logicielles différentes :

- le modèle Cantis pour la plateforme Sol Virtuel (Avignon),
- le modèle Mhydas pour la plateforme OpenFluid (Montpellier),
- le modèle ToyCrop pour la plateforme Record (Toulouse).

Objectifs

Les objectifs de ce stage sont multiples :

- identifier des méthodes adaptées d'analyse de sensibilité pour des modèles à sorties dynamiques, les tester sur les modèles issus des plateformes ;
- sélectionner les plus pertinentes ;
- décrire les résultats de l'application de ces méthodes sur des modèles choisis ;
- et enfin, rédiger des documents pédagogiques destinés aux utilisateurs des plateformes.

Pour atteindre ces objectifs, une phase importante d'interaction avec les modélisateurs des plateformes a été indispensable avec en particulier un séjour de trois jours à Avignon pendant lequel, nous ont été présentées les trois plateformes logicielles et, nous avons présenté les résultats de l'application des méthodes adaptées d'analyse de sensibilité. Ce séjour a permis de soulever de nombreuses

¹Centre Automatisé du Traitement de l'Information

²Informatisation et Utilisation des Modèles dédiés aux Agro-Ecosystèmes

³Méthodes pour l'EXploration Informatique des modèles COmplexes

⁴Mathématiques et Informatique Appliquées du Génome à l'Environnement

⁵Institut National de la Recherche Agronomique

⁶Environnement et Agronomie

⁷Mathématiques et Informatique Appliquées

questions.

Les trois modèles étudiés, Cantis, Mhydas et ToyCrop, sont dits à sorties dynamiques, c'est-à-dire que les variables de sortie de chacun de ces modèles sont dépendantes du temps (une valeur pour plusieurs pas de temps sur la durée de la simulation). Pour de tels modèles, l'analyse de sensibilité est une étape indispensable qui permet de déterminer quels sont les facteurs d'entrée les plus influents sur la ou les sorties.

Nous nous sommes intéressés à deux types d'analyse de sensibilité :

- l'analyse de sensibilité dynamique permettant d'observer l'évolution de l'influence des facteurs d'entrée en fonction du temps,
- l'analyse de sensibilité après réduction de dimension permettant une analyse conjointe des sorties.

L'allure des sorties est très différente selon le modèle. Le premier volet de mon stage est de trouver une méthode d'analyse de sensibilité pertinente, quel que soit le jeu de données sur lequel elle est appliquée, et de définir une règle pour le choix de la méthode adaptée assurant une bonne interprétation des résultats, voire même ajoutant de l'information sur les facteurs d'entrée et le comportement de la sortie dynamique.

Après exécution de ces deux types d'analyse de sensibilité sur les trois modèles, nous avons pu comparer l'intérêt des différentes méthodes d'analyse de sensibilité et discuter de la méthode la plus robuste adaptée au modèle. Lors de mon séjour à Avignon s'est posée la question de la classification. Celle-ci correspond à un deuxième volet plus original que le premier, et pour laquelle les méthodes font encore l'objet de nombreuses recherches.

A l'issue de l'étape d'analyse, j'ai rédigé trois rapports, chacun correspondant à un modèle, dans lesquels j'ai décrit les résultats des deux types d'analyse de sensibilité. Ces trois rapports feront office de corrigé de TPs dont l'énoncé sera rédigé ultérieurement. Enfin, j'ai rédigé deux fiches méthodologiques décrivant de façon succincte mais précise ces deux types d'analyse de sensibilité. Ces documents seront destinés aux utilisateurs des plateformes.

Dans ce rapport de stage, je commence par présenter les différents types d'analyse de sensibilité, notamment celles adaptées pour des modèles à sorties dynamiques. Dans un deuxième temps, je présente une méthode complémentaire d'analyse de sensibilité basée sur une classification préliminaire et un algorithme associé. Enfin, je présente l'application de ces méthodes sur un jeu de données issu du modèle Mhydas de la plateforme OpenFluid en discutant les résultats obtenus.

1 Analyse de sensibilité

Exécuter une analyse de sensibilité sur un modèle a pour but d'identifier les facteurs d'entrée les plus influents sur les sorties d'intérêt, que ce soit en effet direct ou en effet d'interaction avec d'autres facteurs. Les facteurs d'entrée de l'analyse de sensibilité sont le plus souvent des paramètres du modèle dont les valeurs précises sont incertaines. Ce sont aussi, parfois, des variables d'entrée du modèle pouvant varier dans une gamme de valeurs.

1.1 Préambule

Au cours de ce stage, nous nous sommes intéressés à des méthodes d'analyse de sensibilité basées sur la variance. Ces méthodes permettent de déterminer quels facteurs d'entrée expliquent la plus grande part de la variabilité de la sortie. Les résultats sont des indices de sensibilité qui traduisent l'influence des différents facteurs d'entrée et de leurs interactions. Ces indices sont normalisés par la variance de la sortie et ont donc des valeurs comprises dans l'intervalle $[0; 1]$.

1.1.1 Indices de sensibilité de premier ordre et totaux

Soient X_1, \dots, X_K les facteurs d'entrée du modèle sélectionnés pour l'analyse de sensibilité et Y sa sortie. Les indices de sensibilité de premier ordre des $X_i, i \in \{1, \dots, K\}$, quantifient leur influence en tant qu'effet principal. Ces indices sont intéressants non seulement parce qu'ils expliquent généralement une grande part de la variabilité de la sortie mais aussi parce qu'ils sont aisés à calculer (seulement K indices si K facteurs ont été sélectionnés pour l'analyse de sensibilité). L'indice de sensibilité de premier ordre de X_i s'écrit :

$$SI_i = \frac{\text{Var}(\mathbb{E}(Y|X_i))}{\text{Var}(Y)}, \quad i = 1, \dots, K.$$

Les interactions entre les différents facteurs d'entrée ne sont pas à négliger. A la différence des indices de sensibilité de premier ordre, les indices d'interaction sont beaucoup plus difficiles à calculer. En effet, lorsque le nombre K de facteurs d'entrée augmente, le nombre d'interactions peut devenir très élevé : $\frac{K(K-1)}{2}$ interactions doubles et $2^K - K - 1$ interactions en tout. Par ailleurs, comme les indices de sensibilité de premier ordre et d'interaction sont normalisés par la variance de la sortie, leur somme vaut 1. Les indices de sensibilité totaux permettent de regrouper toutes ces interactions. L'indice de sensibilité total de X_i est défini comme la somme de l'indice de sensibilité de premier ordre de X_i et de toutes les interactions entre X_i et les autres facteurs d'entrée considérés dans l'analyse de sensibilité. Nous pouvons aussi le définir comme étant la différence entre l'ensemble des indices de sensibilité (ce qui vaut 1) et la somme des indices de sensibilité de premier ordre de tous les facteurs d'entrée à l'exception de X_i . Un tel indice s'écrit :

$$TSI_i = 1 - SI_{-i} = 1 - \frac{\text{Var}(\mathbb{E}(Y|X_{-i}))}{\text{Var}(Y)}, \quad i = 1, \dots, K,$$

où SI_{-i} désigne l'ensemble des indices de sensibilité de premier ordre de tous les facteurs d'entrée à l'exception de X_i et X_{-i} désigne l'ensemble des facteurs d'entrée à l'exception de X_i .

1.1.2 Méthode d'estimation des indices de sensibilité : eFAST

En général, les indices de sensibilité ne sont pas calculables analytiquement mais il existe plusieurs méthodes d'estimation. Déterminer celle qui est la plus intéressante n'a pas été la priorité au cours de ce stage. Nous allons ici présenter de façon succincte la méthode Extended Fourier Amplitude Sensitivity Test (eFAST, [2]) utilisée pour obtenir les plans d'expérience associés aux jeux de

données des modèles Cantis, Mhydas et ToyCrop. Le plan d'expérience (ou d'échantillonnage) est défini sous la forme de trajectoires discrètes dans le domaine \mathbf{D} des facteurs d'entrée, parcourant le domaine de variation des différents facteurs d'entrée avec des fréquences différentes. Si K correspond au nombre de facteurs d'entrée de l'analyse de sensibilité et N le nombre d'observations du modèle, chaque trajectoire est une suite de scénarios $x_j = (x_{j,1}, \dots, x_{j,K})$, $j \in \{1, \dots, N\}$. Une analyse fréquentielle permet alors d'estimer les indices de sensibilité, en recherchant les facteurs d'entrée dont les variations sont en phase avec celles de la sortie du modèle.

Le plan d'expérience est constitué de N combinaisons de valeurs des facteurs d'entrée. Les N valeurs prises par chaque facteur sont choisies dans le domaine \mathbf{D} leur étant associé. En l'occurrence pendant ce stage, les valeurs sont prises de façon uniforme. C'est à partir de ce plan d'expérience que sont estimés les indices de sensibilité de premier ordre et totaux.

Pour créer le plan d'expérience, nous utilisons la fonction `fast99` du package `sensitivity` :

```
sensitivity_result <- fast99(model, factors, n, q, q.arg, ...)
```

où `model` est une fonction, ou un modèle avec une méthode de prédiction, définissant le modèle à analyser, `factors` est le nombre K de facteurs d'entrée ou un vecteur de caractères donnant leur nom, `n` est la taille de l'échantillon telle que $n.K$ est égal au nombre N de simulations, `q` est un vecteur de noms des fonctions quantiles correspondant à la distribution des facteurs d'entrée et `q.arg` est une liste de paramètres des fonctions quantiles.

1.2 Analyse de sensibilité dynamique

Pour un modèle dynamique, l'analyse de sensibilité dynamique fournit des informations intéressantes sur l'évolution de l'influence des facteurs d'entrée en fonction du temps. La même méthode peut s'appliquer aussi pour des variables de sortie spatiales en une dimension. De façon plus formelle, si les variables de sortie sont Y_1, \dots, Y_T où $t = 1, \dots, T$ représente le temps et Y_t une sortie à l'instant t , alors l'analyse de sensibilité dynamique consiste en T analyses de sensibilité successives sur Y_1, \dots, Y_T . Elle permet donc d'observer comment évolue l'influence des facteurs d'entrée pour chaque Y_t .

1.2.1 Principe de la méthode

La méthode est basée sur le calcul des indices de sensibilité de premier ordre et totaux de chaque facteur d'entrée du modèle pour chaque pas de temps indépendamment. Ceux de premier ordre sont calculés comme suit :

$$SI_i^t = \frac{\text{Var}(\mathbb{E}(Y_t|X_i))}{\text{Var}(Y_t)} \text{ avec } i = 1, \dots, K \text{ et } t = 1, \dots, T.$$

Nous pouvons obtenir de la même manière à partir de l'expression TSI_i du §1.1.1 les indices de sensibilité totaux. La spécificité de cette méthode est d'obtenir une représentation numérique et graphique de ces indices. En ce qui concerne la représentation numérique, une valeur est attribuée aux indices de chaque facteur d'entrée en fonction du temps. Ces valeurs sont comprises dans l'intervalle $[0; 1]$.

1.2.2 Mise en oeuvre dans R

On utilise la fonction `multisensi` du package du même nom :

```
X <- multisensi(design=sensitivity_result,model=Y,reduction=NULL,center=FALSE,
analysis=analysis.sensitivity);
print(X); plot(X,normalized=TRUE,color=heat.colors)
```

Ici, `sensitivity_result` est le plan d'expérience, `Y` est le data.frame correspondant à la sortie du modèle. Pour faire une analyse de sensibilité dynamique, l'argument `reduction=NULL` doit être utilisé. Enfin, nous ne centrons pas nos données et l'argument `analysis=analysis.sensitivity` du package `sensitivity` est utilisé pour obtenir les indices de sensibilité.

La commande `print(X)` permet d'obtenir les valeurs numériques des indices de sensibilité de premier ordre et totaux de chaque facteur d'entrée en fonction du temps.

Enfin, la commande `plot(X,normalized=TRUE,color=heat.colors)`, respectivement avec l'option `normalized=FALSE`, permet d'obtenir le graphe de gauche de la figure 1, respectivement le graphe de droite.

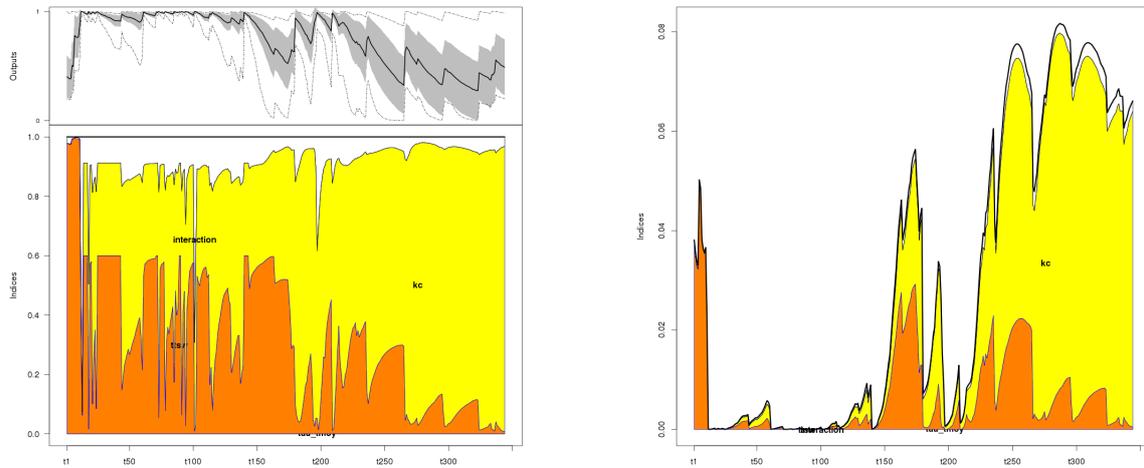


FIG. 1: Analyse de sensibilité dynamique : cas du modèle ToyCrop. Colonne de gauche : indices normalisés. Colonne de droite : indices non-normalisés.

Dans notre application, nous avons utilisé le modèle ToyCrop de la plateforme Record. La sortie du modèle a été calculée sur 366 pas de temps de 1 jour et nous avons mesuré l'influence de trois facteurs d'entrée (`kc`, `ttsw`, et `tau_tsmoy`) sur cette sortie. L'une des principales questions posées par l'analyse de sensibilité dynamique est celle de la normalisation, qui gomme les variations de la variance dans les représentations graphiques. Nous avons donc comparé plusieurs façons de la gérer graphiquement (cf. Figure 1). Après interactions avec les plateformes, il paraît utile de conserver les deux modes de représentation.

Sur la partie supérieure du graphe de gauche, nous pouvons observer certaines valeurs caractéristiques de la distribution empirique (issue des simulations des Y_t) de la variable en fonction du temps : les extrêmes (en pointillé), la moyenne \pm l'écart-type (en gris) et la moyenne (en noir). Il est ainsi possible de voir comment évolue la variabilité de Y_t en fonction du temps.

Sur la partie inférieure du graphe de gauche sont représentés les effets principaux de chaque facteur d'entrée ainsi que la part d'influence de l'ensemble des interactions pour chaque pas de temps. Les indices de sensibilité sont ici normalisés par la variance des Y_t .

Le graphe de droite correspond quant à lui à une représentation des indices de sensibilité non-normalisés par la variance des Y_t .

1.3 Analyse de sensibilité après réduction de dimension

Pour un modèle dynamique, l'analyse de sensibilité après réduction de dimension (basée sur des techniques multivariées) donne une vue plus synthétique de l'impact des facteurs d'entrée sur les variables de sortie que l'analyse de sensibilité dynamique. Elle consiste en la projection des variables de sortie sur une base déterminée par une méthode de réduction de dimension suivie d'une analyse de sensibilité sur les composantes de cette base. Elle permet d'exécuter une analyse conjointe des Y_t . Il existe différentes méthodes de réduction de dimension. Au cours de ce stage, nous nous sommes intéressés à l'analyse en composantes principales (ACP) et à la projection sur base polynomiale, de B-splines et de O-splines.

1.3.1 Principe général

Pour ces quatre méthodes, le principe est le même : on effectue des transformations des variables de sortie (projection sur une base déterminée par la méthode), on les écrit comme une combinaison linéaire des éléments de la base et on exécute alors l'analyse de sensibilité sur les coefficients de la combinaison linéaire. Soit $Y(X, t)$ une variable de sortie dynamique avec X le vecteur des combinaisons des facteurs d'entrée et t le temps. Dans le cas général où l'on centre la variable de sortie, nous pouvons écrire :

$$Y(X, t) = \bar{Y}(t) + \sum_{j=1}^p H_j(X) \cdot L_j(t) + \epsilon(t)$$

où $\bar{Y}(t)$ est la moyenne de la variable de sortie pour chaque pas de temps, p est le nombre de composantes de la base orthonormée retenue par la méthode, $H_j(X)$ correspond aux coefficients dépendants des facteurs d'entrée, $L_j(t)$ correspond aux composantes de la base et $\epsilon(t)$ représente l'erreur d'une telle écriture en ne gardant uniquement que les composantes permettant d'expliquer un certain pourcentage d'inertie. L'inertie totale est proportionnelle à la somme des carrés totale. Soit $SCT = \sum_{t=1}^T \sum_{i=1}^N (Y_{i,t} - \bar{Y}(t))^2$ la somme des carrés totale, où $Y_{i,t}$ est la i -ème observation de la sortie dynamique au temps t , N est le nombre d'observations et T le nombre de pas de temps, et I_{tot} l'inertie totale. Alors nous avons la relation suivante :

$$SCT = N \cdot I_{tot}.$$

En remplaçant $\sum_{i=1}^N (Y_{i,t} - \bar{Y}(t))^2$ par $(N - 1)\text{Var}[Y(X, t)]$ où $\text{Var}[Y(X, t)]$ est la variance empirique sans biais de la sortie dynamique, nous obtenons l'égalité suivante :

$$I_{tot} = \frac{N - 1}{N} \sum_{t=1}^T \text{Var}[Y(X, t)].$$

Nous cherchons une base permettant de bien représenter la variabilité de la sortie, c'est-à-dire une base permettant d'expliquer un pourcentage d'inertie suffisant. En général, nous considérons qu'expliquer 95% d'inertie est équivalent à bien représenter la variabilité de la sortie dans le cas où la projection des variables de sortie est faite sur une base orthonormée.

Les indices de sensibilité de premier ordre sont calculés pour chaque facteur d'entrée sur chacune des composantes de la base :

$$SI_i^j = \frac{\text{Var}(\mathbb{E}(H_j|X_i))}{\text{Var}(H_j)} \text{ avec } i = 1, \dots, K \text{ et } j = 1, \dots, p.$$

Nous pouvons obtenir d'une façon équivalente les indices de sensibilité totaux. Les résultats sont illustrés graphiquement par la représentation de la variabilité des composantes $L_j(t)$ pondérée par $H_j(X)$, par celle des indices de sensibilité sur les premières composantes et par celle des indices de sensibilité généralisés (voir Figure 2 plus bas).

1.3.2 Méthodes de réduction de dimension

Il existe plusieurs bases de réduction de dimension : composantes principales, polynomiale, B-splines, O-splines. Une analyse de sensibilité est alors applicable sur les composantes de chacune de ces bases. Dans l'analyse en composantes principale (ACP), la base est calculée à partir des données, alors que dans les autres méthodes la base est fixée indépendamment des données. L'utilisation d'une méthode ou d'une autre dépendra des objectifs que nous nous fixons. Dans la suite, nous allons décrire de façon succincte le principe de chacune de ces méthodes.

Base de composantes principales L'idée est de traiter la variable de sortie dynamique qui nous intéresse par une analyse en composantes principales (ACP), puis d'exécuter une analyse de sensibilité sur chaque composante principale de la base obtenue. Nous gardons le nombre de composantes principales permettant de bien expliquer la variabilité de la sortie. Réduire les variables lorsqu'elles sont homogènes, en particulier lorsqu'elles ont les mêmes unités de mesure, fausse les résultats de l'analyse de sensibilité. Le centrage est par contre important pour étudier les écarts à la moyenne et non pas les écarts à 0.

Base polynomiale L'idée est de projeter la variable de sortie dynamique qui nous intéresse sur une base polynomiale orthonormée, puis d'exécuter une analyse de sensibilité sur chaque polynôme. Il est nécessaire de préciser quel est le degré maximal que pourront prendre les polynômes de la base. La méthode pour choisir ce degré est la suivante : nous commençons par prendre le degré par défaut qui est le degré 3. Notre objectif est de bien représenter la variabilité de la sortie avec une base de polynôme la plus petite possible. Une telle base peut être obtenue en élevant ou en diminuant la valeur du degré par défaut.

Base de B-splines L'idée est de projeter la variable de sortie dynamique qui nous intéresse sur une base de B-splines, puis d'exécuter une analyse de sensibilité sur chaque B-spline (B-spline, [12]). Soient $N + 1$ noeuds m_i , $i \in \{1, \dots, N\}$, tels que $0 \leq m_0 \leq m_1 \leq \dots \leq m_N \leq 1$. On pose $B_{j,k}$ la j -ème fonction B-spline de degré k . Par récurrence sur le degré inférieur, nous avons :

$$B_{j,0}(m) = \begin{cases} 1 & \text{si } m_j \leq m \leq m_{j+1} \\ 0 & \text{sinon} \end{cases},$$

$$B_{j,k}(m) = \frac{m - m_j}{m_{j+k} - m_j} B_{j,k-1}(m) + \frac{m_{j+k+1} - m}{m_{j+k+1} - m_{j+1}} B_{j+1,k-1}(m).$$

De plus, une B-spline $B_{j,k}(m)$ est à valeurs strictement positives si $m \in [m_j, m_{j+k+1}]$ et à valeurs nulles sinon.

Pour appliquer une analyse de sensibilité sur une base de B-splines, il est nécessaire de préciser le degré et le nombre de noeuds maximal que nous voulons dans la base. Pour choisir le degré et le nombre de noeuds, nous utilisons la méthode suivante : l'objectif est d'atteindre une base de B-splines expliquant le plus grand pourcentage d'inertie avec un nombre satisfaisant de noeuds et un degré maximal pas trop élevé. La part d'inertie expliquée par la base de B-splines n'est pas restreinte à 100% du fait qu'elle n'est pas orthonormée. Par exemple, elle peut parfois dépasser les 200% d'inertie expliquée. L'intérêt de la projection sur base de B-splines est de décomposer

l'intervalle d'indices de temps de la variable de sortie en un nombre fini de noeuds égal à la taille de la base. Chaque B-spline correspond à un sous-intervalle d'indices de temps et l'analyse de sensibilité est exécutée sur chacune de ces B-splines. Il est donc possible d'analyser les indices de sensibilité sur un intervalle d'indices de temps très fin.

Base de O-splines L'idée est de projeter la variable de sortie dynamique qui nous intéresse sur une base de O-splines, puis d'exécuter une analyse de sensibilité sur chaque O-spline. Les O-splines correspondent à des B-splines orthonormalisées. L'objectif est d'atteindre une base de O-spline permettant de bien expliquer la variabilité de la sortie avec un nombre satisfaisant de noeuds et un degré maximal pas trop élevé. A la différence de la base des B-splines, celle des O-splines ne peut pas expliquer plus de 100% d'inertie.

1.3.3 Indices de sensibilité généralisés

Les indices de sensibilité généralisés (GSI) représentent les contributions des facteurs d'entrée sur la dynamique globale de la variable de sortie. Cela permet d'avoir des informations sur l'influence générale des facteurs d'entrée. De plus, il est utile d'étudier les GSI obtenus par les différentes méthodes de réduction de dimension à titre de comparaison. Soit X_i un facteur d'entrée d'un modèle et soit Y sa sortie. Soit p le nombre de composantes de la base obtenue par une méthode de réduction de dimension. On pose SI_i^j , $j \in \{1, \dots, p\}$, l'indice de sensibilité de premier ordre du facteur X_i sur la j -ème composante. Alors l'indice de sensibilité généralisé de premier ordre de X_i s'écrit :

$$GSI_i = \frac{\sum_{j=1}^p SI_i^j \cdot \text{Var}(H_j)}{\sum_{j=1}^p \text{Var}(H_j)}$$

où H_j a été défini au §1.3.1.

Nous pouvons obtenir de la même façon l'indice de sensibilité généralisé total $TGSI_i$ de X_i en remplaçant SI_i^j par TSI_i^j , où TSI_i^j représente l'indice de sensibilité total du facteur X_i sur la j -ème composante.

1.3.4 Mise en oeuvre dans R

Nous utilisons la fonction `multisensi` du package du même nom :

```
X <- multisensi(design=sensitivity_result,model=Y,reduction=S,scale=FALSE,
center=FALSE, analysis=analysis.sensitivity);
summary(X); plot(X)
```

Ici, `sensitivity_result` est le plan d'expérience, `Y` est le data.frame correspondant à la sortie du modèle. Pour faire une analyse de sensibilité après réduction de dimension, l'argument `reduction` doit être utilisé. A la place de `S`, nous pouvons écrire `basis.ACP`, `basis.poly`, `basis.bspline` ou `basis.ospline` pour appliquer une analyse de sensibilité respectivement sur une base de composantes principales, polynomiale, de B-splines et de O-splines. Enfin, nous utilisons l'argument `analysis=analysis.sensitivity` du package `sensitivity` pour obtenir les indices de sensibilité.

La commande `summary(X)` permet d'obtenir les valeurs numériques des indices de sensibilité de premier ordre et totaux de chaque facteur d'entrée sur les composantes de la base ainsi que les indices de sensibilité généralisés.

La commande `plot(X)` permet d'obtenir par exemple la figure 2 dans le cas de l'ACP (les commandes `plot(X,graph=1)` et `plot(X,graph=2)` permettant respectivement d'obtenir le graphe de gauche et le graphe de droite).

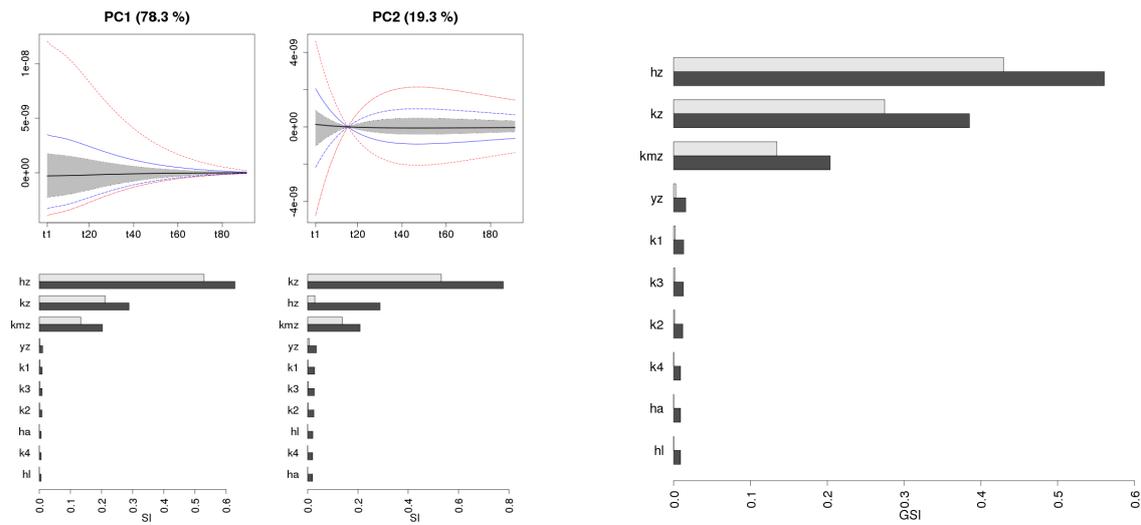


FIG. 2: Colonne de gauche : sortie graphique de l'analyse de sensibilité (Ligne du haut : variabilité des composantes principales $L_j(t)$ pondérée par $H_j(X)$. Ligne du bas : indices de sensibilité sur les premières composantes principales). Colonne de droite : indices de sensibilité généralisés.

Dans cet exemple (Figure 2), nous avons utilisé le modèle Cantis de la plateforme Sol Virtuel. La sortie du modèle a été calculée sur 91 pas de temps de 1 jour et nous avons mesuré l'influence de dix facteurs d'entrée sur cette sortie.

Sur la partie supérieure du graphe de gauche, nous pouvons observer la forme et la variabilité des composantes principales $L_j(t)$ pondérée par $H_j(X)$: les extrêmes (en rouge), les 1/10 et 9/10 percentiles (en bleu), l'écart inter-quartile (en gris) et la médiane (en noir). Les indices de sensibilité de premier ordre (en gris) et totaux (en noir) sur les premières composantes principales sont représentés sur la partie inférieure.

Les indices de sensibilité généralisés de premier ordre et totaux sont représentés sur le graphe de droite.

2 Classification

Appliquer une méthode de classification sur un modèle permet de répartir ses sorties en classes (ou clusters) homogènes afin de cibler des comportements bien distincts. De telles techniques peuvent être intéressantes lorsque les valeurs des facteurs d'entrée conduisent à des dynamiques des Y_t de formes très différentes. Notre objectif est de pouvoir ensuite évaluer quels sont les facteurs d'entrée qui influent le plus sur le comportement de la sortie du modèle.

Les méthodes de classification font encore l'objet de recherche. Elles ne sont pas implémentées dans la package `multisensi`. Cette section correspond donc à un volet plus original de mon rapport que la précédente.

2.1 Qualité d'une classification

L'évaluation de la qualité d'une classification se fait à travers le calcul des inerties intra-cluster et inter-cluster ([10]) définies ci-dessous.

2.1.1 Notations

Soient N le nombre d'observations du modèle à sortie dynamique et T le nombre de pas de temps. On pose $Y_i \in \mathbb{R}^T$, $i \in \{1, \dots, N\}$, la i -ème sortie dynamique du modèle (ou plus généralement à T attributs). De plus, soient K le nombre de clusters obtenus par un algorithme de classification et C_k , $k \in \{1, \dots, K\}$, le k -ième cluster. Enfin, nous introduisons une notion de distance entre deux points $Y_1, Y_2 \in \mathbb{R}^T$ par la distance euclidienne par exemple :

$$d(Y_1, Y_2) = \sqrt{\sum_{t=1}^T (Y_{1,t} - Y_{2,t})^2}.$$

2.1.2 Inerties intra-cluster, inter-cluster et totale

Inertie intra-cluster L'inertie intra-cluster mesure la concentration des points d'un cluster autour de son centre de gravité. Une inertie intra-cluster faible implique une petite dispersion des points autour du centre de gravité. Soient G_k le centre de gravité du cluster C_k et I_k son inertie. Nous avons alors :

$$G_k = \frac{1}{\text{card}(C_k)} \sum_{i \in C_k} Y_i,$$
$$I_k = \sum_{i \in C_k} d^2(Y_i, G_k).$$

L'inertie intra-cluster I_{intra} est la somme des inerties de chaque cluster, soit :

$$I_{\text{intra}} = \sum_{k=1}^K I_k.$$

La partition sera choisie de façon à minimiser cette inertie.

Inertie inter-cluster L'inertie inter-cluster mesure la distance entre les centres des clusters. Une inertie inter-cluster forte implique une bonne séparation des clusters. Soit $\bar{Y} \in \mathbb{R}^T$ le centre de gravité de l'ensemble des observations Y_i . Il s'écrit :

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i.$$

L'inertie inter-cluster I_{inter} s'obtient alors par la formule suivante :

$$I_{\text{inter}} = \sum_{k=1}^K \text{card}(C_k) d^2(G_k, \bar{Y}).$$

La partition sera choisie de façon à maximiser cette inertie.

Inertie totale L'inertie totale I_{totale} des points est la somme de l'inertie intra-cluster et de l'inertie inter-cluster. C'est aussi la somme des carrés totale $\text{SCT} = \sum_{t=1}^T \sum_{i=1}^N (Y_{i,t} - \bar{Y}(t))^2$. Sa valeur est fixe, donc minimiser l'inertie intra-cluster revient à maximiser l'inertie inter-cluster.

2.2 Méthode des K -means

La méthode des K -means (K -moyennes, [10]) permet d'obtenir au moins une bonne partition (mais pas nécessairement la partition optimale) au sens de l'inertie intra-cluster. L'objectif est donc de trouver un minimum local de cette inertie.

2.2.1 Algorithme

L'algorithme des K -means est constitué de quatre étapes.

1ère étape Initialisation (aléatoire ou non) des centres de gravité G_k , $k \in \{1, \dots, K\}$.

2ème étape Affectation de chaque observation Y_i au cluster pour lequel le centre de gravité est le plus proche du point Y_i :

$$C_l \leftarrow Y_i \text{ tel que } l = \underset{k}{\operatorname{argmin}} d(Y_i, G_k).$$

3ème étape Calcul des centres de gravité G_k des K clusters C_k obtenus :

$$G_k = \frac{1}{\text{card}(C_k)} \sum_{i \in C_k} Y_i$$

4ème étape Répétition des étapes 2 et 3 jusqu'à arrêt de l'algorithme ayant lieu :

- soit lorsque deux itérations successives conduisent à une même partition,
- soit lorsque le critère d'arrêt est atteint (par exemple un nombre maximal d'itérations).

2.2.2 Remarques

Si n est le nombre d'itérations de l'algorithme, celui-ci à une complexité de l'ordre de $O(KNn)$. A chaque itération, l'inertie intra-cluster diminue. L'algorithme converge rapidement vers au moins un minimum local de l'inertie intra-cluster. Il faut savoir que des initialisations différentes des centres de gravité dans l'ensemble des Y_i peuvent mener à des partitions différentes. Pour obtenir une partition stable de clusters, nous pouvons exécuter plusieurs segmentations et ne considérer que la meilleure (c'est-à-dire celle d'inertie intra-cluster minimale). Enfin, choisir le nombre de clusters est délicat. Le nombre peut être fixé *a priori*. Nous pouvons aussi chercher la meilleure (à K fixé) partition possible pour différents $K > 1$ et chercher à partir de quel K l'inertie intra-cluster pour chaque partition ne diminue plus que légèrement en fonction du nombre de clusters. Une autre possibilité est d'imposer des contraintes sur la densité des clusters obtenus ([10]).

2.2.3 Mise en oeuvre dans R

Pour exécuter l'algorithme des K -means sur R, nous utilisons la commande suivante :

```
kmeans(Y, centers = K, iter.max = I, nstart = L)
```

où Y est le data.frame correspondant au jeu de données, **centers** le nombre de clusters choisi (ici K) ou alors la position des centres des clusters, **iter.max** le nombre maximal d'itération permis (ici I) et **nstart** le nombre de segmentations (ici L) parmi lesquelles sera choisie la meilleure partition au sens de l'inertie intra-cluster.

2.3 Influence des facteurs d'entrée sur les clusters

L'étape suivante consiste à procéder à une analyse de sensibilité sur les différents clusters obtenus. L'idée est d'évaluer quels sont les facteurs d'entrée qui influent le plus sur le comportement de la sortie du modèle et donc sur l'appartenance aux différents clusters. Les méthodes d'influence des facteurs d'entrée sur les clusters sont peu explorées pour le moment. Dans la suite du rapport, nous allons présenter deux idées intuitives que nous avons eues.

2.3.1 Comparaison des distributions des facteurs d'entrée

Dans mes enquêtes bibliographiques, j'ai trouvé une méthode intéressante proposée par Saltelli et al. (voir [11]). La première idée est de comparer les distributions des facteurs d'entrée conduisant aux clusters. Le test de Kolmogorov-Smirnov ([13]) permet d'estimer si deux échantillons suivent la même loi ou si un échantillon provient bien d'une loi donnée. Si $C = (x_1, \dots, x_n)$ est un échantillon de n variables aléatoires indépendantes, alors nous pouvons définir la fonction de répartition empirique $F_C(x)$ de cet échantillon par la formule suivante :

$$F_C(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i \leq x}, \text{ où } n = \text{card}(C) \text{ et}$$
$$\delta_{x_i \leq x} = \begin{cases} 1 & \text{si } x_i \leq x, \\ 0 & \text{sinon.} \end{cases}$$

Si nous voulons comparer les distributions de deux échantillons C_1 et C_2 , nous mesurons l'écart maximum d_1 qui existe entre la fonction de répartition empirique F_{C_1} de C_1 et la fonction de répartition empirique F_{C_2} de C_2 :

$$d_1 = \max_x |F_{C_1}(x) - F_{C_2}(x)|.$$

Si nous voulons déterminer que l'échantillon C provient bien d'un tirage aléatoire d'une loi de fonction de répartition théorique F , alors nous comparons l'écart maximum d_2 qui existe entre cette dernière et la fonction de répartition empirique F_C de C :

$$d_2 = \max_x |F(x) - F_C(x)|.$$

L'hypothèse nulle H_0 est « les deux échantillons suivent la même distribution » et l'hypothèse alternative H_1 est « les deux échantillons ne suivent pas la même distribution ». Donc sous H_0 , d_1 et d_2 tendent vers 0 quand n tend vers $+\infty$. Rejeter l'hypothèse H_0 implique alors que les deux échantillons C_1 et C_2 ne suivent pas la même loi (cas d_1) ou l'échantillon C ne provient pas d'un tirage aléatoire d'une loi de fonction de répartition théorique F (cas d_2).

Dans notre cas, plusieurs options d'application du test de Kolmogorov-Smirnov sont possibles :

- comparer les distributions des facteurs d'entrée pour chaque cluster deux à deux ;
- comparer la distribution associée à chaque cluster k avec celle de l'ensemble des clusters excepté le k -ième, pour chaque facteur d'entrée indépendamment ;
- comparer la distribution associée à chaque cluster k avec celle de l'ensemble des sorties, pour chaque facteur d'entrée indépendamment.

C'est la deuxième option que nous avons retenu (ce que Saltelli et al. appellent analyse de sensibilité régionale, voir [11]). Saltelli et al. montrent que cette procédure est valide dans le cas où la réponse est binaire et les simulations faites sont du type Monte-Carlo. Les résultats ne s'appliquent pas directement dans notre cas pour au moins deux raisons :

- les simulations faites sont du type eFAST ;
- la réponse catégorielle est issue d'une étape de classification et non pas des simulations initiales de la sortie.

Rejeter H_0 pour un risque α faible implique que X_i est un facteur clé dans la production d'un certain comportement de la sortie. Saltelli et al. proposent plusieurs critères subjectifs pour quantifier l'influence d'un facteur d'entrée X_i . Plus précisément, rejeter H_0 au risque α dans les cas où $\alpha < 1\%$, $\alpha \in [1\%; 10\%]$ et $\alpha > 10\%$ implique respectivement que X_i est critique, important et sans importance.

2.3.2 Analyse de sensibilité sur les distances

La deuxième idée est de créer un nouveau jeu de données, évidemment basé sur le précédent mais aussi sur la notion de distance définie dans le §2.1.1. La variable D_k , de taille N le nombre d'observations, contient la valeur de la distance entre chaque courbe (correspondant à une observation parmi N) et le centre de gravité du cluster k . Par exemple, la distance entre la i -ème courbe et le centre de gravité du cluster k est égale à D_k^i . Nous appliquons alors une analyse de sensibilité sur ces distances. Un facteur d'entrée influence sur les distances entre chaque courbe et le cluster k implique que c'est un facteur clé dans la production d'un certain comportement de la sortie correspondant au cluster k .

3 Application sur le modèle Mhydas (plateforme OpenFluid)

Au cours de ce stage, nous avons appliqué différents types d'analyse de sensibilité et une classification sur des jeux de données provenant des trois modèles Cantis, Mhydas et ToyCrop. Nous allons ici présenter les résultats obtenus par ces méthodes dans le cas du modèle Mhydas de la plateforme OpenFluid. En **Annexe A** et en **Annexe B** sont présentés respectivement le modèle Cantis de Sol Virtuel et le modèle ToyCrop de Record.

3.1 Le modèle

Le modèle Mhydas (Modélisation HYdrologique Des AgroSystèmes) de la plateforme de modélisation/simulation OpenFluid implémente une modélisation hydrologique distribuée qui s'appuie sur un découpage de l'espace en unités de production de surface (Surface Units, SU) et unités d'écoulement linéaires (Reach Segments, RS). Des processus hydrologiques sont simulés sur ces unités, c'est-à-dire partition infiltration-ruissellement et transferts sur les SU, transferts dans les RS. Dans notre cas, nous cherchons à simuler le débit à l'exutoire du bassin versant défini par des parcelles et des fossés (représentation spatialisée) en réponse à un événement pluvieux. Nous nous plaçons dans une configuration simpliste mais réaliste : sous-ensemble d'un bassin versant réel, comprenant l'exutoire. Dans la figure 3 est représenté le domaine spatialisé du bassin versant.

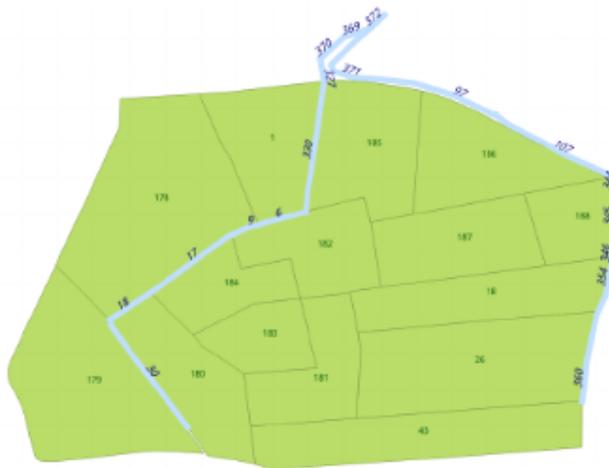


FIG. 3: Représentation spatialisée d'un bassin versant défini par des parcelles et un réseau hydrographique.

Facteurs d'entrée Les facteurs d'entrée du problème d'analyse de sensibilité sont cinq paramètres du modèle :

- **MultiKs** : coefficient multiplicatif de la conductivité à saturation de l'ensemble des parcelles (sans unité).
- **MultiThetaI** : coefficient multiplicatif de l'humidité initiale de l'ensemble des parcelles (sans unité).
- **MeanCellSU** : célérité moyenne de l'onde de crue sur les parcelles (m.s^{-1}).

- **MeanCellRS** : célérité moyenne de l’onde de crue sur le réseau hydrographique (m.s^{-1}).
- **MeanSigma** : diffusivité moyenne de l’onde de crue sur les parcelles et le réseau hydrographique ($\text{m}^2.\text{s}^{-1}$).

Nous pouvons définir les bornes des intervalles d’incertitude de ces cinq facteurs d’entrée :

	MultiKs	MultiThetaI	MeanCellSU	MeanCellRS	MeanSigma
Minimum	0,1	0,25	0,020	0,2	250
Maximum	2,5	1,15	0,080	0,8	750

Variable de sortie La sortie du modèle qui nous intéresse est le **débit à l’exutoire** (en $\text{m}^3.\text{s}^{-1}$) sur 73 pas de temps de 5 minutes. Nous allons travailler sur cette sortie en fonction de deux pluies contrastées (le cas **pluie faible** et le cas **pluie forte**). L’étude de la sortie dans le cas **pluie faible** sera indépendante de celle dans le cas **pluie forte**. Nous décidons d’appliquer la méthode eFAST avec $N = 1500$. Le plan final est constitué de 7500 combinaisons de valeurs des facteurs d’entrée.

La valeur du **débit à l’exutoire** en fonction du cas **pluie faible** est comprise dans le domaine de variation $[0 ; 0,16]$. Pour chaque pas de temps, nous pouvons calculer l’écart-type du **débit à l’exutoire**. 73 écarts-type sont donc obtenus avec les caractéristiques suivantes :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0	1,1e-04	3,7e-04	2,4e-03	2,3e-03	1,7e-02

La valeur du **débit à l’exutoire** en fonction du cas **pluie forte** est comprise dans le domaine de variation $[0 ; 0,75]$. 73 écarts-type sont obtenus avec les caractéristiques suivantes :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0	1,2e-03	3,8e-03	2,1e-02	2e-02	1,4e-01

Dans la figure 4, l’évolution en fonction du temps du **débit à l’exutoire** en fonction des deux années climatiques est représentée.

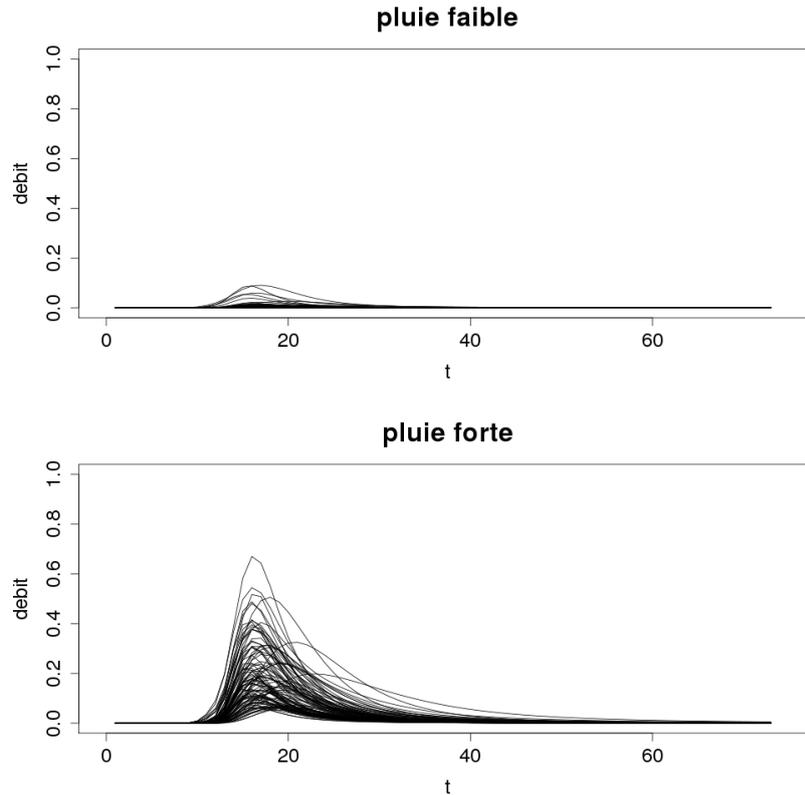


FIG. 4: Evolution en fonction du temps du débit à l'exutoire pour deux années climatiques et 100 jeux de facteurs d'entrée différents.

Les courbes ont le même comportement selon le type de pluie (translation/dilatation de la courbe de débit) du fait qu'il n'y a qu'un seul pic de précipitation. Nous pouvons d'ores et déjà retenir une caractéristique intéressante des sorties : un pic plus ou moins important du débit selon le type de pluie un peu avant le temps 20.

Pour faciliter la lecture dans la suite, il arrivera que l'indice de temps, i , soit remplacé par $5(i - 1)$ afin d'avoir directement la conversion des temps en minutes. Donc le temps t_i correspondra au temps $t = 5(i - 1)$ minutes.

Numériquement, nous avons pu voir que la valeur maximale de la variance dans le cas **pluie faible** et dans le cas **pluie forte** se situe à 75 minutes. De plus, nous avons aussi pu remarquer que la moitié des courbes obtenues ont uniquement des valeurs nulles pour le cas **pluie faible** et non pour le cas **pluie forte**.

3.2 Analyse de sensibilité dynamique

Nous allons observer l'évolution de l'influence des facteurs d'entrée en fonction du temps dans le cas **pluie faible** et **pluie forte** par une analyse de sensibilité dynamique.

3.2.1 Application sur le cas pluie faible

Dans la figure 5 est représenté le graphique des indices de sensibilité normalisés.

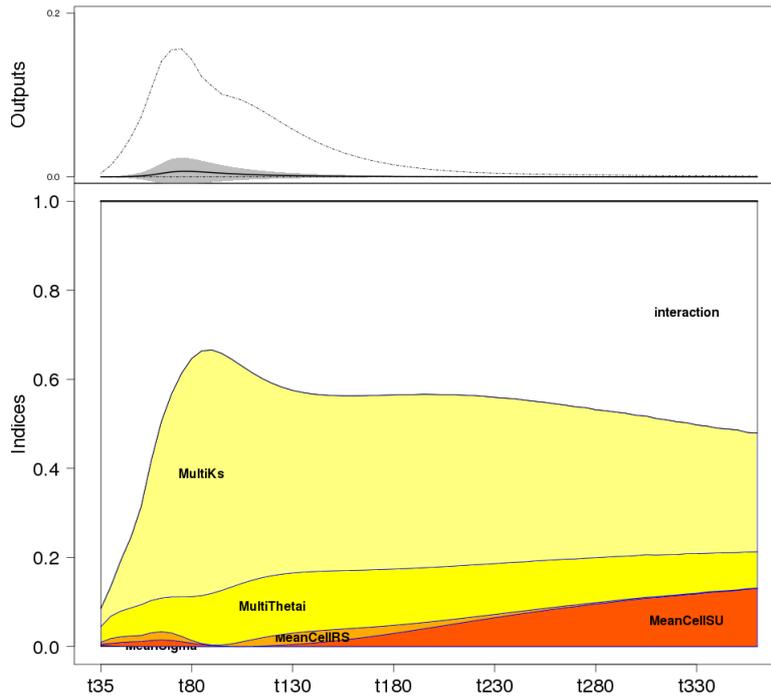


FIG. 5: Graphique des indices de sensibilité : cas pluie faible. Ligne du haut : certaines valeurs caractéristiques de la distribution empirique de la variable en fonction du temps. Ligne du bas : indices de sensibilité principaux en fonction du temps.

Nous voyons que la variance est proportionnelle au signal. **MultiKs** est un facteur d'entrée très influent tout le long de la simulation. Nous pouvons voir qu'à 85 minutes (1h25), il y a un pic au niveau de sa courbe des indices de sensibilité. Il en est de même pour la variance de la sortie 10 minutes avant. Nous en déduisons que **MultiKs** possède une grande influence sur la variance de la sortie. Le fort effet de **MultiKs** est attendu car il impacte directement la partition ruissellement/infiltration. Les autres facteurs d'entrée sont moins influents. L'indice de sensibilité de **MultiThetaI** diminue légèrement en fonction du temps à partir de 85 minutes. L'effet assez régulier dans le temps de **MultiThetaI** est probablement dû au fait qu'il y a peu de pluie. Les facteurs d'entrée **MeanCellRS** et **MeanSigma** possèdent une influence quasiment nulle tout le long de la simulation. **MeanCellSU** possède une influence quasiment nulle au début de la simulation. A partir de 105 minutes (1h45), il devient de plus en plus influent jusqu'à la fin de la simulation. L'importance de ce facteur d'entrée en fin de simulation est à relativiser car la variance de la sortie est quasiment nulle dans le dernier tiers de la simulation. Les effets peu élevés de **MeanCellRS** et **MeanCellSU** sont probablement expliqués par le faible niveau de pluie. La célérité moyenne sur les tronçons a moins d'influence que celle sur les parcelles, probablement en raison de la faible densité du réseau de tronçons. Le faible effet de **MeanSigma** est probablement dû à sa gamme de variation par rapport à la taille du bassin.

Enfin, nous voyons que les interactions jouent un rôle très important dans l'explication de la variabilité de la sortie. Les interactions possèdent souvent plus d'influence que **MultiKs** et toujours plus d'influence que les autres facteurs d'entrée en tant qu'effet principal. Ils ne sont donc pas à négliger. Une des principales questions qui s'est posée lors de mon séjour à Avignon est celle de déterminer quels sont les facteurs d'entrée les plus en interaction.

Dans la figure 6 sont représentés les indices totaux normalisés ainsi que les indices d'interaction normalisés pour chaque facteur d'entrée.

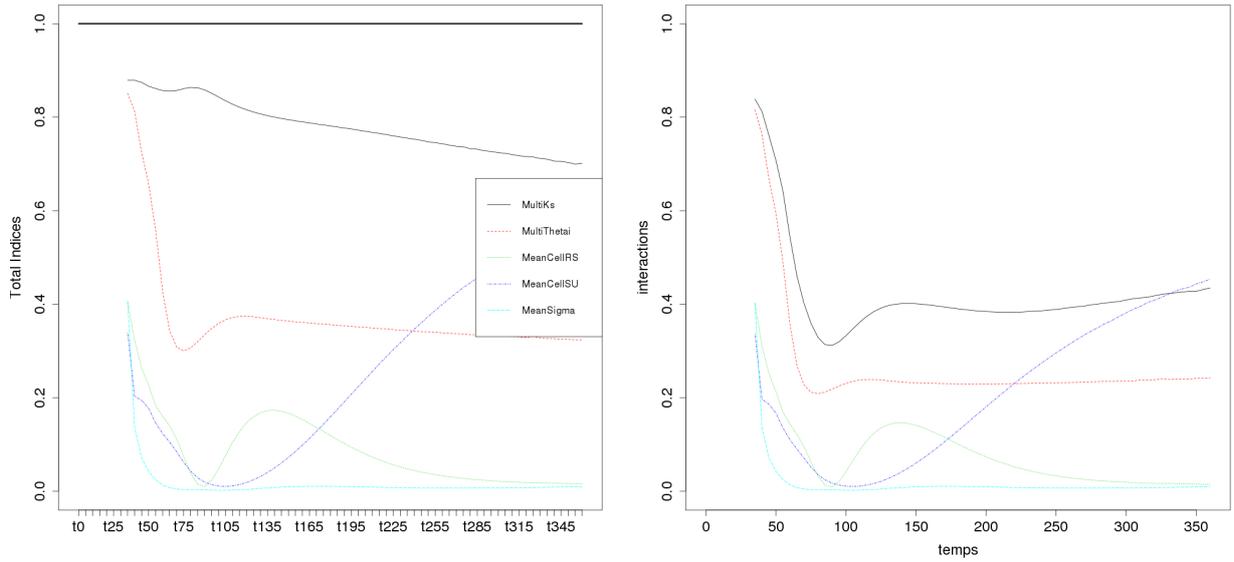


FIG. 6: Indices de sensibilité : cas **pluie faible**. Colonne de gauche : indices totaux normalisés. Colonne de droite : indices d'interactions normalisés.

Nous voyons que ce sont **MultiKs** et **MultiThetai** qui possèdent les indices d'interaction les plus élevés. Les autres facteurs d'entrée ont des valeurs d'indices d'interaction très faibles, excepté **MeanCellSU** en seconde moitié de simulation. Ainsi, nous pouvons en déduire que **MultiKs** est en forte interaction avec **MultiThetai**. Celle-ci est logique : plus le réservoir est plein, moins la vitesse d'infiltration va avoir d'effet. Par ailleurs, la courbe des indices totaux et d'interaction de **MultiThetai** ont la même dynamique : ce facteur d'entrée agit donc principalement par interactions (et avec **MultiKs** plus précisément). Les autres facteurs d'entrée, excepté **MultiKs** qui agit beaucoup en effet principal, agissent aussi principalement par interactions.

3.2.2 Application sur le cas pluie forte

Dans la figure 7 est représenté le graphique des indices de sensibilité normalisés.

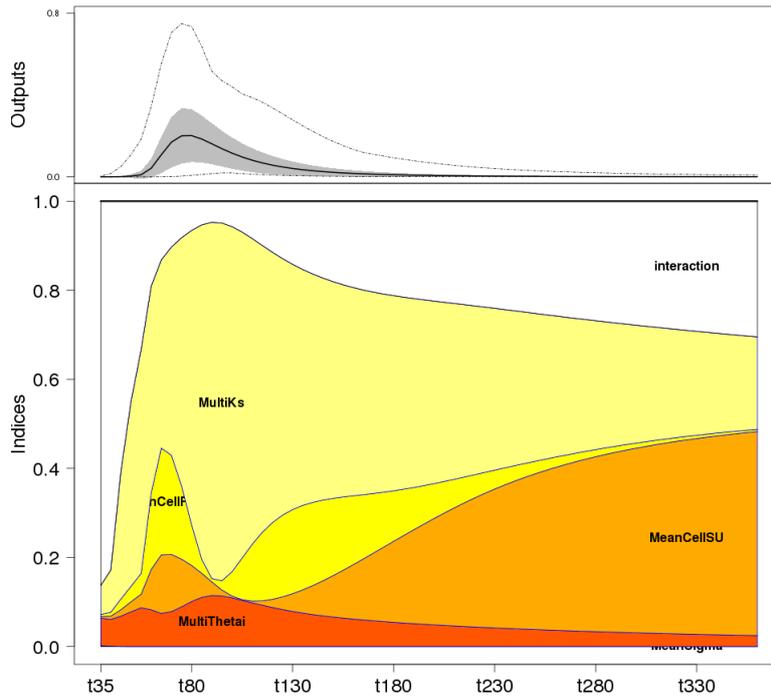


FIG. 7: Graphique des indices de sensibilité : cas **pluie forte**. Ligne du haut : certaines valeurs caractéristiques de la distribution empirique de la variable en fonction du temps. Ligne du bas : indices de sensibilité principaux en fonction du temps.

Ici aussi la variance est proportionnelle au signal. Nous retrouvons le pic de la variance de la sortie à 75 minutes. A 95 minutes (1h35), **MultiKs** est à son pic d'influence et **MultiThetaI** l'est aussi, mais dans une moindre mesure, à 90 minutes (1h30). Pour ces deux facteurs d'entrée, l'influence diminue légèrement avec le temps. Nous voyons que **MeanCellRS** a un pic d'influence juste avant que la variance de la sortie soit à son maximum (à 65 minutes exactement). Ce facteur d'entrée n'est donc pas à négliger à ce moment de la simulation. A 90 minutes, ce facteur d'entrée n'est plus du tout influent. Il gagne alors en influence jusqu'à 135 minutes (2h15), mais pendant ce temps la variance de la sortie diminue, et il possède une influence quasiment nulle en fin de simulation. Comme pour **MeanCellRS**, **MeanCellSU** a un pic d'influence à 65 minutes, bien que ce soit dans une moindre mesure. Il possède la plus grande influence en fin de simulation, mais la variance de la sortie est quasiment nulle à ce moment-là. Les pics d'influence de **MeanCellRS** et **MeanCellSU** ont lieu un peu plus tôt que le pic de variance de la sortie et ceux d'influence de **MultiKs** et **MultiThetaI** jouent sur la quantité d'eau qui arrive.

Contrairement au cas **pluie faible**, les interactions jouent un rôle beaucoup moins important dans l'explication de la variance de la sortie à 100 minutes environ. L'influence des interactions augmente avec le temps mais nous avons vu que la variance de la sortie diminuait aussi avec le temps.

3.3 Analyse de sensibilité après réduction de dimension

Nous avons exécuté quatre méthodes de réduction de dimension sur les sorties du modèle Mhydas. Nous avons pu observer que seule l'ACP offrait des informations supplémentaires à celles obtenues par l'analyse de sensibilité dynamique. Dans le cas des autres modèles, l'analyse de sensibilité par

projection des variables de sortie sur une base polynomiale a aussi été intéressante. L'analyse de sensibilité par projection sur une base de B-splines et de O-splines n'a pas donné d'informations supplémentaires quel que soit le modèle. Ainsi dans la suite, nous allons uniquement présenter les résultats des deux premières méthodes.

3.3.1 Application sur le cas pluie faible

Nous allons traiter la variable de sortie **débit à l'exutoire** dans le cas **pluie faible** par une ACP, puis nous allons exécuter une analyse de sensibilité sur chaque composante principale. Nous allons aussi projeter cette variable de sortie dynamique sur une base polynomiale, puis exécuter une analyse de sensibilité sur chaque polynôme de la base. Nous gardons le nombre de composantes qui suffisent à expliquer 95% de la variabilité de la sortie. Nous ne réduisons pas notre variable de sortie car elle est homogène. Le centrage est par contre important pour l'ACP.

Dans le cas de l'ACP, nous avons décidé d'enlever les 7 premières colonnes car pour ces temps-là la variance de la sortie est nulle. Cela ne change en aucun cas les résultats mais permet d'alléger le coût de calcul et d'améliorer l'esthétique des graphiques. Dans le cas de l'analyse de sensibilité sur base polynomiale, enlever ou non ces colonnes influe sur les résultats. Nous décidons donc de les conserver pour cette méthode.

Composantes de la base Nous avons représenté les composantes de l'ACP et celles de la base polynomiale dans la figure 8 dans le cas **pluie faible**.

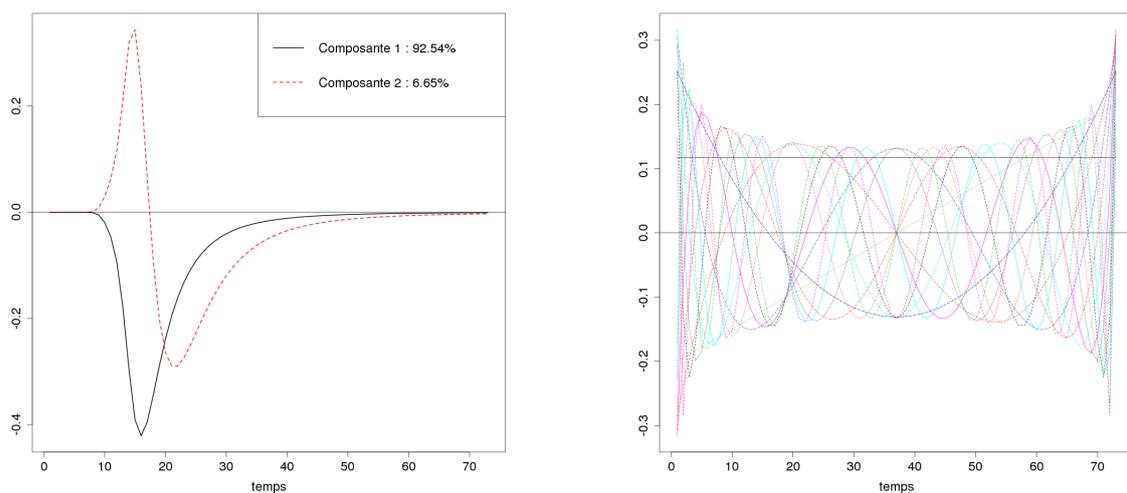


FIG. 8: Représentation des composantes de la base : cas **pluie faible**. Colonne de gauche : base des composantes principales (2 composantes). Colonne de droite : base polynomiale (14 composantes).

Deux composantes suffisent à expliquer 95% d'inertie dans le cas de l'ACP alors qu'il ne faut pas moins de 14 composantes pour la seconde méthode. C'est pourquoi nous préconiserons l'utilisation d'une analyse en composantes principales dans le cas **pluie faible**. En effet, l'évolution en fonction du temps du **débit à l'exutoire** en fonction du cas **pluie faible** n'a rien de polynomiale. Il faut donc beaucoup de polynômes pour arriver à expliquer 95% d'inertie. La projection sur base polynomiale fonctionne beaucoup mieux dans le cas du modèle Cantis de Sol Virtuel, comme nous pouvons le voir en **Annexe C**.

Analyse de sensibilité après ACP Après la réduction de dimension, nous pouvons exécuter une analyse de sensibilité. Dans la figure 9 est représentée la sortie graphique de l'analyse de sensibilité dans le cas de l'ACP.

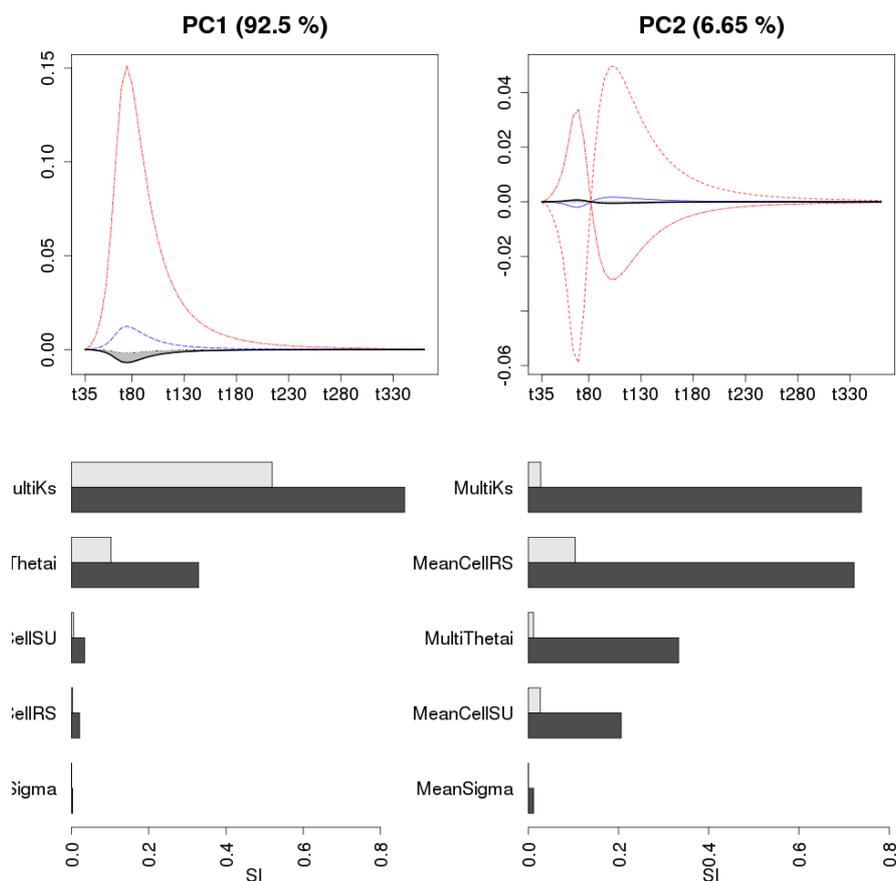


FIG. 9: Sortie graphique de l'analyse de sensibilité par ACP : cas **pluie faible**. Ligne du haut : variabilité des composantes principales $L_j(t)$ pondérée par $H_j(X)$. Ligne du bas : indices de sensibilité sur les premières composantes principales.

La première composante principale explique 92,5% de l'inertie totale. Une augmentation ou une diminution de sa valeur correspond grossièrement à une augmentation ou une diminution de la valeur de la sortie en fin de simulation. La deuxième composante principale explique 6,65% de l'inertie totale. Nous pouvons observer un contraste entre le premier tiers et les deux derniers tiers du temps de la simulation. Sur les deux composantes principales, **MultiKs** est le facteur d'entrée le plus influent si nous prenons en compte les effets principaux et les interactions. **MultiKs** possède la plus grande influence en tant qu'effet principal sur la première composante principale, suivi de **MultiThetai**. L'interaction entre ces deux facteurs d'entrée joue un rôle majeur dans l'explication de la variance de la sortie. Sur la deuxième composante principale, c'est **MeanCellRS** qui possède la plus grande influence en tant qu'effet principal. Il faut retenir que les indices des facteurs d'entrée pour les effets principaux sont assez faibles (excepté chez **MultiKs** sur la première composante principale) comparés aux indices d'interaction. Sur la deuxième composante principale, seules les interactions jouent. Nous retrouvons donc le même résultat que pour l'analyse de sensibilité dynamique.

Analyse de sensibilité après projection sur base polynomiale Dans la figure 10 est représentée la sortie graphique de l'analyse de sensibilité dans le cas de l'analyse de sensibilité sur base polynomiale.

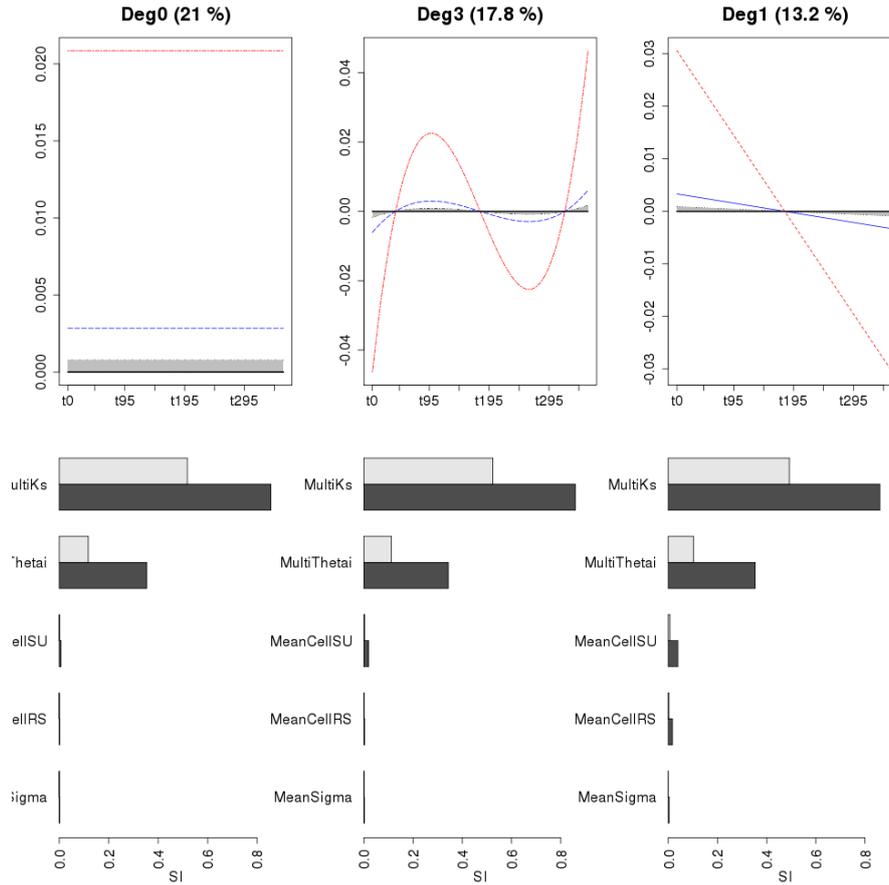


FIG. 10: Sortie graphique de l'analyse de sensibilité sur base polynomiale : cas **pluie faible**. Ligne du haut : variabilité des polynômes $L_j(t)$ pondérée par $H_j(X)$. Ligne du bas : indices de sensibilité sur les polynômes de première importance.

Nous avons représenté les trois degrés qui expliquent la plus grande part d'inertie. Le facteur d'entrée le plus influent sur ces trois degrés est **MultiKs**, que ce soit en tant qu'effet principal ou en interaction avec les autres facteurs d'entrée. L'indice de premier ordre de **MultiThetai** est assez faible sur les trois composantes ainsi que l'indice total. Pour les autres facteurs d'entrée, ces indices sont quasiment nuls.

La somme des pourcentages d'inertie expliqués par les trois premiers polynômes d'importance vaut 52%. Pour avoir 95% d'inertie expliquée, il faudrait ajouter les degrés suivants, ordonnés selon leur contribution à l'inertie :

Deg0	Deg3	Deg1	Deg4	Deg6	Deg7	Deg10
21%	39%	52%	62%	71%	79%	83%
Deg9	Deg13	Deg2	Deg11	Deg14	Deg8	Deg5
86%	88%	90%	92%	93%	95%	96%

Dans le cas où le nombre de polynômes de la base permettant d'atteindre 95% d'inertie expliquée est élevé, on ajoute en complexité du modèle et on perd alors en informations. Mais ne considérer que les premiers polynômes d'importance ajoute de l'erreur dans le modèle du fait qu'ils expliquent un pourcentage d'inertie beaucoup moins important que l'ensemble des polynômes de la base.

Soit Y_{t_i} la variable de sortie dynamique débit à l'exutoire dans le cas pluie faible pour $i \in [1; 73]$. Alors d'après nos résultats, nous pouvons écrire :

$$Y_{t_i} = \sum_{d=0}^{14} c_d t_i^d + \epsilon(t_i)$$

où c_d , $d \in [0; 14]$, sont des fonctions des facteurs d'entrée les plus influents sur les différents degrés et $\epsilon(t_i)$ représente l'erreur d'une telle écriture. Nous pouvons connaître le comportement général des fonctions c_d au niveau de leur signe à l'aide du logiciel R. Nous trouvons alors par exemple que c_0 et c_3 sont des fonctions à valeurs positives ou nulles et que c_1 est une fonction à valeurs négatives ou nulles.

Indices de sensibilité généralisés Dans la figure 11 sont représentés les GSI obtenus par ces deux méthodes.

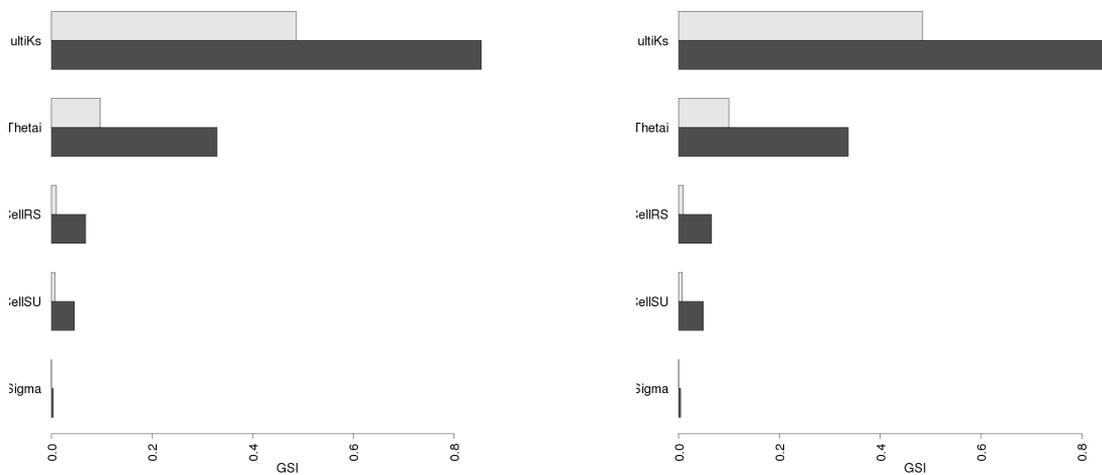


FIG. 11: Indices de sensibilité généralisés : cas pluie faible. Colonne de gauche : ACP. Colonne de droite : analyse de sensibilité sur base polynomiale.

Nous obtenons presque le même classement d'influence pour les deux méthodes. Nous retrouvons le fait que seul **MultiKs** possède une assez grande influence en tant qu'effet principal et que ce sont les interactions, principalement celles qui impliquent **MultiKs**, qui possèdent le plus d'influence sur l'explication de la variabilité de la sortie.

3.3.2 Application sur le cas pluie forte

Nous allons traiter la variable de sortie débit à l'exutoire dans le cas pluie forte par une ACP, puis nous allons exécuter une analyse de sensibilité sur chaque composante principale. Nous allons aussi projeter cette variable de sortie dynamique sur une base polynomiale, puis exécuter une analyse de sensibilité sur chaque polynôme de la base. Comme dans le cas pluie faible, nous gardons le nombre de composantes qui suffisent à expliquer 95% de la variabilité de la sortie et nous ne réduisons pas notre variable de sortie. Enfin, dans le cas de l'ACP, nous la centrons et nous retirons aussi les 7 premières colonnes.

Composantes de la base Nous avons représenté les composantes de l'ACP et celles de la base polynomiale dans la figure 12 pour le cas **pluie forte**.

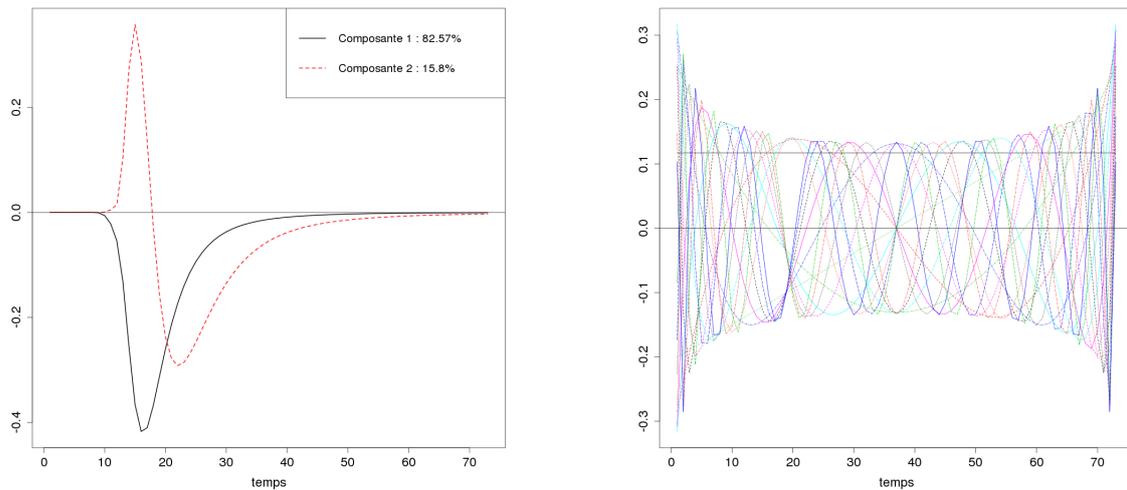


FIG. 12: Représentation des composantes de la base : cas **pluie forte**. Colonne de gauche : base des composantes principales (2 composantes). Colonne de droite : base polynomiale (16 composantes).

Comme pour le cas **pluie faible**, deux composantes suffisent à expliquer 95% d'inertie dans le cas de l'ACP. Il en faut 16 pour la seconde méthode. De nouveau, nous préconiserons l'utilisation d'une analyse en composantes principales dans le cas **pluie forte**.

Analyse de sensibilité après ACP Après la réduction de dimension, nous pouvons exécuter une analyse de sensibilité. Dans la figure 13 est représentée la sortie graphique de l'analyse de sensibilité dans le cas de l'ACP.

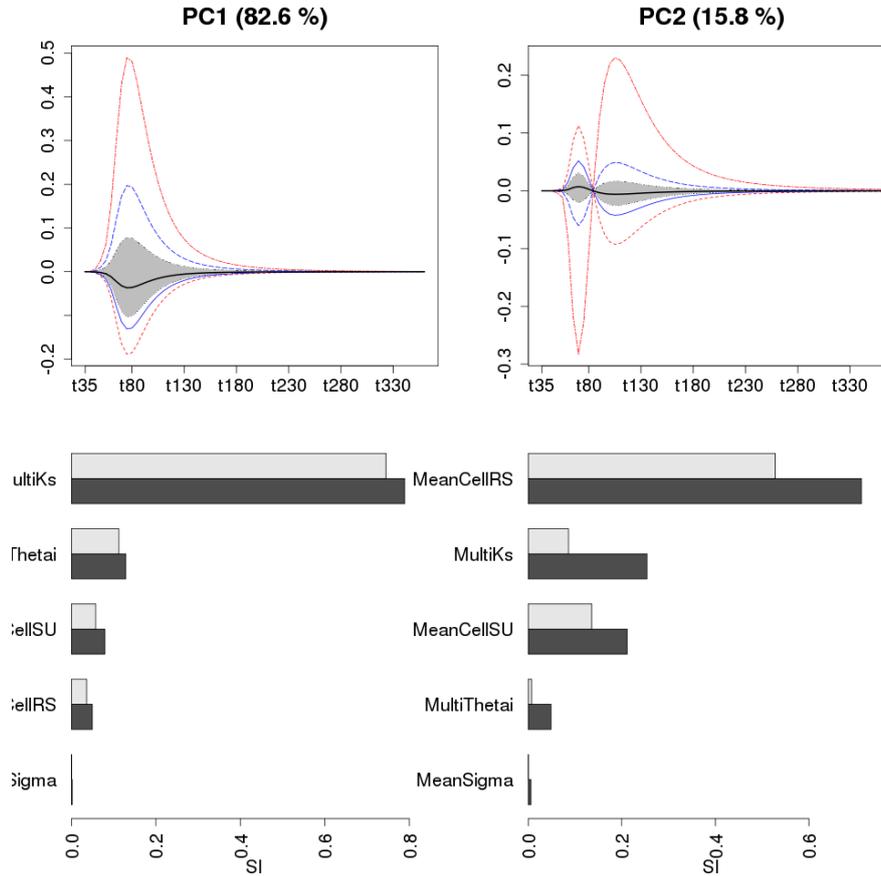


FIG. 13: Sortie graphique de l'analyse de sensibilité par ACP : cas **pluie forte**. Ligne du haut : variabilité des composantes principales $L_j(t)$ pondérée par $H_j(X)$. Ligne du bas : indices de sensibilité sur les premières composantes principales.

Nous retrouvons les mêmes interprétations que dans le cas **pluie faible** pour les courbes obtenues sur les trois composantes principales. Comme pour l'analyse de sensibilité dynamique, les interactions jouent un rôle beaucoup moins important dans le cas **pluie forte** que dans le cas **pluie faible** sur les deux premières composantes principales. A l'inverse, les effets principaux possèdent une plus grande influence. **MultiKs** est le facteur d'entrée de loin le plus influent sur la première composante principale et ce en tant qu'effet principal, suivi de **MultiThetai** dans une bien moindre mesure. Les autres facteurs d'entrée sont très peu influents. Sur la deuxième composante principale, seul **MeanCellRS** est très influent en tant qu'effet principal.

Une autre question s'étant posée lors de mon séjour à Avignon est celle d'étudier l'influence des facteurs d'entrée les plus influents sur les coefficients H_j . Du fait du faible nombre de composantes constituant la base (2 composantes principales) et pour améliorer l'interprétation des résultats de cette méthode, nous allons donc évaluer l'influence des facteurs d'entrée les plus influents sur les H_j (**MultiKs** pour H_1 et **MeanCellRS** pour H_2 dans le cas **pluie forte**). Nous allons décrire les résultats obtenus dans le cas **pluie forte** uniquement (il est évidemment possible de faire de même dans le cas **pluie faible** car la base est aussi constituée de 2 composantes principales). Le deuxième facteur d'entrée le plus influent sur la première composante principale est **MultiThetai** pour lequel la valeur numérique est comprise dans l'intervalle $[0,25;1,15]$. Nous décomposons cet intervalle en quatre sous-intervalles de même longueur, ces derniers étant associés à une couleur et à un type de point : $[0,25;0,475[$ (cercle en noir), $[0,475;0,7[$ (triangle en rouge), $[0,7;0,925[$ (signe plus en vert) et $[0,925;1,15]$ (croix en bleu). Dans la figure 14 est représentée l'influence de **MultiKs** sur H_1 , avec

des couleurs et types de point dépendant du domaine d'appartenance de `MultiThetaI`. Cela permet d'avoir une idée de l'interaction entre ces deux facteurs d'entrée.

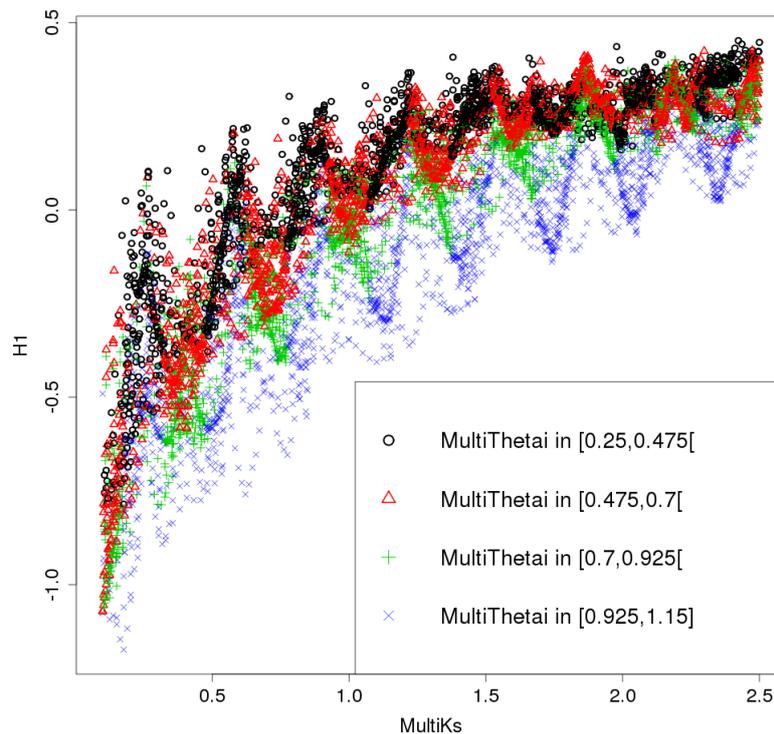


FIG. 14: Influence de `MultiKs` et de `MultiThetaI` sur H_1 .

Augmenter la valeur de `MultiKs` a pour conséquence une augmentation des valeurs du coefficient H_1 . Les oscillations que nous pouvons voir semblent être la conséquence de l'utilisation de la méthode eFAST et non des interactions entre `MultiKs` et `MultiThetaI`. Enfin, plus la valeur de `MultiThetaI` est petite, plus l'amplitude des valeurs de H_1 en fonction de `MultiKs` est petite. Il y a donc une interaction entre ces deux facteurs d'entrée, et celle-ci est évaluable numériquement.

Le deuxième facteur d'entrée le plus influent sur la deuxième composante principale est `MultiKs` pour lequel la valeur numérique est comprise dans l'intervalle $[0,1;2,5]$. Nous décomposons cet intervalle en quatre sous-intervalles de même longueur, ces derniers étant associés à une couleur et à un type de point : $[0,1;0,7[$ (cercle en noir), $[0,7;1,3[$ (triangle en rouge), $[1,3;1,9[$ (signe plus en vert) et $[1,9;2,5]$ (croix en bleu). Dans la figure 15 est représentée l'influence de `MeanCellRS` sur H_2 , avec des couleurs et types de point dépendant du domaine d'appartenance de `MultiKs`.

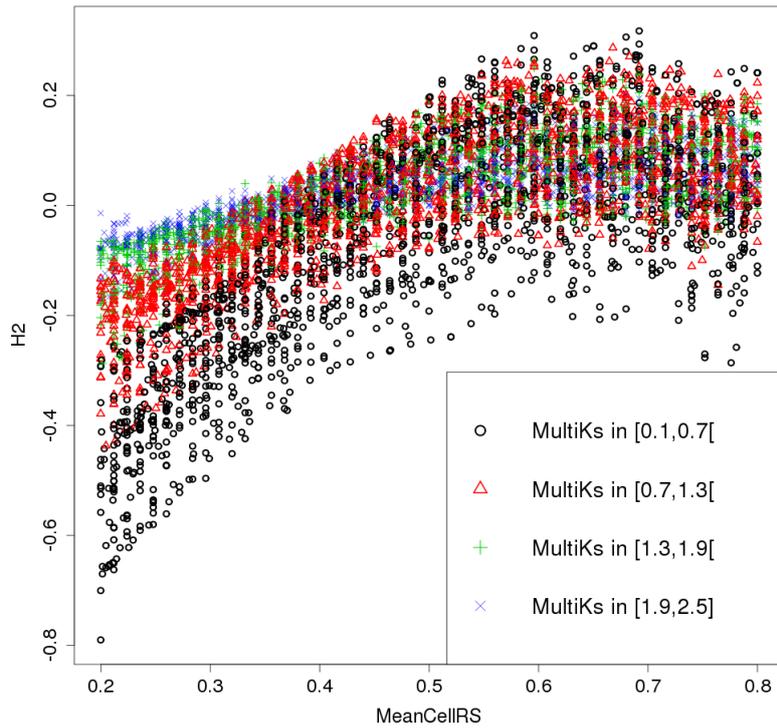


FIG. 15: Influence de MeanCellRS et de MultiKs sur H_2 .

De même, une augmentation de la valeur de MeanCellRS implique une augmentation (assez mineure) des valeurs du coefficient H_2 . L'interaction entre MeanCellRS et MultiKs est moins évidente à observer : il semble que plus la valeur de MultiKs est petite, plus l'amplitude des valeurs de H_2 en fonction de MeanCellRS est grande.

Analyse de sensibilité après projection sur base polynomiale Dans la figure 16 est représentée la sortie graphique de l'analyse de sensibilité dans le cas de l'analyse de sensibilité sur base polynomiale.

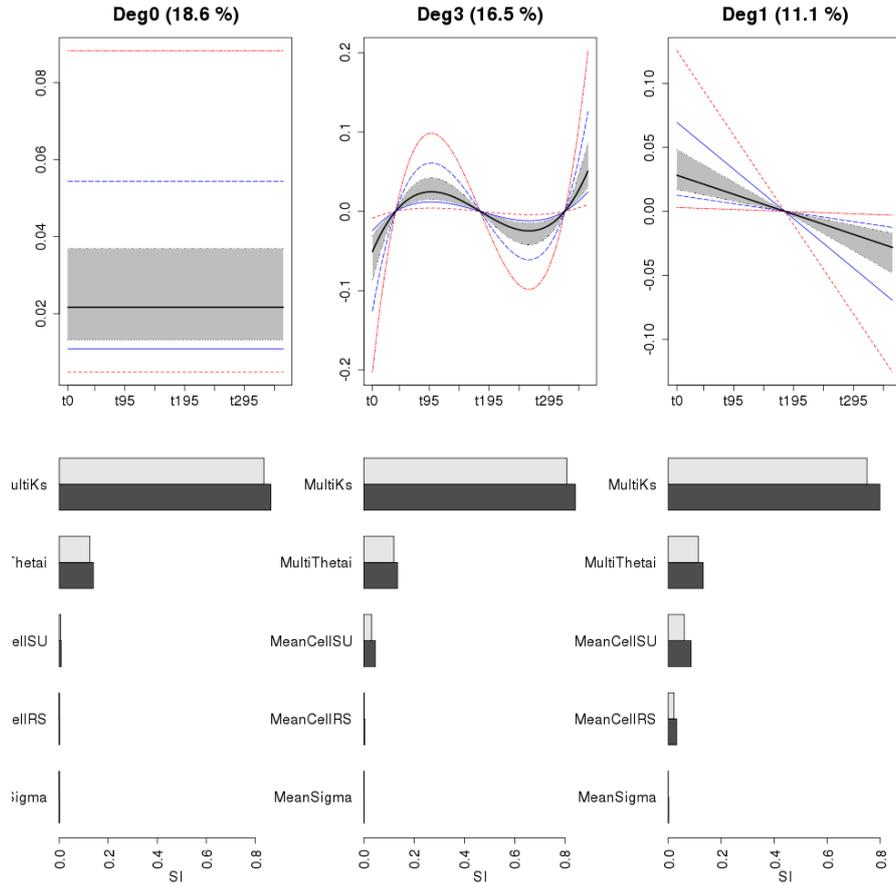


FIG. 16: Sortie graphique de l'analyse de sensibilité sur base polynomiale : cas **pluie forte**. Ligne du haut : variabilité des polynômes $L_j(t)$ pondérée par $H_j(X)$. Ligne du bas : indices de sensibilité sur les polynômes de première importance.

Nous obtenons des résultats très similaires à ceux obtenus pour le cas **pluie faible**. Une fois de plus que les interactions jouent un rôle négligeable dans l'explication de la variabilité de la variable de sortie dynamique **débit à l'exutoire** contrairement aux effets principaux qui prédominent. La somme des pourcentages d'inertie des trois premiers polynômes d'importance est de 46,2%, ce qui est assez faible. Pour avoir 95% d'inertie expliquée, il faudrait ajouter les degrés suivants, ordonnés selon leur contribution à l'inertie :

Deg0	Deg3	Deg1	Deg4	Deg6	Deg7	Deg10	Deg9
19%	35%	46%	55%	64%	71%	76%	80%
Deg2	Deg13	Deg5	Deg8	Deg11	Deg14	Deg17	Deg16
83%	86%	88%	90%	92%	93%	95%	96%

Evaluer l'influence des facteurs d'entrée les plus influents sur les coefficients de la combinaison linéaire n'est pas utile dans ce cas du fait du nombre important de composantes constituant la base (16 polynômes).

Indices de sensibilité généralisés Dans la figure 17 sont représentés les GSI obtenus par ces deux méthodes.

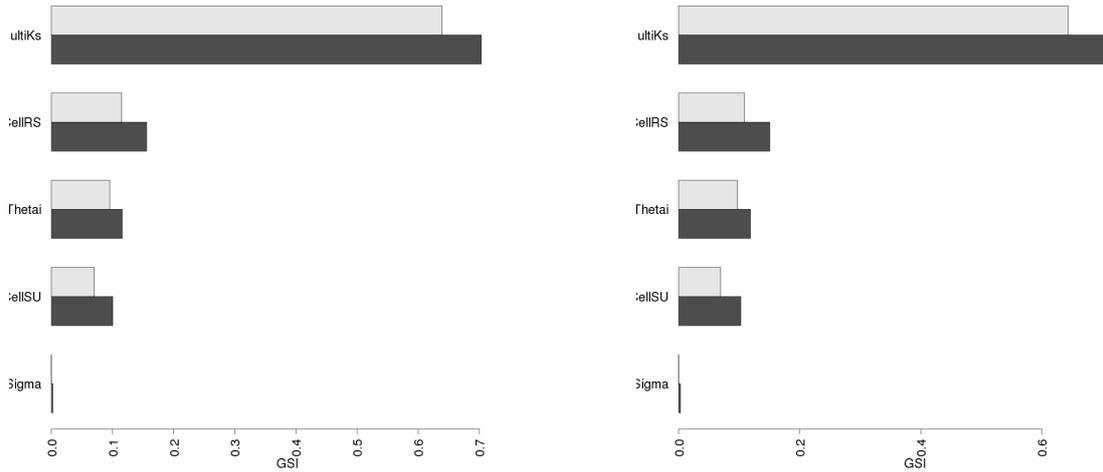


FIG. 17: Indices de sensibilité généralisés : cas **pluie forte**. Colonne de gauche : ACP. Colonne de droite : analyse de sensibilité sur base polynomiale.

Le classement d'influence est sensiblement le même pour ces deux méthodes. Nous retrouvons le fait que **MultiKs** est très influent en tant qu'effet principal et que les interactions possèdent une moins grande influence dans le cas **pluie forte** que dans le cas **pluie faible**.

3.4 Classification

Nous allons nous intéresser à l'application d'une classification dans le cas **pluie forte**.

3.4.1 Partition obtenue par la méthode des K-means

Pour obtenir une bonne partition des courbes du **débit à l'exutoire**, nous avons exécuté l'algorithme des K -means pour des K allant de 2 jusqu'à 10 avec une limite de 20 itérations. Pour chaque K , 30 segmentations sont testées. Nous étudions alors l'inertie intra-cluster obtenue pour chaque K afin de choisir le nombre de clusters. Dans notre cas, avoir beaucoup de clusters n'est pas une bonne chose car nous voulons par la suite évaluer quels sont les facteurs d'entrée qui influent sur un comportement bien différent de la sortie dynamique. En ajoutant le fait qu'à partir de $K = 4$ la décroissance de l'inertie intra-cluster en fonction du nombre de cluster tend à ralentir, nous avons décidé de partitionner les 7500 observations du **débit à l'exutoire** dans le cas **pluie forte** en quatre sous-échantillons. Dans la figure 18 est représentée l'évolution du **débit à l'exutoire** dans le cas **pluie forte** en fonction du temps pour différents jeux de facteurs d'entrée et des couleurs différentes selon l'appartenance aux clusters 1,2,3 ou 4.

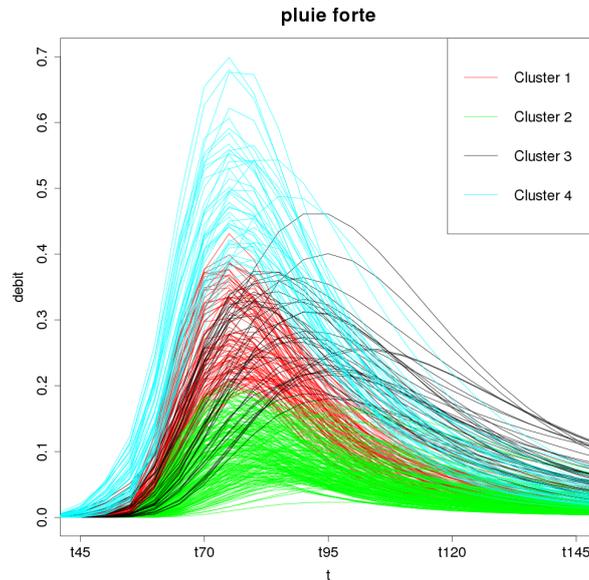


FIG. 18: Evolution en fonction du temps du débit à l'exutoire dans le cas pluie forte pour différents jeux de facteurs d'entrée et des couleurs différentes selon l'appartenance aux clusters 1,2,3 ou 4.

Les clusters 1,2,3 et 4 sont composés respectivement de 1828, 4038, 757 et 877 observations.

Dans le tableau suivant sont définies les bornes des maxima de chaque observation du cluster 1 :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0,18	0,23	0,27	0,28	0,32	0,44

Pour le cluster 2, nous obtenons :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0,02	0,08	0,11	0,12	0,15	0,22

Pour le cluster 3, nous obtenons :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0,17	0,24	0,29	0,29	0,34	0,46

Enfin pour le cluster 4, nous obtenons :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0,35	0,42	0,46	0,48	0,52	0,75

Enfin, nous avons les informations suivantes :

- 71% des observations du cluster 1 atteignent leur valeur maximale à 75 minutes (les 29% restants sont répartis au temps 70 minutes et entre 80 et 85 minutes).
- 63% des observations du cluster 2 atteignent leur valeur maximale entre 75 et 80 minutes (les 37% restants sont répartis au temps 70 minutes et entre 85 et 110 minutes).

- 75% des observations du cluster 3 atteignent leur valeur maximale entre 80 et 90 minutes dont 31% à 85 minutes (les 25% restants sont répartis entre 95 et 110 minutes).
- 67% des observations du cluster 4 atteignent leur valeur maximale à 75 minutes (les 33% restants sont répartis entre 80 et 85 minutes).

Ainsi, nous pouvons affirmer que les courbes des clusters 2 et 4 correspondent respectivement à des valeurs faibles et fortes du débit à l'exutoire en fonction du temps alors que les clusters 1 et 3 correspondent à des valeurs moyennes du débit à l'exutoire mais décalées dans le temps.

3.4.2 Influence des facteurs d'entrée sur les 4 clusters

Nous allons évaluer l'influence des facteurs d'entrée sur les 4 clusters en comparant les distributions des X_i qui leur sont associées et en faisant une analyse de sensibilité sur les distances entre les observations et les centres de gravité des clusters.

Comparaison des distributions des facteurs d'entrée Dans notre cas, il y a cinq facteurs d'entrée et quatre clusters. Il faut donc exécuter quatre tests de Kolmogorov-Smirnov pour chacun des facteurs d'entrée afin de déterminer leur significativité, c'est-à-dire s'ils sont critiques, importants ou sans importance. Soient respectivement $F_{C_1}(X_i)$, $F_{C_2}(X_i)$, $F_{C_3}(X_i)$ et $F_{C_4}(X_i)$ les fonctions de répartition empirique des clusters 1,2,3 et 4 pour un facteur d'entrée X_i . Alors les quatre tests sont les suivants :

- 1er test : $(H_0) F_{C_1}(X_i) = F_{C_2+C_3+C_4}(X_i)$ contre $(H_1) F_{C_1}(X_i) \neq F_{C_2+C_3+C_4}(X_i)$.
- 2ème test : $(H_0) F_{C_2}(X_i) = F_{C_1+C_3+C_4}(X_i)$ contre $(H_1) F_{C_2}(X_i) \neq F_{C_1+C_3+C_4}(X_i)$.
- 3ème test : $(H_0) F_{C_3}(X_i) = F_{C_1+C_2+C_4}(X_i)$ contre $(H_1) F_{C_3}(X_i) \neq F_{C_1+C_2+C_4}(X_i)$.
- 4ème test : $(H_0) F_{C_4}(X_i) = F_{C_1+C_2+C_3}(X_i)$ contre $(H_1) F_{C_4}(X_i) \neq F_{C_1+C_2+C_3}(X_i)$.

1er test (cluster 1) : Tous les facteurs d'entrée excepté **MeanSigma** sont critiques sur la dynamique spécifique du cluster 1 de la sortie (p-value < 2,2e-16). **MeanSigma** est quant à lui important (p-value = 0,04).

2ème test (cluster 2) : On obtient les mêmes résultats à l'exception faite que **MeanSigma** est cette fois-ci sans importance (p-value = 0,96).

3ème test (cluster 3) : **MeanSigma** est sans importance (p-value = 0,56). De nouveau, les autres facteurs d'entrée sont critiques.

4ème test (cluster 4) : Encore une fois **MeanSigma** est sans importance (p-value = 0,16) et les autres facteurs d'entrée sont critiques.

En voyant ces résultats, un modélisateur voudrait savoir quelles sont les valeurs des facteurs d'entrée critiques qui conduisent à quel cluster. Prenons par exemple **MultiKs** qui varie entre 0,1 et 2,5. Dans le tableau suivant sont données certaines valeurs caractéristiques de la distribution de **MultiKs** conduisant à chacun des clusters :

	Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
Cluster 1	0,27	0,72	0,99	1,10	1,35	2,44
Cluster 2	0,46	1,41	1,81	1,76	2,16	2,50
Cluster 3	0,10	0,27	0,45	0,51	0,64	1,46
Cluster 4	0,10	0,18	0,34	0,36	0,46	1,15

On peut observer trois catégories de valeurs de `MultiKs` conduisant aux clusters :

- les petites valeurs de `MultiKs` conduisent aux clusters 3 et 4 ;
- les valeurs moyennes de `MultiKs` conduisent au cluster 1 ;
- les grandes valeurs de `MultiKs` conduisent au cluster 2.

Analyse de sensibilité sur les distances Dans la figure 19 sont représentés les résultats de l'analyse de sensibilité sur les distances entre les courbes et les centres des clusters 1,2,3 et 4.

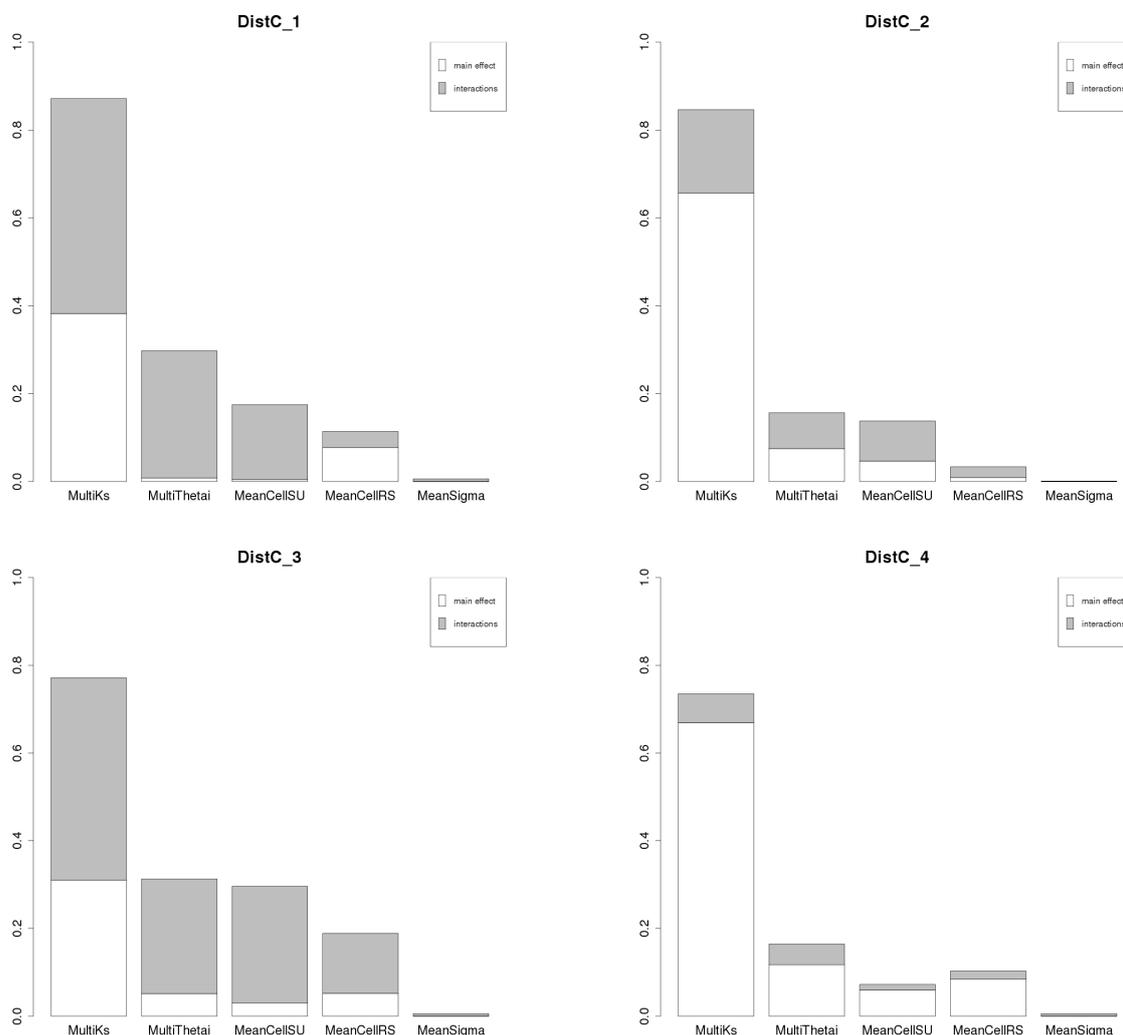


FIG. 19: Résultats de l'analyse de sensibilité sur les distances entre les courbes et les centres des clusters. En haut à gauche : cluster 1. En haut à droite : cluster 2. En bas à gauche : cluster 3. En bas à droite : cluster 4.

En blanc et en gris est représenté respectivement l'effet principal et l'effet d'interaction de chaque facteur. `MultiKs` est de loin le paramètre le plus influent sur les distances entre les courbes et les centres des clusters. Nous avons vu sur la figure 18 que les courbes des clusters 2 et 4 correspondaient à des valeurs faibles ou fortes du débit à l'exutoire en fonction du temps. `MultiKs` est le paramètre le plus influent en tant qu'effet principal sur ces comportements du débit à l'exutoire. Nous avons vu de plus que les clusters 1 et 3 correspondaient à des valeurs moyennes du débit à l'exutoire mais décalées dans le temps. C'est cette fois-ci `MultiKs` principalement en interaction avec `MultiThetaI` qui agissent sur ces comportements du débit à l'exutoire. En effet, nous avons pu voir sur l'analyse de sensibilité dynamique que l'interaction entre ces deux facteurs d'entrée est forte.

Commentaires Les deux méthodes présentées dans ce rapport paraissent intéressantes du point de vue du modélisateur afin de quantifier l'influence des facteurs d'entrée sur les clusters. Cependant, je ne suis pas sûr de la validité des interprétations que l'on peut tirer des résultats apportés par la comparaison des distributions des facteurs d'entrée. Cela est dû en particulier au fait que nous travaillons sur des simulations eFAST et non Monte-Carlo et que la réponse catégorielle n'est pas issue du jeu de données initial mais d'une étape de classification. De plus, je trouve que cette méthode n'apporte pas beaucoup d'informations intéressantes en comparaison des résultats de l'analyse de sensibilité sur les distances.

3.5 Conclusion sur le modèle Mhydas

L'analyse de sensibilité dynamique a offert beaucoup d'informations sur l'influence des facteurs d'entrée pour chaque pas de temps indépendamment. Les sorties graphiques sont lisibles et facilement interprétables par un modélisateur. En ce qui concerne l'analyse de sensibilité après réduction de dimension, l'ACP a donné les résultats les plus intéressants. La projection de la variable de sortie débit à l'exutoire sur une base polynomiale n'offre pas d'informations supplémentaires sur l'influence des facteurs à cause du nombre trop important de composantes de la base.

Afin d'affiner les interprétations obtenues à partir des meilleures méthodes d'analyse de sensibilité pour ce modèle, nous avons imaginé plusieurs outils. Nous avons représenté les indices d'interaction de tous les facteurs d'entrée pour l'analyse de sensibilité dynamique dans le cas **pluie faible** car nous avons pu voir que les interactions jouaient un rôle très important dans l'explication de la variabilité de la sortie. De plus, du fait du faible nombre de composantes principales dans la base obtenue par ACP, nous avons pu évaluer l'influence des facteurs d'entrée sur les coefficients H_j de la combinaison linéaire impliquant les composantes de la base dans le cas **pluie forte**. Enfin, l'application d'une méthode de classification sur le débit à l'exutoire suivie d'une étude de l'influence des facteurs d'entrée sur les clusters obtenus ont permis de déterminer ceux qui sont les plus influents sur des comportements bien distincts de la sortie.

Nous avons vu que `MultiKs`, le coefficient multiplicatif de la conductivité à saturation de l'ensemble des parcelles, est un facteur d'entrée très influent sur la variabilité du débit à l'exutoire dans le cas **pluie faible** et **pluie forte**. Il est fortement en interaction avec `MultiThetaI`, le coefficient multiplicatif de l'humidité initiale de l'ensemble des parcelles, en particulier dans le cas **pluie faible**. Dans le cas **pluie forte**, `MultiKs` en effet principal a une influence sur les valeurs faibles ou fortes du débit à l'exutoire. En interaction avec `MultiThetaI`, il a une influence sur les valeurs moyennes du débit à l'exutoire mais décalées dans le temps. Les facteurs d'entrée `MeanCellRS` et `MeanCellSU` sont très peu influents dans le cas **pluie faible**. Ils gagnent en influence dans le cas **pluie forte**, ce qui est logique. En effet, la célérité moyenne de l'onde de crue sur les parcelles et celle sur le réseau hydrographique impactent sur le temps d'arrivée de l'onde de crue à l'exutoire. Enfin, `MeanSigma`, la diffusivité moyenne de l'onde de crue sur les parcelles et le réseau hydrographique, est un facteur d'entrée négligeable dans l'explication de la variabilité de la

sortie.

L'interaction avec les modélisateurs du modèle Mhydas a été très enrichissante. J'ai pu avoir une vision d'ensemble de ce qu'est la plateforme OpenFluid et de ce qu'elle apporte. De plus, toutes les interactions avec la plateforme m'ont aidé à rédiger le plus clairement et le plus précisément possible le rapport où, sont présentés mes résultats et, faisant office de base au corrigé d'un TP destiné aux utilisateurs de la plateforme.

Conclusion

Les objectifs de ce stage étaient multiples :

- identifier des méthodes adaptées d’analyse de sensibilité pour des modèles à sorties dynamiques, les tester sur les modèles issus des plateformes ;
- sélectionner les plus pertinentes en assurant notamment une bonne qualité d’interprétation des résultats obtenus ;
- décrire les résultats de l’application de ces méthodes sur des modèles choisis ;
- et enfin, rédiger des documents pédagogiques destinés aux utilisateurs des plateformes, non spécialistes de l’analyse de sensibilité.

Quel que soit le jeu de données, l’analyse de sensibilité dynamique a permis d’étudier l’évolution de l’influence des facteurs d’entrée sur la sortie pour chaque pas de temps indépendamment. Les résultats obtenus sont facilement interprétables pour des modélisateurs n’étant pas familiarisé avec l’analyse de sensibilité. Les résultats de l’analyse de sensibilité après réduction de dimension ont été plus délicats à interpréter. L’analyse de sensibilité après ACP a offert beaucoup d’informations quel que soit le jeu de données. L’analyse de sensibilité après projection sur base polynomiale a été intéressante dans le cas du modèle Cantis de Sol Virtuel du fait du faible nombre de composantes constituant la base polynomiale. Ce ne fut pas le cas pour les deux autres modèles. L’analyse de sensibilité après projection sur base de B-splines n’a pas été concluante. En effet, il était délicat de déterminer ce que cette méthode offrait de plus que l’analyse de sensibilité dynamique. De plus, une base de B-splines peut expliquer un pourcentage d’inertie supérieur à 100%. Nous nous sommes dits que cela pouvait fausser les résultats du fait que nous nous intéressons explicitement à des analyses de sensibilité basées sur la variance de la sortie. C’est en comparant les indices de sensibilité généralisés des quatre méthodes que nous avons pu confirmer cela. En effet, les valeurs de ces indices obtenues par la méthode impliquant les B-splines étaient clairement différentes de celles obtenues par les trois autres méthodes. L’analyse de sensibilité après projection sur base de O-splines aurait pu passer outre cette difficulté car une telle base est orthonormée et donc restreinte à expliquer au maximum 100% d’inertie. Cependant, les résultats obtenus nous ont semblé ininterprétables.

A partir de tous les résultats que nous avons obtenus, nous pouvons définir une règle générale pour le choix de la méthode en fonction du jeu de données. Pour étudier l’évolution de l’influence des facteurs d’entrée sur la sortie en fonction du temps, nous préconiserons l’application d’une analyse de sensibilité dynamique. Elle offre beaucoup de résultats facilement interprétables quel que soit le jeu de données. Dans le cas où les interactions jouent un rôle important dans l’explication de la variabilité de la sortie, il est utile de représenter les indices d’interaction de tous les facteurs d’entrée afin de déterminer quels sont ceux qui interagissent le plus entre eux. Les résultats obtenus par l’analyse de sensibilité après réduction de dimension sont difficilement interprétables. Dans le cas où l’analyse de sensibilité après ACP ou projection sur base polynomiale donne une base constituée d’un faible nombre de composantes et représentant bien la variabilité de la sortie, étudier l’influence des facteurs d’entrée sur les coefficients H_j permet d’aider le modélisateur à interpréter ces résultats. Enfin, il est utile d’évaluer quels facteurs d’entrée influent sur quels comportements bien distincts de la sortie. Pour cela, une méthode de classification suivie d’une comparaison des distributions des facteurs d’entrée ou d’une analyse de sensibilité sur les distances entre les courbes des observations et les centres de gravité des clusters est à préconiser.

Des TPs, destinés aux utilisateurs des plateformes, seront rédigés sur la base des trois rapports complets et pédagogiques (un par modèle) que j’ai écrit. J’ai aussi rédigé deux fiches méthodologiques

décrivant de façon succincte mais précise les deux types d'analyse de sensibilité qui permettront aux modélisateurs d'appliquer facilement ces méthodes sans prérequis spécifiques.

Ce stage m'a beaucoup apporté sur de nombreux points. J'ai pu apprendre plusieurs méthodes statistiques de haut niveau et perfectionner mon utilisation du logiciel R. J'ai aussi appris à rechercher des informations sur la méthodologie statistique, que ce soit sur le net ou dans des revues, articles, etc. Enfin j'ai pu améliorer mes qualités de rédaction à l'aide des trois rapports liés aux trois modèles et au rapport de stage que j'ai écrit.

Références

- [1] C. Bidot, H. Monod, M-L. Taupin (2016). A quick guide to multisensi, an R package for multivariate sensitivity analyses. Vignette du package R multisensi, accessible sur le site du CRAN.
URL <https://cran.r-project.org/web/packages/multisensi/vignettes/multisensi-vignette.pdf>.
- [2] R. Faivre, B. Iooss, S. Mahévas, D. Makowski, H. Monod (2013). *Analyse de sensibilité et exploration de modèles. Application aux sciences de la nature et de l'environnement*. Éditions Quae : Versailles.
- [3] J. Baudet (2009). Introduction à l'analyse de sensibilité. Mémoire de stage de l'Université de Bourgogne, INRA : Toulouse.
- [4] F. Boizard (2015). Méthodes d'analyse de sensibilité de modèles pour entrées climatiques. Mémoire de stage d'Agrocampus Ouest, INRA : Toulouse.
- [5] B. Auder (2011). Classification et modélisation de sorties fonctionnelles de codes de calcul. Mémoire de stage de l'Université Paris 6, CEA : Cadarache.
- [6] B. Sudret (2007). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, **93**, 964-979.
- [7] M. Lamboni, D. Makowski, S. Lehuger, B. Gabrielle, H. Monod (2009). Multivariate global sensitivity analysis for dynamic crop models. *Field Crops Research*, **113**, 312-320.
- [8] M. Lamboni, H. Monod, D. Makowski (2011). Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering and System Safety*, **99**, 450-459.
- [9] P. Besse (2014). Exploration statistique multidimensionnelle. In : *WikiStat*. Documentation accessible sur Internet.
URL http://www.math.univ-toulouse.fr/~besse/pub/Explo_stat.pdf.
- [10] G. Gasso, P. Lerray (date inconnue). Clustering. Cours de l'INSA Rouen, département ASI, laboratoire LITIS. Documentation accessible sur Internet.
URL <https://www.iro.umontreal.ca/~mignotte/IFT6150/ComplementCours/clustering.pdf>.
- [11] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto (2004). *Sensitivity analysis in practice*. Wiley : New York.
- [12] Anonyme (2015). *B-spline*. Wikipédia.
URL <https://fr.wikipedia.org/wiki/B-spline>.
- [13] J-Y Baudot (2016). Test d'ajustement de Kolmogorov-Smirnov. In : *Concepts et techniques organisationnelles, descriptives, prédictives et prévisionnelles pour l'entreprise, la finance et l'économie (et fondements mathématiques)*. Documentation accessible sur Internet.
URL <http://www.jybaudot.fr/Inferentielle/kolmogorov.html>.

Annexes

Annexe A - Présentation du modèle Cantis de la plateforme Sol Virtuel

Nous allons présenter les trois jeux de données provenant du modèle Cantis de la plateforme Sol Virtuel sur lesquels nous avons appliqué différents types d'analyse de sensibilité.

Facteurs d'entrée Les facteurs d'entrées du problème d'analyse de sensibilité sont dix paramètres du modèle :

- **k1, k2, k3, k4, kz, kmz** : ce sont des constantes de temps des lois de décomposition des différents compartiments de matière organique.
- **yz** : c'est le taux de carbone assimilé par la biomasse microbienne de type ZYB (le reste est consommé et rejeté sous forme de CO₂).
- **hz, ha, h1** : ce sont des fractions de biomasse humifiées (de la lignine et des biomasses microbiennes).

Nous pouvons définir les bornes des intervalles d'incertitude de ces dix facteurs d'entrée :

	ha	h1	hz	k1	k2	k3	k4	kmz	kz	yz
Min.	0	0	0	0,08	0,032	0,0485	0,0007	1	0,001	0,3
Max.	1	1	1	0,24	0,096	0,1455	0,0021	1000	0,1	0,6

Variables de sortie Les sorties sont :

- **organic matter pools ZYB** : c'est la quantité de biomasse microbienne de type ZYB sur 92 pas de temps de 1 jour.
- **carbon humification cumprof** : c'est la vitesse d'humification du carbone (en kg.m⁻².s⁻¹) sur 91 pas de temps de 1 jour.
- **solution solutes amount N03-** : c'est le stock de nitrates dans le sol (en kg.m⁻²) sur 92 pas de temps de 1 jour.

Nous avons appliqué la méthode eFAST avec $N = 1500$. Le plan final est constitué de 15000 combinaisons de valeurs des facteurs d'entrée.

Dans la figure 20, l'évolution en fonction du temps de la quantité de biomasse microbienne de type ZYB, de la vitesse d'humification du carbone et du stock de nitrates dans le sol pour quinze jeux de facteurs d'entrée différents est représentée.

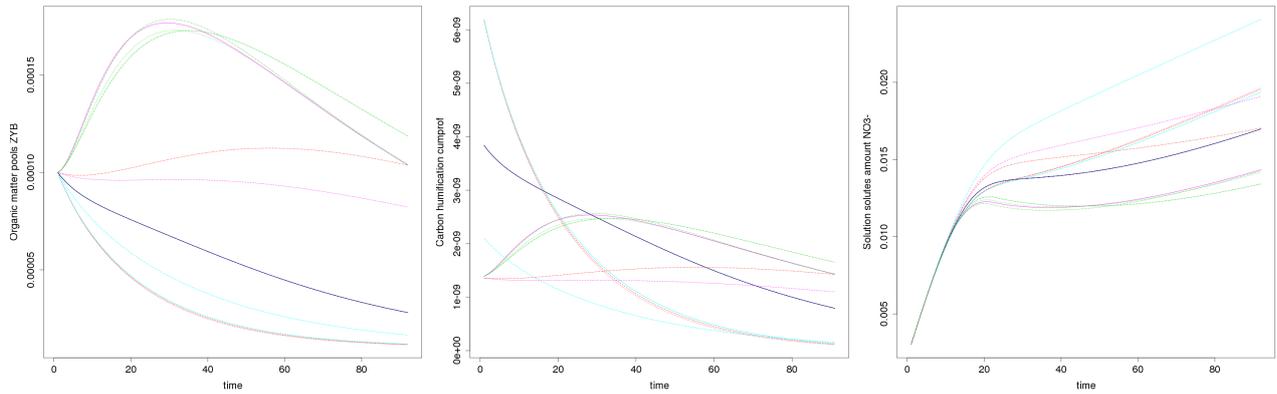


FIG. 20: Evolution en fonction du temps des sorties pour quinze jeux de facteurs d'entrée différents. Colonne de gauche : **organic matter pools ZYB**. Colonne du milieu : **carbon humification cumprof**. Colonne de droite : **solution solutes amount N03-**.

Pour **organic matter pools ZYB**, différents scénarios sont possibles :

- diminution (peu probable car cela est sûrement due à des combinaisons de facteurs d'entrée peu réaliste comme une mortalité forte par rapport à des constantes de temps de décomposition faibles).
- forte augmentation puis diminution.
- peu de variation dans le temps.

La variable de sortie est comprise dans le domaine de variation $[1e-05 ; 2,8e-04]$. 92 écarts-type sont obtenus avec les caractéristiques suivantes :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0	4,5e-05	5,6e-05	4,8e-05	5,8e-05	5,8e-05

carbon humification cumprof représente la quantité de carbone par unité de temps qui passe dans l'humus : comme le taux de carbone décroît avec le temps, la vitesse diminue. Mais le processus peut commencer brutalement et vite décroître ou se mettre en place plus doucement selon la quantité de biomasse microbienne qui humifie le carbone. Cette vitesse va dépendre aussi de la valeur de **hz**, la fraction de carbone humifiée.

La variable de sortie est comprise dans le domaine de variation $[0 ; 1,8e-08]$. 91 écarts-type sont obtenus avec les caractéristiques suivantes :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
5,5e-10	7,1e-10	1e-09	1,3e-09	1,7e-09	3,2e-09

solution solutes amount N03- est le stock d'azote minéral (nitrate) dans le sol. Nous observons une partie commune au début car le stock initial d'ammonium est fort (l'azote est sous deux formes : nitrate ou ammonium qui passe très vite sous forme de nitrate). Le stock croît globalement donc il y a plus de production que de consommation. Quand il n'y a pas de diminution, c'est que la biomasse microbienne n'est pas très importante. Quand il y a diminution, c'est le contraire : la biomasse microbienne se développe vite et a donc des besoins forts. Puis elle s'auto-consomme et

libère ainsi les nitrates consommés.

La variable de sortie est comprise dans le domaine de variation $[-7,6e-06 ; 3e-02]$. 92 écarts-type sont obtenus avec les caractéristiques suivantes :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0	2,5e-03	4e-03	3,5e-03	4,9e-03	5,4e-03

Annexe B - Présentation du modèle ToyCrop de la plateforme Record

Nous nous intéressons ici au modèle ToyCrop de la plateforme Record : rendement avec contraintes hydrique et thermique. C'est un modèle à quatre paramètres. L'objectif de ToyCrop est de simuler la dynamique de stress hydrique d'une culture par un bilan hydrique impliquant quelques flux simples.

Paramètres Les facteurs d'entrée du problème d'analyse de sensibilité sont trois paramètres du modèle :

- **tau_tmoy** : seuil de sensibilité au stress thermique (unité °C).
- **kc** : coefficient qui pilote le niveau du flux de transpiration de la plante relativement à une plante de référence (sans unité).
- **ttsw** : taille du réservoir d'eau transpirable dans le sol (unité mm).

Nous pouvons définir les bornes des intervalles d'incertitude de ces trois facteurs d'entrée :

	tau_tmoy	kc	ttsw
Min.	20	0,2	50
Max.	30	0,8	250

Variables de sortie Nous allons étudier deux sorties indépendantes l'une de l'autre en fonction de deux années climatiques assez contrastées (climat humide et climat sec) :

- la **biomasse accumulée** sur 100 pas de temps de 1 jour (du jour 180 au jour 279).
- la **fraction d'eau transpirable** (comprise dans l'intervalle [0,1] où 0 signifie « aucun stress » et 1 signifie « stress ») sur 366 pas de temps de un jour.

L'étude de ces variables de sortie en fonction du climat humide est indépendante de celle en fonction du climat sec. Nous avons appliqué la méthode eFAST avec $N = 1500$. Le plan final est constitué de 4500 combinaisons de valeurs des facteurs d'entrée pour les deux sorties dynamiques en fonction des deux années climatiques.

La **biomasse accumulée** en fonction du climat humide est comprise dans le domaine de variation [463 ;210070]. Pour chaque pas de temps, nous pouvons calculer l'écart-type de la variable de sortie. 100 écarts-type sont donc obtenus avec les caractéristiques suivantes :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
192	6010	6196	7923	8754	23710

La **biomasse accumulée** en fonction du climat sec est comprise dans le domaine de variation [2460 ;232550]. 100 écarts-type sont obtenus avec les caractéristiques suivantes :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0	5496	23720	22650	37750	48420

La **fraction d'eau transpirable** en fonction du climat humide est comprise dans le domaine de variation [3,4e-04 ;1]. 366 écarts-type sont obtenus avec les caractéristiques suivantes :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
0	2,4e-02	9,6e-02	1,3e-01	2,5e-01	2,9e-01

La **fraction d'eau transpirable** en fonction du climat sec est comprise dans le domaine de variation $[3e-04;1]$. 366 écarts-type sont obtenus avec les caractéristiques suivantes :

Min.	1er Qu.	Médiane	Moy.	3ème Qu.	Max.
1,6e-02	1,3e-01	1,8e-01	1,6e-01	2e-01	2,5e-01

Dans la figure 21 est représentée l'évolution de la **biomasse accumulée** en fonction des deux années climatiques pour soixante jeux de facteurs d'entrée différents.

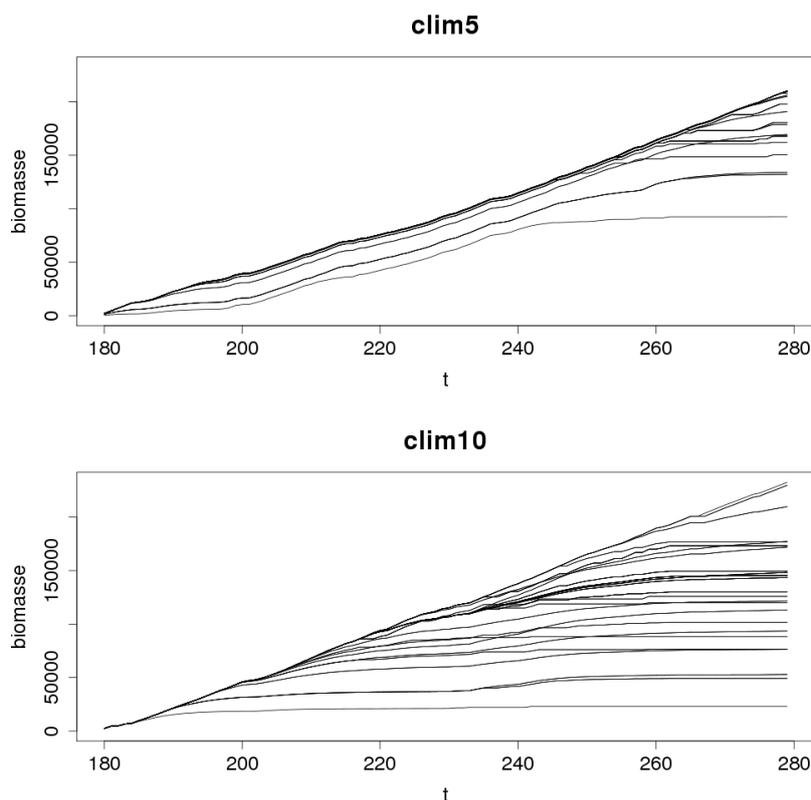


FIG. 21: Evolution en fonction du temps de la **biomasse accumulée** pour deux années climatiques et soixante jeux de facteurs d'entrée différents.

Nous pouvons remarquer que la variance de la sortie est plus grande pour le climat sec que pour le climat humide, et que celle-ci commence à augmenter de façon significative plus tôt pour le climat sec que pour le climat humide.

Dans la figure 22 est représentée l'évolution de la **fraction d'eau transpirable** en fonction des deux années climatiques pour soixante jeux de facteurs d'entrée différents.

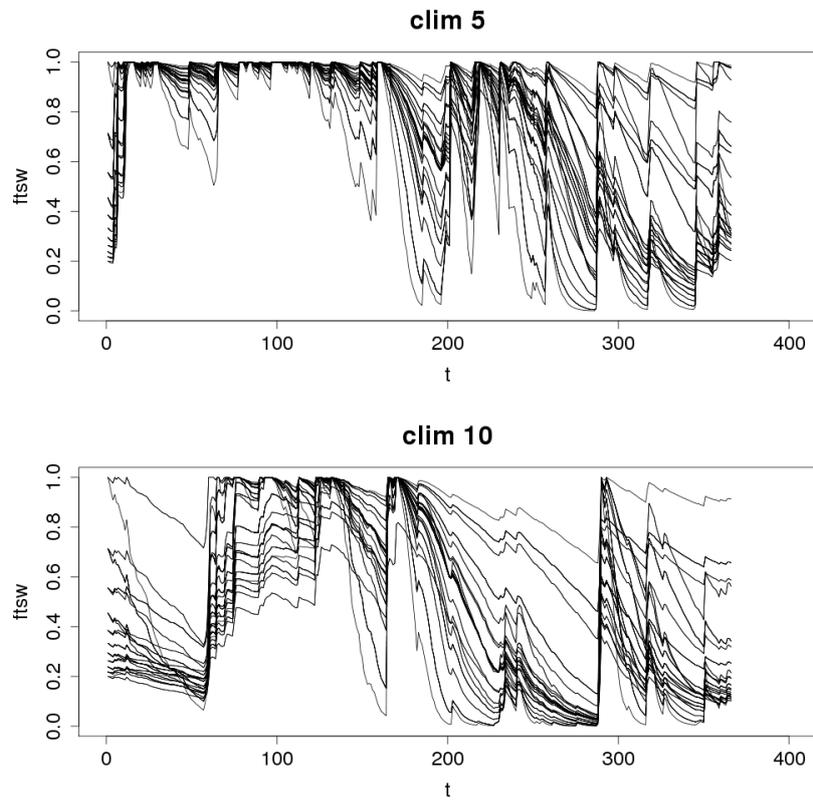


FIG. 22: Evolution en fonction du temps de la fraction d'eau transpirable pour deux années climatiques et soixante jeux de facteurs d'entrée différents.

Nous remarquons aisément sur ce graphique le contraste entre les deux années climatiques : la variance de la variable de sortie est plus élevée pour le climat sec que pour le climat humide.

Annexe C - Base polynomiale dans le cas du modèle Cantis de Sol Virtuel

Nous allons projeter les variables de sortie `organic matter pools ZYB`, `carbon humification cumprof` et `solution solutes amount N03-` sur une base polynomiale, puis nous allons exécuter une analyse de sensibilité sur chaque polynôme. Nous choisissons comme degré maximal le degré 2 pour `carbon humification cumprof` et le degré 1 pour `solution solutes amount N03-` et `organic matter pools ZYB`. Dans la figure 23 est représentée la base des polynômes retenue par la méthode pour les trois variables.

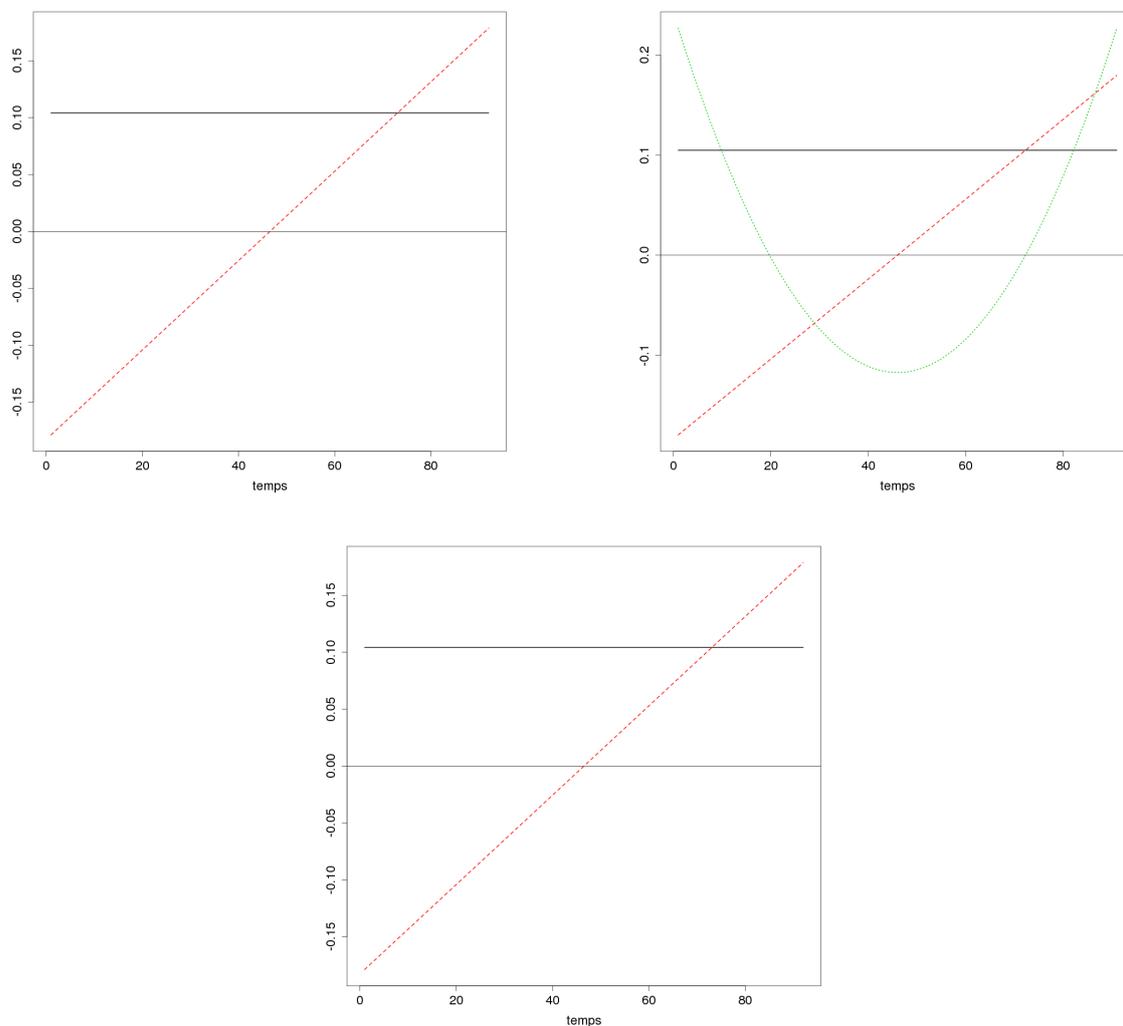


FIG. 23: Représentation de la base des polynômes. Colonne de gauche : `organic matter pools ZYB`. Colonne du milieu : `solution solutes amount N03-`. Colonne de droite : `carbon humification cumprof`.

Peu de polynômes ont été retenus afin de constituer la base complète expliquant 95% d'inertie pour les trois variables de sortie. Ainsi, il est intéressant d'étudier l'influence des facteurs d'entrée sur les coefficients H_j . Nous prenons comme exemple le cas de `organic matter pools ZYB`.

Influence des facteurs d'entrée sur les coefficients de la combinaison linéaire : cas de `organic matter pools ZYB` Nous allons évaluer l'influence des facteurs d'entrée les plus influents

sur les coefficients H_j : kz et kmz . Nous avons pu choisir ces deux facteurs d'entrée comme étant les plus influents du modèle après avoir appliqué une analyse de sensibilité sur chacun des polynômes. Dans la figure 24 est représentée l'influence de kz sur H_1 et H_2 .

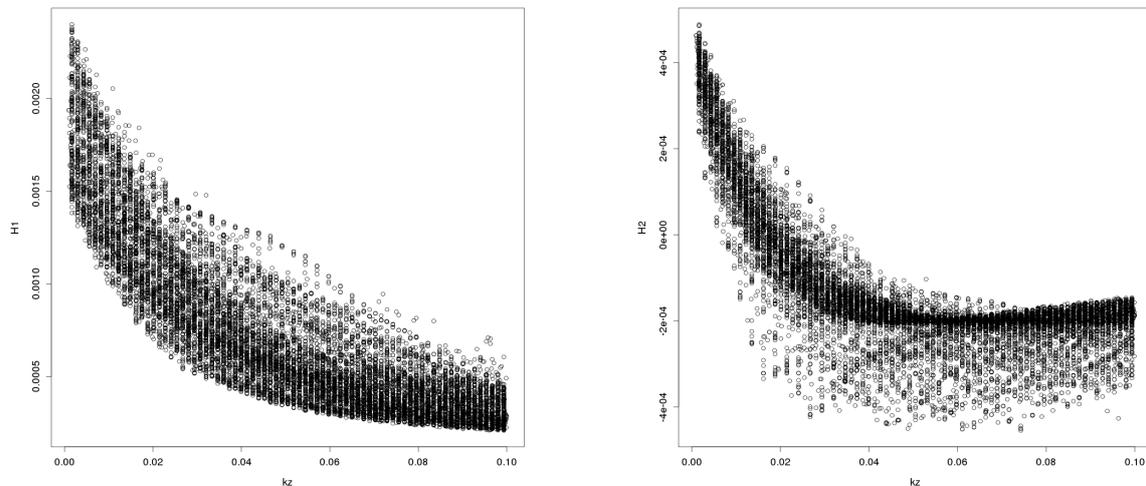


FIG. 24: Influence de kz sur les H_j . Colonne de gauche : H_1 . Colonne de droite : H_2 .

Augmenter la valeur de kz a pour conséquence une diminution des valeurs des coefficients H_1 et H_2 . Si nous relient le fait que H_1 est associé au polynôme constant (toujours positif) et que H_2 est relié au polynôme linéaire (négatif dans la première moitié de la simulation et positif dans la seconde moitié), nous pouvons dire qu'une augmentation de la valeur de kz implique une diminution de la valeur de **organic matter pools ZYB** dans la seconde moitié de la simulation. Dans la figure 25 est représentée l'influence de kmz sur H_1 et H_2 .

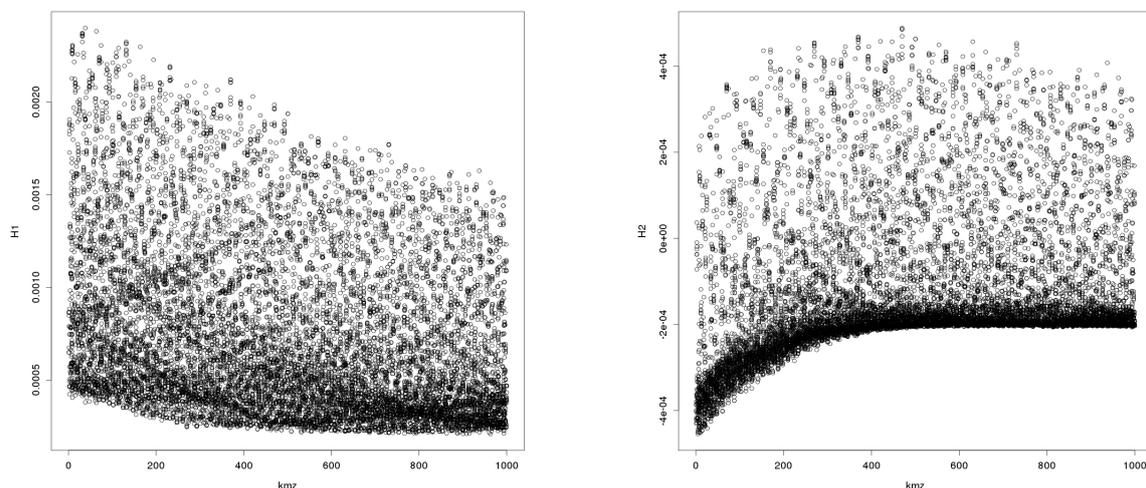


FIG. 25: Influence de kmz sur les H_j . Colonne de gauche : H_1 . Colonne de droite : H_2 .

Il n'est pas aisé d'évaluer l'influence de kmz sur H_1 : il semble qu'une augmentation de la valeur de kmz correspond grossièrement à une diminution de la valeur de H_1 . Sur H_2 , nous pouvons voir

qu'une augmentation de la valeur de \mathbf{kmz} correspond grossièrement à une augmentation de la valeur de H_2 . Cela est à moduler avec le fait qu'il y a beaucoup de points pour lesquels il n'est pas aisé d'observer un tel comportement.