



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université Amar Telidji- Laghouat

FACULTÉ: GENIE CIVIL ET D'ARCHITECTURE

DÉPARTEMENT : GENIE CIVIL

MÉMOIRE DE MASTER

Présenté par : KEMITI AbdelWahab

DOMAINE : Sciences et Techniques

FILIERE : Hydraulique

OPTION : Ressources Hydrauliques

Thème

**Analyse de sensibilité globale d'un modèle mathématique
pour l'estimation des paramètres Hydrodynamiques d'une
nappe d'eau souterraine**

Jury de soutenance :

Nom et Prénom	Grade	Qualité
Djehiche Abdelkader	Pr.	Président
Chettih Mohamed	Pr.	Rapporteur
Tadj Walid	MCA	Examineur

Juin : 2023

Remerciements

Nos remerciements les plus sincères s'adressent à :

· Monsieur M.chettih mon encadreur

Pour avoir accepté de m'encadrer, et dont les conseils,
la disponibilité et l'orientation ont rendu possible la réalisation de
ce mémoire.

· Les membres de jury qui m'ont fait l'honneur d'accepter le
jugement de mon travail.

· Corps enseignant et toute l'équipe pédagogique du Département
Génie civil

· Tous qui contribués de près ou de loin à l'élaboration de ce travail.

Dédicace

Je dédie ce projet :

A ma chère mère,

A mon cher père,

*Qui n'ont jamais cessé, de formuler des prières à mon égard, de me
soutenir*

et de m'épauler pour que je puisse atteindre mes objectifs.

A mes frères

A mon cher grand-père,

Qui je souhaite une bonne santé.

Sommaire

Introduction générale.....	1
I. Analyse de sensibilité des Modèles	3
I.1. Introduction :.....	3
I.2. Terminologie :.....	3
I.3. Modèles :.....	4
I.4. Modèles et incertitudes :	6
I.5. Propagation des incertitudes :	7
I.6. Méthodes pour la propagation d'incertitude :.....	7
I.6.1 La Méthode des moments :.....	7
I.6.2 Les méthodes d'échantillonnage :.....	8
I.6.3 La méthode de surface de réponse :.....	9
I.7. Analyse de Sensibilité :.....	10
I.7.1 Introduction :.....	10
I.7.2 Analyse documentaire :.....	10
I.7.3 Introduction à l'Analyse de sensibilité :	11
I.8. Conclusion :	16
II. Calcul Numérique des Indices de Sobol.....	17
II.1. Introduction :.....	17
II.2. Indices de Sobol :	17
II.3. Estimation des Indices de Sobol :	23
II.3.1 Algorithme de la méthode FAST :.....	23
II.3.2 Décomposition en Polynômes de Chaos :	24
II.3.3 Organigrammes de la méthode FAST :.....	27
II.3.4 Organigrammes du Polynômes de Chaos :.....	29
II.4. Conclusion :	31
III. Applications à quelques problèmes d'Hydraulique souterraine	32
III.1. Introduction.....	32
III.2. Transport de soluté en milieu poreux	32
III.3. Essais de pompage :	36
III.3.1 Nappe captive :.....	37

III.3.2	Nappe libre :	39
III.3.3	Nappe semi-captive :	43
III.3.4	Pompage en régime transitoire et fonction de puits :	44
III.4.	Résultats et Discussions :	46
III.4.1	Utilisation de la méthode FAST :	46
III.4.1.1	Résultats du Modèle de la nappe captive :	47
III.4.1.2	Résultats du Modèle de la nappe libre :	49
III.4.2	Utilisation de la méthode du Polynômes de Chaos :	50
III.4.2.1	Résultats de Transport de soluté en milieu poreux:	50
III.4.2.2	Résultats de Nappe semi-captive :	53
III.4.2.3	Résultats de Pompage en régime transitoire et fonction de puits :	56
Conclusion générale		61
Références Bibliographiques.....		63

Liste des figures

Fig. 1:Schéma simplifié d'un modèle dans le cadre d'une analyse de sensibilité.....	5
Fig. 2:Modélisation selon Rosen (1991)	5
Fig. 3:Organigramme simplifié du programme de la méthode FAST	28
Fig. 4:Etapes de calcul des indices de sensibilité de Sobol par décomposition du chaos polynômial..	29
Fig. 5:Organigramme simplifié du programme de la méthode du chaos polynomial.....	30
Fig. 6:Fonction d'erreur (erf) et fonction d'erreur complémentaire (erfc)	36
Fig. 7:Solution de l'équation de transport (Solution analytique calculée avec MATLAB)	36
Fig. 8:Vue schématique en coupe d'un puits pompant dans un aquifère confiné.....	38
Fig. 9:Abaissement de la hauteur piézométrique de l'eau souterraine dans un aquifère captif (confiné) dû au pompage (formule de Thiem)	39
Fig. 10:Vue schématique en coupe d'un puits pompant dans un aquifère libre	40
Fig. 11: Abaissement de la hauteur piézométrique de l'eau souterraine dans un aquifère liber.....	42
Fig. 12: Vue schématique en coupe d'un puits pompant dans un aquifère semi-captif.....	43
Fig. 13:Abaissement de la hauteur piézométrique de l'eau souterraine dans un aquifère semi- captif dû au pompage, d'après de Glee (1930)	44
Fig. 14:Abaissement régulier de la hauteur piézométrique des eaux souterraines dans un aquifère confiné, un aquifère semi- captif et un aquifère non confiné	44
Fig. 15:Abaissement instable de la hauteur piézométrique des eaux souterraines dans un aquifère confiné dû au pompage selon Theis ; paramètres donnés dans le texte, pour les temps	45
Fig. 16:Abaissement de la hauteur piézométrique de l'eau souterraine dans un aquifère captif dû au pompage, d'après la formule de Theis.	46
Fig. 17:Histogramme montre les paramètres qui affectent le plus le modèle la nappe captive	48
Fig. 18:Evolution de la variance totale en fonction du Nombre d'échantillon	49
Fig. 19: Evolution de l'erreur sur la variance totale.	49
Fig. 20: Histogramme montre les paramètres qui affectent le plus le modèle la nappe captive	50
Fig. 21:surface de réponse des paramètres d'entrée pour le modèle Transport de soluté en milieu poreux	52
Fig. 22:Histogramme montre les paramètres qui affectent le plus le modèle Transport de soluté en milieu poreux.....	53
Fig. 23:surface de réponse des paramètres d'entrée pour le modèle Nappe semi-captive.....	55
Fig. 24:Histogramme montre les paramètres qui affectent le plus le modèle Nappe semi-captive. ...	56
Fig. 25: surface de réponse des paramètres d'entrée pour le modèle Pompage en régime transitoire	58
Fig. 26:Histogramme montre les paramètres qui affectent le plus le modèle Pompage en régime transitoire	59

Liste des Tableaux :

Tab. 1:: Pertinence d'un paramètre d'entrée à partir de son indice de sensibilité globale.....	47
Tab. 2: Pertinence d'un paramètre d'entrée à partir de son indice de sensibilité globale.....	47
Tab. 3 : Résultats de l'analyse de sensibilité pour les paramètres du modèle la nappe captive	47
Tab. 4: : Résultats de l'analyse de sensibilité pour les paramètres du modèle la nappe libre.	49
Tab. 5: : Résultats de l'analyse de sensibilité pour les paramètres du modèle Transport de soluté en milieu poreux.....	50
Tab. 6: Résultats de l'analyse de sensibilité pour les paramètres du modèle Nappe semi-captive. ...	53
Tab. 7:: Résultats de l'analyse de sensibilité pour les paramètres du Pompage en régime transitoire.	56

Resumé :

Les progrès des connaissances sur le fonctionnement des systèmes aquifères ont conduit à la création et à l'utilisation de modèles mathématiques complexes destinés à décrire des données de terrain et expérimentales. Il est essentiel d'évaluer la capacité du modèle à représenter avec précision les données afin d'obtenir des estimations réalistes des paramètres du modèle. L'analyse de sensibilité joue un rôle important dans la détermination des paramètres qui ont un impact significatif sur la sortie du modèle et dans l'évaluation de sa qualité

Cette thèse porte sur l'étude de l'analyse de sensibilité globale et des indicateurs de Sobol. Les indicateurs Sobol sont couramment utilisés pour mesurer la sensibilité des paramètres. La deuxième section décrit les techniques numériques utilisées pour calculer ces indices et explique comment ils peuvent être interprétés afin d'identifier les paramètres qui influencent le plus les résultats du modèle. L'analyse polynomiale du chaos et FAST sont utilisées pour estimer l'influence des paramètres sur la sortie des modèles numériques utilisés. Le calcul des indices de Sobol fournit des informations précieuses sur la contribution de chaque paramètre à la variance des résultats.

Enfin les applications de l'analyse de sensibilité globale dans le domaine de l'hydraulique souterraine. Elle met en œuvre les concepts et les méthodes présentés dans les parties précédentes pour résoudre des problèmes spécifiques liés à l'estimation des paramètres hydrodynamiques des nappes d'eau souterraines. Cette partie présente des études de cas dans lesquelles l'analyse de sensibilité est utilisée pour prendre en compte différentes contraintes et incertitudes, et pour obtenir des estimations précises des paramètres hydrodynamiques. Les résultats obtenus sont discutés en mettant en évidence les implications pratiques de l'analyse de sensibilité dans le domaine de l'hydraulique souterraine.

Abstract :

Advances in our knowledge of how aquifer systems work have led to the creation and use of complex mathematical models to describe field and experimental data. It is essential to assess the model's ability to accurately represent the data in order to obtain realistic estimates of model parameters. Sensitivity analysis plays an important role in determining which parameters have a significant impact on the model's output, and in assessing its quality.

This thesis focuses on the study of global sensitivity analysis and Sobol indicators. Sobol indicators are commonly used to measure parameter sensitivity. The second section describes the numerical techniques used to calculate these indices and explains how they can be interpreted to identify the parameters that most influence model results. Polynomial chaos analysis and FAST are used to estimate the influence of parameters on the output of the numerical models used. The calculation of Sobol indices provides valuable information on the contribution of each parameter to the variance of the results.

Finally, applications of global sensitivity analysis in the field of groundwater hydraulics. It applies the concepts and methods presented in the previous sections to specific problems linked to the estimation of hydrodynamic parameters in groundwater. This section presents case studies in which sensitivity analysis is used to take into account various constraints and uncertainties, and to obtain accurate estimates of hydrodynamic parameters. The results obtained are discussed, and the practical implications of sensitivity analysis in the field of groundwater hydraulics are discussed.

ملخص :

أدى التقدم في معرفتنا بكيفية عمل أنظمة الخزان الجوفي إلى إنشاء واستخدام نماذج رياضية معقدة لوصف البيانات الميدانية والتجريبية. من الضروري تقييم قدرة النموذج على تمثيل البيانات بدقة من أجل الحصول على تقديرات واقعية لمعاملات النموذج. يلعب تحليل الحساسية دوراً مهماً في تحديد المعلمات التي لها تأثير كبير على مخرجات النموذج ، وفي تقييم جودته.

تتناول هذه الرسالة دراسة تحليل الحساسية العالمية ومؤشرات سوبول. تُستخدم مؤشرات سوبول بشكل شائع لقياس حساسية المعلمة. يصف القسم الثاني التقنيات العددية المستخدمة لحساب هذه المؤشرات ويشرح كيف يمكن تفسيرها من أجل تحديد المعلمات الأكثر تأثيراً على نتائج النموذج. يتم استخدام تحليل الفوضى متعدد الحدود واختبار حساسية سعة فوربيه لتقدير تأثير المعلمات على مخرجات النماذج العددية المستخدمة . يوفر حساب مؤشرات سوبول معلومات قيمة حول مساهمة كل معلمة في تباين النتائج.

أخيراً ، نطبق تحليل الحساسية العالمية في مجال الهيدروليكا الجوفية. نقوم بتنفيذ المفاهيم والطرق المقدمة في الأجزاء السابقة لحل مشاكل محددة تتعلق بتقدير المعلمات الهيدروديناميكية للمياه الجوفية. يقدم هذا القسم الأخير دراسات الحالة التي يتم فيها استخدام تحليل الحساسية لمراعاة القيود وأوجه عدم اليقين المختلفة ، وللحصول على تقديرات دقيقة للمعاملات الهيدروديناميكية. تمت مناقشة النتائج التي تم الحصول عليها ، وتناقش الآثار العملية لتحليل الحساسية في مجال المياه الجوفية.

Introduction Générale

La compréhension du fonctionnement des systèmes aquifères est devenue un enjeu majeur compte tenu de la problématique de la gestion des ressources en eau. La modélisation des nappes d'eau souterraine repose sur des équations aux dérivées partielles décrivant l'évolution spatio-temporelle de certaines grandeurs physiques.

De nos jours, ces équations peuvent être résolues par des méthodes numériques robustes donnant des résultats précis même pour des milieux naturels présentant une grande hétérogénéité spatiale. Toutefois, la modélisation des nappes d'eau souterraine est par essence incertaine : d'une part les données disponibles permettant la caractérisation du système aquifère sont limitées car souvent coûteuses ; d'autre part les données quand elles sont disponibles sont souvent entachées d'erreurs ; enfin, les modèles mathématiques sont eux-mêmes incertains car ils ne sont qu'une approximation de la réalité. L'ensemble de ces incertitudes influence de manière considérable les résultats de la modélisation. Ces sources d'incertitudes doivent être quantifiées et leur effet sur la sortie du modèle étudié. Ainsi, les prédictions du modèle doivent non seulement inclure les résultats les plus probables mais également les incertitudes associées.

Ainsi, l'analyse de sensibilité constitue un ensemble de méthodes qui permettent d'explorer la relation entre les entrées et les sorties du modèle et de quantifier l'impact de chaque variable d'entrée sur la variable de sortie d'intérêt. A ce titre, l'analyse de sensibilité globale est souvent préférée car elle fournit une compréhension plus complète du fonctionnement du modèle et des variables d'entrée les plus importantes.

Il existe plusieurs méthodes pour effectuer une analyse de sensibilité globale, notamment des méthodes basées sur la variance, des méthodes de dépistage et des méthodes indépendantes du moment.

L'objectif de ce travail est d'effectuer une analyse de sensibilité globale à l'aide des indices de Sobol pour mesurer l'influence relative des paramètres incertains sur la sortie de quelques problèmes d'hydraulique souterraine.

Ainsi, nous proposons d'utiliser les tests de sensibilité à l'amplitude de Fourier et la décomposition en polynôme de chaos pour estimer les indices de Sobol et classer les paramètres des modèles.

Ce manuscrit comprend trois chapitres :

Dans le premier chapitre, nous explorons en détail les principaux termes et définitions requis pour les sujets sélectionnés, nous introduisons également les fondements théoriques de l'analyse de l'incertitude, ses sources, l'analyse de sensibilité, ses objectifs et ses méthodes.

(Dans le deuxième chapitre : nous avons présenté les indicateurs Sobol et leurs méthodes de calcul, Ce sont des outils puissants pour mesurer la sensibilité de modèles complexes à de nombreux paramètres ,et expliqué en détail Les méthodes adoptées dans ce travail dans notre travail FAST et chaos polynomial, et mis les organigrammes et étapes pour effectuer l'analyse de sensibilité pour ces deux méthodes.

Le troisième chapitre est consacré à la description des concepts de base des différents modèles physiques et mathématiques régissant fonctionnement des systèmes aquifères.Ce chapitre met en pratique les concepts et méthodes présentés et en appliquant l'analyse de sensibilité globale et les indices de Sobol à des problèmes concrets de hydraulique souterraine. En évaluant l'impact des paramètres hydrodynamiques sur les résultats du modèle, ce chapitre a permis de mieux cibler les efforts d'estimation des paramètres et de prendre des décisions éclairées sur la gestion des ressources en eaux souterraines.

Analyse de sensibilité des Modèles

Chapitre 1

I. Analyse de sensibilité des Modèles

I.1. Introduction :

L'analyse de sensibilité joue un rôle important dans la compréhension et l'évaluation de modèles et de complexes d'utilisation dans de nombreux domaines de la science et des techniques. Il permet d'explorer l'effet des différences d'entrée sur la sortie du modèle, en identifiant les facteurs les plus influences et en quantifiant les contributions de chacun. Dans ce chapitre, nous étudions en détail la terminologie clé requise pour les sujets retenus. Le chapitre a commencé par introduire les termes et définitions clés. La section suivante explique l'importance des incertitudes dans la modélisation informatique et donne un aperçu des différentes sources d'incertitudes des modèles. Nous fournissons ensuite des Explications sur l'analyse de sensibilité, ses objectifs et ses méthodes, et continuons dans les chapitres suivants sur la façon de calculer.

I.2. Terminologie :

Cette section établit la terminologie essentielle et les définitions nécessaires en ce qui concerne l'analyse de sensibilité. D'une manière très générale, l'analyse de sensibilité peut être définie comme suit :

L'analyse de sensibilité est la quantification des contributions relatives des données d'entrée à la sortie d'un modèle en déterminant comment les variations des données d'entrée affectent la sortie observée.

Cette définition nécessite la clarification des trois termes : modèle, entrée et sortie (Brandstätter, 2021).

- *Un modèle est une représentation simplifiée et instructive d'un système.*
- *Une sortie de modèle est une réponse, une donnée ou une information produite par un modèle.*
- *Une entrée de modèle est tout ce qui peut entraîner une variation de la sortie du modèle*

Cette définition simple du modèle couvre les modèles physiques, tels que les expériences et les prototypes de produits, mais aussi les modèles abstraits tels que les équations mathématiques et les codes informatiques.

Les sorties de modèle comprennent les quantités mesurées lors d'une expérience, ou pour les modèles mathématiques, toute fonction objective de la réponse du modèle.

Pour une analyse de sensibilité significative, il est donc primordial de déterminer quelles sorties doivent être étudiées. Naturellement, ce choix dépend du contexte et de l'objectif de l'analyse. De nombreuses décisions doivent être prises lors de la configuration du modèle, notamment le type et la structure du modèle, les paramètres, les conditions limites et initiales, la résolution, la méthode de discrétisation et bien d'autres encore (Saltelli et al., 2008 ; Brandstätter, 2021). Tous ces choix sont associés à une certaine incertitude, sont donc des hypothèses et peuvent en général être considérés comme des entrées du modèle qui, si elles varient, produisent une valeur de sortie différente.

L'analyse de sensibilité peut être également définie comme étant l'étude de la manière dont l'incertitude dans la sortie d'un modèle peut être répartie entre différentes sources d'incertitude dans l'entrée du modèle (Saltelli et al., 2004).

En d'autres termes c'est une méthode utilisée pour comprendre comment les résultats d'un modèle changent en réponse à des changements dans les paramètres ou les variables d'entrée. Elle est largement utilisée dans divers domaines, notamment les sciences physiques, les sciences de la terre, la biologie, l'économie, l'ingénierie et la finance (Brandstätter, 2021).

Une classification très importante des méthodes d'analyse de sensibilité implique la distinction entre les méthodes locales et les méthodes globales.

- **Analyse de sensibilité locale :**

L'analyse de sensibilité locale se concentre sur la variation de la sortie du modèle lorsque les entrées sont perturbées autour d'une valeur nominale.

L'analyse de sensibilité locale est souvent réalisée en analysant la dérivée de la sortie par rapport aux entrées individuelles. De cette manière, l'analyse de sensibilité locale se retrouve régulièrement dans le contexte de l'optimisation, des méthodes adjointes et de calage de modèles déterministes. Cette approche présente toutefois de sérieuses limitations en ce qui concerne le traitement des modèles non linéaires. Par conséquent, les modèles non linéaires nécessitent une perspective globale. Alors que l'analyse de sensibilité locale limite l'analyse à un seul point d'intérêt dans l'espace d'entrée du modèle, les méthodes globales couvrent l'ensemble de l'espace d'entrée.

Dans ce travail, on définira l'analyse de sensibilité globale selon (Saltelli et al., 2008) comme suit :

- **Analyse de sensibilité globale :**

L'analyse de sensibilité globale est l'étude de la manière dont l'incertitude des résultats du modèle peut être répartie entre les contributions des incertitudes des données d'entrée du modèle.

I.3. Modèles :

A la lumière de la définition donnée précédemment, plusieurs questions liées aux modèles restent pertinentes. Parmi lesquelles :

- Le modèle fonctionne-t-il correctement ?
- Répond-il de manière appropriée aux variations des paramètres ?
- Quels processus sont prédominant dans le système étudié ?
- Quel est le poids des incertitudes des différents paramètres d'entrée sur la simulation des sorties ?

La figure 01, simplifie la conception d'un modèle qui peut être expérimental, abstrait ou simplement un code informatique. Bien entendu, lors de la conceptualisation du modèle on devrait bien spécifier les variables d'entrée explicatives ou de forçage, la structure du modèle, les paramètres, les conditions limites et initiales, ...

L'analyse de sensibilité permet d'étudier l'effet respectif de ces entrées du modèle sur ses variables de sortie constituant les variables d'intérêt. Cependant, une analyse de sensibilité ne peut toutefois analyser qu'un sous-ensemble de toutes les entrées. Les entrées de ce sous-ensemble d'intérêt sont généralement appelées facteurs d'entrée.

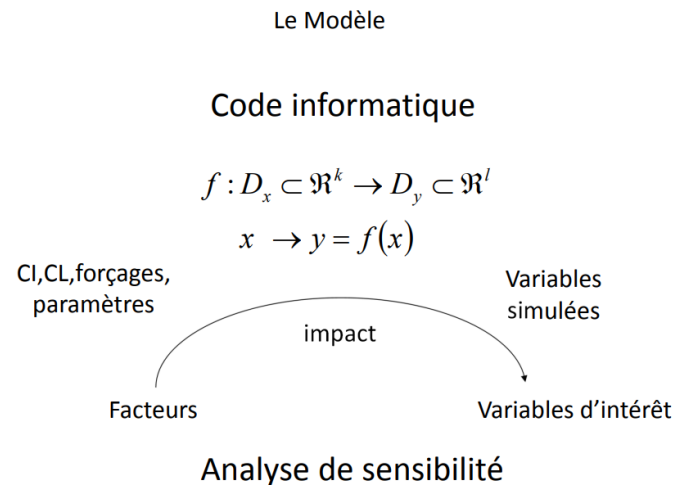


Fig. 1: Schéma simplifié d'un modèle dans le cadre d'une analyse de sensibilité.

La figure 02 ci-dessous tirée des travaux de Robert (Rosen, 1991) présente une vision de la modélisation qui peut aider à illustrer le rôle de l'analyse de sensibilité dans le processus scientifique. Sur la gauche du diagramme de Rosen se trouve le "monde", c'est-à-dire le système qui fait l'objet de recherche. Le système, qu'il soit naturel ou artificiel, est régi par des règles qu'on souhaite découvrir. À cette fin, on suppose des hypothèses sur un ensemble de structures dans un modèle (représenté dans la partie droite de la figure).

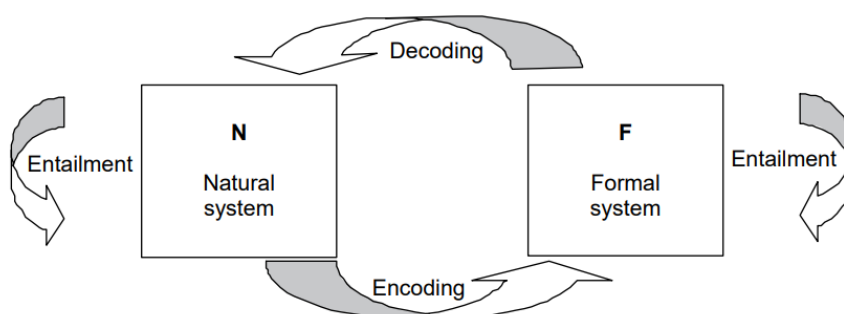


Fig. 2: Modélisation selon Rosen (1991)

I.4. Modèles et incertitudes :

Ce qui rend la modélisation et la recherche scientifique en général si pénibles, c'est l'incertitude. L'incertitude n'est pas un accident de la méthode scientifique, mais sa substance.

Les différentes sources d'incertitude peuvent être classées en trois catégories :

- **L'incertitude structurelle :**

Cette incertitude surgit au moment du passage du phénomène réel au modèle mathématique. En effet, les modèles mathématiques sont une approximation de la réalité. Cette approximation est basée sur plusieurs hypothèses simplificatrices qui constituent des sources d'incertitude.

- **L'incertitude numérique :**

Les modèles mathématiques en Hydrogéologie sont souvent décrits par des équations aux dérivées partielles traduisant la variation dans le temps et/ou dans l'espace des échanges de quantité de mouvement, de masse ou d'énergie. Pour ces modèles, il n'existe généralement pas de solution analytique et nous avons recours à des solutions numériques. Ces solutions contiennent des imprécisions inhérentes à toute simulation numérique. Ces imprécisions sont dues au fait que nous approchons un système continu par un autre discret (dans l'espace et/ou dans le temps). Ces sources d'incertitude ne peuvent être éliminées complètement, mais seulement réduites en utilisant des discrétisations spatiales et temporelles définies à plus petites échelles. Notons que pour les transferts en milieux poreux, l'échelle de discrétisation spatiale ne doit pas descendre en dessous de l'échelle de validité des équations (échelle du Volume Élémentaire Représentatif). D'autres sources d'incertitudes viennent se rajouter à celles dues à la discrétisation telles que les incertitudes liées aux erreurs de séparation d'opérateurs ou aux erreurs de résolution des systèmes non linéaires.

- **L'incertitude paramétrique**

Les modèles contiennent souvent un nombre important de paramètres présentant des incertitudes importantes liées à leur estimation, tels que les constantes physiques, les valeurs initiales, les conditions aux limites, la géométrie initiale,.... On distingue alors l'incertitude aléatoire de l'incertitude épistémique. L'incertitude aléatoire provient de la variabilité naturelle des paramètres d'entrée du modèle. On l'explique parfois par l'incertitude responsable de l'obtention de résultats différents lorsque l'on répète plusieurs fois dans des conditions identiques une expérience.

Alors que l'incertitude épistémique découle du manque de connaissance, telle que l'estimation des paramètres par le biais d'un échantillon ne contenant pas assez de mesures, ou un échantillon non représentatif. Cette incertitude peut être réduite en recueillant des informations supplémentaires. Les modèles d'écoulement et de transport dans les milieux poreux sont souvent caractérisés par une forte variabilité spatiale et temporelle des processus et des incertitudes sur les paramètres d'entrée. Connaître l'influence des différents paramètres d'entrée et quantifier l'impact de leur incertitude sur la réponse du modèle constitue alors l'objet de la "propagation des incertitudes" (Saltelli et al., 2008).

I.5. Propagation des incertitudes :

Après avoir identifiées et caractérisées les différentes sources d'incertitude, elles doivent être propagées dans le modèle numérique, afin d'évaluer leur impact sur la variabilité de la sortie du modèle. A ce titre, il est nécessaire de choisir une méthode pour explorer le modèle numérique.

Les approches probabilistes dont les paramètres incertains sont représentés par des variables aléatoires ou des champs stochastiques constituent les techniques les plus utilisées (Sobol, 1967 ; Faure, 1982). Ces approches probabilistes ont reçu un intérêt particulier depuis les années 70 et ont été largement utilisées dans la littérature (Fajraoui, 2014). Avec ces approches, les incertitudes liées aux entrées sont décrites par des distributions de probabilité. L'analyse d'incertitude comporte deux étapes :

- La détermination des densités de probabilité des paramètres d'entrée : Ceci est accompli à l'aide des techniques d'estimation statistique ou à partir de jugements d'experts. Les techniques d'estimation statistique permettent l'estimation des distributions de probabilité à partir des données disponibles ou par le biais d'un échantillon informatif. Dans les cas où les données disponibles sont limitées et /ou bien non informative, un jugement d'expert fournit les informations sur la distribution de probabilité d'entrée. Ainsi, par exemple, une distribution uniforme est sélectionnée si une seule gamme de valeurs possibles pour une entrée est disponible, mais qu'aucune valeur préférentielle susceptible de se produire n'est favorisée.
- La propagation de l'incertitude à travers le modèle : il s'agit de propager l'incertitude des paramètres dans le modèle et de caractériser la variabilité induite de la sortie du modèle. Cette dernière sera donc caractérisée par une densité de probabilité qui est dépendante de la variabilité des paramètres d'entrée.

I.6. Méthodes pour la propagation d'incertitude :

Plusieurs méthodes de propagation d'incertitude ont été proposées dans la littérature (Fajraoui, 2014). Elles font généralement partie des trois grandes classes de méthodes qui seront décrit ci-dessous : la méthode des moments, les méthodes d'échantillonnage et la méthode de surface de réponse.

On considère la sortie scalaire Y d'un modèle représentée par une fonction mathématique $f(x)$ où les entrées scalaires $x = \{x_1, \dots, x_n\}$ sont définies sur l'hypercube unité $\Omega_n = [0,1]^n$.

I.6.1 La Méthode des moments :

L'objectif est d'obtenir la moyenne et la variance de $Y = f(x)$ à partir de la connaissance de l'incertitude sur les paramètres d'entrée. Si toutes les variables sont continues, les deux premiers moments sont définis comme suit :

$$\mu_Y = E[f(x)] = \int_{\Omega_n} f(x) p_x(x) dx \quad (1)$$

$$\sigma_Y^2 = E[(f(x) - \mu_Y)^2] = \int_{\Omega_n} (f(x) - \mu_Y)^2 p_x(x) dx \quad (2)$$

où p_x est la fonction de probabilité conjointe correspondant aux distributions des paramètres d'entrée. Ces deux moments peuvent être évalués à l'aide de plusieurs méthodes comme par exemple la méthode de quadrature ou la méthode de perturbation (Fajraoui, 2014).

1.6.2 Les méthodes d'échantillonnage :

Dans cette section on se contentera de deux méthodes d'échantillonnage les plus utilisées, la méthode de Monté Carlo et la méthode du Latin Hypercube Sampling.

• La méthode de Monté Carlo :

La méthode de MC a été largement utilisée dans l'analyse d'incertitude depuis les années 1940. Pour un problème contenant n paramètres incertains $(x_1, \dots, x_n)$, le principe de base de la méthode est le suivant :

- Générer un échantillon de N réalisations du vecteur $(x_1, \dots, x_n)$ à partir de leur densité de probabilité jointe ;
- Calculer la sortie du modèle déterministe pour chaque jeu de paramètre d'entrée ;
- Calculer les statistiques de la distribution de la sortie : moyenne, variance, densité de probabilité, ...

Ainsi, Il est possible d'estimer l'espérance et la variance de la sortie Y en utilisant les expressions suivantes :

$$E[Y] \approx \frac{1}{N} \sum_{i=1}^N f(x^i) \quad (3)$$

$$Var[Y] \approx \frac{1}{N-1} \sum_{i=1}^N (f(x^i) - E[x])^2 \quad (4)$$

La méthode de Monté Carlo est généralement applicable à tous les problèmes d'analyse d'incertitude. L'inconvénient majeur de la méthode est qu'elle est très gourmande en temps de calcul car elle nécessite un nombre important d'échantillons pour garantir une estimation précise des statistiques du modèle.

• La méthode d'échantillonnage par hypercube latin :

Plusieurs autres techniques d'échantillonnage, utilisant un nombre d'échantillons moindre que celui requis par la méthode Monté Carlo, ont été développées dans la littérature (Fajraoui, 2014). La technique d'échantillonnage par hypercube latin, développée par (McKay et al ,2000), est l'une des techniques largement utilisée dans la littérature. L'échantillonnage par hypercube latin est une méthode statistique pour l'échantillonnage quasi-aléatoire, à partir d'une loi de probabilité à plusieurs variables, inspirée de la méthode de Monte-Carlo. Cette technique permet de générer des échantillons avec une meilleure couverture (plus uniforme)

du domaine de variation des paramètres que la méthode Monté Carlo. En effet, les échantillons générés par la méthode Monté Carlo ont tendance à se regrouper autour de la région de plus forte probabilité et très peu de points seront échantillonnés dans des régions à faible probabilité.

La technique utilise un schéma de simulation par stratification pour balayer tout l'espace des entrées aléatoires du modèle. Le principe de la méthode est le suivant :

- Diviser la fonction de répartition de chaque variable aléatoire d'entrée en N intervalles équidistants, avec N la taille de l'échantillon fixée à priori. Ainsi, le domaine de variation de chaque paramètre est stratifié en intervalles disjoints et équiprobables.
- Pour chaque paramètre d'entrée, on échantillonne aléatoirement une seule fois dans chacun de ses sous-intervalles, de tel sorte que toute la gamme de variation du paramètre est explorée. On obtient ainsi, N valeurs pour chaque paramètre, affectée d'un nombre qui correspond au numéro du sous-intervalle. Ces numéros sont ensuite permutés aléatoirement pour donner un nouvel ordre de l'échantillon.
- Les valeurs ainsi obtenues pour chaque variable sont alors associés dans l'ordre pour obtenir N n-uplets de valeurs.

Cela donne un échantillon qui est bien réparti sur toute la gamme de variation de chaque entrée, ce qui peut donner une distribution de probabilité plus précise de la sortie. La méthode suppose que l'échantillonnage est effectué d'une façon indépendante pour chaque paramètre.

I.6.3 La méthode de surface de réponse :

Les méthodes basées sur l'échantillonnage requièrent souvent un grand nombre de simulations. Ces méthodes sont donc très gourmandes en temps de calcul surtout si une seule simulation avec le modèle direct nécessite plusieurs heures de calcul (Fajraoui, 2014). Par conséquent des alternatives ont été développées en utilisant des métamodèles ou des modèles de substitution.

Parmi ces alternatives, la méthode de surface de réponse cherche à caractériser la relation entre les paramètres incertains d'entrée et la sortie du modèle. L'idée est de décomposer la sortie en composantes déterministes et composantes stochastiques sous la forme :

$$f(x) = \sum_{i=1}^P f_i \psi_i(x) \quad (5)$$

Par la suite, il suffit de :

- Construire un plan d'expérience (plusieurs jeux de paramètres d'entrée) ;
- Evaluer le modèle direct pour ces jeux de paramètres ;
- Ajuster le métamodèle avec les données du modèle direct.

- Le métamodèle ainsi construit est ensuite utilisé pour remplacer le modèle direct et toutes les statistiques liées à l'analyse d'incertitude sont calculées à l'aide du métamodèle. Les inconvénients de cette approche incluent la difficulté à construire une bonne approximation de la surface de réponse en présence de discontinuités et/ou de non-linéarités.

I.7. Analyse de Sensibilité :

I.7.1 Introduction :

Au début de ce chapitre, on a défini sommairement l'analyse de sensibilité et donné la principale classification des méthodes d'analyse de sensibilité, dans cette section, on présentera un bref résumé des principes fondamentaux de l'analyse de sensibilité, ainsi qu'un bref aperçu des méthodes existantes. Par la suite, on tentera de présenter la dérivation plus ou moins détaillée de la méthode d'analyse de sensibilité globale basée sur la variance, appelée indices de Sobol.

I.7.2 Analyse documentaire :

Dans cette section, une liste d'excellents manuels est présentée d'une façon très brève et très objective. Les méthodes locales et hybrides locales-globales ont été examinées en détail dans la monographie en deux volumes de (Cacuci,2003) et (Cacuci et al.,2005). Le livre de (Satelli et al , 2008) constitue certainement le manuel de référence sur l'analyse de sensibilité globale d'où on a puisé les principales définitions et les différentes méthodes pour ce mémoire. Une introduction courte, concise et très accessible aux méthodes locales et globales se trouve dans les chapitres 14 et 15, respectivement, de l'ouvrage de (Smith,2013). De même, la partie IV du très complet "Handbook of Uncertainty Quantification" de (Ghanem et al, 2017) offre une vue d'ensemble récente des approches bien établies de l'analyse de sensibilité locale et globale. (Iooss et Lemaître,2015) se concentrent sur les méthodes globales et fournissent une classification et un diagramme de décision qui peuvent aider à trouver une méthode appropriée en fonction de l'objectif de l'étude et des propriétés du modèle. Une introduction mathématique approfondie à la quantification de l'incertitude et à l'analyse de sensibilité peut être trouvée dans le livre de (Sullivan,2015). Enfin, le chapitre 7 de (Santner et al,2003) présente l'analyse de sensibilité dans le contexte de la conception et de l'analyse d'expériences informatiques

Il existe également de nombreux articles de synthèse de très bonne qualité sur l'analyse de sensibilité. Dans (Borgonovo et Plischke,2016), les auteurs présentent un examen de l'analyse de sensibilité locale et globale dans le contexte de la recherche opérationnelle. (Wentworth et al, 2016) montrent de manière structurée et accessible, comment les méthodes d'analyse de sensibilité globale peuvent être utilisées pour la sélection des paramètres dans les modèles biologiques et physiques, avec un exemple de modèle dynamique du VIH. De même, (Eck et al, 2016) proposent un "Guide de la quantification de l'incertitude et de l'analyse de sensibilité dans les applications cardiovasculaires". (Pianosi et al, 2016) présentent un examen systématique des méthodes d'analyse de sensibilité, y compris une classification informative.

Une vue d'ensemble récente des progiciels disponibles est proposée par (Douglas-Smith et al, 2020). (Qian et Mahdi,2020) ont écrit une revue complète mais concise des méthodes d'analyse de sensibilité en soulignant leur importance pour les sciences biomédicales. Plus récemment, (Razavi et al, 2021) ont présenté un manifeste pour l'analyse de sensibilité globale en tant que "discipline essentielle pour la modélisation des systèmes". Les auteurs offrent non seulement une excellente vue d'ensemble de l'état de l'art, mais aussi des perspectives d'avenir.

I.7.3 Introduction à l'Analyse de sensibilité :

Comme décrit précédemment, un grand nombre d'excellents textes introduisent les concepts de base de l'analyse de sensibilité. Le contenu et la structure sont assez similaires dans la plupart des textes, ce qui peut s'expliquer par le fait qu'il n'y a qu'un nombre limité d'approches différentes que l'on peut adopter pour introduire les concepts fondamentaux et les quantités essentielles de manière cohérente. Le contenu de ce paragraphe est toutefois compilé à partir de nombreux textes cités précédemment.

- **Objectifs de l'analyse de sensibilité :**

Les raisons de réaliser une analyse de sensibilité sont diverses. Comme pour toute étude, il est essentiel de bien comprendre les informations que l'on souhaite obtenir en réalisant une analyse de sensibilité (Borgonovo et Plischke, 2016). Les objectifs d'une analyse de sensibilité peuvent être les suivants (Saltelli et al., 2008 ; Smith, 2013 ; Iooss et Saltelli, 2017 ; Iooss et Prieur, 2019 ; Razavi et al., 2021) :

- L'analyse de sensibilité permet de déterminer si les résultats du modèle sont robustes ou vulnérables aux perturbations des données d'entrée ;
- Elle peut également aider à explorer les causalités : comment les combinaisons et les interactions entre les hypothèses concurrentes ou les divers paramètres d'entrée influencent-elles un modèle ?
- Elle peut soutenir le processus de prise de décision en situation d'incertitude ;
- Elle permet l'identification des paramètres non influents du modèle et de les fixer à n'importe quelle valeur nominale dans leur plage de variabilité sans réduction significative de l'incertitude du résultat ;
- Elle permet la hiérarchisation des paramètres à l'aide de l'analyse des priorités qui vise à identifier les paramètres les plus importants. Ce sont ceux qui, s'ils étaient fixés, entraîneraient une réduction significative de l'incertitude des résultats ;
- Elle permet une cartographie des paramètres, dans ce cas, il peut y avoir une région d'intérêt particulier dans l'espace de sortie. On peut alors vouloir savoir s'il existe des entrées ou des combinaisons d'entrées qui sont responsables des sorties dans ce domaine d'intérêt particulier.
-
- L'analyse de sensibilité permet aussi la réduction de la variance, ceci est valable pour certaines applications de modèles, où il existe des valeurs seuils prédéfinies pour l'incertitude de la sortie ;
- Lors du calage du modèle, l'analyse de sensibilité peut aider à identifier les résultats du modèle qui sont sensibles aux paramètres qui doivent être estimés. Si les données ne sont disponibles que pour un résultat insensible, on peut s'attendre à ce qu'une estimation fiable des valeurs des paramètres soit difficile. Il est également possible d'appliquer les méthodes d'analyse de sensibilité directement à la fonction objective ou à la phase de calage.

- **Aperçu des méthodes existantes :**

En réalité, il existe une myriade de méthodes d'analyse de sensibilité et toutes les méthodes ne conviennent pas de la même manière à chacun des objectifs présentés précédemment. Dans ce paragraphe, on présente une vue d'ensemble des classes de méthodes bien établies. Par conséquent, seule l'idée conceptuelle des méthodes, leurs avantages et leurs limites seront exposés sans aucun détail mathématique.

- **Approches Locales**

- **Approche Individuelle**

La façon la plus simple d'étudier la sensibilité de la sortie d'un modèle est peut-être de faire varier une entrée après l'autre et d'observer le changement dans la sortie. L'approche est intuitive et formellement décrite par le maintien de toutes les entrées à leur valeur nominale, à l'exception d'une seule qui est modifiée d'une fraction de sa valeur nominale (Qian et Mahdi, 2020). La variation fractionnelle de la sortie qui en résulte peut être utilisée pour vérifier la sensibilité de la sortie à l'unique entrée modifiée. Les mesures quantitatives de sensibilité basées sur cette approche fournissent à la fois la direction et l'ampleur du changement qui peuvent, à leur tour, être utilisées pour classer les paramètres en fonction de leur impact (Borgonovo et Plischke, 2016). Parfois, la méthode est utilisée pour une évaluation qualitative uniquement, par exemple en traçant la sortie du modèle pour la valeur nominale par rapport à la valeur modifiée. Au fond, la méthode est très étroitement liée à ce que l'on appelle les études de paramètres dans la modélisation informatique. La simplicité de la méthode est un grand avantage, mais sa nature locale, c'est-à-dire la référence à une valeur nominale, est un facteur très limitant. En changeant un seul paramètre à la fois, il est impossible de détecter les interactions entre les paramètres, c'est pourquoi les méthodes "un à la fois" ne conviennent qu'aux modèles linéaires. Une façon naturelle d'étendre l'approche pourrait être de répéter le processus pour différents changements fractionnaires. Bien que cela augmente certainement le contenu de l'information, cela n'aide pas à surmonter la restriction de la méthode aux modèles linéaires.

- **Approche basée sur les dérivées**

L'archétype d'une mesure de sensibilité locale est la première dérivée partielle de la production par rapport à l'une de ses entrées, évaluée à une valeur nominale d'intérêt. Par définition, les dérivées partielles donnent une idée de l'ampleur de la variation de la production due à une petite variation (infinitésimale) d'un intrant, les autres étant maintenus constants. Une valeur élevée implique qu'une petite perturbation de l'entrée par rapport à sa valeur nominale entraîne un changement important de la sortie et que la sortie peut être considérée comme sensible à cette entrée. De nombreux chercheurs sont familiarisés avec cette notion de sensibilité. Les applications de la méthode sont multiples. Par exemple, l'analyse de fiabilité et l'évaluation des risques s'appuient traditionnellement sur cette méthode. Elle est également étroitement liée à l'étalonnage des modèles (déterministes), à l'assimilation des données et, en général, à l'optimisation. La difficulté d'application de la méthode réside dans le calcul des dérivées, en particulier dans le cas d'un système de grande taille avec de nombreuses entrées. En général, il existe plusieurs façons d'aborder cette question. Si elles sont correctement formulées, les approches "une à la fois" décrites ci-dessus peuvent être considérées comme une approche naïve. Une approche plus raffinée consisterait à utiliser des schémas d'approximation numérique tels que les différences finies. Bien que cette approche soit simple, son coût de calcul est souvent prohibitif. Si la base de code le permet, la différenciation automatique peut être utile dans ce cas (Hwang et al., 1997). Une méthode complètement différente mais élégante, en particulier pour les cas de très grandes dimensions, consiste à recourir à des méthodes adjointes (Cacuci, 2003).

- **Approches Globales**

Les approches globales sont presque exclusivement basées sur des échantillons. Elles nécessitent un échantillon entrée-sortie où chaque point d'échantillonnage est constitué d'un ensemble de valeurs d'entrée et de la valeur de sortie correspondante. Comme pour toute approche basée sur un échantillon, la valeur informative de l'échantillon est déterminante pour le succès.

Étant donné que l'on s'intéresse à une perspective globale, les points d'échantillonnage doivent couvrir suffisamment l'ensemble de l'espace d'entrée. Il existe de nombreuses stratégies d'échantillonnage différentes qui tentent de répondre à cette exigence de manière optimale. Une stratégie d'échantillonnage très générale et donc populaire est la méthode de Monte Carlo (Robert & Castella, 1999). Toutefois, des approches telles que les plans d'expériences en cube latin (Mckay et al, 2000) ; Sheikholeslami et Razavi ,2017) et les séquences quasi-aléatoires à faible discrédance telles que les séquences de (Sobol,1967 ;sobol,2011) se sont avérées à plusieurs reprises plus performantes que les méthodes de Monte Carlo classiques dans les applications d'analyse de sensibilité (Kucherenko et al ,2015 ;Piano et al,2021).

○ Méthodes visuelles

Les outils graphiques sont des méthodes de base mais essentielles pour l'analyse de sensibilité globale. Sur la base d'un échantillon d'entrées-sorties donné, les méthodes visuelles peuvent être facilement appliquées. Dans ce qui suit, les deux méthodes les plus courantes sont présentées. ci-après, les deux méthodes les plus courantes sont présentées (Pianosi et al , 2016 ; Qian et Mahdi,2020).

Un ensemble de diagrammes de dispersion entre la sortie du modèle et chaque entrée peut aider à déterminer la nature des relations entrée-sortie, par exemple, si la sortie dépend linéairement ou non linéairement d'une entrée (Looss et Prieur , 2019). Les diagrammes de dispersion peuvent également servir à distinguer qualitativement les paramètres influents de ceux qui ne le sont pas (Smith , 2013) . Les diagrammes de dispersion marginaux simples ne peuvent toutefois pas détecter les interactions entre les entrées (Cacuci et al ,2005). À cette fin, des diagrammes de coordonnées parallèles peuvent être utilisés (Looss et Lemaitre,2015) Dans ces diagrammes, chaque entrée est représentée par un axe vertical, ainsi que par un axe vertical supplémentaire pour la sortie. Ainsi, pour chaque entrée et chaque sortie, il existe un axe de coordonnées parallèles et verticales, ce qui donne la disposition des noms. Un point d'échantillonnage est représenté par une ligne reliant ses valeurs d'entrée à la valeur de sortie correspondante. En limitant l'axe de sortie, par exemple, aux petites valeurs, on peut voir quelle combinaison de valeurs d'entrée est responsable de cette région de sortie, ce qui révèle des effets d'interaction potentiels. Pour un exemple de tracé de coordonnées parallèles, voir la figure 7 du document C. Les tracés de coordonnées parallèles sont parfois également appelés tracés en toile d'araignée, mais ils ne doivent pas être confondus avec les tracés en toile d'araignée connus dans le cadre de l'analyse des systèmes dynamiques.

Le plus grand avantage des méthodes visuelles est certainement que leur application est souvent simple et ne nécessite que peu de connaissances préalables. Elles conviennent à la fois aux modèles linéaires et non linéaires et sont relativement bon marché, en particulier si un échantillon d'entrées-sorties existant peut être exploité. L'inconvénient des méthodes visuelles est leur nature purement qualitative. La quantité d'informations peut être écrasante, en particulier pour un grand nombre d'entrées. C'est pourquoi les méthodes visuelles requièrent une certaine expérience en matière d'interprétation (Stantner et al , 2018). La possibilité d'ajuster dynamiquement la représentation, par exemple la disposition des coordonnées dans un diagramme de coordonnées parallèles, peut être très utile à cet égard. Pour un exemple, voir le matériel supplémentaire du document C, (Brandstaeter et al ,2018) disponible en ligne.

○ Méthodes de dépistage

Des méthodes de dépistage ont été mises au point pour parvenir à une fixation efficace des paramètres dans un cadre global avec un grand nombre d'entrées. La raison d'être de ces approches est que la pratique a montré que la sortie n'est souvent sensible qu'à un petit nombre d'entrées.

L'objectif des méthodes de dépistage est de séparer efficacement les paramètres influents des paramètres non influents. Dans ce contexte, l'efficacité se réfère au nombre d'évaluations de modèles nécessaires en équilibre avec des hypothèses réalistes de complexité du modèle, c'est-à-dire le nombre d'entrées étudiées. Dans une étape suivante, des méthodes plus détaillées et plus coûteuses peuvent être appliquées pour une analyse approfondie des paramètres influents restants.

Il existe de nombreuses méthodes de dépistage différentes, chacune correspondant à un cas d'utilisation particulier (Woods et Lewis, 2017). La méthode de Morris (Morris, 1991) est une méthode très courante. Le filtrage de Morris permet de distinguer trois catégories d'entrées : Le premier groupe est constitué d'intrants non influents aux effets négligeables, le deuxième groupe contient des intrants qui ont des effets linéaires importants mais qui ne sont impliqués dans aucune interaction, et le troisième groupe est constitué d'intrants influents qui ont des effets non linéaires importants, avec ou sans effets d'interaction. Le classement est basé sur la moyenne de mesures de différences grossières appelées effets élémentaires. Les effets élémentaires ne sont rien d'autre que des approximations grossières des dérivées par différence finie. Morris en a déduit deux mesures de sensibilité pour chaque paramètre. Premièrement, la moyenne de la valeur absolue des effets élémentaires est une mesure de l'influence globale : plus la moyenne est grande, plus l'influence est importante. Deuxièmement, l'écart-type des effets élémentaires est une mesure de la non-linéarité ou des effets d'interaction : plus l'écart-type est élevé, moins l'hypothèse de linéarité est vraie. L'efficacité de la méthode résulte de l'évaluation intelligente des effets élémentaires individuels. La nature globale résulte de la distribution aléatoire des points d'évaluation dans l'espace d'entrée et du calcul de la moyenne qui s'ensuit.

La méthode de Morris pour le filtrage des paramètres est une méthode populaire car elle présente un moyen efficace sur le plan informatique de classer globalement l'importance des paramètres pour les grands modèles non linéaires. En général, la méthode renvoie des mesures de sensibilité semi-quantitatives, car leurs valeurs absolues sont difficiles à interpréter. Par conséquent, le classement des paramètres peut être difficile : le seuil qui sépare les paramètres influents des paramètres non influents dépend du problème. Néanmoins, il est souvent possible de trouver une ligne de démarcation claire dans la pratique, en particulier pour les grands modèles (Qian et Mahdi, 2020).

Naturellement, la méthode présente également quelques inconvénients. Pour les modèles très coûteux avec des temps de solution de plusieurs heures à plusieurs jours, elle peut s'avérer trop onéreuse. Plus important encore, la méthode a des difficultés à traiter les interactions et la non-linéarité en détail. Si le screening de Morris permet de détecter la présence de non-linéarité et d'interactions, il est impossible de faire la distinction entre les deux. Par conséquent, si des interactions sont présentes, il n'est pas possible de détecter les paramètres qui y participent.

○ Méthodes basées sur la variance

En supposant que les entrées sont des variables aléatoires avec des distributions de probabilité, la sortie du modèle est également une variable aléatoire dont la distribution résulte de la propagation des distributions d'entrée à travers le modèle. Les méthodes qui analysent les propriétés de la distribution de la sortie pour mener des études de sensibilité sont appelées méthodes basées sur la distribution en général. Les méthodes d'analyse de sensibilité basées sur la variance, qui reposent sur l'analyse de la variance de la distribution de la sortie en particulier, constituent une catégorie particulière. Elles quantifient la manière dont les entrées individuelles et leurs interactions contribuent à la variance résultante (Razavi et al., 2021).

La méthode la plus courante, la méthode de Sobol, peut être comprise par la décomposition fonctionnelle de l'analyse de la variance (ANOVA) (Sobol, 1990 ; Homma et Saltelli, 1996). Cette décomposition de la variance de sortie aboutit à une somme où chaque terme peut être directement lié à des entrées individuelles et à des combinaisons d'entrées. Les mesures de sensibilité, appelées indices de Sobol, sont alors considérées comme les fractions de la variance totale de la production invoquées par les intrants ou leurs interactions. Par conséquent, les indices sont liés entre zéro et un, ce qui donne immédiatement lieu à leur notion d'importance. Les valeurs proches de un (01) indiquent une forte influence, tandis que les valeurs proches de zéro caractérisent des combinaisons de paramètres sans influence. De cette manière, les indices de Sobol permettent une comparaison quantitative rigoureuse de l'influence globale de paramètres spécifiques et de toutes les combinaisons possibles.

Les indices de Sobol ont des propriétés convaincantes. Les deux plus grands avantages sont peut-être leur nature véritablement globale et leur facilité d'interprétation. C'est la seule méthode présentée jusqu'à présent qui convienne à la fois à la fixation des paramètres et à la hiérarchisation des paramètres pour les modèles non linéaires et non monotones. Un avantage connexe découle du fait que la méthode fournit naturellement une estimation de la distribution de sortie. Par conséquent, si elle est correctement appliquée, la méthode comprend intrinsèquement une quantification de l'incertitude, à condition que la variance soit acceptée comme la mesure appropriée de l'incertitude. La version de base de la méthode repose sur l'hypothèse que les entrées sont statistiquement indépendantes. Il s'agit manifestement d'une limitation puisque, dans la réalité, on peut s'attendre à ce que de nombreux paramètres d'entrée soient dépendants. Si la définition des indices eux-mêmes reste valable, l'interprétabilité est perdue en cas d'entrées dépendantes, ce qui réduit considérablement leur valeur informative. En outre, la quantité d'informations détaillées contenues dans les indices de Sobol a un prix. La charge de calcul de la méthode est extrême si l'on tente une réalisation brute. Dans la pratique, des approches plus raisonnables équilibrant le coût de calcul et le contenu informatif ont été développées. Néanmoins, l'application à des modèles complexes et coûteux reste une tâche difficile et nécessite des implémentations logicielles sophistiquées et une utilisation efficace des ressources informatiques, des approches de modélisation de substitution, ou les deux.

○ Méthodes indépendantes du moment

Parfois, l'incertitude de la distribution de sortie n'est pas correctement décrite par la variance (Looss et Lemaitre, 2015). C'est par exemple le cas dans le cadre de la cartographie factorielle, où une région spécifique de l'espace de sortie présente un intérêt particulier. À cette fin, il existe des méthodes basées sur la distribution qui s'abstiennent de se concentrer sur des moments spécifiques - comme la variance, l'asymétrie ou l'aplatissement de la distribution de sortie. Ces méthodes sont dites indépendantes des moments. En général, elles mesurent la différence entre les variants des distributions de sortie par rapport à la distribution de sortie inconditionnelle. Elle est comparée aux distributions conditionnelles de la production qui résultent de la fixation d'un ou de plusieurs intrants (Razavi et al, 2021). La méthode la plus courante est l'analyse de sensibilité régionale ou le filtrage de Monte Carlo (Young et al 1978 ; Sarrazin et al, 2016). D'autres variantes sont basées sur la zone comprise entre les deux distributions (Borgonovo, 2007), l'entropie ou le principe d'information mutuelle (Park et Ahn, 1994 ; Lin et al, 2006), les tests statistiques (Pianosi et Wagener, 2015) ou les échantillons génériques d'entrées-sorties (Pianosi et Wagener, 2018).

- **Autres méthodes**

Les méthodes présentées ci-dessus ne représentent qu'une petite partie de l'ensemble des méthodes disponibles pour l'analyse de sensibilité. Pour une vue d'ensemble approfondie, le lecteur intéressé est invité à se reporter à la littérature présentée à la section 3.1. Néanmoins, la liste ci-dessus est complétée par certaines des méthodes globales alternatives les plus importantes : les méthodes basées sur la régression, comme les coefficients de corrélation de Pearson et de Spearman ou les coefficients de régression standard reposent sur l'ajustement d'un modèle linéaire à travers l'échantillon d'entrées-sorties. Les mesures de sensibilité globales basées sur les dérivés constituent une méthode efficace liée aux indices de Morris et de Sobol (Kucherenko et al, 2009 ; Sobol et Kucherenko , 2009). Récemment, (Razavi et Gupta, 2016 ; Shekoleslami et Razavi , 2022) ont proposé une nouvelle classe de méthodes appelées méthodes basées sur le variogramme pour faire le lien entre les méthodes basées sur les dérivés et les méthodes basées sur la distribution. Enfin, les effets de Shapley sont basés sur la théorie des jeux et ont été proposés comme généralisation des indices de Sobol à des modèles avec des entrées dépendantes (Looss et Prieur , 2019 ; Owen , 2014 ; Owen et Prieur , 2017).

I.8. Conclusion :

Dans ce chapitre, nous avons exploré l'importance de l'analyse de sensibilité dans l'évaluation et la compréhension des modèles complexes. Nous avons exposé deux approches principales : l'analyse de sensibilité locale et l'analyse de sensibilité globale.

L'analyse de sensibilité locale nous permet d'identifier les paramètres les plus influents et de quantifier leur effet sur la sortie du modèle. En comprenant les relations entre les entrées et les sorties, nous sommes en mesure de cibler les paramètres critiques et d'améliorer les performances du modèle.

D'autre part, l'analyse de sensibilité globale permet de prendre en compte les interactions complexes entre facteurs et de quantifier l'influence de chacun. En regardant les variations simultanées des paramètres, on peut comprendre l'impact global de ces variations sur la sortie du modèle.

Nous avons également abordé la question des incertitudes inhérentes aux modèles. Les incertitudes peuvent provenir de diverses sources, telles que des données d'entrée incertaines, des paramètres inconnus ou des estimations numériques. La caractérisation et la représentation de ces incertitudes sont essentielles pour prendre des décisions éclairées et évaluer la fiabilité des résultats du modèle.

Aussi, nous avons examiné la prévalence des incertitudes au cours des différentes étapes de l'analyse. Comprendre comment les incertitudes se propagent nous permet de mieux évaluer la solidité des résultats et de prendre des mesures pour réduire les incertitudes si nécessaire.

En combinant ces différentes techniques d'analyse de sensibilité et de modélisation de l'incertitude et de la diffusion, nous pouvons acquérir une compréhension approfondie des modèles et de leur comportement. Cela nous permet de prendre des décisions éclairées, d'améliorer la confiance dans les résultats du modèle et d'identifier les domaines où des améliorations peuvent être apportées.

Dans ce travail, nous avons mené une analyse de sensibilité globale plus en détail où nous reviendrons sur la méthode de Sobol en détail dans le chapitre suivant en présentant les principales définitions et les méthodes de calcul.

Calcul Numérique des Indices de Sobol

Chapitre 2

II. Calcul Numérique des Indices de Sobol

II.1. Introduction :

À ce jour, de nombreuses études de sensibilité des paramètres reposent encore sur des méthodes locales dans lesquelles la valeur d'un paramètre est modifiée après l'autre et le changement qui en résulte dans la sortie du modèle est analysé. Cette approche est toutefois inadéquate pour les modèles non linéaires étudiés en situation d'incertitude. Les méthodes locales sous-estiment largement l'espace des paramètres en raison de la dimensionnalité et ne parviennent pas à détecter des interactions potentiellement cruciales entre les paramètres (Saltelli et al , 2019 ; Saltelli et al,2010) L'analyse de sensibilité globale basée sur la variance ne souffre pas de ces limitations. Elle échantillonne l'espace d'entrée complet et quantifie l'influence globale des paramètres sur la sortie du modèle en attribuant le degré d'incertitude de la sortie aux paramètres individuels et à leurs combinaisons.

Ce paragraphe présente le cadre d'analyse de sensibilité basé sur la variance, appelé méthode de Sobol ou simplement indices de Sobol. Cette méthode a été développée à l'origine par le mathématicien Russe Ilya Meyerovich Sobol en 1990. Depuis lors, elle a fait l'objet d'une attention considérable, ce qui a permis des avancées méthodologiques majeures (Iooss et Prieur,2019 ;Homma et Saltelli,1996 ; Sobol et Kucherenko,2009 ; Kucherenko et al 2017) ainsi que le développement de schémas d'évaluation informatique efficaces (Qian et al ,2018 ;Saltelli,2002 ; Kucherenko et Songe ,2017). La méthode de Sobol est très générale et les domaines d'application potentiels sont variés, mais les sciences de l'environnement et l'évaluation des risques ont été parmi les premiers à l'adopter (Pianosi et al , 2016 ; Saltelli et Tarantola , 2002).

II.2. Indices de Sobol :

- **Définition :**

Les indices de Sobol dans ce paragraphe sont dérivés en tant que mesures de sensibilité globales basées sur la variance pour des modèles génériques non linéaires (Saltelli et al ,2008 ;Wentwork et al ,2016 ;Sobol ,2001). Nous commençons par présenter quelques notions et concepts essentiels.

Soit $x = \{x_1, x_2, \dots, x_n\} \in \Omega \subseteq \mathbb{R}^n$ désigne un vecteur aléatoire continu composé de n variables aléatoires à valeurs réelles x_i . Alors, $x_{\sim i}$ représente le vecteur aléatoire dérivé de x en supprimant la composante x_i . Ainsi :

$$\mathbf{x}_{\sim i} = \{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n\} \quad (6)$$

La valeur d'une fonction scalaire f de $\mathbf{x}$ est à nouveau une variable aléatoire. La valeur attendue de cette variable aléatoire est notée :

$$\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})] = \int_{\Omega} f(\mathbf{x}) p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}, \quad (7)$$

où $p_{\mathbf{x}}(\mathbf{x})$ est la fonction de densité de probabilité de $\mathbf{x}$. La variance de $f(\mathbf{x})$ est définie comme :

$$\mathbb{V}_{\mathbf{x}}[f(\mathbf{x})] = \mathbb{E}_{\mathbf{x}}[(f(\mathbf{x}) - \mathbb{E}_{\mathbf{x}}[f(\mathbf{x})])^2] = \int_{\Omega} (f(\mathbf{x}) - \mathbb{E}_{\mathbf{x}}[f(\mathbf{x})])^2 p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}. \quad (8)$$

Une fonction de carré intégrable f doit décrire le modèle générique non linéaire d'intérêt. f a n entrées indépendantes $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ et une sortie scalaire y telle que :

$$f : K^n = [0,1]^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto y = f(\mathbf{x}) \quad (9)$$

Les entrées $\mathbf{x}$ sont supposées être distribuées uniformément dans l'hypercube unitaire à n dimensions $K^n \subset \mathbb{R}^n$ sans perte de généralité :

$$x_i \sim U(0,1) \quad (10)$$

où $U(0,1)$ désigne la distribution de probabilité uniforme continue sur l'intervalle $[0,1]$. La fonction de densité de probabilité conjointe peut être donnée comme suit :

$$p_{\mathbf{x}}(\mathbf{x}) = \prod_{i=1}^n p_{x_i}(x_i) = 1, \quad (11)$$

en raison de l'indépendance des entrées x_i . Sur la base de ces hypothèses, les mesures de sensibilité globales suivantes sont introduites, comme proposé à l'origine par (Saltelli et al, 2010 ; Sobol , 2001) :

▪ **Indice de sensibilité de Sobol de premier ordre :**

L'indice de sensibilité de Sobol de premier ordre, également appelé effet principal S_i du paramètre x_i est défini comme :

$$S_i = \frac{\mathbb{V}_{x_i} [\mathbb{E}_{\mathbf{x}_{\sim i}}[y | x_i]]}{\mathbb{V}_{\mathbf{x}}[y]} \quad (12)$$

où $\mathbb{V}_{\mathbf{x}}[y] > 0$ est la variance totale et inconditionnelle de y sur tous les paramètres d'entrée et $\mathbb{E}_{\mathbf{x}_{\sim i}}[y | x_i]$ est la valeur conditionnelle attendue de y compte tenu de la valeur du paramètre x_i .

L'indice de premier ordre S_i décrit la fraction de la variance de y due à l'incertitude du seul paramètre x_i . En d'autres termes, il représente le montant par lequel la variance de y pourrait être diminuée si la valeur précise de x_i était connue.

▪ Indice de sensibilité de Sobol total

L'indice de sensibilité Sobol total S_i^T du paramètre x_i est défini comme :

$$S_i^T = 1 - \frac{\mathbb{V}_{x_{\sim i}} [\mathbb{E}_{x_i} [y | x_{\sim i}]]}{\mathbb{V}_x [y]} = \frac{\mathbb{E}_{x_{\sim i}} [\mathbb{V}_{x_i} [y | x_{\sim i}]]}{\mathbb{V}_x [y]} \quad (13)$$

où $\mathbb{V}_x [y] > 0$ est la variance totale et inconditionnelle de y sur tous les paramètres d'entrée. $\mathbb{E}_{x_i} [y | x_{\sim i}]$ est l'espérance conditionnelle et $\mathbb{V}_{x_i} [y | x_{\sim i}]$ la variance conditionnelle de y compte tenu des valeurs de tous les paramètres sauf x_i . L'indice de Sobol total est également appelé l'effet total.

L'indice total de Sobol présente la quantité moyenne de la variance de sortie restante si l'on connaît la valeur exacte de tous les paramètres sauf x_i .

En d'autres termes, il donne une mesure de l'importance globale du paramètre x_i pour la variance de la sortie y , c'est-à-dire l'importance de x_i , son indice de premier ordre, et en plus de chaque interaction d'ordre supérieur de x_i .

En résumé, les indices de Sobol de premier ordre S_i quantifient la sensibilité de la sortie à chaque paramètre seul, et les indices de Sobol totaux S_i^T quantifient en outre la sensibilité de la sortie à toutes les interactions dans lesquelles chaque paramètre est impliqué.

Il existe également des indices d'ordre supérieur qui quantifient l'influence de termes d'interaction spécifiques mais ceux-ci sont rarement utilisés en pratique. Le calcul des deux types d'indices ci-dessus présente un bon équilibre entre le coût de calcul et les connaissances acquises.

• Interprétation :

Une interprétation instructive des indices de Sobol peut être obtenue sur la base de la décomposition additive de la fonction modèle f appelée représentation ANOVA de f .

Dans ce qui suit, des sous-ensembles arbitraires de $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ pour $n \in \mathbb{N}^+$ seront sélectionnés.

Les paramètres sélectionnés sont notés $\mathbf{x}_{i_1 i_2 \dots i_s} = \{x_{i_1}, x_{i_2}, \dots, x_{i_s}\}$, où un ensemble d'indices ordonnés $I = \{i_1, i_2, \dots, i_s\}$ de cardinal $s \leq n$ avec $1 \leq i_1 < i_2 < \dots < i_s \leq n$ a été introduit. L'ensemble d'indices I est un sous-ensemble de tous les n indices tels que $I \subseteq \{1, 2, \dots, n\}$. Trois classes spéciales de I de cardinalité 1, 2 et 3 sont respectivement les ensembles d'index $I = \{i\}$, $I = \{i, j\}$ et $I = \{i, j, k\}$. Un exemple pour $n = 4$ et $s = 3$ est l'ensemble $\{1, 3, 4\}$ qui donne $\mathbf{x}_{134} = \{x_1, x_3, x_4\}$. Avec cette notation, la représentation ANOVA de f peut être définie.

▪ **Représentation par analyse de variance (ANOVA) de f**

Une représentation d'une fonction f comme :

$$f(x_1, x_2, \dots, x_n) = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{i=1}^{n-1} \sum_{j>i}^n f_{ij}(x_i, x_j) + \sum_{i=1}^{n-2} \sum_{j>i}^{n-1} \sum_{k>j}^n f_{ijk}(x_i, x_j, x_k) + \dots + f_{12\dots n} \quad (14)$$

est appelée représentation par analyse de variance (ANOVA) ou représentation de modèle à haute dimension (HDMR pour High-Dimensional Model Representation), si

$$f_0 = \text{const.} \quad (15)$$

et

$$\int_0^1 f_{i_1 i_2 \dots i_s}(x_{i_1}, x_{i_2}, \dots, x_{i_s}) dx_{i_k} = 0, \forall i_k \in \{i_1, i_2, \dots, i_s\} \quad (16)$$

où $f_{i_1 i_2 \dots i_s}(x_{i_1}, x_{i_2}, \dots, x_{i_s})$ est une sommation arbitraire du membre droit de l'équation (14) autre que f_0 .

Éq. (16) indique que pour chaque sommation dans l'Eq. (14) les intégrales par rapport à l'une de ses variables s'annulent. Rappelons que le modèle f devait être intégrable de carré sur l'hypercube unitaire K^n . Dans ce cas, on peut prouver que l'Eq. (14) existe et est unique (Sobol, 1993). Dans ce qui suit, l'hypothèse d'indépendance des paramètres et l'hypothèse de paramètres uniformément distribués dans l'hypercube unitaire sont exploitées. Ces hypothèses permettent d'éviter l'énoncé explicite des fonctions de densité de probabilité $p_{x_i}(x_i) = 1$ dans la plupart des intégrales ci-dessous.

On peut par exemple écrire $\int_0^1 f(x_i) p_{x_i}(x_i) dx_i = \int_0^1 f(x_i) dx_i$.

Deux propriétés importantes découlent de **Représentation par (ANOVA) de f** Tout d'abord, f_0 peut être identifié comme la valeur attendue de la sortie du modèle $y = f(x)$

$$f_0 = \int_{K^n} f(\mathbf{x}) d\mathbf{x} = \mathbb{E}_x[y] \quad (17)$$

Deuxièmement, deux sommations différentes de l'équation. (14) sont orthogonales, tels que :

$$\int_{K^n} f_{i_1 i_2 \dots i_s}(x_{i_1}, x_{i_2}, \dots, x_{i_s}) f_{j_1 j_2 \dots j_l}(x_{j_1}, x_{j_2}, \dots, x_{j_l}) d\mathbf{x} = 0 \quad (18)$$

avec des ensembles d'index ordonnés $I = \{i_1, i_2, \dots, i_s\}, J = \{j_1, j_2, \dots, j_l\}, s, l \leq n$ et $I \neq J$.

Alors que les termes de l'Eq. (14) peut en théorie être dérivé, en pratique on n'a pas besoin de calculer les sommations pour utiliser les indices Eq. (12) et Éq. (13). En effet on n'a même pas besoin de connaître la forme de $f(x)$. Il suffit de pouvoir l'évaluer ponctuellement (Homma et Saltelli, 1996).

Néanmoins, il est très instructif de regarder le calcul des sommes individuelles. Par souci de concision, raccourcissement de la notation en utilisant $f_{i_1 i_2 \dots i_s}(x_{i_1}, x_{i_2}, \dots, x_{i_s}) = f_{i_1 i_2 \dots i_s}$ et $dx_{j_1} dx_{j_2} \dots dx_{j_{n-s}} = d\mathbf{x}_{\sim i_1 i_2 \dots i_s}$ est introduit. Ici, les ensembles d'indices ordonnés $I = \{i_1, i_2, \dots, i_s\}$ avec $1 \leq i_1 < i_2 < \dots < i_s \leq n$ et $J = \{j_1, j_2, \dots, j_{n-s}\}$ avec $1 \leq j_1 < j_2 < \dots < j_{n-s} \leq n$ sont disjoints tels que $I \cap J = \emptyset$ et $I \cup J = \{1, 2, \dots, n\}$. Ensuite, on peut calculer les sommations individuelles dans l'équation. (14) par le schéma récursif suivant :

$$\begin{aligned}
 \int_{K^n} f(\mathbf{x}) d\mathbf{x} &= f_0 = \mathbb{E}_x[y] \\
 \int_{K^{n-1}} f(\mathbf{x}) d\mathbf{x}_{\sim i_1} &= f_0 + f_{i_1} \\
 \int_{K^{n-2}} f(\mathbf{x}) d\mathbf{x}_{\sim i_1 i_2} &= f_0 + f_{i_1} + f_{i_2} + f_{i_1 i_2} \\
 &\vdots \\
 \int_{K^{n-s}} f(\mathbf{x}) d\mathbf{x}_{\sim i_1 i_2 \dots i_s} &= f_0 + \sum_{i=i_1}^{i_s} f_i + \sum_{i=i_1}^{i_s-1} \sum_{j>i}^{i_s} f_{ij} + \sum_{i=i_1}^{i_s-2} \sum_{j>i}^{i_s-1} \sum_{k>j}^{i_s} f_{ijk} + \dots + f_{i_1 i_2 \dots i_s} \\
 &\vdots \\
 f(\mathbf{x}) &= f_0 + \sum_{i=1}^n f_i + \sum_{i=1}^{n-1} \sum_{j>i}^n f_{ij} + \sum_{i=1}^{n-2} \sum_{j>i}^{n-1} \sum_{k>j}^n f_{ijk} + \dots + f_{12 \dots n}
 \end{aligned} \tag{19}$$

Résoudre chaque ligne dans Eq. (19) pour la somme la plus élevée :

$$f_{i_1 i_2 \dots i_s} = \int_{K^{n-s}} f(\mathbf{x}) d\mathbf{x}_{\sim i_1 i_2 \dots i_s} - f_0 - \sum_{i=i_1}^{i_s} f_i - \sum_{i=i_1}^{i_s-1} \sum_{j>i}^{i_s} f_{ij} - \sum_{i=i_1}^{i_s-2} \sum_{j>i}^{i_s-1} \sum_{k>j}^{i_s} f_{ijk} - \dots \tag{20}$$

conduit à l'interprétation instructive souhaitée. $f_{i_1 i_2 \dots i_s}$ est la partie de $f(\mathbf{x})$ qui dépend uniquement des termes d'interaction entre les $\{x_{i_1}, x_{i_2}, \dots, x_{i_s}\}$. Toutes les variables restantes sont intégrées et les termes d'ordre inférieur sont soustraits.

La représentation ANOVA Eq. (14) peut être utilisé pour décomposer la variance de $f(\mathbf{x})$:

$$\mathbb{V}_x[f(\mathbf{x})] = \int_{K^n} f^2(\mathbf{x}) d\mathbf{x} - (\mathbb{E}_x[f(\mathbf{x})])^2 = \int_{K^n} f^2(\mathbf{x}) d\mathbf{x} - f_0^2 \tag{21}$$

Brancher l'éq. (14) avec la propriété d'orthogonalité Eq. (18) en Eq. (21) aboutit à la formulation de la décomposition de la variance de base :

$$\mathbb{V}_x[f(\mathbf{x})] = \sum_{i=1}^n \mathbb{V}_{x_i}[f_i] + \sum_{i=1}^{n-1} \sum_{j>i}^n \mathbb{V}_{x_i x_j}[f_{ij}] + \sum_{i=1}^{n-2} \sum_{j>i}^{n-1} \sum_{k>j}^n \mathbb{V}_{x_i x_j x_k}[f_{ijk}] + \dots + \mathbb{V}_x[f_{12 \dots n}] \tag{22}$$

où les sommations individuelles dans l'Eq. (22) :

$$\mathbb{V}_{x_{i_1}x_{i_2}\dots x_{i_s}}[f_{i_1i_2\dots i_s}] = \int_{K^{n-s}} f_{i_1i_2\dots i_s}^2 dx_{i_1} dx_{i_2} \dots dx_{i_s} \quad (23)$$

peuvent être identifiées comme les variances des sommations dans Eq. (14). En rappelant l'interprétation de $f_{i_1i_2\dots i_s}$ ci-dessus, cela signifie que $\mathbb{V}_{x_{i_1}x_{i_2}\dots x_{i_s}}[f_{i_1i_2\dots i_s}]$ décrit la contribution de la variance attribuée uniquement à l'interaction entre $\{x_{i_1}, x_{i_2}, \dots, x_{i_s}\}$. En raison de l'éq. (16), l'espérance de $f_{i_1i_2\dots i_s}$ est 0. Enfin, l'Eq. (22) propose la définition des indices généraux de sensibilité suivants.

▪ **Indice de sensibilité de Sobol d'ordre s**

L'indice de sensibilité Sobol d'ordre s de la fonction f, donné par Eq. (9), est défini comme :

$$S_{i_1i_2\dots i_s} = \frac{\mathbb{V}_{x_{i_1}x_{i_2}\dots x_{i_s}}[f_{i_1i_2\dots i_s}]}{\mathbb{V}_x[f(\mathbf{x})]} \quad (24)$$

où $\mathbb{V}_x[f(\mathbf{x})] > 0$ est la variance totale et inconditionnelle de $y = f(\mathbf{x})$ sur tous les paramètres d'entrée et $\mathbb{V}_{x_{i_1}x_{i_2}\dots x_{i_s}}[f_{i_1i_2\dots i_s}]$ est la contribution de la variance attribuée uniquement à la interaction entre $\{x_{i_1}, x_{i_2}, \dots, x_{i_s}\}$.

L'indice de sensibilité de Sobol d'ordre s $S_{i_1i_2\dots i_s}$ quantifie la quantité de variance de sortie causée par les interactions entre les s paramètres $\{x_{i_1}, x_{i_2}, \dots, x_{i_s}\}$ en excluant toute interaction potentielle d'ordre inférieur entre ces paramètres.

En raison de l'éq. (22) la propriété essentielle suivante est vérifiée :

$$\sum_{i=1}^n S_i + \sum_{i=1}^{n-1} \sum_{j>i}^n S_{ij} + \sum_{i=1}^{n-2} \sum_{j>i}^{n-1} \sum_{k>j}^n S_{ijk} + \dots + S_{12\dots n} = 1 \quad (25)$$

L'indice de sensibilité totale S_i^T tel que défini dans **Indice de sensibilité de Sobol total** découle de l'Eq. (24) à :

$$S_i^T = S_i + \sum_{\substack{j=1 \\ j \neq i}}^n S_{ij} + \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{\substack{k>j \\ k \neq i}}^n S_{ijk} + \dots + S_{12\dots n} \quad (26)$$

c'est-à-dire que S_i^T est la somme de tous les indices de sensibilité associés à x_i . Aussi, les indices de Sobol du premier ordre définis par **Indice de sensibilité de Sobol de premier ordre** et **Indice de sensibilité de Sobol d'ordre s** sont égaux : en général, $\mathbb{V}_x[x + a] = \mathbb{V}_x[x]$. Il s'ensuit :

$$\mathbb{V}_{x_i}[f_i] = \mathbb{V}_{x_i} \left[\mathbb{E}_{x_{\sim i}}[y \mid x_i] \right], \text{ "où" } f_i = \mathbb{E}_{x_{\sim i}}[y \mid x_i] - f_0 \quad (\text{voir Éq. (19)}).$$

II.3. Estimation des Indices de Sobol :

Le calcul exact des indices de sensibilité de Sobol, n'est pas toujours possible. En effet, les lois des paramètres d'entrée, ainsi que l'expression de la sortie du modèle f peuvent être extrêmement complexes, et les intégrales exprimant les variances et les variances d'espérances conditionnelles qui interviennent dans la définition des indices peuvent s'avérer incalculables analytiquement. Il faut donc avoir recours à des méthodes d'approximation numérique, qui sont au demeurant suffisantes pour les applications pratiques. Nous présenterons dans ce chapitre principalement deux méthodes différentes d'approximation numérique des indices de Sobol utilisées pour estimer ces indices.

Nous nous contenterons de présenter les estimateurs basés sur la méthode FAST et la méthode de développement sur chaos polynômial qui sont les estimateurs sur lesquels nous nous sommes centrés durant ce travail.

II.3.1 Algorithme de la méthode FAST :

Dans les années 70, le **Test de Sensibilité d'Amplitude de Fourier** (FAST, pour *Fourier Amplitude Sensitivity Test*) a été introduit par le biais d'une série de publications (Cukier et al., 1973 ; Schaibly et Shuler, 1973 ; Cukier et al., 1975 ; Cukier et al., 1978). Initialement développée pour l'étude de la cinétique des réactions chimiques, cette méthode s'est depuis répandue dans de nombreux domaines tels que le nucléaire (Lu et Mohanty, 2001), la thermique du bâtiment (Mara et Tarantola, 2008) et la biologie marine (Marino et al., 2008), subissant ainsi diverses évolutions. Parmi ces évolutions, deux sont particulièrement notables : l'extension de la méthode pour estimer l'effet total d'un facteur (EFAST, Saltelli et al., 1999) et son couplage avec une technique de randomisation appelée Random Balance Design (RBD, Tarantola et al., 2006 ; Xu et Gertner, 2008 ; Mara, 2009). FAST appartient à la famille des méthodes d'analyse de sensibilité basées sur la variance, qui permettent d'évaluer l'influence des variables d'entrée sur la variance de la sortie.

Commençons par supposer que $X = (X_1, \dots, X_p)$ suit une loi uniforme sur $[0; 1]^p$. Lorsque les paramètres d'entrée sont indépendants, il est toujours possible (au moins en théorie) de se ramener à ce cas, par composition avec la fonction de répartition vectorielle $F : \mathbb{R}^p \rightarrow [0; 1]^p$ donnée par :

$$F(x_1, \dots, x_p) = (P(X_1 \leq x_1), \dots, P(X_p \leq x_p)). \quad (27)$$

Pour tout p -uplet $k \in \mathbb{Z}^p$, notons $c_k(f)$ le $k^{\text{ème}}$ coefficient de Fourier de f , donné par :

$$c_k(f) = \int_{[0;1]^p} f(X) \exp(-2i\pi X \cdot k) dX \quad (28)$$

Pour tout $i = 1, \dots, p$, le $i^{\text{ème}}$ indice de Sobol s'exprime en fonction des coefficients de Fourier :

$$S_i = \frac{\sum_{l \in \mathbb{Z}^*} |c_{l\delta_i}(f)|^2}{\sum_{k \in \mathbb{Z}^d, k \neq 0} |c_k(f)|^2} \quad (29)$$

où δ_i est le vecteur de $\mathbb{Z}^p$ dont la j ème composante vaut :

$$(\delta_i) = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{sinon.} \end{cases} \quad (30)$$

Choisissons un vecteur de fréquences $\omega = (\omega_1, \dots, \omega_p) \in \mathbb{N}^p$, et notons $x(t)$ le vecteur de $\mathbb{R}^p$ dont la j ème composante est donnée par :

$$(x(t))_j = \frac{1}{\pi} \arcsin(\sin(\omega_j t)) + \frac{1}{2}. \quad (31)$$

Notons, pour $l \in \mathbb{Z}$:

$$c_l^{\text{ERG}}(f) = \int_0^1 f \circ x(t) \exp(-2i\pi l t) dt \quad (32)$$

Les auteurs de FAST proposent l'approximation suivante des coefficients de Fourier de f :

$$c_k(f) \approx c_{k \cdot \omega}^{\text{ERG}}(f), \quad (33)$$

où $k \cdot \omega$ est le produit scalaire Euclidien de k avec ω .

Cette approximation, qui permet de remplacer une intégrale p -dimensionnelle par une intégrale unidimensionnelle, est basée sur le théorème ergodique de (Weyl, 1938), et n'est pertinente que si le vecteur ω est bien choisi, par exemple à l'aide du critère d'absence d'interférences donné dans (Cukier et al, 1973).

L'étape suivante est d'approcher l'intégrale apparaissant dans $c_l^{\text{ERG}}(f)$ par la méthode des rectangles à $N > 0$ points, ce qui donne lieu à l'approximation de $c_l^{\text{ERG}}(f)$:

$$c_l^{\text{FAST}}(f) = \frac{1}{n} \sum_{j=0}^{N-1} f \circ x\left(\frac{j}{N}\right) \exp\left(-2i\pi l \frac{j}{N}\right) \quad (34)$$

Enfin, les sommes apparaissant dans (29) sont tronquées, ce qui donne l'approximation S_i^{FAST} suivante :

$$S_i^{\text{FAST}} = \frac{\sum_{k=1}^{n-1} |c_{k w_i}^{\text{FAST}}(f)|^2}{\sum_{l=1}^{n-1} |c_l^{\text{FAST}}(f)|^2} \quad (35)$$

II.3.2 Décomposition en Polynômes de Chaos :

La méthode de décomposition en polynômes de chaos ou encore du développement du chaos polynômial est une approche probabiliste utilisée pour développer une fonction qui peut être sommée en carré sur une base de polynômes orthogonaux deux à deux, tels que les polynômes d'Hermite ou de Legendre. Cette méthode s'appuie sur la théorie du Chaos Homogène introduite par Wiener en 1938, ainsi que sur les travaux de Cameron et Martin en 1947. L'utilisation des polynômes du chaos (PC) dans la simulation numérique des processus stochastiques a été réintroduite par Ghanem et Spanos en 1991, ce qui a renouvelé l'intérêt pour cette approche.

Dans cette hypothèse, nous considérons que le vecteur d'entrées $X = (X_1, \dots, X_p)$ suit une distribution normale multivariée $\mathcal{N}_p(0, I_p)$, où I_p représente la matrice identité de taille P .

- **Polynômes de Chaos :**

Nous introduisons la famille de polynômes orthogonaux de Hermite $(\psi_n)_{n \in \mathbb{N}}$, tels que $\deg \psi_n = n$, et

$$\int_{\mathbb{R}} \psi_n(x) \psi_m(x) \omega(x) dx = \begin{cases} 1 & \text{si } n = m \\ 0 & \text{sinon} \end{cases} \quad (36)$$

pour tous n et m entiers, où la fonction de poids $\omega(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ est la densité d'une variable aléatoire normale centrée réduite. Etant donné un multi-indice $i = (i_1, \dots, i_p) \in \mathbb{N}^p$, on note :

$$\psi_i(X) = \psi_{i_1}(x_1) \cdots \psi_{i_p}(x_p) \quad (37)$$

pour une collection infinie de réels $(y_i)_{i \in \mathbb{N}^p} \in \mathbb{N}^p$ appelés coefficients de Y dans sa décomposition en polynômes de chaos, la sommation sur le multi-indice i étant convergente dans L^2 .

Le polynôme de Hermite à p variables d'ordre i .

Comme $Y \in L^2(\Omega)$ est X -mesurable, le théorème de (Cameron et Martin, 1947; Sudret, 2008) affirme que Y peut s'écrire sous la forme d'une décomposition en polynômes de chaos (de Hermite) :

$$Y = \sum_{i \in \mathbb{N}^p} y_i \psi_i(X) \quad (38)$$

L'intérêt de la décomposition en polynômes de chaos est que les indices de Sobol peuvent s'exprimer en fonction des coefficients de Y dans cette décomposition, à l'aide d'une formule très similaire à celle utilisée pour la méthode FAST :

$$S_i = \frac{\sum_{j=1}^{\infty} y_{I(j,i)}^2}{\sum_{j \in \mathbb{N}^p, j \neq 0} y_j^2}, \quad (39)$$

où

$$I(j, i) = (\underbrace{0, \dots, 0}_{i-1}, j, \underbrace{0, \dots, 0}_{p-i}) \quad (40)$$

est le multi-indice de $\mathbb{N}^p$ contenant j à la i ème place, et 0 aux autres places. Le lien entre cette formule et la méthode FAST est précisé dans (Tissot et Prieur, 2012), l'idée sous-jacente étant l'utilisation de la tensorisation d'une base orthonormée de $L^2(P_{X_i})$.

Pour être évaluées numériquement, les sommes de (39) doivent être tronquées à un ordre $P \in \mathbb{N}^*$:

$$S_i \approx \frac{\sum_{j=1}^P y_{I(j,i)}^2}{\sum_{j \in \mathbb{N}^p, 0 < |j| \leq P} y_j^2}, \quad (41)$$

où, pour un multi-indice $j = (j_1, \dots, j_p) \in \mathbb{N}^*$,

$$\square = \square 1 + \dots + \square \square. \quad (42)$$

Tout revient donc à approcher les $\frac{(p+P)!}{p!P!} - 1$ coefficients y_j (pour $0 < |j| < P$) par $\tilde{y}_j$, ce qui donne lieu à l'approximation de S_i par S_i^{PC} définie par :

$$S_i^{PC} = \frac{\sum_{j=1}^P \tilde{y}_{I(j,i)}^2}{\sum_{j \in \mathbb{N}^p, 0 < |j| \leq P} \tilde{y}_j^2} \quad (43)$$

Les paragraphes suivants sont consacrés à trois méthodes de calcul des $\tilde{y}_j$.

- **Calcul par projection.**

Les propriétés d'orthogonalité des polynômes de Hermite permettent d'écrire :

$$\forall j \in \mathbb{N}^p, y_j = \int_{\mathbb{R}^p} Y \psi_j w_d \quad (44)$$

et cette intégrale multi-dimensionnelle peut ensuite être approchée par diverses méthodes de quadrature, tant déterministes que stochastiques.

- **Calcul par régression.**

Dans l'approche par régression, présentée dans (Sudret, 2008), on se donne un plan d'expérience $\Xi = \{X^{(1)}, \dots, X^{(N)}\} \subset \mathbb{R}^d$, on calcule les sorties du modèle sur ce plan : $Y(X^{(1)}), \dots, Y(X^{(N)})$, puis on estime les coefficients de la décomposition en polynômes de chaos tronquée par moindres carrés :

$$(\tilde{y}_i)_{i \in \mathbb{N}^d, |i| \leq P} = \underset{(z_i)_{i \in \mathbb{N}^d, |i| \leq P}}{\operatorname{argmin}} \sum_{n=1}^N \left(Y(X^{(n)}) - \sum_{|i| \leq P} z_i \psi_i(X^{(n)}) \right)^2 \quad (45)$$

Pour que ce problème soit bien posé, la taille du plan d'expérience (N) doit être supérieure ou égale au nombre de coefficients à estimer. Lorsque ces deux quantités sont égales, la régression est une interpolation et la méthode porte alors le nom de collocation.

Pour le choix du plan d'expérience, contient des stratégies basées sur les racines de polynômes de Hermite.

Le problème (45) peut également être rendu bien posé lorsque le plan d'expérience contient moins de points que de coefficients à estimer par une méthode de régularisation L^1 . Ceci est intéressant car le nombre de coefficients à estimer croît rapidement avec le nombre d'entrées p et l'ordre de troncature P . On peut ainsi être amené à résoudre (45) avec la contrainte supplémentaire :

$$\sum_{i \in \mathbb{N}^d, |i| \leq P} |z_i| \leq s \quad (46)$$

où S est un paramètre de régularisation à choisir. Plus s est petit, et plus les solutions sont « creuses » (c'est à dire que beaucoup des $\hat{y}_j$ estimés sont nuls). Ce problème de moindres carrés contraints peut être résolu à l'aide d'un algorithme *Least Angle Regression* (Blatman , 2009).

- **Calcul par une méthode intrusive.**

Par rapport aux autres méthodes présentées, qui traitaient le modèle comme une boîte noire à laquelle on peut faire appel pour obtenir les sorties évaluées sur différents jeux de paramètres d'entrée, les méthodes intrusives type Galerkin font des hypothèses fortes sur le « type » de sortie considérée.

Nous esquissons seulement le principe de la méthode, et renvoyons à (Chomen et Spanos , 2003 ; Deb et al , 2001) pour plus de détails.

Dans cette approche, on suppose que Y est une fonctionnelle ϕ d'une fonction $u = u(\cdot, X)$ satisfaisant une équation aux dérivées partielles possédant de "bonnes propriétés" (par exemple, l'ellipticité et la linéarité). L'idée consiste à injecter la décomposition de u en polynômes de chaos :

$$u(\cdot, X) = \sum_{j \in \mathbb{N}^p} u_j(\cdot) \psi_j(X) \quad (47)$$

dans l'équation aux dérivées partielles afin d'écrire un système (infini) d'EDP portant sur les fonctions u_j . Ce système est alors tronqué afin de ne conserver qu'un nombre fini de $u_j(\cdot)$ à calculer ; puis le système d'EDP obtenu est discrétisé afin d'obtenir un système fini d'équations portant sur des inconnues réelles.

La résolution de ce système permet dans certains cas (le cas le plus simple étant le cas où ϕ est linéaire) d'estimer les coefficients de la décomposition en polynômes de chaos de .

Signalons que le système fini d'équations à résoudre se révèle être souvent de grande taille, et que des méthodes spécifiques de réduction de dimension, telle que la décomposition spectrale généralisée (GSD(Noy et al ,2009) peuvent être utilisées afin d'accélérer sa résolution numérique.

II.3.3 Organigrammes de la méthode FAST :

Les étapes de base pour effectuer une analyse de sensibilité à l'aide de la méthode FAST peuvent être résumées par l'organigramme indiqué par la figure 3. L'organigramme décrivant comment analyser la sensibilité globale d'un modèle commence par la création d'un nouveau projet (pro_Create). Ensuite, pour chaque variable d'entrée x_i , une nouvelle entrée avec ses caractéristiques doit être ajoutée au projet en utilisant la fonction pro_Ad-dInput.

Ensuite, le modèle, qui est la fonction , est défini dans le projet à l'aide de la fonction pro_SetModel(). Ce modèle prend les valeurs des variables d'entrée et génère une sortie correspondante.

Le nombre d'échantillons pour la simulation Monte-Carlo quasi-aléatoire est défini sur N. Cela signifie que le modèle sera évalué à N points d'échantillonnage différents pour obtenir une estimation robuste de la sensibilité des variables d'entrée.

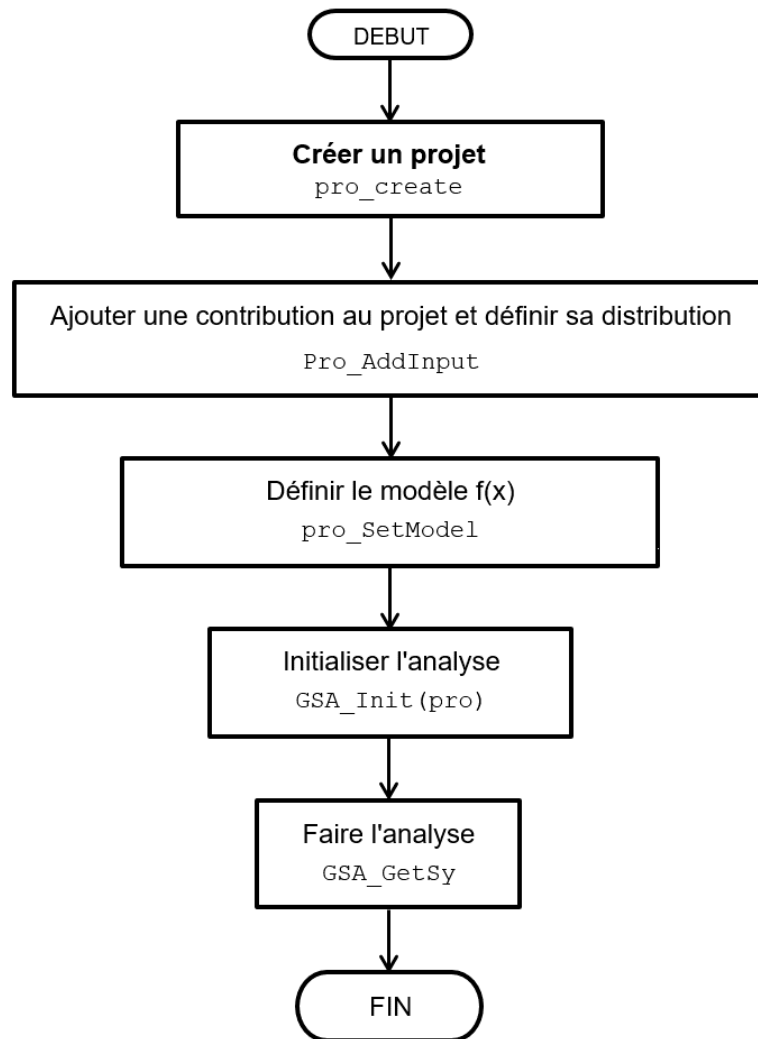


Fig. 3: Organigramme simplifié du programme de la méthode FAST

Le projet est ensuite initialisé en calculant le modèle aux points d'échantillonnage à l'aide de la fonction `GSA_Init()`. Cela signifie que le modèle f sera évalué pour chaque combinaison de valeurs des variables d'entrée.

Ensuite, les coefficients de sensibilité globale du premier ordre sont calculés à l'aide de l'algorithme FAST avec la fonction `GSA_FAST_GetSi()`. Ces coefficients mesurent l'influence individuelle de chaque variable d'entrée sur la sortie du modèle.

En outre, le coefficient de sensibilité globale du premier ordre pour la deuxième variable d'entrée est spécifiquement calculé et comparé à la valeur précédemment obtenue par l'algorithme FAST. Cela permet de vérifier si les résultats sont cohérents.

Enfin, le coefficient de sensibilité globale totale pour la première variable d'entrée est calculé à l'aide de la fonction `GSA_GetTotalSy()`. Ce coefficient mesure l'influence globale de la variable d'entrée sur la sortie du modèle, en prenant en compte les interactions avec les autres variables.

II.3.4 Organigrammes du Polynômes de Chaos :

Les étapes de calcul utilisées pour la méthode de décomposition en polynômes de chaos peuvent être simplifiées par l'organigramme suivant (Fig. 4) :

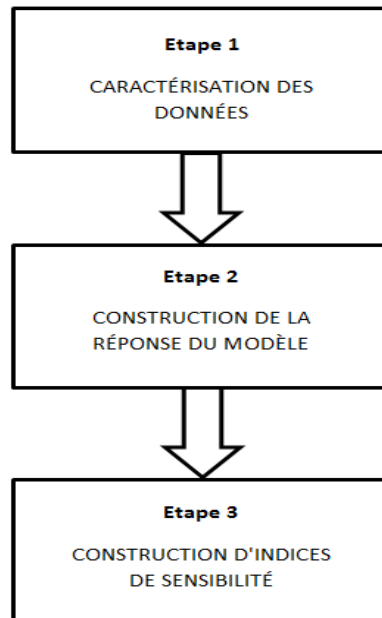


Fig. 4:Etapes de calcul des indices de sensibilité de Sobol par décomposition du chaos polynômial

Les calculs s'effectuent en trois étapes principales :

1) Première étape :

Caractérisation des données : Elle permet d'évaluer les moments statistiques des paramètres d'entrée ;

2) Deuxième étape :

Construction de la réponse du modèle : Elle permet d'approcher la surface de réponse du modèle par un polynôme.

3) Troisième étape :

Construction d'indices de sensibilité : Elle permet d'évaluer efficacement les mesures de sensibilité à partir du polynôme.

La procédure de calcul qui sera effectuée par la suite peut être schématisé par l'organigramme détaillé de la figure (5) .

L'analyse de sensibilité globale du modèle commence par l'initialisation des paramètres incertains, qui sont des données d'entrée pour le modèle physique. Par la suite, un objet "**PC**" est créé, qui est la classe contenant l'outil d'expansion polynomiale. La méthode de construction de la base du polynôme est déterminée, ainsi que le degré d'expansion.

La prochaine étape construit un ensemble de données d'entraînement (**Training Input**) à partir d'une méthode appelée "**Probabilistic Collocation Method**" (PCM), qui est une technique de points de collocation pour résoudre les équations différentielles stochastiques.

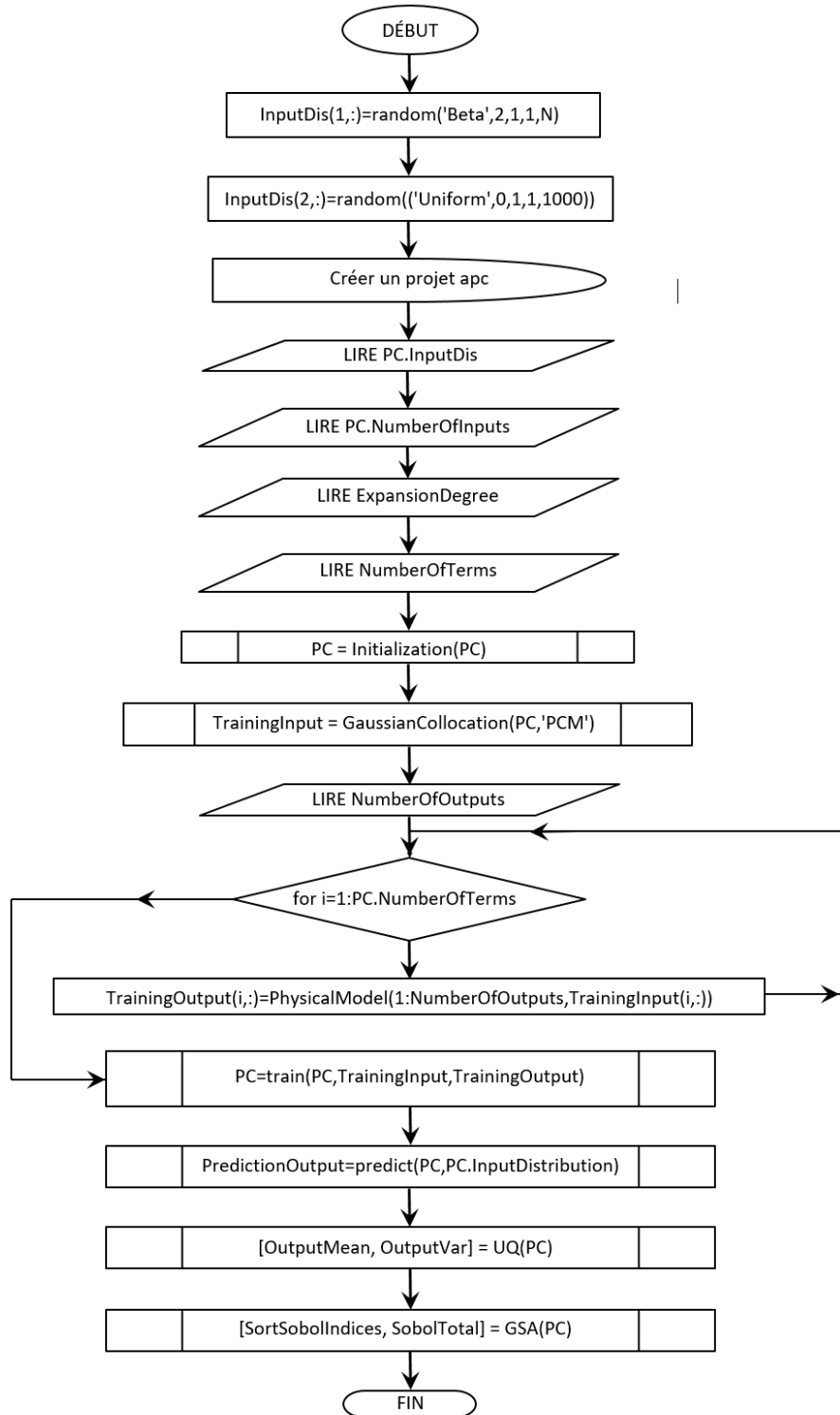


Fig. 5: Organigramme simplifié du programme de la méthode du chaos polynomial.

La boucle qui s'ensuit, calcule les sorties correspondantes à chaque point d'entraînement à partir du modèle physique, Le modèle du chaos polynômial est entraîné à l'aide de l'ensemble d'entrée et de sortie (**Training Input et Training Output**) pour prédire les sorties du modèle physique inconnu à partir des données d'entrée.

La sortie (**Prediction Output**) est ensuite prédite à partir des entrées incertaines en utilisant la méthode "predict". A ce titre, les moyennes et les variances de la sortie sont calculées à l'aide de la méthode "UQ".

Enfin, la méthode « **GSA** » est utilisée pour effectuer une analyse de sensibilité complète de chaque entrée sur la sortie de la fonction de réponse.

II.4. Conclusion :

En définitive, plusieurs méthodes existent pour estimer les indices de sensibilité de Sobol. Toutefois, les deux méthodes présentées dans ce chapitre constituent deux classes différentes pour estimer les indices de Sobol.

La méthode FAST repose sur l'idée que les variations des paramètres d'entrée se traduisent par des variations dans les fréquences des oscillations de la sortie du modèle. En utilisant la transformation de Fourier, puis évalue l'effet de chaque paramètre d'entrée sur ces composantes. Ainsi, la série de Fourier tronquée qui en découle peut être assimilée à un émulateur. La méthode permet d'évaluer efficacement la sensibilité des résultats du modèle aux variations des paramètres d'entrée. En décomposant l'espace des paramètres d'entrée en bandes de fréquence à l'aide de l'analyse de Fourier, la méthode FAST quantifie la contribution de chaque paramètre à la variabilité de la sortie. Elle fournit un classement des paramètres d'entrée en fonction de leur importance dans l'incertitude de la sortie. Il s'agit d'une méthode informatiquement efficace qui peut traiter un grand nombre de paramètres, ce qui en fait un outil utile pour la compréhension et l'optimisation de modèles complexes. Cependant, la méthode n'est souvent utilisée que pour estimer des intégrales, ce qui est le propre même des méthodes basées sur l'échantillonnage.

La méthode de décomposition du chaos polynomial est utilisée dans la quantification de l'incertitude pour approximer le comportement d'un système stochastique. Elle fournit un moyen systématique de représenter et de propager les incertitudes dans un modèle en décomposant les variables stochastiques en un ensemble de polynômes orthogonaux. La méthode commence par la définition d'un ensemble de polynômes orthogonaux, tels que les polynômes de Legendre ou d'Hermite, qui présentent des propriétés mathématiques intéressantes. Ces polynômes sont choisis en fonction de la distribution de probabilité des variables stochastiques du système. Les variables stochastiques sont ensuite développées en termes de ces polynômes, formant une expansion du chaos polynomial. L'expansion polynomiale du chaos permet de représenter les variables aléatoires comme des fonctions déterministes, ce qui simplifie l'analyse du système. Dans l'ensemble, la méthode de décomposition du chaos polynomial constitue une approche flexible et efficace pour aborder l'incertitude dans les modèles mathématiques, offrant un équilibre entre la précision et le coût de calcul.

Applications à quelques problèmes d'Hydraulique souterraine

Chapitre 3

III. Applications à quelques problèmes d'Hydraulique souterraine

III.1. Introduction

Dans ce chapitre, nous aborderons quelques modèles hydrauliques où nous tenterons d'appliquer une analyse de sensibilité globale pour calculer les indices de Sobol. Nous décrirons dans un premier temps ces modèles, puis nous présenterons quelques illustrations à titre de rappel. Les calculs seront effectués sur la plateforme Matlab en utilisant quelques fonctions comme celle de Bessel type 2 et d'ordre 0, et la fonction $W(u)$ nommée l'intégrale exponentielle $E_1(u)$ disponibles dans la librairie Matlab. La méthode du Test de Sensibilité d'Amplitude de Fourier (FAST) et la méthode de décomposition en polynômes de chaos seront appliquées aux problèmes hydrauliques choisis. Nous avons opté pour un modèle de transport de soluté et quelques modèles utilisés pour l'interprétation des essais de pompage. En phase finale de chapitre, nous présenterons les résultats obtenus à l'aide des deux techniques de l'analyse de sensibilité globale et nous commenterons les principaux résultats.

III.2. Transport de soluté en milieu poreux

Lorsque la Convection et la Diffusion/Dispersion sont prises en compte, le vecteur de flux dans la direction x est la somme des deux contributions :

$$j_x = -D \frac{\partial c}{\partial x} + vc \quad (49)$$

où dans la Diffusivité D , différentes contributions doivent être prises en compte : de l'échelle moléculaire, de la tortuosité ou de la dispersion à l'échelle régionale. Des formules similaires peuvent être énoncées pour les composantes du flux dans les directions y et z . En notation vectorielle, les résultats :

$$j = -D \nabla c + vc \quad (50)$$

Ici, le coefficient est écrit sous forme de matrice afin de tenir compte du cas général, comme décrit ci-dessus. Il est maintenant temps d'utiliser ce résultat pour remplacer les termes de flux dans l'équation de conservation de la masse (57). Le résultat en une dimension est le suivant :

$$\theta \frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \theta \left(D \frac{\partial c}{\partial x} - v c \right) + q \quad (51)$$

Dans le cas d'une vitesse constante, on obtient la formulation la plus courante de l'équation de transport:

$$\theta \frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left(\theta D \frac{\partial c}{\partial x} \right) - \theta v \frac{\partial c}{\partial x} + q \quad (52)$$

Dans le cas d'une constante D , la formulation équivalente est la suivante :

$$\theta \frac{\partial c}{\partial t} = \theta D \frac{\partial^2 c}{\partial x^2} - \theta v \frac{\partial c}{\partial x} + q \quad (53)$$

Une solution analytique pour le transport de soluté (53) avec des coefficients constants et $q = 0$ a été donnée par Ogata & Banks (1961). Elle est indépendante de θ et se lit comme suit :

$$c(x, t) = \frac{cin}{2} \left(\operatorname{erfc} \left(\frac{x - vt}{2\sqrt{Dt}} \right) + \exp \left(\frac{v}{D} x \right) \operatorname{erfc} \left(\frac{x + vt}{2\sqrt{Dt}} \right) \right) \quad (54)$$

erfc désigne la fonction d'erreur complémentaire qui est définie comme suit :

$$\operatorname{erfc}(\xi) = 1 - \frac{2}{\sqrt{\pi}} \int_0^{\xi} \exp(-\zeta^2) d\zeta \quad (55)$$

erfc est lié à la fonction d'erreur "erf" :

$$\operatorname{erfc}(\xi) := 1 - \operatorname{erf}(\xi) \text{ mit } \operatorname{erf}(\xi) := \frac{2}{\sqrt{\pi}} \int_0^{\xi} \exp(-\zeta^2) d\zeta \quad (56)$$

Les deux fonctions peuvent être appelées directement par une command **MATLAB**, égale à l'abréviation utilisée ci-dessus. La séquence de commandes :

```
x = [-4:0.04:4]; plot(x, erf(x), '--r', x, erfc(x), 'b');
```

```
legend ('erf','erfc',2); grid;
```

illustre les graphiques des deux fonctions :

Dans **MATLAB**, d'autres fonctions liées à la fonction d'erreur sont implémentées. Les détails peuvent être obtenus en utilisant le mot-clé "error function" dans le système d'aide de **MATLAB**.

La solution 1D pour le transport, donnée dans (54), est valable dans le demi-espace infini $x \geq 0$ pour la condition initiale

$$c(x, t = 0) = 0$$

et la condition limite :

$$c(x = 0, t) = c_{in} \quad c(x = \infty, t) = 0$$

Par exemple, lorsque le composant n'est pas présent au moment de la simulation initiale et qu'il est introduit dans le système par un afflux de concentration c_{in} à la position $x = 0$. La deuxième condition limite concerne le comportement du système en un point infiniment éloigné de la limite d'afflux, où la concentration reste constante. Pour l'application de la formule dans la pratique, il s'ensuit que la limite d'écoulement est éloignée pendant tous les instants qui nous intéressent. Lorsque le front s'approche de la limite d'écoulement, la solution d'Ogata-Banks n'est plus valable.

La généralisation de la formule pour une condition initiale non nulle $c(x = 0, t) = c_0$ est donnée par :

$$c(x, t) = c_0 + \frac{c_{in} - c_0}{2} \left(\operatorname{erfc} \left(\frac{x - vt}{2\sqrt{Dt}} \right) + \exp \left(\frac{v}{D} x \right) \operatorname{erfc} \left(\frac{x + vt}{2\sqrt{Dt}} \right) \right) \quad (57)$$

Dans **MATLAB** (Wexler 1992), la formule est calculée dans un fichier M "analtrans.m", qui est décrit ci-après. Les premières données d'entrée sont spécifiées :

```
T = 1;           % temps maximum
L = 1;           % longueur maximale
v = 1;           % vitesse
D = 0.1;         % diffusivité / dispersion
c0 = 0;          % valeur initiale
cin = 1;         % valeur d'entrée
M = 50;          % nombre de pas de temps
N = 50;          % nombre de nœuds
```

Avant de mettre en œuvre la formule, certaines variables auxiliaires sont introduites :

```
e = ones (1,N);           % vecteur de lignes
t = linspace (T/M,T,M);   % discrétisation du temps
x = linspace (0,L,N);     % discrétisation de l'espace
c = c0*e;                 % distribution initiale
```

où e est un vecteur de ligne de longueur N . Dans un autre vecteur de ligne t , les instants pour lesquels la formule doit être évaluée sont rassemblés. Il y a M instants, l'instant initial $t = 0$ n'étant pas compté. Le vecteur de ligne x contient les positions auxquelles la formule est évaluée. Il y a N positions, sans position de départ ni de fin. Dans le vecteur ligne c , la condition initiale à $t = 0$ de la concentration aux positions x est donnée.

Pendant l'exécution du programme, le vecteur c devient une matrice, pour laquelle une nouvelle ligne est créée pour chacun des instants donnés par le vecteur t . Cela se produit dans la boucle suivante, dans laquelle la formule est mise en œuvre :

```
for i=1:size(tt,2)
    h=1/(2.*sqrt(D*tt(i)));
    c=[c;c0+(cin-c0)*0.5*(erfc(h*(x-e*v*tt(i)))...
        +exp((v/D)*x).*erfc(h*(x+e*v*tt(i))))];
end
```

L'indice de la boucle est i , le temps actuel dans la boucle est $t(i)$. La fonction `size` détermine le nombre de colonnes d'un vecteur, ici le nombre d'entrées du vecteur ligne t . h est le coefficient $1/2\sqrt{Dt}$, qui apparaît entre parenthèses. L'expression longue des troisième et quatrième lignes correspond à la formule d'Ogata-Banks. Notez que x est un vecteur. Pour ajouter ou soustraire le terme vt (voir (57)), il faut s'assurer que le terme est également un vecteur : cela se fait par multiplication avec le vecteur un e . Notez également que la multiplication du facteur "exp" et du facteur "erfc" dans le dernier terme de l'expression doit être effectuée par élément : c'est pourquoi l'opérateur de multiplication `.*` est nécessaire.

Lorsque la commande **MATLAB**. `plot` est appliquée à une matrice, les valeurs des colonnes sont tracées en fonction des numéros de ligne. A titre d'exercice, vérifiez que la commande

```
plot([1 2 ; 2 6 ; 7 8; 8 1 ])
```

produit deux graphiques de fonctions basés sur quatre nœuds chacun. Pour obtenir un tracé des lignes en fonction des nombres de colonnes, il faut utiliser la commande pour la matrice transposée. La commande finale

```
plot (c');
```

produit le graphique illustré à la figure 7 :

Dans le cas d'une diffusion pure, la solution d'Ogata et Banks se simplifie sous la forme suivante :

$$c(x, t) = c_{in} \operatorname{erfc} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (58)$$

La solution sans dimension c/c_{in} est identique à la fonction d'erreur complémentaire avec l'argument sans dimension $\xi = x/2\sqrt{Dt}$. L'évaluation peut être effectuée facilement de manière classique : si l'on s'intéresse à la concentration à l'endroit x et au temps t , il est commode de calculer ξ et d'utiliser la valeur obtenue dans le graphique de la fonction d'erreur (Fig. 6) pour obtenir la valeur fonctionnelle correspondante. Cette dernière doit être multipliée par c_{in} pour obtenir la valeur c souhaitée.

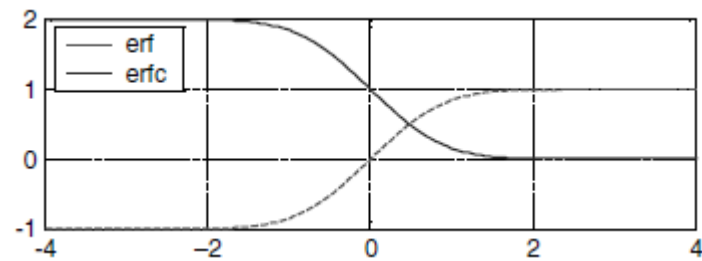


Fig. 6: Fonction d'erreur (erf) et fonction d'erreur complémentaire (erfc)

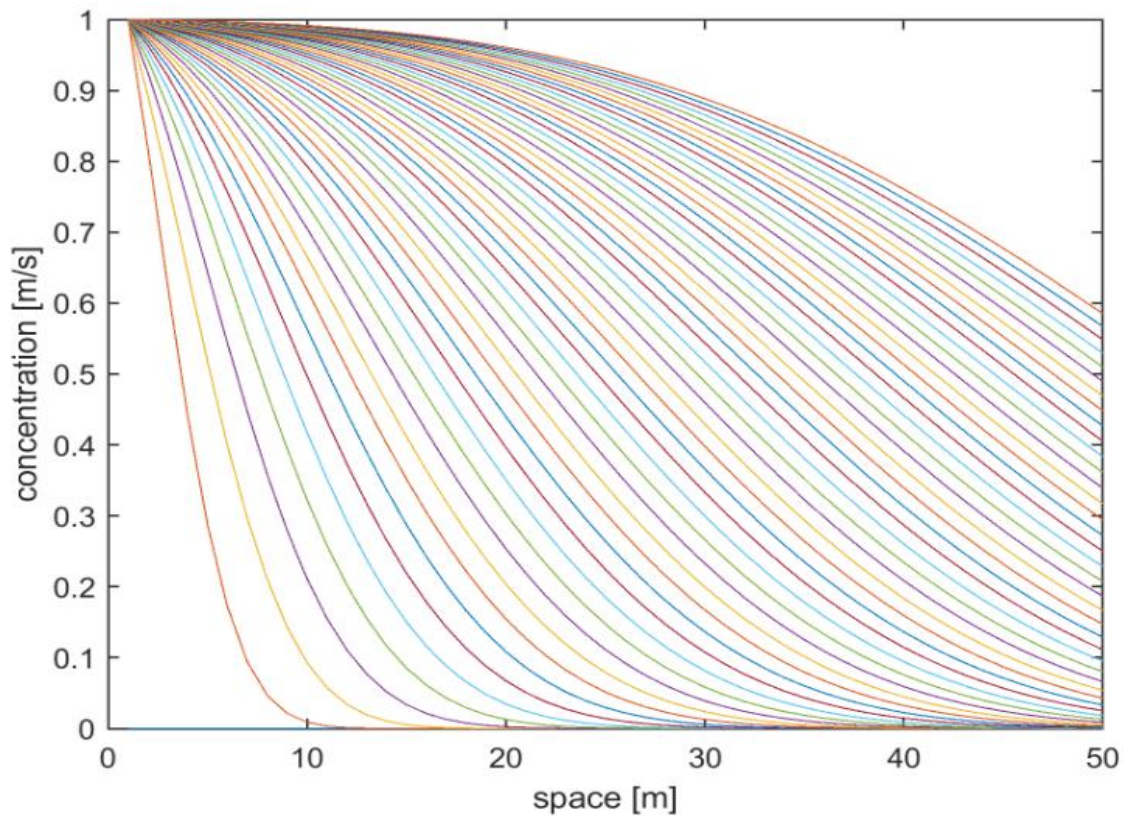


Fig. 7: Solution de l'équation de transport (Solution analytique calculée avec MATLAB)

III.3. Essais de pompage :

Les aquifères constituent une ressource inestimable en eau. Les eaux souterraines sont largement utilisées dans de nombreuses régions du monde pour des besoins industriels, municipaux et l'approvisionnement en eau potable. Dans de nombreux endroits, les eaux souterraines représentent la totalité, soit 100 %, de l'approvisionnement public en eau. Bien que la composition chimique des eaux souterraines puisse varier considérablement, leur qualité est souvent conforme aux normes les plus rigoureuses.

On extrait l'eau souterraine soit par des puits individuels, soit par des galeries comprenant plusieurs puits. Lorsque les puits sont situés à proximité, le niveau de la nappe phréatique peut diminuer en fonction du type d'aquifère. Dans tous les cas, la hauteur piézométrique diminue, ce phénomène sera expliqué plus en détail par la suite.

Il est essentiel de réaliser des études environnementales en lien avec l'extraction des eaux souterraines pour plusieurs raisons. L'organisme responsable de l'exploitation du puits doit se préoccuper de maximiser le rendement durablement. Le rabattement de la nappe phréatique lui-même revêt une importance écologique, car il peut affecter les écosystèmes du bassin versant du puits. Les zones humides, par exemple, sont des systèmes sensibles qui réagissent de manière très sensible aux variations de la nappe phréatique souterraine ou de surface.

La préservation de la qualité de l'eau est un autre aspect essentiel des systèmes de prélèvement d'eau. Si l'eau pompée ne satisfait pas aux normes de qualité requises, il est crucial de comprendre le bassin versant et les voies de circulation afin de prendre des mesures préventives visant à éviter ou à réduire la migration de l'eau contaminée vers les installations de pompage. L'utilisation judicieuse de puits de recharge, positionnés de manière appropriée et rechargeant l'aquifère à un taux adéquat, peut empêcher une telle migration.

Dans cette section, nous étudions les fluctuations de la hauteur piézométrique h à proximité d'un seul puits de pompage. Des solutions analytiques existent pour h en fonction de la distance par rapport au centre du puits r , et elles peuvent être calculées facilement à l'aide du logiciel **MATLAB**. Toutefois, dans ces calculs, on néglige les autres facteurs qui influencent l'écoulement des eaux souterraines, tels que l'écoulement de base. Des situations plus complexes seront abordées dans les deux sections suivantes.

III.3.1 Nappe captive :

Un aquifère captif est une couche d'eau souterraine perméable située entre deux couches imperméables (aquitards), comme le montre la figure 8. Dans des situations idéales, qui sont traitées dans ce chapitre, les eaux souterraines s'écoulent dans une couche perméable, l'aquifère, de tous les côtés radialement vers un puits de pompage installé. On suppose que la situation est totalement égale dans toutes les directions radiales, ce qui permet d'utiliser le rayon r comme variable d'espace.

On suppose également qu'il n'y a pas de différences dans la direction verticale : le puits est blindé sur l'ensemble de l'aquifère et il n'y a pas de différences concernant les propriétés hydrauliques à l'intérieur de la couche perméable. L'aquifère reste saturé en eau, c'est-à-dire qu'il n'y a pas de parties qui s'assèchent à cause du pompage.

Dans la situation idéalisée de la figure 8, l'aquifère est caractérisé par une épaisseur H [m] et une transmissivité T [m²/s]. Le paramètre de transmissivité représente la conductivité hydraulique K du matériau poreux et l'épaisseur de l'aquifère H :

$$T = K.H \quad (59)$$

T augmente avec l'épaisseur ; T est plus élevé pour les aquifères plus perméables. On suppose que le puits prélève de l'eau à un taux constant Q [m³/s], ce qui permet de décrire l'écoulement de l'eau souterraine en régime permanent. La variable pertinente pour l'analyse de l'écoulement des eaux souterraines est la hauteur piézométrique h , qui varie en fonction de la distance r par rapport à la position du puits. La hauteur piézométrique est la variable clé de l'écoulement, quantifiant la hauteur de la nappe phréatique au-dessus d'un niveau de référence mesuré par un piézomètre. Un piézomètre est un tuyau ouvert aux deux extrémités et qui pénètre dans l'aquifère par l'extrémité inférieure. h diminue si l'on s'approche du puits et peut être calculée en utilisant la formule de Thiem (1906) :

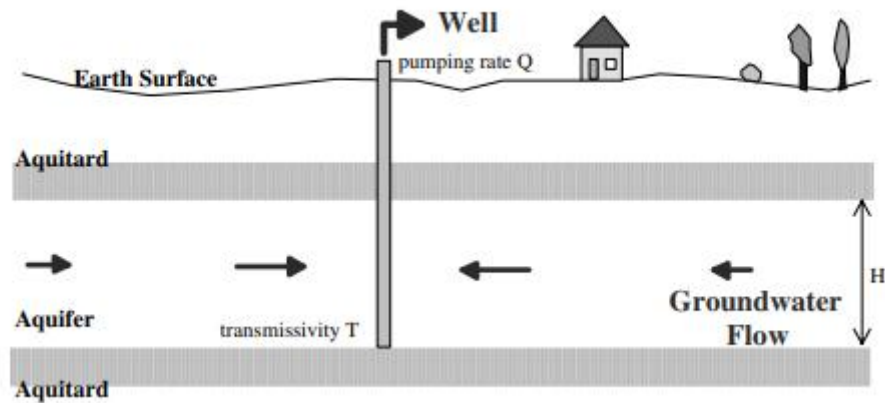


Fig. 8: Vue schématique en coupe d'un puits pompant dans un aquifère confiné

$$h(r) = h_0 + \frac{Q}{2\pi T} \log\left(\frac{r}{r_0}\right) \quad (60)$$

avec: h_0 hauteur piézométrique au-dessus de la base au rayon r_0 [m]

Q débit de pompage [m^3/s]

T transmissivité de l'aquifère [m^2/s]

r_0 rayon [m]

Nous décrivons ici une courte séquence de commandes permettant de calculer les valeurs de la hauteur piézométrique à proximité du puits. Tout d'abord, les valeurs des paramètres d'entrée doivent être données :

```
h0 = 5;
T = 5.e-6;
Q = 1.e-4;
r0 = 0.1;
```

Le vecteur de rayon r est ensuite spécifié. Pour chaque rayon du vecteur, l'abaissement de la hauteur piézométrique h est calculé :

```
r = [0.1:0.1:20];
```

La commande suivante lance le calcul des valeurs souhaitées selon la formule (60) :

$$h = h_0 + (Q/(2 * \pi * T)) * \log(r/r_0); \quad (61)$$

Dans **MATLAB**, $\log$ est le logarithme naturel de la base e . Dans le vecteur h , nous trouvons maintenant toutes les hauteurs piézométriques pour les rayons, donnés dans le vecteur r .

Pour une illustration, il est préférable d'utiliser un graphique qui est produit le plus facilement par la commande :

```
plot(r,h);
ylabel ('piezometric head [m]'); xlabel ('distance [m]');
```

Le résultat est illustré à la figure 9.

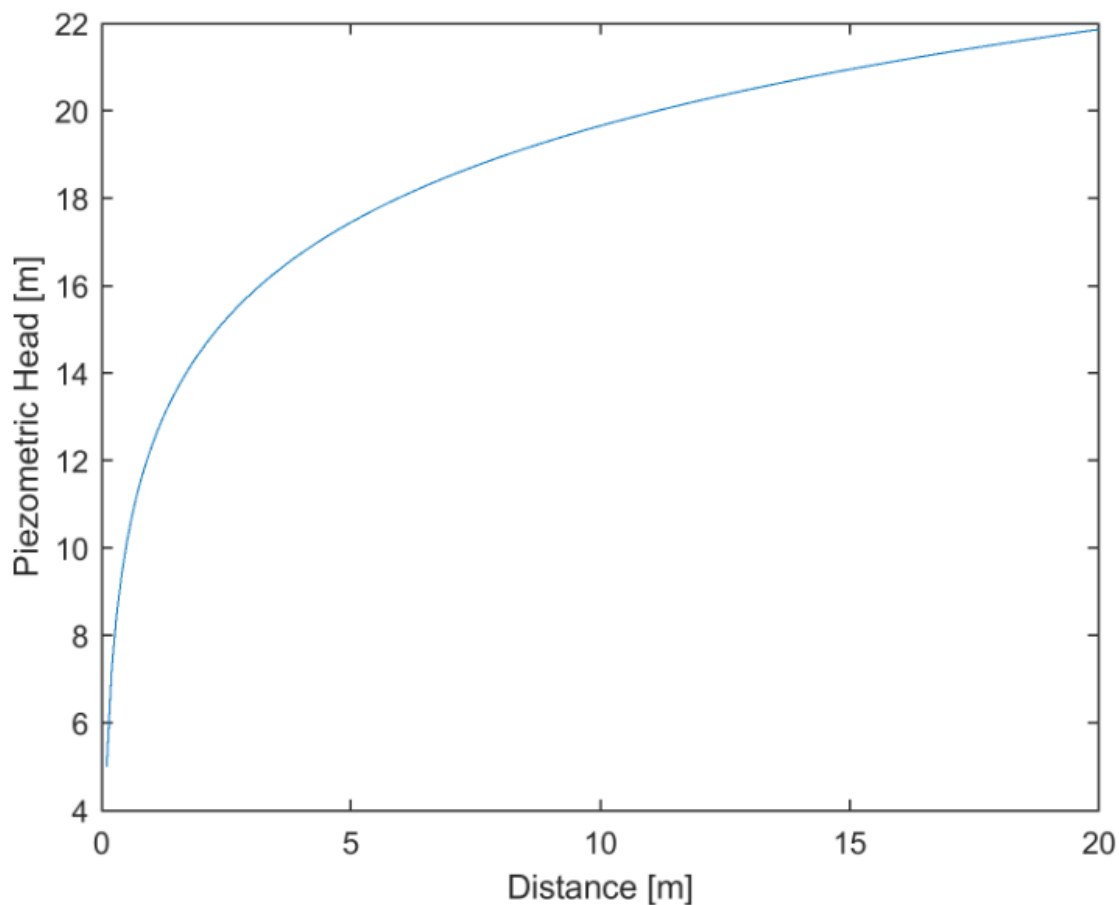


Fig. 9: Abaissement de la hauteur piézométrique de l'eau souterraine dans un aquifère captif (confiné) dû au pompage (formule de Thiem)

III.3.2 Nappe libre :

Contrairement à un aquifère captif, un aquifère libre (également appelé aquifère phréatique) n'est pas limité par une couche imperméable située au-dessus. La limite supérieure d'un aquifère non confiné est donnée par la nappe phréatique. Entre la nappe phréatique et la surface de la terre se trouve la zone non saturée, où l'espace poreux à l'intérieur du matériau poreux est rempli d'eau et d'air.

Dans l'aquifère, seule l'eau circule dans l'espace poreux. La situation est représentée schématiquement par la Fig. 10.

Si elle est mesurée par rapport à la base de l'aquifère, la variable h est l'épaisseur saturée en eau de l'aquifère, c'est-à-dire la distance entre la position de la nappe phréatique et la base de l'aquifère situé en dessous. Contrairement à l'aquifère captif, dans l'aquifère libre, la hauteur piézométrique h représente la position de la nappe phréatique.

La formule suivante adapte l'équation de Thiem (55) à la situation d'un aquifère non confiné. Elle est dérivée de l'encadré 12.1 et fournit la hauteur piézométrique h à la distance r d'un puits :

$$h^2(r) = h_0^2 + \frac{Q}{\pi K} \log\left(\frac{r}{r_0}\right) \quad (62)$$

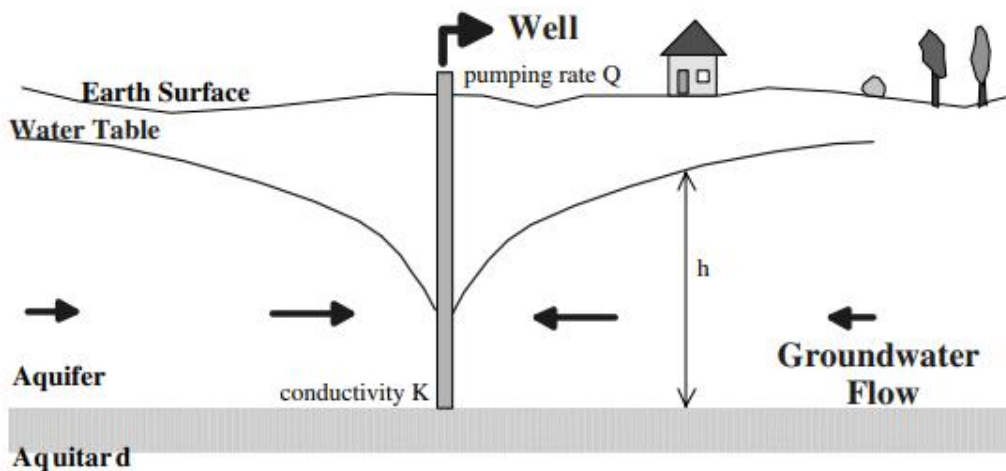


Fig. 10: Vue schématique en coupe d'un puits pompant dans un aquifère libre

avec:

- h_0 niveau de l'eau dans le puits [m]
- Q débit de pompage [m^3/s]
- K conductivité hydraulique de l'aquifère [m/s]
- r_0 rayon du puits [m]

Dans l'aquifère confiné, l'écoulement horizontal vers un puits en régime permanent doit satisfaire à l'équation de conservation du volume :

$$2\pi r H v_r = Q \quad (63)$$

pour tous les rayons r avec une vitesse v_r dépendant du rayon, une profondeur d'aquifère H et un taux de pompage Q . La loi de Darcy s'applique :

$$v_r = K \frac{\partial h}{\partial r} \quad (64)$$

Les deux équations combinées donnent une équation différentielle pour $h(r)$:

$$r \frac{\partial h}{\partial r} = \frac{Q}{2\pi T} \quad (65)$$

avec $T = KH$. Le côté droit étant une constante, l'équation différentielle peut également s'écrire comme suit :

$$\frac{\partial}{\partial r} \left(r \frac{\partial h}{\partial r} \right) = 0 \quad (66)$$

Afin d'obtenir une formule de solution, nous procédons à une reformulation de l'équation :

$$\frac{\partial h}{\partial r} = \frac{Q}{2\pi T} \frac{1}{r} \quad (67)$$

La solution peut être obtenue simplement par intégration :

$$h = \frac{Q}{2\pi T} \log(r) + C \quad (68)$$

avec la constante d'intégration C . Si la tête h_0 à une position r_0 est donnée, la constante d'intégration peut être déterminée :

$$C = h_0 - \frac{Q}{2\pi T} \log(r_0) \quad (69)$$

Dans la situation non confinée, on commence par analogie avec le principe de conservation du volume: Voir l'équation (63)

Au lieu de la hauteur totale de la couche d'eau souterraine, il faut considérer la hauteur de la nappe phréatique h au-dessus de la base. L'utilisation de la loi de Darcy, comme indiqué ci-dessus, donne :

$$r h \frac{\partial h}{\partial r} = \frac{Q}{2\pi K} \quad (70)$$

Où

$$r \frac{\partial h^2}{\partial r} = \frac{Q}{\pi K} \quad (71)$$

avec la solution générale :

$$h^2 = \frac{Q}{\pi K} \log(r) + C \quad (72)$$

Comme ci-dessus, la connaissance d'une paire (r0,h0) permet de déterminer la constante d'intégration C et la formule (62) en résulte.

h varie en fonction du rayon r, comme expliqué ci-dessus. Nous calculons les valeurs changeantes dans une courte séquence de commandes. Dans la fenêtre de commande **MATLAB®**, spécifiez d'abord le nouveau paramètre d'entrée :

```
K = 1.e-4;
```

La ligne suivante lance le calcul du vecteur des hauteurs piézométriques :

$$h = \text{sqrt}(h0 * h0 + (Q/(pi * K)) * \log(r/r0)); \quad (73)$$

sqrt désigne la racine carrée. Avec la commande suivante, les résultats sont affichés sous la forme

```
plot (r,h, '--g');
```

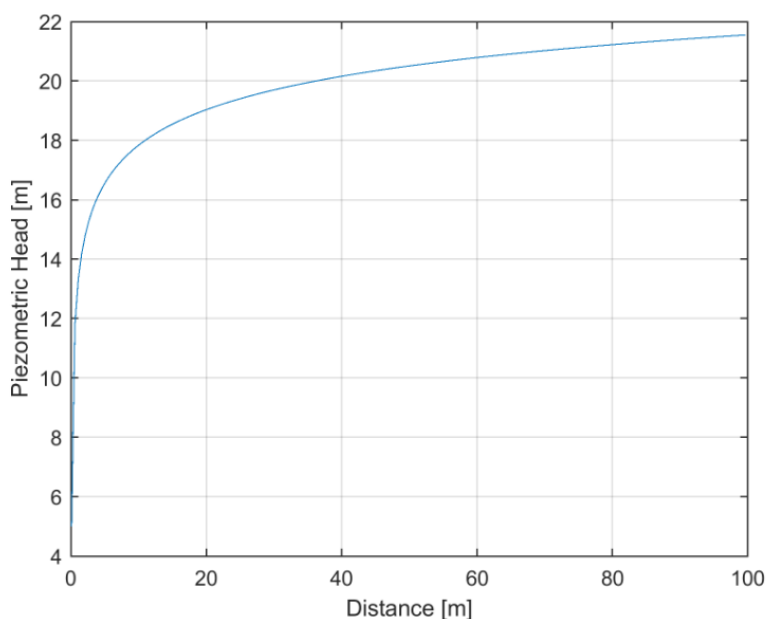


Fig. 11: Abaissement de la hauteur piézométrique de l'eau souterraine dans un aquifère libéré

III.3.3 Nappe semi-captive :

La présence d'une couche semi-perméable recouvre un aquifère, permettant ainsi à l'eau pompée de provenir à la fois de l'aquifère lui-même et des couches supérieures qui sont connectées par cette couche semi-perméable.

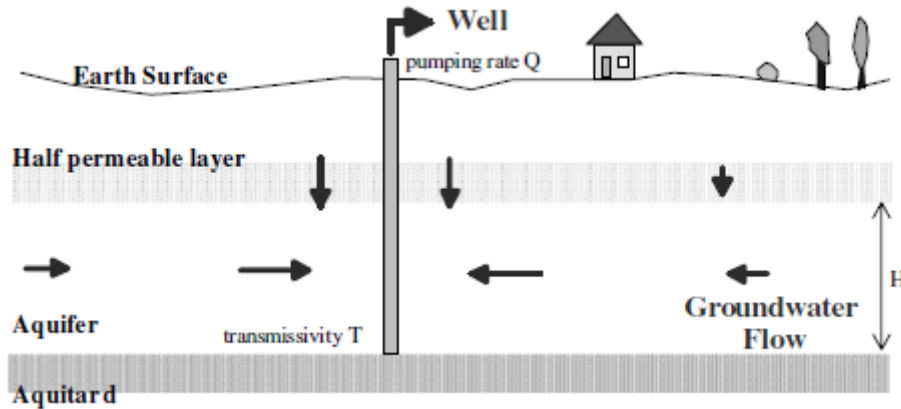


Fig. 12.: Vue schématique en coupe d'un puits pompant dans un aquifère semi-captif.

Dans le cas d'un aquifère épais partiellement confiné, de Glee (1930) a développé une formule qui décrit les diminutions de la hauteur piézométrique à une distance r d'un puits :

$$s(r) = \frac{Q}{2\pi T} K_0 \left(\frac{r}{\sqrt{Tc}} \right) \quad (74)$$

avec :

Q	débit de pompage [m ³ /s]
T	transmissivité [m ² /s]
c	résistance de la couche semi-perméable [s]
K ₀	fonction de Bessel modifiée

Dans **MATLAB**, introduisez le nouveau paramètre c :

```
c = 1.e7;
```

et de calculer le vecteur des rabattements (drawdowns) en fonction de :

$$s = (Q/(2 * pi * T)) * bessellk(0, r/sqrt(T * c)); \quad (75)$$

La variante correcte de la fonction de Bessel (ici bessellk) et les paramètres sont indiqués dans l'aide **MATLAB**. Le résultat est affiché à l'aide de la commande plot :

```
plot (r,-s);
```

La figure 13 illustre la sortie graphique. La figure 14 compare les rabattements d'un aquifère captif, d'un aquifère libre et d'un aquifère semi-captif. Le rabattement pour la situation de semi-captive se situe entre le rabattement pour l'aquifère confiné et l'aquifère non confiné. L'utilisateur peut facilement trouver des valeurs de paramètres pour lesquelles ce résultat raisonnable n'est pas vrai. La raison de cette incompatibilité apparente est que les trois formules sont valables dans des conditions différentes. La formule de de Glee est dérivée pour le demi-espace situé sous la couche semi-perméable, c'est-à-dire en supposant que l'aquifère est trop étendu, ce qui rend la valeur de son épaisseur non pertinente.

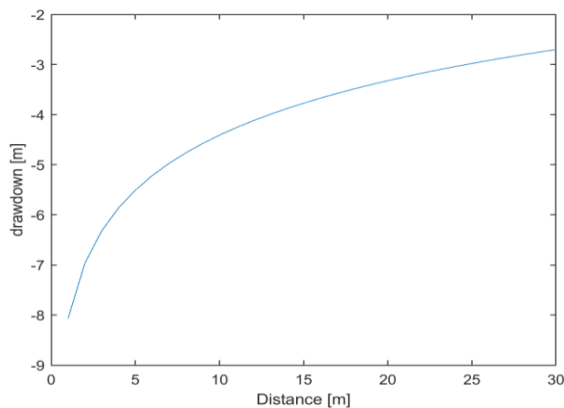


Fig. 13: Abaissement de la hauteur piézométrique de l'eau souterraine dans un aquifère semi- captif dû au pompage, d'après de Glee (1930)

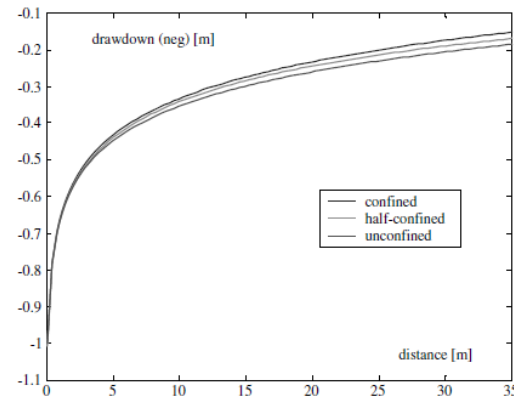


Fig. 14: Abaissement régulier de la hauteur piézométrique des eaux souterraines dans un aquifère confiné, un aquifère semi- captif et un aquifère non confiné

III.3.4 Pompage en régime transitoire et fonction de puits :

Dans un aquifère confiné, le rabattement de la hauteur piézométrique s est donné par la formule de Theis (1983). s est fonction de la distance du puits r et du temps t après le début du pompage :

$$s(r) = \frac{Q}{4\pi t} W\left(\frac{Sr^2}{4Tt}\right) \quad (76)$$

avec:

Q	débit de pompage [m3/s]
T	transmissivité de l'aquifère [m2/s]
S	coefficient de stockage de l'aquifère ou coefficient d'emmagasinement

Dans la formule apparaît la fonction W , avec un argument qui est généralement abrégé en u . $W(u)$ est si important pour le rabattement du puits qu'il est appelé fonction du puits dans la littérature correspondante. Dans la littérature mathématique, la même fonction est appelée intégrale exponentielle et est généralement désignée par $E1(u)$. Ce nom s'explique par la définition de $E1(u)$:

$$W(u) = E_1(u) = \int_u^{\infty} \frac{\exp(-\varsigma)}{\varsigma} d\varsigma \quad (77)$$

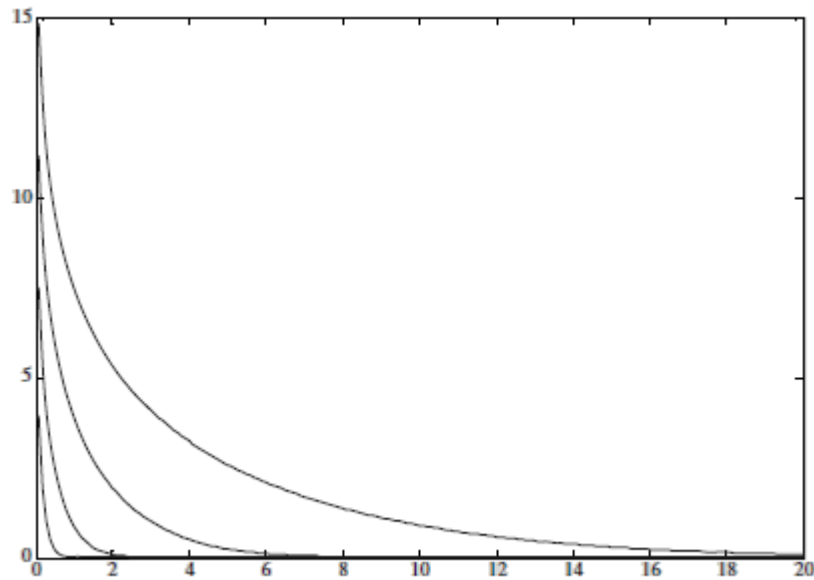


Fig. 15: Abaissement instable de la hauteur piézométrique des eaux souterraines dans un aquifère confiné dû au pompage selon Theis ; paramètres donnés dans le texte, pour les temps $t = 10^3, 10^4, 10^5$ et 10^6

La figure 15 illustre les rabattements obtenus à quatre moments différents.

Dans **MATLAB**, la fonction well peut être trouvée sous sa notation mathématique. Elle est appelée par expint. Pour calculer la formule donnée ci-dessus, spécifiez le paramètre de stockage comme nouveau paramètre :

```
S = 0.1;
```

Spécifiez ensuite la durée pour laquelle le drawdown doit être calculé :

```
t = [1000000];
```

Calculez et tracez les résultats à l'aide des commandes suivantes :

$$s = (Q/(4 * \pi * T)) * \expint(S * r.^2/(4 * T * t)); \quad (78)$$

```
plot (r,-s);
```

La figure 16 représente la variation du rabattement croissant en fonction de la distance :

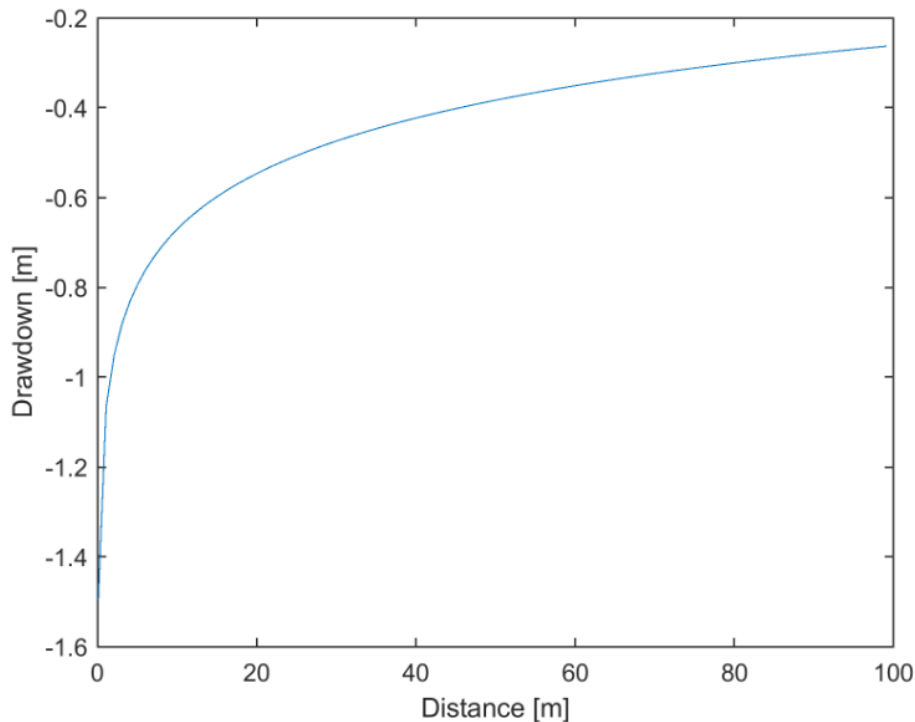


Fig. 16: Abaissement de la hauteur piézométrique de l'eau souterraine dans un aquifère captif dû au pompage, d'après la formule de Theis.

III.4. Résultats et Discussions :

Dans cette section, nous appliquerons le cadre présenté dans les chapitres précédents aux modèles mathématiques que nous avons expliqués dans les sections II.3 et II.2.

Nous utiliserons d'abord la méthode de FAST sur les modèles à nappe captive et à nappe libre, puis nous utiliserons la méthode de Polynômes de Chaos sur le modèle de Transport de soluté en milieu poreux, le modèle à Nappe semi-captive et au modèle de Pompage en régime transitoire et la fonction de puits, et nous déterminerons l'effet des paramètres d'entrée sur leurs sorties.

III.4.1 Utilisation de la méthode FAST :

L'approche utilisée par Sobol' (1990) pour le calcul des indices de sensibilité est basée sur la décomposition ANOVA et permet de calculer également la sensibilité des paramètres d'entrée couplés. L'idée est de calculer les variances des termes dans la décomposition ANOVA.

Le tableau 1 résume les informations au-delà des indices de sensibilité globaux et le tableau 2 classifie de manière heuristique les paramètres d'entrée.

Tab. 1:: Pertinence d'un paramètre d'entrée à partir de son indice de sensibilité globale.**Résumé de l'importance des indices de sensibilité globale.**

$(S1 = 1) \wedge (Stot = 1) \leftrightarrow$ le modèle $f(X)$ ne dépend pas des entrées $x2$
$(S1 = 0) \wedge (Stot = 0) \leftrightarrow$ le modèle $f(X)$ ne dépend pas des entrées $x1$
$S1 \rightarrow 1 \leftrightarrow$ l'ensemble des entrées $x1$ joue un rôle décisif dans la sortie du modèle $f(X)$
$S_a^{Tot} \rightarrow 1 \leftrightarrow$ l'ensemble des entrées $x1$ n'est pas important dans la sortie du modèle $f(X)$
$S1 = Stot \leftrightarrow$ correspond à l'absence d'interactions entre les entrées $x1$ et toute autre variable

Tab. 2: Pertinence d'un paramètre d'entrée à partir de son indice de sensibilité globale.**Pertinence d'un paramètre d'entrée à partir de son indice de sensibilité globale.**

Très important	$0.8 \leq S1 \leq Stot \leq 1$
Important	$0.5 \leq S1 \leq Stot \leq 0.8$
Sans importance	$0.3 \leq S1 \leq Stot \leq 0.5$
Sans importance	Sans objet $0 \leq S1 \leq Stot \leq 0.3$

avec : Stot : variance totale
 S1 : variances partielles de paramètre 1

III.4.1.1 Résultats du Modèle de la nappe captive :

Le tableau 3 présente les résultats obtenus pour l'analyse de sensibilité globale du modèle de la nappe captive pour mesurer l'effet des paramètres $x1$ (transmissivité) et $x2$ (distance).

Tab. 3 : Résultats de l'analyse de sensibilité pour les paramètres du modèle la nappe captive

N	t	x1	x2	Stot	S1	eStot	eS1
1000	0.153	0.569	0.0352	0.986	0.975	0.0005	0.198
4000	0.211	0.569	0.0354	0.997	0.969	0.00042	0.097
8000	0.2817	0.569	0.0354	0.988	0.969	0.0001	0.0686
10000	0.295	0.571	0.0351	0.991	0.969	0.00009	0.061
14000	0.3215	0.573	0.0352	0.994	0.964	0.000047	0.051
18000	0.324	0.571	0.0351	0.993	0.965	0.000045	0.045
100000	0.91	0.569	0.0354	0.9926	0.966	0.000024	0.019
140000	1.21	0.569	0.0354	0.992	0.966	0.000021	0.016
180000	1.50	0.569	0.0354	0.992	0.966	0.000018	0.014
1000000	7.42	0.569	0.0354	0.992	0.966	0.000007	0.006
10000000	72.97	0.569	0.0354	0.992	0.966	0.000002	0.0019
100000000	726	0.569	0.0354	0.992	0.966	0.0000007	0.0006

avec :

N : nombre d'échantillons

t : Le temps de calcul (s)

$x(1,2,n)$: coefficients de sensibilité

$eStot$: erreur de variance totale ($Stot$)

$es1$: erreur de variance partielle ($S1$)

La figure 17 montre les paramètres qui affectent le plus le modèle la nappe captive

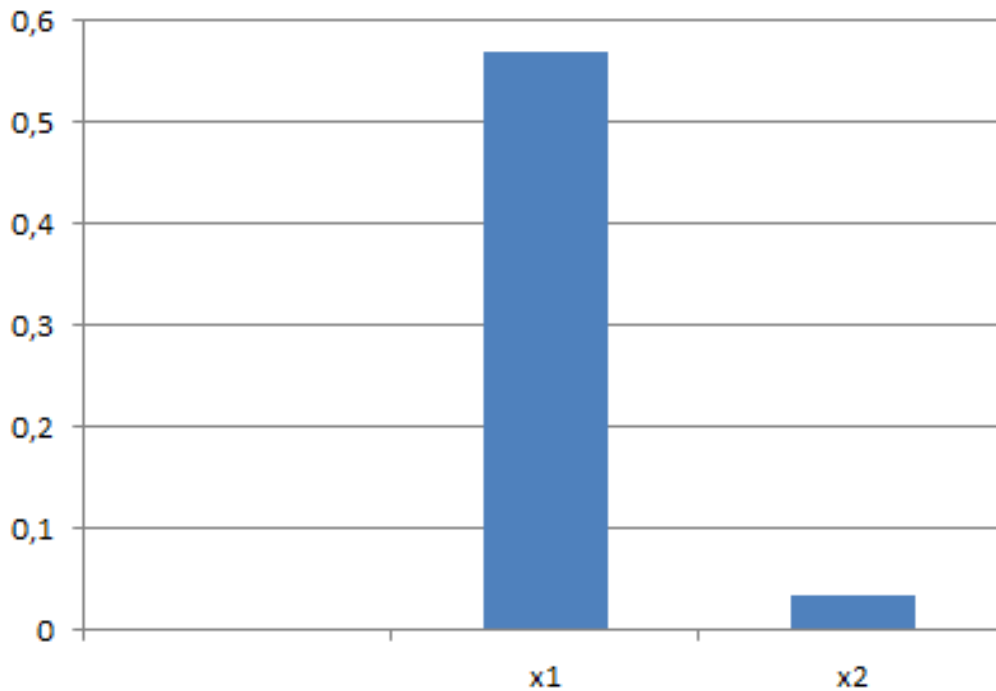


Fig. 17: Histogramme montre les paramètres qui affectent le plus le modèle la nappe captive

D'après les tableaux 1,2,3 et la figure 17, nous constatons que $x1$ est plus efficace que $x2$ sur le plan importance, nous devons prendre en compte à l'opposé $x2$ comme sans objet nous voyons également qu'il n'y a pas d'interaction entre $x1$ et toute autre variable.

La figure 18 montre l'évolution de la variance totale en fonction du nombre d'échantillon, elle montre une parfaite stabilité de la variance à partir de 100000 échantillon. On constate également sur la figure 19, la nette diminution de l'erreur sur la variance totale avec le nombre d'échantillon ce qui consolide l'analyse de sensibilité.

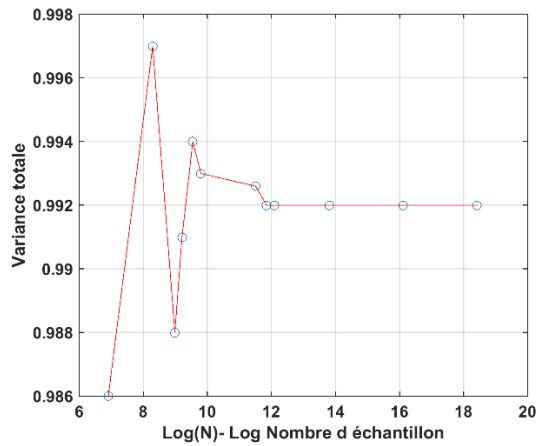


Fig. 18: Evolution de la variance totale en fonction du Nombre d'échantillon

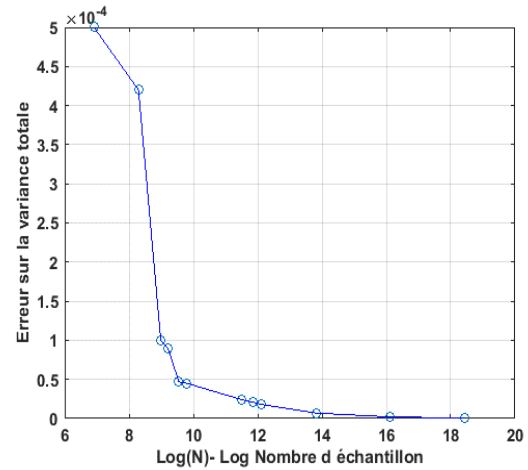


Fig. 19: Evolution de l'erreur sur la variance totale.

III.4.1.2 Résultats du Modèle de la nappe libre :

Les résultats de l'analyse de sensibilité globale à l'aide de la méthode FAST sont résumés dans le tableau 4. x1 (débit de pompage) ,x2(conductivité hydraulique de la nappe) et x3 (distance)

Tab. 4: : Résultats de l'analyse de sensibilité pour les paramètres du modèle la nappe libre.

N	t	x1	x2	x3	Stot	S1	eStot	eS1
1000	0.184	0.1953	0.5183	0.0417	0.2672	0.187	0.083	0.0060
4000	0.22	0.1953	0.5186	0.0419	0.2788	0.192	0.039	0.0032
8000	0.29	0.1954	0.5188	0.042	0.2869	0.189	0.0246	0.0022
10000	0.316	0.1954	0.5188	0.0417	0.288	0.1885	0.0236	0.002
12000	0.338	0.1953	0.5188	0.0418	0.2879	0.186	0.0214	0.0017
16000	0.374	0.1954	0.521	0.0417	0.2876	0.1876	0.0185	0.0016
100000	1.4	0.1953	0.5208	0.0417	0.2862	0.188	0.0074	0.0006
140000	1.8	0.1953	0.521	0.0417	0.2862	0.1883	0.0063	0.0005
180000	2.36	0.1953	0.521	0.0417	0.286	0.188	0.0055	0.00048
1000000	12	0.1953	0.521	0.0417	0.286	0.188	0.0023	0.0002
10000000	116.53	0.1953	0.521	0.0417	0.286	0.188	0.00021	0.000024
100000000	1542	0.1953	0.521	0.0417	0.286	0.188	0.000065	0.000007

La figure 20 illustre l'importance des paramètres qui affectent le plus le modèle de la nappe libre.

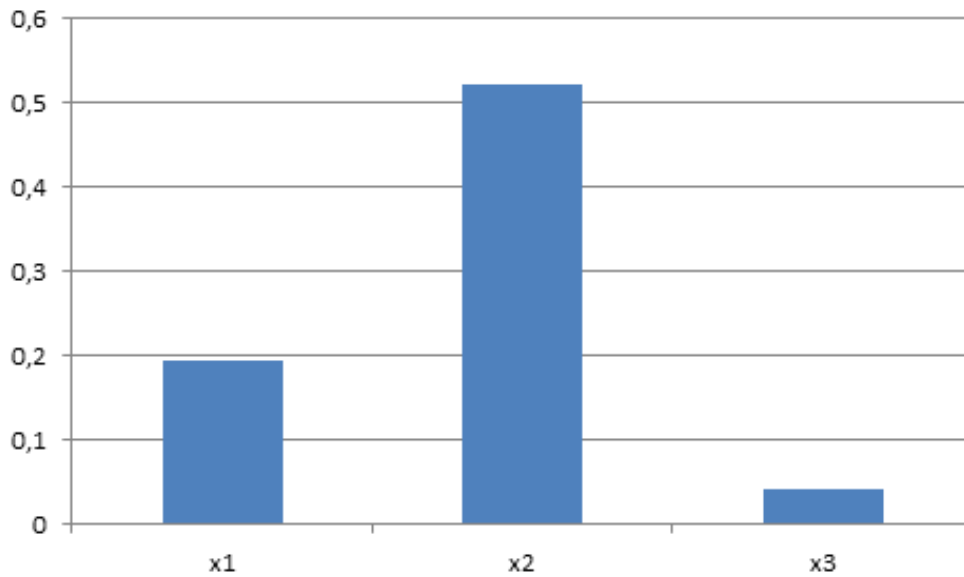


Fig. 20: Histogramme montre les paramètres qui affectent le plus le modèle la nappe captive

La figure 20 montre les paramètres d'entrée les plus influents sur les sorties du modèle. Nous voyons que x1 et x3 sont inférieurs à 0,3, selon le tableau 2, les paramètres sont classés sans importance. Par contre, le paramètre x2 semble stabilisé rapidement et sa valeur de l'ordre de 0.52 le classe comme important.

III.4.2 Utilisation de la méthode du Polynômes de Chaos :

III.4.2.1 Résultats de Transport de soluté en milieu poreux:

Le tableau 5 présente les principaux résultats de l'analyse de sensibilité des paramètres d'entrée x1 (Dispersion) et x2 (vitesse) pour le modèle Transport de soluté en milieu poreux. Des résultats complémentaires sont présentés ci-dessous. Nous constatons que x2 (Vitesse) est classée comme paramètre très important.

Tab. 5: : Résultats de l'analyse de sensibilité pour les paramètres du modèle Transport de soluté en milieu poreux

N	t	x1	x2	OutputMean	OutputVar :
1000	0.12	0.0002	0.9997	0.1857	0.02187
4000	0.168	0.00048	0.9995	0.1799	0.0209
8000	0.227	0.00041	0.99959	0.1768	0.02
10000	0.2594	0.00041	0.99959	0.1809	0.02
14000	0.31	0.00028	0.9997	0.1797	0.02
18000	0.368	0.00049	0.9995	0.18	0.02
120000	1.77	0.00044	0.9995	0.1798	0.02
180000	2.62	0.00045	0.9995	0.1799	0.02
1000000	14.08	0.00044	0.9995	0.1797	0.02
10000000	139.21	0.000442	0.9995	0.1796	0.02
100000000	3449	0.000441	0.9995	0.1797	0.02

SortSobolIndices :

0.9219	0	2.0000
0.0776	0	2.0000
0.0003	1.0000	0
0.0003	1.0000	0
0.0000	1.0000	2.0000
0.0000	1.0000	2.0000

TrainingInput :

0.722123437056550	0.670634128493896
0.722123437056550	0.330625524295189
0.414667952309245	0.670634128493896
0.722123437056550	0.930779670193380
0.414667952309245	0.330625524295189
0.943473375731451	0.670634128493896
0.722123437056550	0.0685425525971980
0.414667952309245	0.930779670193380
0.943473375731451	0.330625524295189
0.137400205869924	0.670634128493896

TrainingOutput :

0.230954914149933
 0.0695397831644622
 0.231189714789600
 0.438525494780387
 0.0704746430474601
 0.230908522147608
 0.0279144180604326
 0.438571352239410
 0.0693042247478982

0.243051416489299

aPC.OrthonormalBasis :

`val(:, :, 1) =`

1.0000	0	0	0	0
-2.8192	4.2377	0	0	0
5.1937	-20.7471	17.2773	0	0
-7.8973	59.2872	-118.7081	69.2950	0
10.8193	-131.2843	462.0638	-618.1090	278.7206

`val(:, :, 2) =`

1.0000	0	0	0	0
-1.7342	3.4601	0	0	0
2.2129	-13.3481	13.3755	0	0
-2.6110	31.6165	-79.1094	52.7433	0
2.9772	-60.0785	270.6800	-421.0514	210.4644

La figure 21 montre la surface de réponse des paramètres d'entrée x1 (dispersivité) et x2 (la vitesse) pour le modèle Transport de soluté en milieu poreux.

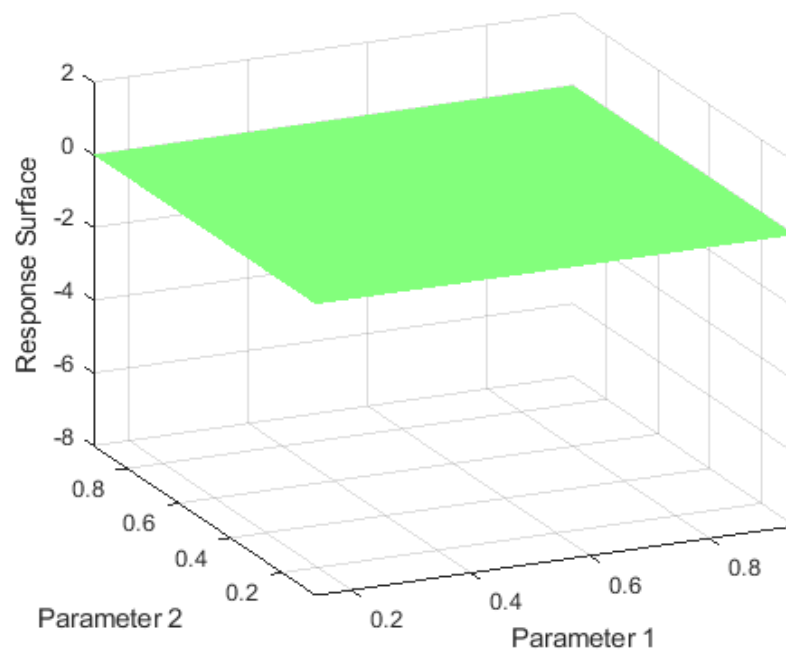


Fig. 21: surface de réponse des paramètres d'entrée pour le modèle Transport de soluté en milieu poreux

La figure 22 montre le paramètre qui affectent le plus le modèle de c est celui de la vitesse. Ce résultat tout à fait compatible avec la réalité car dans l'équation de transport de soluté le mécanisme de transfert par convection est plus important que la Dispersion et même l'interaction.

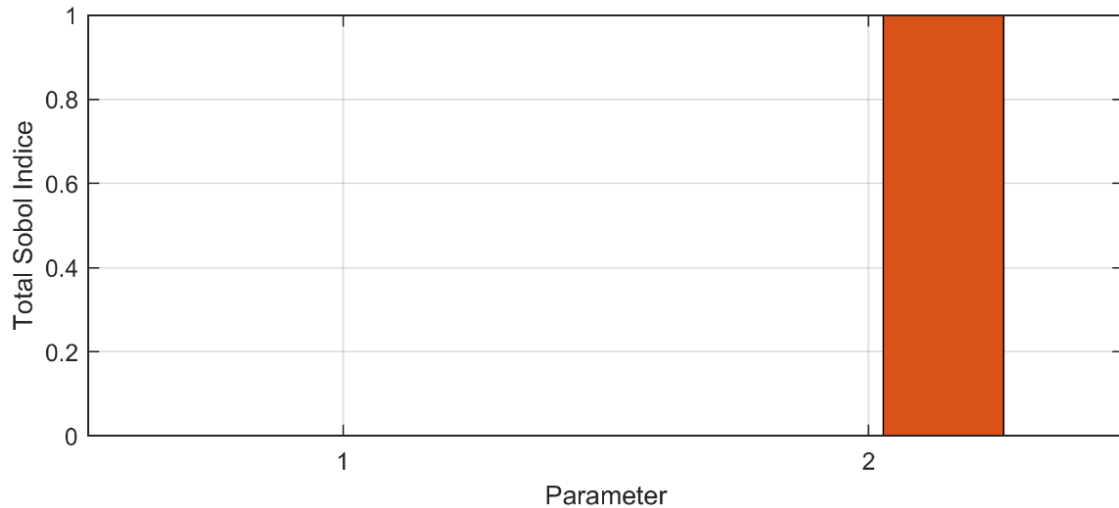


Fig. 22:Histogramme montre les paramètres qui affectent le plus le modèle Transport de soluté en milieu poreux.

D'après le tableau 5 et la figure 22, nous remarquons que x2 a une grande influence sur le modèle, nous devons prendre en compte cette donnée dans l'estimation du paramètre. Cependant, le deuxième paramètre x1 est classé sans objet.

III.4.2.2 Résultats de Nappe semi-captive :

Le tableau 6 présente les résultats de l'analyse de sensibilité des paramètres d'entrée x1 (transmissivité) et x2 (résistance de la couche semi-perméable) pour le modèle Nappe semi-captive

Tab. 6: Résultats de l'analyse de sensibilité pour les paramètres du modèle Nappe semi-captive.

N	T	x1	x2	OutputMean	OutputVar
1000	0.116	0.0658	0.983	2.220e-20	2.054e-39
2000	0.13	0.0658	0.978	2.33e-20	2.35e-39
6000	0.158	0.0684	0.982	1.91e-20	3.22e-39
10000	0.186	0.0657	0.983	2.58e-20	2.89e-39
12000	0.247	0.0731	0.9764	2.35e-20	2.87e-39
16000	0.272	0.0654	0.9833	2.44e-20	2.65e-39
100000	0.33	0.0701	0.982	2.62e-20	2.96e-39
1000000	13.96	0.0681	0.982	2.58e-20	2.85e-39
10000000	138.5	0.0692	0.9821	2.55e-20	2.80e-39
100000000	4413.62	0.0691	0.982	2.56e-20	2.82e-39

SortSobolIndices :

0.645935324853850	0	2
0.383236372140146	0	2
0.0513974204141653	1	2
0.0482383505009488	1	2
0.0167227382138123	1	0
0.00113676122866760	1	0

TrainingInput :

0.723159551485595	0.330010792940020
0.723159551485595	0.669977261231225
0.416398372759289	0.330010792940020
0.723159551485595	0.0694199809409889
0.416398372759289	0.669977261231225
0.942888285086446	0.330010792940020
0.723159551485595	0.930575448128687
0.416398372759289	0.0694199809409889
0.942888285086446	0.669977261231225
0.139747851950059	0.330010792940020

TrainingOutput :

1.50173066644425e-30
 1.60411483045103e-22
 7.58204765982503e-39
 3.37562733242012e-62
 2.72909627154676e-28
 2.53187300281102e-27
 1.19226329609064e-19
 1.68275312904352e-80
 2.78029699077742e-20

5.16765887831476e-64

aPC.OrthonormalBasis :

val(:, :, 1) =

1.0000	0	0	0	0
-2.8285	4.2426	0	0	0
5.1970	-20.7869	17.3217	0	0
-8.0005	60.0016	-120.0026	70.0014	0
11.1787	-134.1515	469.5442	-626.0702	281.7352

val(:, :, 2) =

1.0000	0	0	0	0
-1.7320	3.4643	0	0	0
2.2358	-13.4153	13.4160	0	0
-2.6453	31.7463	-79.3672	52.9120	0
2.9994	-59.9951	269.9864	-419.9854	209.9944

La figure 23 montre la surface de réponse des paramètres d'entrée x1 (transmissivité) et x2 (résistance de la couche semi-perméable) pour le modèle Nappe semi-captive.

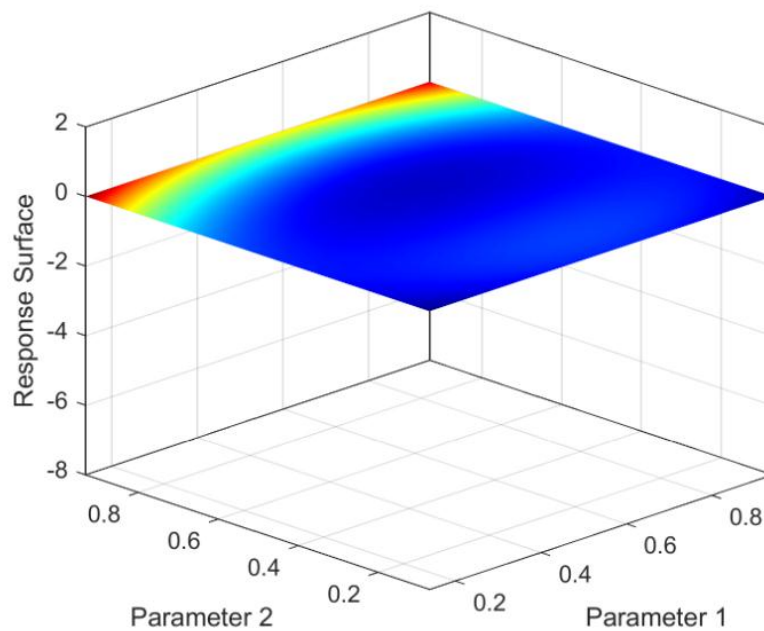


Fig. 23: surface de réponse des paramètres d'entrée pour le modèle Nappe semi-captive

La figure 24 montre les paramètres qui affectent le plus le modèle de Nappe semi-captive

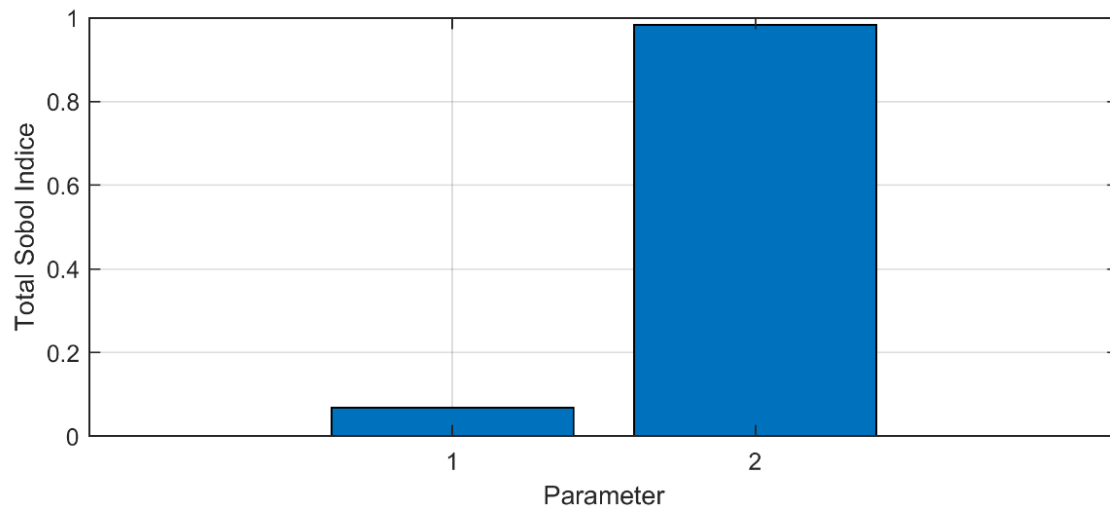


Fig. 24: Histogramme montre les paramètres qui affectent le plus le modèle Nappe semi-captive.

À travers les résultats présentés dans le tableau 6 et la figure 24, nous notons l'importance du paramètre x2 et la possibilité d'un impact important sur la sortie du modèle, et de celui-ci il faut être prudent lors de sa mesure et de son estimation, contrairement à x1, qui peut être ignoré et sa valeur fixée dans un intervalle spécifique.

III.4.2.3 Résultats de Pompage en régime transitoire et fonction de puits :

Le tableau 7 présente les résultats de l'analyse de sensibilité des paramètres d'entrée x1 (transmissivité) et x2 (coefficient de stockage de l'aquifère) pour le modèle Pompage en régime transitoire :

Tab. 7:: Résultats de l'analyse de sensibilité pour les paramètres du Pompage en régime transitoire.

N	t	x1	x2	OutputMean	OutputVar
1000	0.12	0.9759	0.028	0.024	0.0002
2000	0.15	0.778	0.0246	0.0243	0.0002
6000	0.19	0.9773	0.0264	0.024	0.0002
10000	0.15	0.981	0.0247	0.024	0.0002
12000	0.28	0.977	0.0263	0.025	0.0002
16000	0.34	0.976	0.0277	0.024	0.0002
100000	1.51	0.975	0.0277	0.024	0.0002
1000000	13.95	0.976	0.0276	0.024	0.0002
10000000	138.83	0.976	0.0277	0.024	0.0002
100000000	3585.51	0.976	0.0277	0.024	0.0002

SortSobolIndices :

0.743870965459756	1	0
0.280587812465435	1	0
0.0209541246129842	0	2
0.00344861600865001	0	2
0.00220846813558607	1	2
0.00194115274971118	1	2

TrainingInput :

0.723147608258900	0.670000448038609
0.723147608258900	0.330029480784694
0.416403826103472	0.670000448038609
0.723147608258900	0.930571285157516
0.416403826103472	0.330029480784694
0.942896882471956	0.670000448038609
0.723147608258900	0.0694344955288359
0.416403826103472	0.930571285157516
0.942896882471956	0.330029480784694
0.139768491862731	0.670000448038609

TrainingOutput :

0.0173840467777982
 0.0189422377513768
 0.0280809332339221
 0.0166611967318485
 0.0307866674641943
 0.0137803636561543
 0.0223727618936417
 0.0268258240605889
 0.0149754480041378

0.0712371161918560

aPC.OrthonormalBasis :

`val(:, :, 1) =`

1.0000	0	0	0	0
-2.8286	4.2429	0	0	0
5.1975	-20.7879	17.3226	0	0
-8.0021	60.0075	-120.0086	70.0031	0
11.1797	-134.1522	469.5297	-626.0400	281.7187

`val(:, :, 2) =`

1.0000	0	0	0	0
-1.7323	3.4643	0	0	0
2.2362	-13.4167	13.4168	0	0
-2.6460	31.7506	-79.3736	52.9147	0
3.0001	-60.0007	269.9932	-419.9789	209.9857

La figure 25 montre la surface de réponse des paramètres d'entrée x1 (transmissivité) et x2 (coefficient de stockage de l'aquifère) pour le modèle Pompage en régime transitoire.

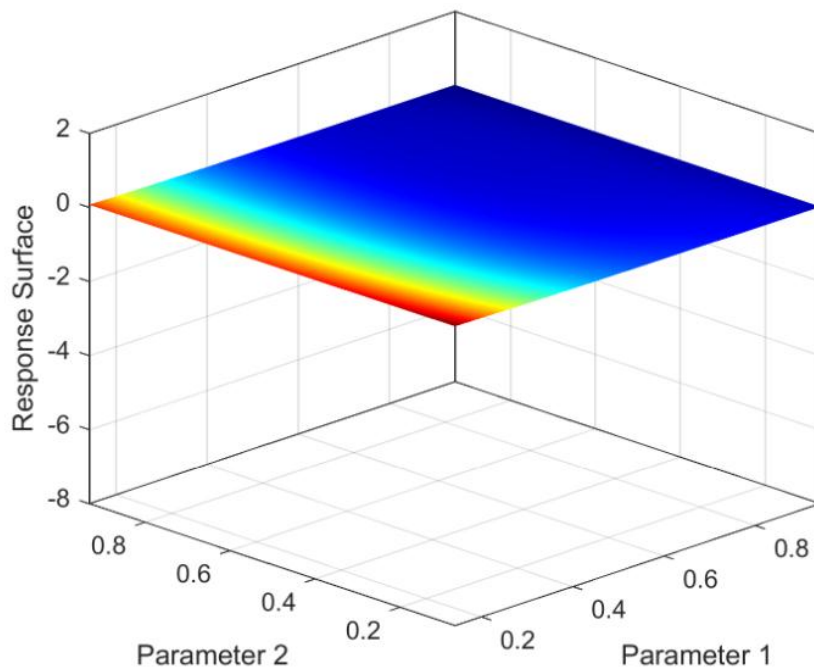


Fig. 25: surface de réponse des paramètres d'entrée pour le modèle Pompage en régime transitoire

La figure 26 montre le paramètre x_1 qui affecte le plus le modèle de l'essai de pompage en régime transitoire et qui est classé dans la catégorie du paramètre très important.

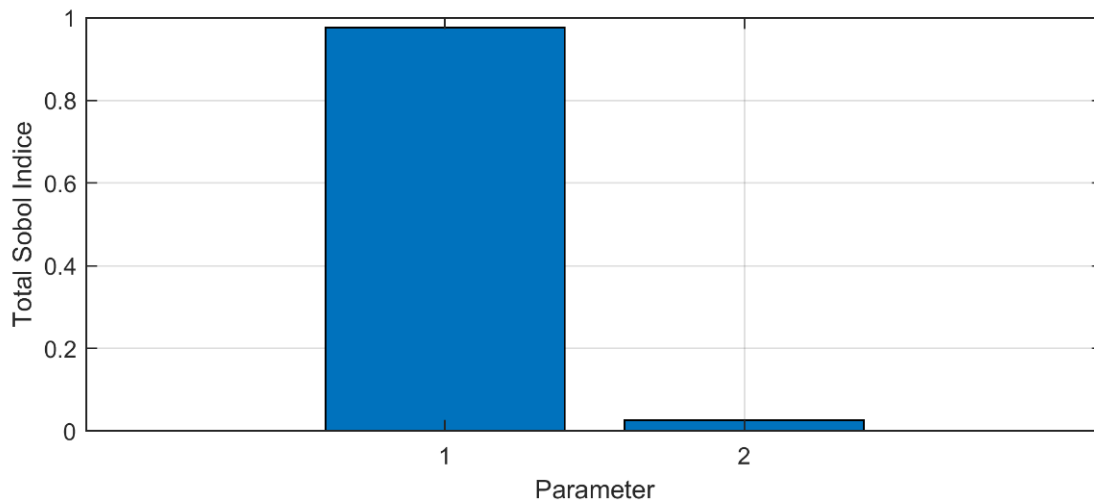


Fig. 26: Histogramme montre les paramètres qui affectent le plus le modèle Pompage en régime transitoire

D'après les résultats présentés dans le tableau 7 et la figure 26, nous voyons que la transmissivité a la plus grande part dans la variation totale, contrairement à le coefficient de stockage de l'aquifère, qui était presque inexistante .

III.5. Conclusion

Dans ce chapitre on pratique les concepts et méthodes introduits précédemment en appliquant l'analyse de sensibilité globale et les indicateurs de Sobol à des problèmes concrets d'hydraulique souterraine.

Dans le modèle de la nappe captive, nous avons constaté que la transmissivité avait un effet significatif sur la sortie du modèle, contrairement à la distance qui n'avait aucun effet.

Dans le modèle de la nappe libre, nous avons vu que le débit de pompage et la distance ne sont pas importantes, contrairement à la conductivité, dont la valeur était supérieure à 0,5, ce qui indique son importance selon le tableau 2.

Comme pour le modèle Transport de soluté en milieu poreux, nous avons constaté que la vitesse a un effet très significatif sur la sortie, et la dispersion n'est pas applicable

Nous avons également noté que résistance de la couche semi-captive avait un impact plus important sur les sorties du modèle de la nappe semi captive, et que l'effet de la transmissivité était quasi inexistant.

Après examen des résultats du modèle de l'essai de pompage en régime transitoire, transmissivité le coefficient de stockage de l'aquifère avait la plus grande partie dans la variance totale, contrairement à le coefficient de stockage de l'aquifère, qui était presque négligeable.

Cette analyse permet de négliger des paramètres dont l'effet est négligeable en limitant leur valeur et en réduisant ainsi la dimensionnalité de l'espace des paramètres. Le modélisateur peut alors se concentrer sur les paramètres les plus pertinents et tenter de réduire l'incertitude.

Conclusion Générale

La réalisation de cette étude sur l'analyse de sensibilité globale des modèles hydrodynamiques pour l'estimation des paramètres des eaux souterraines a permis d'approfondir notre compréhension des facteurs affectant le comportement de la nappe. En analysant la sensibilité des paramètres dans cette thèse, la méthode d'analyse de sensibilité globale, basée sur la décomposition de la variance, est privilégiée car elle est indépendante de la nature du modèle. Dans ce cadre, l'effet des paramètres incertains sur la sortie du modèle est évalué à travers les indices de Sobol. Pour calculer ces indices, deux méthodes d'expansion du chaos polynomial et 2 sont adoptées

Le premier chapitre a souligné l'importance de l'analyse de sensibilité des modèles. Utilisation de méthodes telles que l'analyse de la variance et la méthode de dérivation

Dans le deuxième chapitre, nous nous sommes concentrés sur les indicateurs Sobol et leur calcul, en particulier les deux méthodes adoptées dans nos travaux, qui sont des outils largement utilisés pour estimer la sensibilité de modèles complexes à de nombreux paramètres. Les indicateurs Sobol fournissent des mesures numériques de l'importance relative de chaque paramètre, ainsi que des interactions entre les paramètres. En comprenant ces interactions, nous sommes en mesure de mieux orienter nos efforts d'estimation et de réduire l'incertitude associée aux paramètres du modèle.

Le troisième chapitre a appliqué les concepts et méthodes présentés dans les chapitres précédents. Nous avons appliqué l'analyse de sensibilité globale et les indices de Sobol à des problèmes réels d'hydraulique des eaux souterraines. Nous avons présenté 5 expériences :

L'étude menée sur l'analyse de sensibilité globale des modèles hydrodynamiques pour l'estimation des paramètres des eaux souterraines a permis d'approfondir notre compréhension des facteurs qui influencent le comportement de la nappe. En utilisant la méthode d'analyse de sensibilité globale basée sur la décomposition de la variance, nous avons privilégié une approche indépendante de la nature du modèle. Pour évaluer l'effet des paramètres incertains sur la sortie du modèle, nous avons utilisé les indices de Sobol. Deux méthodes, l'expansion du chaos polynomial et Fast, ont été adoptées pour calculer ces indices.

Le premier chapitre a souligné l'importance de l'analyse de sensibilité des modèles, en utilisant des méthodes telles que l'analyse de la variance et la méthode de dérivation.

Dans le deuxième chapitre, nous nous sommes concentrés sur les indicateurs de Sobol et leur calcul, en mettant en avant les deux méthodes que nous avons utilisées dans notre étude. Ces indicateurs sont largement utilisés pour estimer la sensibilité des modèles complexes à de nombreux paramètres. Les indices de Sobol fournissent des mesures numériques de l'importance relative de chaque paramètre, ainsi que des interactions entre les paramètres. En comprenant ces interactions, nous sommes en mesure de mieux orienter nos efforts d'estimation et de réduire l'incertitude associée aux paramètres du modèle.

Le troisième chapitre a appliqué les concepts et méthodes présentés précédemment à des problèmes réels d'hydraulique des eaux souterraines. Nous avons réalisé cinq expériences, en utilisant la méthode Fast pour analyser la sensibilité des modèles de nappe captive et de nappe libre. Cette méthode a donné de bons résultats, notamment après avoir augmenté le nombre d'échantillons, ce qui a permis d'obtenir une variance totale stable et une réduction nette de l'erreur.

Dans le modèle de nappe captive, nous avons constaté que la transmissivité avait un fort impact sur la sortie du modèle, tandis que la distance n'avait aucun effet. De plus, nous avons conclu qu'il n'y avait pas d'interaction entre la transmissivité et les autres variables.

Pour le modèle de nappe libre, les paramètres du débit de pompage et de la distance ont été classés comme non significatifs. En revanche, nous avons observé l'importance de la vitesse sur la sortie du modèle, et comment sa valeur se stabilisait rapidement.

La deuxième méthode que nous avons utilisée était pour analyser la sensibilité des modèles de transport de soluté en milieu poreux, de nappe semi-captive, et pour l'essai de pompage en régime transitoire.

Dans le modèle de transport de soluté en milieu poreux, nous avons observé que la vitesse avait un impact significatif sur le modèle, et nous devons prendre en compte cette donnée lors de l'estimation du paramètre. Cependant, le deuxième paramètre, la dispersion, n'est pas applicable.

Les résultats obtenus dans le modèle semi-captif de nappe mettent en évidence l'importance du paramètre résistance de la couche semi-perméable et son potentiel d'avoir un impact significatif sur la sortie du modèle. Il est donc essentiel de prendre des précautions lors de sa mesure et de son estimation. En revanche, le paramètre transmissivité peut être négligé et sa valeur peut être fixée dans un intervalle spécifique.

la transmissivité a contribué de manière significative à la variance globale du modèle de l'essai de pompage en régime transitoire, tandis que le coefficient de stockage de l'aquifère était pratiquement insignifiant.

Cette analyse nous a permis de déterminer dans quelle mesure chaque paramètre affecte les résultats du modèle. Cela permet d'identifier les principaux facteurs affectant les résultats auxquels il convient d'accorder une plus grande attention, d'améliorer leurs estimations et de réduire la certitude des résultats finaux. Cela nous permet également de comprendre les interactions entre les transactions. Cela aide à comprendre les relations complexes entre les variables, y compris prendre des décisions plus intelligentes.

En conclusion, cette étude sur l'analyse de sensibilité globale des modèles hydrodynamiques pour estimer les paramètres des eaux souterraines est un outil précieux pour améliorer notre compréhension de ces systèmes complexes. En identifiant les paramètres les plus influents et en évaluant leur impact sur les résultats du modèle, nous pouvons mieux orienter nos efforts d'estimation et de gestion. Les connaissances et les méthodes développées dans cette thèse peuvent être appliquées dans divers domaines de l'hydrologie et contribuer à une meilleure gestion des ressources en eaux souterraines pour assurer leur durabilité à long terme.

Références Bibliographiques

- Blatman. Chaos polynomial creux et adaptatif pour la propagation d'incertitudes et l'analyse de sensibilité. PhD thesis, Université Blaise Pascal - Clermont-Ferrand II, 2009.
- Borgonovo, 'A new uncertainty importance measure', *Reliability Engineering & System Safety*, vol. 92, no. 6, pp. 771–784, 2007. DOI: 10.1016/j.ress.2006.04.015.
- Borgonovo, E., & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248(3), 869-887..
- Brandstaeter, A. Gizzi, S. L. Fuchs, A. M. Gebauer, R. C. Aydin and C. J. Cyron 'Computational model of gastric motility with active-strain electromechanics', *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 98, no. 12, pp. 2177–2197, 2018. DOI: 10.1002/zamm.201800166.
- Brandstätter, S. (2021). *Global Sensitivity Analysis for Models of Active Biomechanical Systems* (Doctoral dissertation, Technische Universität München).
- Cacuci D. G. (2003). Sensitivity and Uncertainty Analysis, Volume I. Chapman and Hall/CRC, 2003. DOI: 10.1201/9780203498798.
- Cacuci D. G., Ionescu-Bujor M. and Navon I. N. (2005). Sensitivity and Uncertainty Analysis, Volume II. CRC Press, 2005. DOI: 10.1201/9780203483572.
- Cameron and WT Martin. The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals. *Annals of Mathematics*, 48(2) :385–392, 1947.
- Crestaux et al. Polynomial chaos expansion for sensitivity analysis. *Reliability engineering & System Safety*, 94(7) :1161–1172, 2009.
- Cukier R.I., J.H. Schaibly et K.E. Shuler, (1975): Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations. *Journal of chemical Physics*; 63: 1140-1149.
- Cukier R.I., R.I. Levine et K.E. Shuler, (1978) : Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics* ; 26 :1-42.
- Cukier, CM Fortuin, K.E. Shuler, AG Petschek, and JH. Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i theory. *The Journal of Chemical Physics*, 59(8) :3873, 1973.
- Deb, I.M. Babuka, and J.T. Oden. Solution of stochastic partial differential equations using Galerkin finite element techniques. *Computer Methods in Applied Mechanics and Engineering*, 190(48) :6359– 6372, 2001.
- Douglas-Smith, D., Iwanaga, T., Croke, B. F., & Jakeman, A. J. (2020). Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques. *Environmental Modelling & Software*, 124, 104588.
- Eck, V. G., Donders, W. P., Sturdy, J., Feinberg, J., Delhaas, T., Hellevik, L. R., & Huberts, W. (2016). A guide to uncertainty quantification and sensitivity analysis for cardiovascular applications. *International journal for numerical methods in biomedical engineering*, 32(8), e02755.

- Fajraoui, N. (2014). *Analyse de sensibilité globale et polynômes de chaos pour l'estimation des paramètres: Application aux transferts en milieu poreux* (Doctoral dissertation, Strasbourg).
- Faure, H. (1982). Discrépance de suites associées à un système de numération (en dimension s). *Acta arithmetica*, 41(4), 337-351.
- Ghanem and P.D. Spanos. *Stochastic finite elements : a spectral approach*. Dover Pubns, 2003.
- Ghanem, R. G., et Spanos, P. D., (1991): *Stochastic finite elements, a spectral approach*. Berlin: Springer.
- Ghanem, R., Higdon, D., & Owhadi, H. (Eds.). (2017). *Handbook of uncertainty quantification*. New York: Springer.
- Homma and A. Saltelli, 'Importance measures in global sensitivity analysis of nonlinear models', *Reliability Engineering and System Safety*, vol. 52, no. 1, pp. 1–17, 1996. DOI: 10.1016/0951-8320(96)00002-6.
- Homma and A. Saltelli, 'Importance measures in global sensitivity analysis of nonlinear models', *Reliability Engineering and System Safety*, vol. 52, no. 1, pp. 1–17, 1996. DOI: 10.1016/0951-8320(96)00002-6.
- Hwang, D., Byun, D. W., & Odman, M. T. (1997). An automatic differentiation technique for sensitivity analysis of numerical advection schemes in air quality models. *Atmospheric Environment*, 31(6), 879-888.
- Iooss and C. Prieur, 'Shapley effects for sensitivity analysis with correlated inputs: comparisons with Sobol' indices, numerical estimation and applications', *International Journal for Uncertainty Quantification*, vol. 9, no. 5, pp. 493–514, 2019. DOI: 10.1615 /Int.J.UncertaintyQuantification.2019028372.
- Iooss, B., & Lemaître, P. (2015). A review on global sensitivity analysis methods. *Uncertainty management in simulation-optimization of complex systems: algorithms and applications*, 101-122.
- Iooss, B., & Prieur, C. (2019). Shapley effects for sensitivity analysis with correlated inputs: comparisons with Sobol' indices, numerical estimation and applications. *International Journal for Uncertainty Quantification*, 9(5).
- Iooss, B., & Saltelli, A. (2017). Introduction to sensitivity analysis. In *Handbook of uncertainty quantification* (pp. 1103-1122). Springer, Cham.
- Kucherenko and S. Song, 'Different numerical estimators for main effect global sensitivity indices', *Reliability Engineering and System Safety*, vol. 165, no. February, pp. 222– 238, 2017. DOI: 10.1016/j.ress.2017.04.003.
- Kucherenko, D. Albrecht and A. Saltelli, 'Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques', arXiv e-prints, arXiv:1505.02350, 2015. [Online]. Available: <http://arxiv.org/abs/1505.02350>.
- Kucherenko, M. Rodriguez-Fernandez, C. Pantelides and N. Shah, 'Monte Carlo evaluation of derivative-based global sensitivity measures', *Reliability Engineering&System Safety*, vol. 94, no. 7, pp. 1135–1148, 2009. DOI: 10.1016/j.ress.2008.05.006.
- Kucherenko, O. V. Klymenko and N. Shah, 'Sobol' indices for problems defined in non-rectangular domains', *Reliability Engineering and System Safety*, vol. 167, pp. 218– 231, 2017. DOI: 10.1016/j.ress.2017.06.001.

- Liu, W. Chen and A. Sudjianto, '*Relative Entropy Based Method for Probabilistic Sensitivity Analysis in Engineering Design*', Journal of Mechanical Design, vol. 128, no. 2, p. 326, 2006. DOI: 10.1115/1.2159025.
- Lo Piano, F. Ferretti, A. Puy, D. Albrecht and A. Saltelli, '*Variance-based sensitivity analysis: The quest for better estimators and designs between explorativity and economy*', Reliability Engineering and System Safety, vol. 206, p. 107300, 2021. DOI: 10.1016/j.ress.2020.107300.
- Lu Y. et S. Mohanty, (2001) : Sensitivity analysis of a complex proposed geologic waste disposal system using the Fourier Amplitude Sensitivity method. Reliability Engineering and System Safety; 72 :275-291.
- Mara T. A. et S. Tarantola, (2008) : Application of global sensitivity analysis of model output to building thermal simulations. Journal of Building Simulation; 1 :290-302.
- Mara T. A., (2009): Extension of the rbd-fast method to the computation of global sensitivity indices. Reliability Engineering and System Safety; 94 :1274-1281.
- Marino S., I. B. Hogue, C. J. Ray et D. E Kirschner, (2008) : A methodology for performing global uncertainty and sensitivity analysis in systems biology. Journal of Theoretical Biology ; 254(1) :178-196.
- McKay, M. D., Beckman, R. J., & Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1), 55-61.
- Morris, '*Factorial Sampling Plans for Preliminary Computational Experiments*', Technometrics, vol. 33, no. 2, pp. 161–174, 1991. DOI: 10.1080/00401706.1991.10484804.
- Nouy et al. Generalized spectral decomposition for stochastic nonlinear problems. Journal of Computational Physics, 228(1) :202–235, 2009.
- Owen and C. Prieur, '*On Shapley Value for Measuring Importance of Dependent Inputs*', SIAM/ASA Journal on Uncertainty Quantification, vol. 5, no. 1, pp. 986–1002, 2017. DOI: 10.1137/16M1097717.
- Owen, '*Sobol' Indices and Shapley Value*', SIAM/ASA Journal on Uncertainty Quantification, vol. 2, no. 1, pp. 245–251, 2014. DOI: 10.1137/130936233.
- Park and K.-I. Ahn, '*A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment*', Reliability Engineering & System Safety, vol. 46, no. 3, pp. 253–261, 1994. DOI: 10.1016/0951-8320(94)90119-8.
- Pianosi and T. Wagener, '*Distribution-based sensitivity analysis from a generic inputoutput sample*', Environmental Modelling & Software, vol. 108, pp. 197–207, 2018. DOI: 10.1016/j.envsoft.2018.07.019.
- Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, 79, 214-232.
- Pianosi, K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson and T. Wagener, '*Sensitivity analysis of environmental models: A systematic review with practical workflow*', Environmental Modelling & Software, vol. 79, pp. 214–232, 2016. DOI: 10.1016/j.envsoft.2016.02.008.

- Qian, B. Peherstorfer, D. O'Malley, V. V. Vesselinov and K. Willcox, 'Multifidelity monte carlo estimation of variance and sensitivity indices', *SIAM-ASA Journal on Uncertainty Quantification*, vol. 6, no. 2, pp. 683–706, 2018. DOI: 10.1137/17M1151006.
- Qian, G., & Mahdi, A. (2020). Sensitivity analysis methods in the biomedical sciences. *Mathematical biosciences*, 323, 108306.
- Razavi and H. V. Gupta, 'A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory', *Water Resources Research*, vol. 52, no. 1, pp. 423–439, 2016. DOI: 10.1002/2015WR017558.
- Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., ... & Maier, H. R. (2021). The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, 137, 104954.
- Robert, C. P., Casella, G., & Casella, G. (1999). *Monte Carlo statistical methods*. New York: Springer.
- Rosen, R. (1991). *Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life*. Columbia University Press.
- Saltelli A., Tarantola S., Chan K., (1999) : A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*; 41:39-56.
- Saltelli and S. Tarantola, 'On the Relative Importance of Input Factors in Mathematical Models', *Journal of the American Statistical Association*, vol. 97, no. 459, pp. 702– 709, 2002. DOI: 10.1198/016214502388618447.
- Saltelli, 'Making best use of model evaluations to compute sensitivity indices', *Computer Physics Communications*, vol. 145, no. 2, pp. 280–297, 2002. DOI: 10.1016/S0010-4655(02)00280-1.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., & Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models* (Vol. 1). New York: Wiley.
- Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li and Q. Wu, 'Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices', *Environmental Modelling and Software*, vol. 114, pp. 29–39, 2019. DOI: 10.1016/j.envsoft.2019.01.012.
- Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola, *Global Sensitivity Analysis. The Primer*. Chichester, UK: John Wiley & Sons, Ltd, 2008, pp. 1–292. DOI: 10.1002/9780470725184.
- Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola, 'Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index', *Computer Physics Communications*, vol. 181, no. 2, pp. 259–270, 2010. DOI: 10.1016/j.cpc.2009.09.018.
- Santner, B. J. Williams and W. I. Notz, *The Design and Analysis of Computer Experiments*, ser. *Springer Series in Statistics*. New York, NY: Springer New York, 2018, pp. 1–436. DOI: 10.1007/978-1-4939-8847-1.

- Santner, B. J. Williams and W. I. Notz, *The Design and Analysis of Computer Experiments*, ser. Springer Series in Statistics. New York, NY: Springer New York, 2018, pp. 1–436. DOI: 10.1007/978-1-4939-8847-1.
- Santner, T. J., Williams, B. J., Notz, W. I., & Williams, B. J. (2003). *The design and analysis of computer experiments* (Vol. 1). New York: Springer.
- Sarrazin, F. Pianosi and T. Wagener, 'Global Sensitivity Analysis of environmental models: Convergence and validation', *Environmental Modelling and Software*, vol. 79, pp. 135–152, 2016. DOI: 10.1016/j.envsoft.2016.02.005.
- Sheikholeslami .R and S. Razavi, 'Progressive Latin Hypercube Sampling: An efficient approach for robust sampling-based analysis of environmental models', *Environmental Modelling & Software*, vol. 93, pp. 109–126, 2017. DOI: 10.1016/j.envsoft.2017.03.010.
- Sheikholeslami and S. Razavi, 'A Fresh Look at Variography: Measuring Dependence and Possible Sensitivities Across Geophysical Systems From Any Given Data', *Geophysical Research Letters*, vol. 47, no. 20, 2020. DOI: 10.1029/2020GL089829.
- Smith, R. C. (2013). *Uncertainty quantification: theory, implementation, and applications* (Vol. 12). Siam.
- Sobol, 'Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates', *Mathematics and Computers in Simulation*, vol. 55, no. 1-3, pp. 271–280, 2001. DOI: 10.1016/S0378-4754(00)00270-6.
- Sobol, 'On sensitivity estimation for nonlinear mathematical models', *Matematicheskoe Modelirovanie*, vol. 2, no. 1, pp. 112–118, 1990.
- Sobol', I. M., Asotsky, D., Kreinin, A., & Kucherenko, S. (2011). Construction and comparison of high-dimensional Sobol'generators. *Wilmott*, 2011(56), 64-79.
- Sobol, I. Y. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4), 784-802.
- Sobol' and S. Kucherenko, 'Derivative based global sensitivity measures and their link with global sensitivity indices', *Mathematics and Computers in Simulation*, vol. 79, no. 10, pp. 3009–3017, 2009. DOI: 10.1016/j.matcom.2009.01.023.
- Sobol' and S. Kucherenko, 'Derivative based global sensitivity measures and their link with global sensitivity indices', *Mathematics and Computers in Simulation*, vol. 79, no. 10, pp. 3009–3017, 2009. DOI: 10.1016/j.matcom.2009.01.023.
- Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7) :964–979, 2008.
- Sullivan, T. J. (2015). *Introduction to uncertainty quantification* (Vol. 63). Springer.
- Tarantola, S., D. Gatelli et T. A Mara, (2006): Random balance designs for the estimation of first-order sensitivity indices. *Reliability Engineering and System Safety* ; 91 :717-727.

- Tissot and C. Prieur. Variance-based sensitivity analysis using harmonic analysis. Submitted, 2012.
- Wentworth, M. T., Smith, R. C., & Banks, H. T. (2016). Parameter selection and verification techniques based on global sensitivity analysis illustrated for an HIV model. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1), 266-297.
- Weyl. Mean motion. *American Journal of Mathematics*, 60(4) :889–896, 1938.
- Wiener N., (1938): The homogeneous chaos. *American Journal of Mathematics*; 60:897-936.
- Woods and S. M. Lewis, '*Design of Experiments for Screening*', in *Handbook of Uncertainty Quantification*, R. Ghanem, H. Owhadi and D. Higdon, Eds., Cham: Springer International Publishing, 2017, ch. 33, pp. 1143–1185. DOI: 10.1007/978-3-319-12385-1_33.
- Xu C. et G. Z. Gertner., (2008): A general first-order global sensitivity analysis method. *Reliability Engineering and System Safety*, 93:1060-1071.
- Young, G. M. Hornberger and R. C. Spear, '*Modeling badly defined systems: some further thought*', in *Proceedings SIMSIG Conference*, Canberra, 1978, pp. 24–32.