



RÉPUBLIQUE DU BÉNIN

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE  
LA RECHERCHE SCIENTIFIQUE



UNIVERSITÉ NATIONALE DES SCIENCES, TECHNOLOGIES, INGÉNIERIE ET  
MATHÉMATIQUES (UNSTIM) D'ABOMEY

ECOLE NATIONALE SUPÉRIEURE DE GÉNIE MATHÉMATIQUE ET MODÉLISATION

## Mémoire

pour l'obtention du diplôme d'Ingénieur de conception

Spécialité : Mathématiques Appliquées et Modélisation

Option : Modélisation et Simulation Numérique (MSN)

---

# Comparaison des méthodes BME (Bayesian Maximum Entropy) et HASM(High Accuracy Surface Modeling) pour la modélisation en géostatistique

---

Présenté par :  
Christophe WAVOEKE

Sous la supervision de :  
Prof Eric ALAMOU  
*Enseignant-Chercheur à l'ENSTP/ UNSTIM d'Abomey*  
*Maître de Conférences des Universités du CAMES*

Sous l'encadrement de :  
**Dr AGBANGBA Codjo Emile**  
*Enseignant-Chercheur à EPAC/ UAC*  
*Maître-Assistant des Universités du CAMES*

**Soutenu à Abomey, le ..... Octobre 2024**

### Composition du jury

Président : **MA. AGBANGBA C. Emile**, Enseignant-chercheur à l'EPAC/UAC  
Membres :  
1- **Dr. ....**, Enseignant-chercheur à la FAST/UNSTIM  
2- **Dr. ....**, Enseignant-chercheur à l'ENSGMM

Année académique 2023 - 2024

**4-ère Promotion**

L'Ecole Nationale Supérieure de Génie Mathématique et Modélisation (ENSGMM) n'entend donner ni approbation ni improbation aux opinions émises dans ce document. Les raisonnements et arguments reviennent donc à l'auteur.

---

---

## APPROBATION

---

Nous certifions que le présent mémoire a été réalisé par l'auteur et est exempt de tout plagiat. Il est arrivé à terme et peut être soutenu devant un jury.

Abomey, le ..... 2024

### Signatures

Encadreur :

Superviseur :

**Dr (MA) AGBANGBA Emile C.**

**Dr (Prof) ALAMOU ERIC**

---

---

## DEDICACE

---

« *Dédicace* »

À Dieu le Père et le Fils et le Saint Esprit ;

À mon cher Papa, Iréné WAVOEKE ;

À ma chère maman, Madeleine ADJAVEMEDE ;

À mon frère, Miguel WAVOEKE ;

À mes sœurs, Carine WAVOEKE et Rose WAVOEKE ;

À tous mes camarades de la quatrième Promotion de l'INSPEI ou de l'ENSGMM.

---

---

## REMERCIEMENTS

---

Je souhaite avant tout exprimer ma profonde gratitude envers l'Éternel Dieu Tout-Puissant. Je tiens à adresser mes sincères remerciements :

- à Monsieur le recteur de l'UNSTIM, le Professeur **Joachim Djimon GBENOU**, pour sa vision pragmatique et son ouverture d'esprit;
- au Coordonnateur, le Docteur (MC) **MOUSSA DJIBRIL Aliou** , qui assure la coordination de l'ENSGMM;
- À mon encadrant, le Docteur **AGBANGBA Codjo Emile**, pour sa disponibilité, la qualité de son suivi presque quotidien, et pour avoir été bien plus qu'un simple encadrant.
- À mon tuteur de stage, le Docteur **ADETOLA Jamal** , pour sa disponibilité et ses sacrifices en faveur du bon déroulement des travaux de recherche.
- À mon superviseur, le professeur **ALAMOU Eric**, pour m'avoir offert les meilleures conditions pour mener à bien les travaux de recherche au Laboratoire;
- À mon aîné **Jacques AGON**, pour sa précieuse contribution et son soutien dans le travail;
- Aux **autorités et enseignants** de l'École Nationale Supérieure de Génie Mathématique et Modélisation (ENSGMM) pour la qualité de leur formation;
- Au Docteur (MC) **Christian AKOWANOU** , le Directeur de l'Institut National Supérieur des Classes Préparatoires aux Études d'Ingénieur (INSPEI), pour leur soutien à la première promotion;
- À mes **camarades** de la première promotion de l'ENSGMM ou de l'INSPEI pour leur solidarité;
- À mon père, **Iréné WAVOEKE**, qui représente pour moi un modèle de vie et une source de motivation;
- À ma mère, **Madeleine ADJAVEMEDE**, et à mes soeurs Rose , Carine pour leur amour qu'elles me témoignent;
- À mes frères **Miguel , Crépin , Achille et Roland**, pour leur soutien tout au long de ma vie;

---

---

## LISTE DES ABRÉVIATIONS

---

<b>BME</b>	<b>B</b> ayesian <b>M</b> aximum <b>E</b> ntropy
<b>DF</b>	<b>D</b> ifférences <b>F</b> inies
<b>EDP</b>	<b>E</b> quation aux <b>D</b> érivées <b>P</b> artielles
<b>ENSGMM</b>	<b>E</b> cole <b>N</b> ationale <b>S</b> upérieure de <b>G</b> énie <b>M</b> athématique et <b>M</b> odélisation
<b>HASM</b>	<b>H</b> igh <b>A</b> ccuracy <b>S</b> urface <b>M</b> odeling
<b>IB</b>	<b>I</b> nférence <b>B</b> ayesian
<b>INSPEI</b>	<b>I</b> nstitut <b>N</b> ational <b>S</b> upérieur des <b>C</b> lasses <b>P</b> réparatoires aux <b>E</b> tudes d'Ingénieur
<b>MAE</b>	<b>M</b> ean <b>A</b> bsolute <b>E</b> rror
<b>PDF</b>	<b>P</b> robability <b>D</b> ensity <b>F</b> onction
<b>PERMANOVA</b>	<b>P</b> ermutational <b>M</b> ultivariate <b>A</b> Analysis <b>O</b> f <b>V</b> ariance
<b>ME</b>	<b>M</b> aximum <b>E</b> ntropy
<b>RMSE</b>	<b>R</b> oot <b>M</b> ean <b>S</b> quare <b>E</b> rror
<b>SD</b>	<b>S</b> patial <b>D</b> ependence
<b>UNSTIM</b>	<b>U</b> niversité <b>N</b> ationale des <b>S</b> ciences, <b>T</b> echnologies, <b>I</b> ngénierie et <b>M</b> athématiques

---



---

## LISTE DES TABLEAUX

---

3.1	Classification des IDS pour les modèles par (SEIDEL et OLIVEIRA, 2016) . . .	32
3.2	Paramètres du variogramme pour la dépendance spatiale faible en fonction de la taille de l'échantillon . . . . .	33
3.3	Paramètres du variogramme pour la dépendance spatiale modérée en fonction de la taille de l'échantillon . . . . .	33
3.4	Paramètres du variogramme pour la dépendance spatiale forte en fonction de la taille de l'échantillon . . . . .	34
3.5	Paramètres pour le calcul du BME . . . . .	35
3.6	Paramètres pour le calcul de la méthode HASM . . . . .	36
4.1	Average value of MAE and RMSE determined from BME . . . . .	41
4.2	Average value of MAE and RMSE determined from HASM . . . . .	42
4.3	PERMANOVA results . . . . .	43
4.4	Effet de la taille et niveau de dépendance sur la prédiction de BME : PERMANOVA results . . . . .	44
4.5	Effet de la taille et niveau de dépendance sur la prédiction de HASM : PERMANOVA results . . . . .	45

---



---

## TABLE DES FIGURES

---

2.1	BME scheme . . . . .	8
2.2	Maillage domaine . . . . .	20
3.1	Illustration des Coefficients d'Asymétrie ( symétrique, asymétrie positive et asymétrie négative ). . . . .	31
3.2	Simulation Flowchart . . . . .	39
4.1	Boxplots de MAE_BME des niveaux de dépendances en fonction des tailles	44
4.2	Boxplots de MAE_HASM des niveaux de dépendances en fonction des tailles	45
4.3	Boxplots de MAE_HASM des niveaux de dépendances en fonction des tailles	46
4.4	Boxplots de MAE_HASM des niveaux de dépendances en fonction des tailles	47
4.5	Boxplots de la donnée suivant le SD et le skew . . . . .	54
4.6	Plot 3D d'un attribut de la donnée . . . . .	55
4.7	Plot 3D de l'attribut prédit . . . . .	56
4.8	Plot scatter d'un attribut de la donnée . . . . .	57
4.9	Plot scatter de l'attribut prédit . . . . .	58



---

---

## RÉSUMÉ

---

Cette étude se concentre sur la comparaison des méthodes BME et HASM pour la modélisation en géostatistique. La méthode BME, fondée sur l'entropie maximale bayésienne, se distingue par sa capacité à intégrer des données et à gérer l'incertitude de manière robuste. En comparaison, HASM est conçu pour offrir une haute précision dans la modélisation des surfaces grâce à des techniques avancées et à une approche de modélisation améliorée. À travers une analyse comparative, nous avons évalué la performance de ces deux méthodes en utilisant des métriques telles que l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne (MAE). Les résultats montrent que le BME est significatif en termes d'erreur et de cohérence avec les données réelles, tout comme le HASM, qui offre une bonne prédiction des attributs. Cependant, les résultats indiquent que le HASM, avec un MAE de 1.139815, est légèrement supérieur au BME, qui présente un MAE de 1.182439, en termes de critère d'erreur. L'étude fournit des exemples concrets de l'application de BME et HASM avec des données simulées suivant divers paramètres (taille de données, degré d'asymétrie et niveau de dépendance spatiale), démontrant comment ces méthodes modernes peuvent améliorer la qualité de la modélisation géostatistique. Cette comparaison met en lumière les avantages et les limites de chaque méthode.

**Keywords :** géostatistique , significatif , attributs , critère d'erreur

---

---

## ABSTRACT

---

This study focuses on the comparison of the BME and HASM methods for geostatistical modeling. The BME method, based on Bayesian maximum entropy, stands out for its ability to integrate data and manage uncertainty robustly. In comparison, HASM is designed to provide high accuracy in surface modeling through advanced techniques and an improved modeling approach. Through a comparative analysis, we evaluated the performance of these two methods using metrics such as root mean square error (RMSE) and mean absolute error (MAE). The results indicate that BME is significant in terms of error and consistency with actual data, as is HASM, which provides good predictions of attributes. However, the results show that HASM, with a MAE of 1.139815, is slightly superior to BME, which has a MAE of 1.182439, in terms of error criteria. The study provides concrete examples of the application of BME and HASM with simulated data following various parameters (data size, degree of asymmetry, and level of spatial dependence), demonstrating how these modern methods can enhance the quality of geostatistical modeling. This comparison highlights the advantages and limitations of each method.

**Keywords :** geostatistical modeling , significant , attributes ,error criteria

# Table des matières

<b>DEDICACE</b>	<b>iii</b>
<b>Remerciements</b>	<b>iv</b>
<b>Résumé</b>	<b>viii</b>
<b>Abstract</b>	<b>ix</b>
<b>1 Introduction Générale</b>	<b>1</b>
1.1 Contexte et justification de l'étude . . . . .	1
1.2 But de la recherche . . . . .	2
1.3 Problème et questions de recherche . . . . .	2
1.4 Objectifs du mémoire . . . . .	3
1.5 Limites de l'étude . . . . .	3
1.6 Résultats attendus de l'étude et cibles . . . . .	4
1.7 Nouveauté de l'étude . . . . .	4
1.8 Grandes divisions du mémoire . . . . .	5
<b>2 Cadre Théorique et Conceptuel</b>	<b>6</b>
2.1 Revue de littérature . . . . .	6
2.2 La méthode de BME (Bayesian Maximum Entropy) . . . . .	8
2.3 La méthode de HASM (High Accuracy Surface Modeling) . . . . .	16
<b>3 Matériel et méthodes</b>	<b>29</b>
3.1 Matériel / Outils . . . . .	29
3.2 Méthodes . . . . .	29
<b>4 Résultats et Discussions</b>	<b>40</b>
4.1 Résultats des analyses . . . . .	40
4.2 Discussion . . . . .	47

---

## INTRODUCTION GÉNÉRALE

---

### 1.1 Contexte et justification de l'étude

La modélisation et l'interpolation spatiale ou spatio-temporelle jouent un rôle important dans plusieurs domaines de la science de l'environnement, incluant la géographie (SOURIS, 2019), la géologie (ABDENNOUR, 2021); (RATA, 2010), la biologie (LI, 2015); (GUEROUI, 2014), la météorologie (LEBEL et al., 1996); (LOUVET et al., 2011); (COMAN, 2008) et la conception de réseaux de capteurs (DAHLSTRÖM, 2003); (POUPRY, 2023).

Par exemple, en géographie, (SOURIS, 2019) présente les objectifs, principes, méthodes et outils de l'analyse spatiale et des systèmes d'information géographique (SIG) dans le contexte de la santé, fournissant une introduction pratique à l'analyse spatiale et spatio-temporelle en épidémiologie et géographie de la santé. En géologie, (ABDENNOUR, 2021) examine la variabilité spatio-temporelle de la salinisation des sols dans le périmètre irrigué du Ziban en Algérie en utilisant des techniques géostatistiques et de télédétection pour analyser la dynamique de la salinisation. En biologie, (GUEROUI, 2014) analyse la qualité physico-chimique et microbiologique des eaux souterraines de l'aquifère superficiel de la plaine de Tamoulouka en Algérie, en identifiant les paramètres chimiques et bactériologiques, les sources de pollution, et propose des mesures de gestion pour la protection des ressources en eau souterraine. En météorologie, (LEBEL et al., 1996) se concentrent sur la distribution spatio-temporelle des précipitations au Sahel en Afrique de l'Ouest en utilisant des données de l'expérience EPSAT-Niger. Dans le domaine des réseaux de capteurs, (POUPRY, 2023) propose un système de surveillance de la qualité de l'air pour les vallées des Gaves, dans une thèse de doctorat à l'Institut National Polytechnique de Toulouse.

La géostatistique étant une branche des statistiques appliquées aux sciences de la Terre, qui se concentre sur l'analyse et la modélisation des données spatiales. Elle n'est donc pas épargnée dans l'utilisation des méthodes de prédiction spatiale. La précision des modèles géostatistiques est déterminante pour de nombreuses applications, allant de la gestion des ressources naturelles à la planification urbaine. Dans ce contexte, le choix des méthodes de modélisation joue un rôle fondamental dans la qualité des prédictions spatiales et la compréhension des phénomènes étudiés.

Ces domaines utilisent des méthodes avancées de modélisation et d'interpolation pour

comprendre et prédire des phénomènes considérés comme des variables continues en tout point de l'espace, au-delà des valeurs échantillonnées. Les méthodes classiques de géostatistique pour la modélisation et l'interpolation incluent la Trend Surface Analysis (TSA), qui ajuste une surface de tendance par régression sur les coordonnées spatiales; l'Inverse Distance Weighted (IDW), qui utilise une fonction de pondération inverse à la distance; le Triangulated Irregular Network (TIN), qui calcule les valeurs à l'intérieur de triangles par une fonction linéaire; le Kriging, qui minimise le résidu moyen et la variance des erreurs; et les Splines, qui utilisent des splines-bases cubiques univariées pour simuler des surfaces.

Les nouvelles méthodes, telles que le Bayesian Maximum Entropy (BME) et le High Accuracy Surface Modeling (HASM), ont révolutionné ces techniques classiques grâce à leur fiabilité, performance et précision accrues dans l'estimation et la prédiction des variables en points discrétisés. De nombreuses études ont montré la supériorité des méthodes modernes par rapport aux techniques classiques de modélisation.

À travers cette étude comparative, nous cherchons à identifier les avantages et les faiblesses de chaque approche, à comprendre comment elles se comportent sous différentes conditions et à fournir des recommandations basées sur des résultats empiriques. Cette analyse approfondie est essentielle pour optimiser les pratiques de modélisation en géostatistique et améliorer la précision des prévisions dans divers domaines d'application.

## 1.2 But de la recherche

Le choix et l'utilisation des méthodes d'interpolation adéquates sont essentiels pour obtenir des résultats précis et pertinents dans la modélisation spatiale. Pour répondre à ces défis, il est nécessaire d'évaluer et de comparer les différentes techniques disponibles afin d'optimiser la qualité des modèles géostatistiques.

Ce mémoire a pour but de comparer les méthodes Bayesian Maximum Entropy (BME) et High Accuracy Surface Modeling (HASM) dans le domaine de la modélisation géostatistique. En examinant la performance de ces méthodes pour prédire des variables spatiales, cette recherche vise à déterminer laquelle est la plus efficace et la mieux adaptée aux divers contextes d'application. L'étude entend enrichir les connaissances académiques sur les méthodes géostatistiques tout en proposant des recommandations pratiques pour améliorer la précision et la fiabilité des modèles dans divers domaines d'application.

## 1.3 Problème et questions de recherche

Ne serait-il pas appréciable, dans le contexte agricole du Bénin, de mettre au point des techniques d'interpolation géostatistique capables non seulement de prédire avec précision la répartition des nutriments et la concentration des composés chimiques dans les sols, mais aussi d'évaluer leur fertilité de manière efficace? Notre recherche vise à comparer les méthodes avancées telles que Bayesian Maximum Entropy (BME) et High Accuracy Surface Modeling (HASM) pour identifier celles qui offrent les meilleures performances et sont les mieux adaptées aux conditions spécifiques des terres agricoles béninoises.

Nous proposons de répondre comme pour exemple à la question suivante : Comment utiliser les techniques d'interpolation pour obtenir des prévisions précises sur les

nutriments des engrais et la fertilité des sols, tout en tenant compte des particularités locales? En fournissant des recommandations pratiques pour l'application de ces méthodes, cette étude cherche à améliorer la gestion des sols et à promouvoir une agriculture plus productive et durable au Bénin.

## 1.4 Objectifs du mémoire

### 1.4.1 Objectif général

Bien que les méthodes de modélisation en géostatistique aient fait l'objet de nombreuses études, aucune recherche scientifique n'a encore comparé directement les deux techniques modernes, à savoir Bayesian Maximum Entropy (BME) et High Accuracy Surface Modeling (HASM). Notre défi et objectif est donc de réaliser une étude comparative approfondie de ces deux méthodes, dans le but de contribuer à la recherche scientifique et d'explorer les bénéfices pratiques de ces approches.

### 1.4.2 Objectifs spécifiques

Les objectifs spécifiques de cette recherche sont :

**Simuler des ensembles de données variés :** Générer des données en variant plusieurs paramètres clés, tels que la taille des données, le degré d'asymétrie, et le niveau de dépendance spatiale. Cela permettra de créer des bases de données diversifiées qui reflètent différentes conditions possibles.

**Appliquer BME et HASM sur les données simulées :** Appliquer les méthodes Bayesian Maximum Entropy (BME) et High Accuracy Surface Modeling (HASM) sur les données simulées pour analyser leur performance en termes de robustesse et de précision dans des contextes variés.

**Comparer la performance des méthodes :** Mesurer et comparer les performances des méthodes en utilisant des critères d'erreur spécifiques. Cette comparaison visera à identifier la méthode offrant les meilleures performances globales pour la modélisation géostatistique.

**Analyser les avantages et les limites :** Examiner les avantages et les limites de chaque méthode dans différents scénarios de simulation. Cette analyse inclura une évaluation de la facilité d'application, de la précision des prédictions, et de la capacité à traiter des données diverses.

**Formuler des recommandations :** Proposer des recommandations basées sur les résultats de l'étude comparative pour guider les praticiens et les chercheurs dans le choix de la méthode la plus appropriée pour leurs besoins spécifiques en géostatistique.

## 1.5 Limites de l'étude

Les limites de l'étude incluent plusieurs aspects vitaux :

1. **Limitation des informations :** La comparaison des méthodes BME et HASM nécessite leur application sur les mêmes bases de données. Cependant, la méthode BME

intègre non seulement les données dures (hard data), mais aussi les données souples (soft data). En revanche, la méthode HASM se limite uniquement aux données dures. Cette différence peut influencer les résultats de la comparaison.

2. **Limitation de dimension** : Le BME est adapté pour des modélisations spatiales et spatio-temporelles, tandis que HASM est une méthode déterministe qui ne prend pas en compte la dimension temporelle. Cela limite notre étude à une modélisation spatiale uniquement, sans considérer les aspects spatio-temporels.
3. **Nécessité des conditions aux bords** : Les contraintes de prédiction de HASM nécessitent des conditions aux bords du domaine de surface, ce qui est moins pertinent pour le BME. Cette contrainte impacte la base de données utilisée et la manière dont les prédictions sont effectuées.
4. **Régularité du pas de discrétisation** : La méthode HASM est souvent basée sur la résolution d'équations différentielles partielles (EDP) en utilisant des méthodes de différences finies pour la discrétisation du domaine. Cette approche impose l'utilisation de domaines rectangulaires ou carrés avec des pas de discrétisation réguliers, ce qui peut restreindre la flexibilité dans la modélisation par rapport à des domaines plus irréguliers.

## 1.6 Résultats attendus de l'étude et cibles

Les résultats attendus de l'étude incluent :

1. **Validation de la robustesse et de la capacité de prédiction de la méthode BME** : La méthode BME ayant déjà démontré son efficacité, elle sera confirmée par les résultats de cette étude.
2. **Évaluation de la performance de la méthode HASM** : Étant une technique nouvelle, il est espéré que la méthode HASM surpassera le BME en termes de performance. L'étude vise à déterminer si HASM apporte une amélioration significative à la recherche scientifique en comparaison avec BME.

**Participants et bénéficiaires de l'étude :**

- ✓ **Participants** : Les chercheurs en géostatistique et en modélisation spatiale qui souhaitent comparer et comprendre les performances des méthodes BME et HASM.
- ✓ **Bénéficiaires** : La communauté scientifique et les praticiens dans les domaines de la géostatistique, de l'analyse spatiale, et des applications liées, qui pourront bénéficier de nouvelles connaissances et de techniques améliorées pour la modélisation et l'interpolation spatiale.

## 1.7 Nouveauté de l'étude

Malgré les diverses études sur les méthodes de modélisation en géostatistique, aucune recherche scientifique n'a encore comparé directement les deux méthodes modernes, à savoir BME et HASM. L'objectif de ce document est de fournir une clarification approfondie sur la comparaison entre ces deux approches.

## 1.8 Grandes divisions du mémoire

La suite du document sera structurée de la manière suivante :

- **Cadre Théorique et Conceptuel** : Revue de la littérature sur les méthodes BME et HASM.
- **Matériel et Méthodes** : Outils de programmation et langages utilisés, ainsi que les différentes méthodes de simulation des données.
- **Résultats et Discussions** : Prédiction et comparaison des deux méthodes.
- **Conclusion et Perspectives** : Synthèse des résultats et orientations futures.



---

## CADRE THÉORIQUE ET CONCEPTUEL

---

### 2.1 Revue de littérature

Ces dernières années, les méthodes BME et HASM ont été appliquées pour la modélisation, prédiction des données en géostatistique. L'estimation des polluants atmosphériques, tels que l'exposition aux particules fines et à l'ozone troposphérique, a été étudiée par RAMOS (2017). Par ailleurs, BAYAT, NASSERI et NASER (2014) ont travaillé sur l'estimation des précipitations dans un grand bassin. La spatialisation de la teneur en eau du sol a été explorée par HENINE et al. (2015), tandis que GONGNET et al. (2022) ont prédit la distribution spatiale de la matière organique du sol. En outre, RABOULI et al. (2021) ont étudié la spatialisation de la conductivité hydraulique, et VIGNAUX et al. (1998) ont cartographié la distribution des poissons à petite échelle. (DELANG et al., 2021) propose une méthode permettant d'estimer les concentrations d'ozone à une résolution mondiale élevée pour la période 1990-2017, en combinant les données d'observation avec des modèles atmosphériques en utilisant les approches M3Fusion et Bayesian Maximum Entropy (BME). L'étude de (MODIS, VATALIS et SACHANIDIS, 2013) démontre l'application de l'approche Bayesian Maximum Entropy (BME) pour l'évaluation et la cartographie du risque de pollution des sols par des métaux lourds dans une région minière de lignite en Grèce, intégrant des données spatio-temporelles pour une meilleure gestion de la réhabilitation environnementale. (PUANGTHONGTHUB, 2006) décrit une modélisation spatio-temporelle de la distribution des particules PM10 en Thaïlande, utilisant la méthode Bayesian Maximum Entropy (BME) dans le but de caractériser cette distribution, d'identifier les zones non conformes aux normes, et d'optimiser le réseau de mesure. (RAMOS, 2017) explore la pollution atmosphérique liée aux PM2.5 et à l'ozone, en cherchant à évaluer de nouvelles approches géostatistiques telles que le krigeage avec dérive externe (KED) et le modèle bayésien d'entropie maximale (BME) pour une meilleure estimation de leur distribution spatio-temporelle à Montréal. (YUE, CHEN et LI, 2010) illustre l'application réussie de la méthode HASM adaptative (HASM-AM) pour la simulation de la surface d'élévation de la table de Dongzhi en Chine, sur une résolution de 10m x 10m, en calculant uniquement 870 000 mailles grâce à des raffinements locaux, comparé à 27,24 millions pour le HASM non adaptatif. (ZHAO et al., 2014) met en évidence l'utilisation réussie de HASM pour la modélisation de surface des modèles numériques de terrain et pour la modélisation des écosystèmes terrestres, améliorant ainsi la précision de

l'interpolation. (YUE et al., 2013) rapporte l'application de la méthode multi-grille HASM (HASM-MG) pour la modélisation spatiale des précipitations annuelles moyennes en Chine entre 1971 et 2000, permettant de réduire les temps de calcul par rapport à HASM-PCG, tout en maintenant une précision acceptable, en particulier dans les zones planes. (YUE et al., 2015) résume l'application de la méthode HASM dans divers domaines, tels que la modélisation numérique de terrain, le remplissage de données manquantes SRTM, la simulation du changement climatique, ainsi que la modélisation des sols et des propriétés écosystémiques, offrant des résultats plus précis que les méthodes conventionnelles. De nombreuses études ont montré la supériorité des méthodes modernes par rapport aux techniques classiques de modélisation. (GAO et al., 2022) ont démontré que le BME surpassait le Kriging Ordinaire (OK) dans l'interpolation des données de température de surface de la mer (SST), offrant le biais moyen le plus faible et le coefficient de corrélation le plus élevé. (CHRISTAKOS et LI, 1998) affirme que le BME est une approche plus rigoureuse que le Kriging pour intégrer les données "soft" et ne fait aucune hypothèse sur la linéarité, la normalité des lois de probabilité ou l'homogénéité spatiale, contrairement au Kriging. (ADAM-POUPART et al., 2014) ont révélé que le modèle BME combiné au Land-Use Regression (LUR) était le modèle prédictif le plus performant avec le plus faible RMSE, suivi par le modèle LUR et le modèle Kriging BME. Les résultats de (BAYAT, NASSERI et NASER, 2014) ont validé la supériorité du BME sur le Kriging Ordinaire par validation croisée. (KANG, CHEN et JIA, 2024) ont montré que le BME corrigeait les prédictions erronées pour les tremblements de terre en Japon et aux États-Unis, offrant une meilleure cohérence avec les résultats de ShakeMap comparée à l>IDW et au Kriging, et surmontait les erreurs de sous-estimation lors des faibles magnitudes.

(KARAMI et al., 2019) ont démontré que le BME et le Kriging pouvaient estimer la profondeur d'affouillement de manière plus précise que l>IDW. Les recherches de Yue (2011) et (ZHAO et YUE, 2021) ont montré que les méthodes HASM surpassaient les approches classiques (TSA, IDW, TIN, Kriging, Splines) en termes de MAE, MRE et RMSE. Les tests numériques ont montré que HASM présentait une erreur quadratique moyenne nettement inférieure à celle des méthodes classiques et résolvait les problèmes d'erreur rencontrés avec les modèles numériques de terrain. (CHEN, LI et YUE, 2013) ont constaté que HASM-SA était généralement plus précis et rapide que les méthodes courantes d'interpolation, telles que l>IDW, le Kriging et les différentes versions des Splines (RSpline, TPS, ANUDEM). (SHI et al., 2011) ont révélé que le HASM combiné avec le Land-Use (HASM\_LU) avait une performance supérieure à celle du HASM seul, du Kriging Ordinaire avec Land Use, du Kriging Stratifié et du Regression-Kriging avec un Generalized Linear Model, avec un biais d'estimation plus faible et une meilleure efficacité prédictive.

Enfin, (VIHOTOGBÉ et al., 2022) a examiné l'impact de la taille de l'échantillon, de l'asymétrie et de la dépendance spatiale sur la robustesse du BME, en utilisant des données avec différents niveaux d'asymétrie et de dépendance spatiale. (GONGNET et al., 2023) ont étudié les effets des caractéristiques des données sur la performance du BME. (WANG et al., 2023) ont présenté un cadre BME pour les modèles probabilistes, et (KANG et al., 2021) ont abordé l'utilisation du BME pour des problèmes d'analyse spatiale spécifiques, en tenant compte des caractéristiques des données. Cette recherche intégrera ces paramètres pour comparer les méthodes BME et HASM de manière exhaustive.

## 2.2 La méthode de BME (Bayesian Maximum Entropy)

La théorie du Bayesian Maximum Entropy (BME) en géostatistique spatiotemporelle se concentre sur la modélisation, l'estimation et la cartographie des attributs naturels tels que les variables physiques, les processus écologiques, les paramètres de santé et les indicateurs sociaux (CHRISTAKOS et LI, 1998).

La méthode BME est une méthode d'interpolation non linéaire précise qui combine la théorie de l'incertitude de l'entropie maximale avec l'inférence bayésienne. Elle se divise en trois étapes :

1. **Prior Stage** : Cette étape consiste à déduire la Fonction de Densité de Probabilité (PDF) à priori à partir des connaissances générales (GK) de la variable physique. Cette distribution à priori est censée suivre une distribution gaussienne. Cette étape correspond à la Maximisation de l'Entropie (ME) dans le but de trouver une PDF à priori.
2. **Meta Prior Stage** : Les données, c'est-à-dire les données dures (hard) et les données molles (soft), sont collectées, arrangées et éventuellement transformées pour fournir des informations spécifiques afin de déterminer la PDF à posteriori.
3. **Posterior Stage** : Cette étape, correspondant à l'Inférence Bayésienne (IB), met à jour la PDF à priori à partir des informations obtenues lors du Meta Prior Stage.

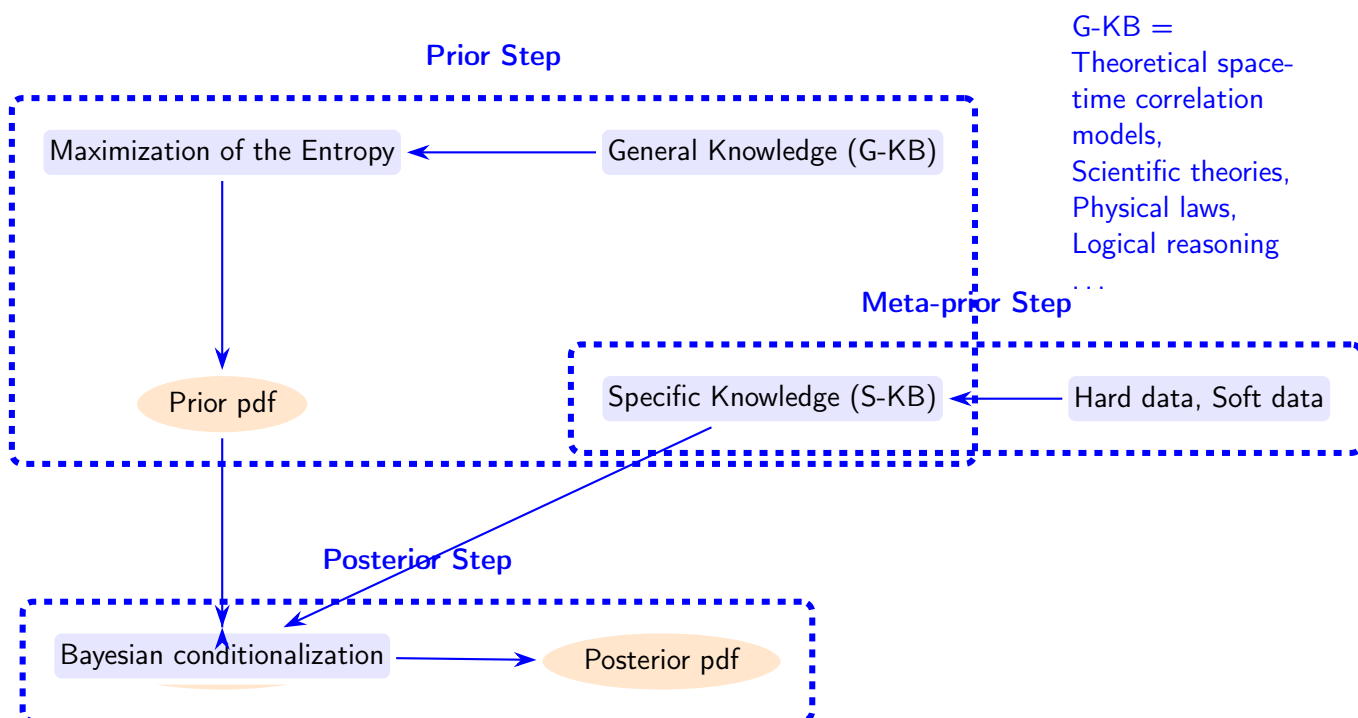


FIGURE 2.1 – BME scheme

### 2.2.1 Maximum Entropy

#### Généralité sur ME

La partie entropie maximale de la méthode BME a pour objectif de quantifier l'incertitude maximale des données observées. Elle cherche à identifier, parmi toutes les distributions de probabilités pouvant représenter ces données, celle qui a la plus grande entropie au sens de Shannon HE et CHRISTAKOS, 2023 ; SOIZE, 2008. Cette distribution d'entropie maximale est déterminée en respectant les contraintes imposées par les caractéristiques des données, telles que leur moyenne ou leur variance et d'autres moments d'ordre supérieur, mais sans incorporer d'autres informations a priori (CHRISTAKOS et LI, 1998). Il s'agit donc de la distribution la moins informative possible tout en étant compatible avec les observations. En choisissant la distribution ayant la plus forte entropie parmi celles satisfaisant aux contraintes des données, on sélectionne celle contenant la quantité d'information ajoutée de manière arbitraire la plus faible. Cette distribution d'entropie maximale, qui modélise l'incertitude inhérente au phénomène observé de manière neutre et objective, sera ensuite utilisée comme probabilité a priori dans l'étape d'inférence bayésienne.

D'après ((GUIASU et SHENITZER, 1985 ; CHRISTAKOS et LI, 1998 ; HE et CHRISTAKOS, 2023)), on a :

Pour une distribution de probabilité discrète  $p$  sur l'ensemble dénombrable  $\{x_1, x_2, \dots\}$ , avec  $p_i = p(x_i)$ , l'entropie de  $p$  est définie comme suit :

$$h(p) = - \sum_{i \geq 1} p_i \ln(p_i) \quad (2.1)$$

Pour une fonction de densité de probabilité continue  $p(x)$  sur un intervalle  $I$ , son entropie est définie comme suit :

$$h(p) = - \int_I p(x) \ln(p(x)) dx \quad (2.2)$$

Au cas où  $p = 0$ , on pose  $0 \cdot \ln(0) = 0$ .

Cette définition de l'entropie, introduite par Shannon, est considérée dans notre contexte probabiliste comme une mesure de l'information portée par  $p$ . Une entropie plus élevée correspond à moins d'information (plus d'incertitude, ou plus de manque d'information).

Le concept d'entropie maximale est utilisé dans divers domaines, tels que la reconnaissance de formes, la thermodynamique, la théorie de l'information, les statistiques et les séries temporelles (KULLBACK, 1968 ; BURG, 1972 ; SHORE et JOHNSON, 1980 ; JAYNES, 1982).

Considérons un ensemble fini  $\{x_1, x_2, \dots\}$ . Si  $p(x_1) = 1$  tandis que  $p(x_j) = 0$  pour  $j > 1$ , alors :

$$h(p) = -1 \cdot \ln(1) = 0$$

Dans ce cas, les statistiques gouvernées par  $p$  produisent presque sûrement un seul résultat possible,  $x_1$ . Nous avons une connaissance complète de ce qui va se passer. En revanche, si  $p$  est la fonction de densité uniforme, ou  $p(x_j) = \frac{1}{n}$  pour tous les  $j$ , alors :

$$h(p) = \ln(n)$$

L'entropie de la densité gaussienne sur  $\mathbb{R}$  avec une moyenne  $\mu$  et une variance  $\sigma^2$  est :

$$-\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right) dx = \frac{1}{2} (1 + \ln(2\pi\sigma^2)) \quad (2.3)$$

La moyenne  $\mu$  n'entre pas dans la formule finale, donc toutes les gaussiennes avec un  $\sigma$  commun ont la même entropie.

L'entropie de la densité exponentielle sur  $(0, \infty)$  avec une moyenne  $\lambda$  est :

$$-\int_0^\infty \frac{1}{\lambda} e^{-x/\lambda} \left(-\ln(\lambda) - \frac{x}{\lambda}\right) dx = 1 + \ln(\lambda) \quad (2.4)$$

Cette entropie devient négative pour de petites valeurs de  $\lambda$ .

Notre défi réside plutôt dans le processus contraire, en utilisant la maximisation de l'entropie pour trouver la fonction de densité de probabilité (PDF). Soit  $X(s)$  un champ aléatoire représentant une quantité physique distribuée spatialement, et supposons que les valeurs observées  $x_i$  de cette quantité soient disponibles aux points  $s_i$ ,  $i = 1, 2, \dots, m$ , dans l'espace. Soit  $\hat{X}(s_k)$  l'estimateur de  $X(s_k)$  en un point  $s_k$ ,  $k \neq i$ , où aucune observation n'est disponible.

Soit  $\chi_{map} = (\chi_{hard}, \chi_{soft}, s_k)$  où les  $\chi_{hard}$  sont relatifs aux données quantitatives ponctuelles et les  $\chi_{soft}$  sont relatifs aux données quantitatives groupées en intervalles.

$$\chi_{hard} = (s_1, s_2, \dots, s_m)$$

Autrement, on peut écrire :  $\chi_{map} = (\chi_{data}, s_k)$  avec  $\chi_{data} = (\chi_{hard}, \chi_{soft})$ , les données disponibles de Méta Prior Stage, c'est-à-dire les données hard et soft.

Donc,

$$\chi_{map} = (s_1, s_2, \dots, s_m, \chi_{soft}, s_k)$$

Désignons par  $T_G(\chi_{map})$  la fonction de densité de probabilité conjointe des variables aléatoires associées  $X(s_1), X(s_2), \dots, X(s_k)$  avant l'observation des données  $x_i$ .

Comme  $T_G(\chi_{map})$  est la fonction de densité de probabilité, on a alors par définition :

$$\int T_G(\chi_{map}) d\chi_{map} = 1 \quad (2.5)$$

Étant donné que  $\chi_{map}$  est l'ensemble des variables aléatoires,  $\chi_{map} = (s_1, s_2, \dots, s_m, \chi_{soft}, s_k)$ , donc l'intégrale (2.5) est une imbrication de plusieurs intégrales suivant chaque composante de  $\chi_{map}$ . Ces intégrations imbriquées sont effectuées sur les intervalles des variables aléatoires mentionnées ci-dessus. Ces intervalles seront supposés varier de  $-\infty$  à  $+\infty$ . L'équation (5) est la contrainte fondamentale de normalisation.

La densité de probabilité a priori  $T_G(\chi_{map})$  doit être dérivée au moyen d'un processus d'estimation prenant en compte des contraintes physiques. Ces contraintes représentent soit les connaissances a priori que l'on possède sur la grandeur physique à estimer, soit les propriétés spécifiques que l'on souhaite intégrer dans l'estimateur  $\hat{X}(s_k)$ .

D'après Shannon (1948), l'information contenue dans l'ensemble  $\{X(s_1), X(s_2), \dots, X(s_k)\}$  peut être quantifiée par :

$$\text{Info}(X(s_1), X(s_2), \dots, X(s_k)) = -\ln(T_G(\chi_{map})) \quad (2.6)$$

Selon l'équation (2.6) :

La diminution de  $T_G(\chi_{map})$ , de sorte que  $T_G(\chi_1) < T_G(\chi_2)$ , implique  $\text{Info}(\chi_1) > \text{Info}(\chi_2)$ , plus  $\{X(s_1), X(s_2), \dots, X(s_k)\}$  contient d'information, moins sa survenue est probable.

On considère une forte information a priori (ou de manière équivalente, une faible probabilité a priori) concernant la variabilité spatiale de la grandeur physique étudiée. Dans un contexte stochastique, cela peut être atteint en considérant la valeur attendue de la mesure d'information appropriée et en cherchant à la maximiser.

L'idée est de modéliser une bonne connaissance a priori de la répartition spatiale de la grandeur physique, ce qui se traduit mathématiquement par une faible probabilité a priori. Pour formaliser cela dans le cadre stochastique, on s'intéressera à la valeur espérée de la mesure d'information utilisée, et on cherchera à la maximiser afin de refléter un haut niveau d'information a priori.

En relation avec la mesure d'information (2.6), l'information attendue considérée sera l'espérance de l'information (2.6) :

$$E[\text{Info}(X(s_1), X(s_2), \dots, X(s_k))] = E[-\ln(T_G(\chi_{map}))]$$

$$E[\text{Info}(X(s_1), X(s_2), \dots, X(s_k))] = - \int \ln(T_G(\chi_{map})) T_G(\chi_{map}) d\chi_{map}$$

Soit  $h(T_G)$  la mesure de cette information :

$$h(T_G) = - \int \ln(T_G(\chi_{map})) T_G(\chi_{map}) d\chi_{map} \quad (2.7)$$

La fonction  $h(T_G)$  correspond à l'entropie de Shannon de la grandeur physique estimée. Elle permet de quantifier le degré d'incertitude contenu dans la densité de probabilité a priori  $T_G(\chi_{map})$ .

En cherchant à maximiser l'information attendue par rapport à cette densité de probabilité a priori, sous des contraintes liées aux connaissances physiques du système et aux propriétés souhaitées de l'estimateur, on vise en réalité à maximiser l'entropie associée. Cela permet d'obtenir une densité de probabilité a priori reflétant au mieux le manque d'information initial sur la répartition spatiale de la grandeur physique, tout en respectant les contraintes du problème.

Les estimateurs  $\hat{X}(s_k)$  sont définis statistiquement en termes d'espérances basées sur le champ aléatoire spatial  $X(s)$ . Il est donc judicieux de définir des contraintes physiques a priori compatibles avec cette formulation, sous la forme d'expressions mathématiques similaires impliquant des espérances de fonctions de  $X(s)$ .

$$E(g_v) = \int g_v(\chi_{map}) T_G(\chi_{map}) d\chi_{map} \quad (2.8)$$

En supposant que nous avons  $N$  contraintes où  $g_v(\chi_{map})$ ,  $v = \{1, 2, \dots, N\}$  sont des fonctions appropriées de  $X(s)$ .

Si  $v = 0$ , on a :

$$g_0(\chi_{map}) = 1, \text{ de sorte que } E[g_0] = 1$$

Cette dernière définition définit la contrainte de normalisation (2.5).

### Estimation de la PDF à priori

Formulons le problème d'estimation comme suit : Soit  $X(s)$  un champ aléatoire spatial. Nous cherchons à trouver des estimateurs  $\hat{X}(s_k)$  de  $X(s_k)$  en des points  $s_k$  à partir des données  $x_i$ ,  $i = 1, 2, \dots, m$  (avec  $i \neq k$ ), de telle sorte que :

L'entropie  $h(p)$  du modèle a priori de l'équation (2.7) est maximisée par rapport à  $T_G(\chi_{map})$ , sous la contrainte de normalisation (2.5) et les contraintes physiques de l'équation (2.8).

Ainsi, on définit le Lagrangien de la formule de l'entropie maximale sous les contraintes. La maximisation de ce Lagrangien permet de trouver la distribution de probabilité correspondante :

$$T_G(\chi_{map}) = \frac{1}{A} \exp \left( \sum_{v=0}^N \gamma_v g_v(\chi_{map}) \right) \quad (2.9)$$

### Preuve

Soit  $L(T_G)$  le Lagrangien associé à l'entropie  $h(p)$ . Maximiser  $h(p)$  revient à minimiser  $-h(p)$  sous les contraintes (2.5) et (2.8). On peut exprimer le Lagrangien associé à l'entropie  $h(p)$  comme suit :

$$L(T_G) = \int \ln(T_G(\chi_{map})) T_G(\chi_{map}) d\chi_{map} - \left( \sum_{v=1}^N \gamma_v \int g_v(\chi_{map}) T_G(\chi_{map}) d\chi_{map} - E(g_v) \right) \quad (2.10)$$

où  $g_0(\chi_{map}) = 1$ , et  $\gamma_v$  pour  $v \in \{0, 1, \dots, N\}$  sont les multiplicateurs de Lagrange associés à la contrainte de normalisation  $v = 0$  et aux autres contraintes  $v \neq 0$ .

L'entropie atteint son maximum lorsque la dérivée fonctionnelle s'annule :

Étant donné que  $L(T_G)$  est l'imbrication de plusieurs intégrales sur  $\mathbb{R}$ , alors on ajoute une constante  $\beta$ .

$$\frac{\delta L(T_G)}{\delta T_G} = 0 \implies (\ln(T_G) + 1) - \sum_{v=0}^N \gamma_v g_v - \beta = 0$$

$$\frac{\delta L(T_G)}{\delta T_G} = 0 \implies \ln(T_G) = \beta - 1 + \sum_{v=0}^N \gamma_v g_v$$

$$\frac{\delta L(T_G)}{\delta T_G} = 0 \implies T_G = \exp(\beta - 1) \cdot \exp \left( \sum_{v=0}^N \gamma_v g_v \right) \quad (2.11)$$

Posons  $T_G$  sous la forme :

$$T_G = \frac{1}{A} \exp \left( \sum_{v=0}^N \gamma_v g_v(\chi_{\text{map}}) \right) \quad (2.12)$$

D'après la contrainte de normalisation (équation 5), on a :

$$\int T_G(\chi_{\text{map}}) d\chi_{\text{map}} = 1$$

$$\int T_G(\chi_{\text{map}}) d\chi_{\text{map}} = 1 \implies \int \frac{1}{A} \exp \left( \sum_{v=0}^N \gamma_v g_v(\chi_{\text{map}}) \right) d\chi_{\text{map}} = 1 \quad (2.13)$$

$$\int T_G(\chi_{\text{map}}) d\chi_{\text{map}} = 1 \implies A = \int \exp \left( \sum_{v=0}^N \gamma_v g_v(\chi_{\text{map}}) \right) d\chi_{\text{map}} \quad (2.14)$$

D'où :

$$T_G(\chi_{\text{map}}) = \frac{1}{A} \exp \left( \sum_{v=0}^N \gamma_v g_v(\chi_{\text{map}}) \right) \quad \text{avec}$$

$$A = \int \exp \left( \sum_{v=0}^N \gamma_v g_v(\chi_{\text{map}}) \right) d\chi_{\text{map}} \quad (2.15)$$

Comme  $g_0 = 1$  (ne dépend pas de  $\chi_{\text{map}}$ ), en multipliant maintenant l'équation (2.15) par  $\exp(-\gamma_0)$ , on obtient :

$$\exp(-\gamma_0) = \frac{1}{A} \int \exp \left( \sum_{v=1}^N \gamma_v g_v(\chi_{\text{map}}) \right) d\chi_{\text{map}} \quad (2.16)$$

Par conséquent :

$$\gamma_0 = -\ln \left( \frac{1}{A} \int \exp \left( \sum_{v=1}^N \gamma_v g_v(\chi_{\text{map}}) \right) d\chi_{\text{map}} \right) \quad (2.17)$$

Cela exprime  $\gamma_0$  en fonction des autres multiplicateurs de Lagrange  $\gamma_v$ ,  $v \in \{1, 2, \dots, N\}$ . Puisque  $\gamma_0$  est maintenant considéré comme une fonction de  $\gamma_v$ ,  $v \in \{1, 2, \dots, N\}$ , on différencie l'équation (2.16) par rapport à un quelconque multiplicateur de Lagrange  $\gamma_i$  :

$$-\frac{\partial \gamma_0}{\partial \gamma_i} \exp(-\gamma_0) = \frac{1}{A} \int g_i(\chi_{\text{map}}) \exp \left( \sum_{v=1}^N \gamma_v g_v(\chi_{\text{map}}) \right) d\chi_{\text{map}}$$

et en multipliant par  $\exp(\gamma_0)$  :

$$-\frac{\partial \gamma_0}{\partial \gamma_i} = \frac{1}{A} \int g_i(\chi_{\text{map}}) \exp \left( \sum_{v=0}^N \gamma_v g_v(\chi_{\text{map}}) \right) d\chi_{\text{map}} \quad (2.18)$$



D'après l'équation (2.8), on a :

$$E(g_\nu) = \int g_\nu(\chi_{\text{map}}) T_G(\chi_{\text{map}}) d\chi_{\text{map}}$$

En remplaçant la fonction de densité de probabilité de l'équation (2.9) dans l'équation (2.8), on a :

$$E(g_\nu) = \frac{1}{A} \int g_\nu(\chi_{\text{map}}) \exp\left(\sum_{\nu=0}^N \gamma_\nu g_\nu(\chi_{\text{map}})\right) d\chi_{\text{map}}$$

D'où :

$$-\frac{\partial \gamma_0}{\partial \gamma_i} = E(g_i) \quad (2.19)$$

Le système d'équations (2.18) et (2.19) peut être résolu par rapport aux inconnues  $\gamma_\nu$ ,  $\nu \in \{1, 2, \dots, N\}$ . Les solutions seront insérées dans la fonction densité de probabilité (2.9) pour obtenir la forme requise du modèle a priori  $T_G(\chi_{\text{map}})$ .

## 2.2.2 Inférence Bayésienne

### Généralité su IB

La partie inférence bayésienne de la méthode BME a pour but de mettre à jour la distribution de probabilité obtenue par entropie maximale avec les nouvelles données observées (SCHMIDT, GEORGE et WOOD, 1999). En s'appuyant sur le théorème de (BAYES, 1958), elle calcule la distribution a posteriori à partir de la distribution a priori fournie par l'entropie maximale et des nouvelles observations. Cette distribution a posteriori représente la révision des probabilités des différentes hypothèses à la lumière des nouvelles données. À chaque ajout d'observations, l'inférence bayésienne affine la distribution en la recalculant, permettant d'améliorer progressivement l'estimation. Par cette mise à jour successive intégrant les nouvelles mesures, la méthode BME fournit une estimation probabiliste de plus en plus précise du processus étudié.

Jusqu'à présent, on a considéré la probabilité a priori  $T_G(\chi_{\text{map}})$  qui inclut un modèle a priori sur la relation entre  $X(s_k)$  et les  $X(s_i)$ . Ce modèle se rapporte à nos connaissances sur la variabilité spatiale avant de prendre en compte des mesures spécifiques de la grandeur physique.

### Estimation de la PDF à posteriori

En considérant les points suivants :

- $A$  et  $B$  sont deux événements;
- $P(A)$  et  $P(B)$  sont les probabilités des deux événements;
- $P(A \cap B)$  est la probabilité que les deux événements se réalisent;
- $P(A | B)$  est la probabilité conditionnelle que l'événement  $A$  se réalise étant donné que l'événement  $B$  s'est réalisé.

En vertu du théorème de Bayes (BAYES, 1958), on a :

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$

Comme  $\chi_{map} = (\chi_{data}, s_k)$  avec  $\chi_{data} = (\chi_{hard}, \chi_{soft})$ , la probabilité a posteriori

$$T_G^*(s_k | \chi_{data}) = \frac{T_G(\chi_{map})}{T_G(\chi_{data})} \quad (2.20)$$

avec  $\chi_{map} = (\chi_{data}, s_k)$ . Soit  $B_x(s_k)$  la fonction de Bayes associée à (2.20), on a :

$$B_x(s_k) = \ln(T_G^*(s_k | \chi_{data}))$$

La fonction de Bayes  $B_x(s_k)$  est maximisée par rapport à  $x_k$ . Cette valeur maximale sera la valeur attendue pour l'estimateur  $\hat{X}(s_k)$  en le point  $s_k$ .

ou équivalent :

$$B_x(s_k) = \ln(T_G(\chi_{map})) - \ln(T_G(\chi_{data})) \quad (2.21)$$

La probabilité a posteriori  $T_G^*(s_k | \chi_{data})$ , ou la fonction de Bayes  $B_x(s_k)$ , doit être maximisée par rapport à  $x_k$ . Dans ce cas,  $x_k$  est considéré comme un paramètre de la probabilité a posteriori ou de la fonction de Bayes du champ aléatoire spatial sous-jacent.

### Estimation de la variable physique d'étude

Pour estimer  $x_k$ , nous allons maximiser la fonction de Bayes par rapport à  $x_k$ , autrement dit :

$$\frac{\partial B_x(s_k)}{\partial x_k} \Big|_{x_k = \hat{x}_k} = 0$$

D'après l'équation (2.21), si :

$$\frac{\partial B_x(s_k)}{\partial x_k} \Big|_{x_k = \hat{x}_k} = 0$$

alors

$$\frac{\partial \ln(T_G(\chi_{map}))}{\partial x_k} \Big|_{x_k = \hat{x}_k} = 0 \quad (2.22)$$

où  $\hat{x}_k$  est la valeur de l'estimateur  $\hat{X}(s_k)$  en le point  $s_k$ . En utilisant la densité de probabilité finale trouvée avec ses paramètres et en maximisant l'équation (2.22), on obtient l'équation :

$$\sum_{v=0}^N \gamma_v \frac{\partial g_v(\chi_{map})}{\partial x_k} \Big|_{x_k = \hat{x}_k} = 0 \quad (2.23)$$

L'équation (2.23) est résolue par rapport à l'estimateur  $\hat{X}(s_k) = \hat{x}_k$ .

De la même manière que dans l'équation (2.19), on peut écrire :

$$-\frac{\partial^2 \gamma_0}{\partial \gamma_i^2} = E(g_i^2) \quad (2.24)$$

$$-\frac{\partial^2 \gamma_0}{\partial \gamma_i \partial \gamma_\nu} = E(g_i g_\nu) \quad (2.25)$$

Dans le cas où la fonction de densité de probabilité a priori est connue, l'équation (2.22) est utilisée pour estimer la variable physique. L'équation du maximum de vraisemblance est ainsi :

$$\frac{d \ln(T_G(\chi_{\text{data}} | s_k))}{dx_k} \Big|_{x_k = \hat{x}_k} = \frac{d \ln(T_G(\chi_{\text{data}}))}{dx_k} - \frac{d \ln(T_G(s_k))}{dx_k} \Big|_{x_k = \hat{x}_k} = 0 \quad (2.26)$$

## 2.3 La méthode de HASM (High Accuracy Surface Modeling)

### 2.3.1 Introduction au HASM

L'approche de modélisation de surfaces par équations aux dérivées partielles (EDP) suscite un intérêt croissant depuis ses débuts dans les années 1970. Cette technique consiste à représenter les surfaces comme solutions d'EDP. Elle présente l'avantage de pouvoir faire varier la forme de la surface modélisée en changeant les conditions aux limites et paramètres de l'équation (PROTOPOESCU, SANTORO et DOCKERY, 1989).

Depuis, de nombreuses études se sont penchées sur le potentiel de cette approche EDP. (PASADAS et RODRÍGUEZ, 2009) ont notamment souligné ses atouts pour le mélange de surfaces, la modélisation de formes libres ou les spécifications fonctionnelles. Grâce à sa flexibilité géométrique supérieure aux méthodes d'interpolation, elle s'est révélée utile dans divers domaines de recherche.

Parmi ces méthodes basées sur les EDP, la technique de modélisation de surfaces à haute précision (High Accuracy Surface Modeling) a démontré sa capacité à représenter des surfaces complexes avec une grande fidélité géométrique (YUE et al., 2010). En s'appuyant sur une résolution numérique avancée des EDP, elle offre une alternative prometteuse pour la modélisation de formes de haute qualité.

La méthode High Accuracy Surface Modeling (HASM) est basée sur le théorème fondamental des surfaces et a été développée pour résoudre les problèmes liés à la modélisation de surface (YUE et al., 2010; YUE et WANG, 2010). L'objectif de la modélisation de surfaces est de décrire la variation spatiale d'une variable donnée. Dans la plupart des procédures numériques pour résoudre des équations aux dérivées partielles, le problème est d'abord discrétisé en choisissant des équations algébriques sur un espace d'approximation à dimension finie. Ensuite, une méthode numérique est conçue pour résoudre ce système d'équations discrètes. De nombreuses études ont été menées pour montrer que la méthode HASM est plus précise que les méthodes classiques. Il a été montré que la précision de la simulation de HASM est plus élevée que celle des méthodes d'interpolation classiques (krigeage, méthodes d'interpolation de distance inverse, Splines, etc).

### 2.3.2 Formulation des EDP basiques de HASM

Selon le théorème fondamental des surfaces, une surface est entièrement définie par ses coefficients fondamentaux du premier et du second ordre. Les coefficients du premier ordre décrivent comment la surface hérite des propriétés géométriques intrinsèques de

la surface telles que les longueurs de courbes, les angles entre vecteurs tangents ou les aires de régions (TOPONOGOV, 2006). Les coefficients du second ordre reflètent quant à eux les déformations locales de la surface par rapport à son plan tangent en chaque point (LISEIKIN, 2006). Bien que la pente, l'aspect ou la courbure soient importants pour décrire localement une surface (EVANS, 1968), selon la géométrie différentielle ces éléments ne déterminent que les lignes de hachures et non la surface elle-même. Celle-ci ne peut être entièrement définie que par ses formes fondamentales du premier et du second ordre (TOPONOGOV, 2006).

Ainsi, les deux groupes de coefficients fondamentaux permettent de caractériser complètement la géométrie d'une surface de manière intrinsèque, indépendamment de sa configuration dans l'espace. Cette approche différentielle est à la base de la modélisation géométrique des surfaces.

(ZHAO et YUE, 2021) nous donnent les deux formes fondamentales d'une surface.

### La première forme fondamentale d'une surface

Considérons une surface  $S$  définie par les paramètres :  $t(x, y) = (x, y, T(x, y))$  et  $O(x, y)$  et  $O'(x + \Delta x, y + \Delta y)$  deux points adjacents de  $S$ . Leur distance intrinsèque  $O'O$  s'exprime en fonction des variations infinitésimales  $\Delta x$  et  $\Delta y$  des paramètres :

$$O'O = t(x + \Delta x, y + \Delta y) - t(x, y) = t_x \Delta x + t_y \Delta y + \dots$$

Lorsque  $O$  est infiniment proche de  $O'$ , les termes d'ordre supérieur à 2 en  $\Delta x$  et  $\Delta y$  peuvent être omis, et  $dt \approx t_x dx + t_y dy$ . Dans ce cas,  $ds$ , la partie principale de la norme du vecteur  $OO'$ , est définie comme la distance entre les deux points infiniment adjacents sur la surface  $S$ .

$$ds^2 = |dt|^2 = dt \cdot dt = (t_x dx + t_y dy) \cdot (t_x dx + t_y dy)$$

$$ds^2 = E(dx)^2 + 2F dx dy + G(dy)^2$$

En posant :

$$\begin{aligned} E &= t_x t_x, \quad F = t_x t_y, \quad G = t_y t_y \\ ds^2 &= E(dx)^2 + 2F dx dy + G(dy)^2 \end{aligned} \quad (2.27)$$

L'équation  $ds^2 = E(dx)^2 + 2F dx dy + G(dy)^2$  correspond à la première forme fondamentale d'une surface  $S$ . Les coefficients  $E$ ,  $F$  et  $G$  sont appelés coefficients fondamentaux de première espèce. La quantité  $ds^2 = |dt|^2 = dt \cdot dt$  étant une forme quadratique définie positive, cela implique  $E > 0$ ,  $G > 0$  et  $EG - F^2 > 0$ . Les coefficients de la première forme fondamentale sont des quantités qui définissent les propriétés géométriques intrinsèques de la surface (la longueur d'une courbe, l'angle entre deux courbes, l'aire d'une région ou la courbure d'une géodésique), indépendantes de sa place dans l'espace.

### La deuxième forme fondamentale d'une surface

Pour étudier le degré de courbure en un point  $O$  de la surface, on calcule la distance verticale  $\delta$  entre le point  $O'$  (proche de  $O$ ) et le plan tangent en  $O$ .

$$O'O = t(x + \Delta x, y + \Delta y) - t(x, y) = t_x \Delta x + t_y \Delta y + \frac{1}{2} (t_{xx}(\Delta x)^2 + 2t_{xy} \Delta x \Delta y + t_{yy}(\Delta y)^2) + \dots$$

$$\omega = O'O \Delta n = \frac{1}{2} (t_{xx} \Delta n (\Delta x)^2 + 2t_{xy} \Delta n \Delta x \Delta y + t_{yy} \Delta n (\Delta y)^2) + \dots$$

$n$  est le vecteur normal à  $S$ .

Étant donné que  $r_x \cdot n = 0$  et  $r_y \cdot n = 0$ , les dérivées par rapport à  $x$  et  $y$  sont prises des deux côtés.

$$r_{xx} \cdot n + r_x \cdot n_x = 0,$$

$$r_{xy} \cdot n + r_x \cdot n_y = 0,$$

$$r_{yx} \cdot n + r_y \cdot n_x = 0,$$

$$r_{yy} \cdot n + r_y \cdot n_y = 0$$

En posant :

$$L = t_{xx} \cdot \Delta n, \quad M = t_{xy} \cdot \Delta n, \quad N = t_{yy} \cdot \Delta n$$

En omettant les autres termes ,

$$2\omega \approx L(\Delta x)^2 + 2M\Delta x \Delta y + N(\Delta y)^2$$

Quand  $O'$  est très proche de  $O$ , la partie significative de  $2\omega$  est :

$$\alpha = L(dx)^2 + 2M dx dy + N(dy)^2 \quad (2.28)$$

Cette équation est considérée comme la seconde forme fondamentale d'une surface, où  $L$ ,  $M$ ,  $N$  sont les coefficients de la seconde forme fondamentale et sont également appelés les coefficients fondamentaux de second ordre. La seconde forme fondamentale décrit la forme de la surface et reflète la variation locale de courbure de la surface.

### 2.3.3 Théorème fondamental des surfaces

D'après le théorème principal de la théorie des surfaces (Su et Hu, 1979), si les coefficients fondamentaux de première et seconde espèce  $E, F, G, L, M, N$  de la surface sont symétriques, et si  $E, F, G$  sont définis positifs (satisfont au système d'équations de Gauss-Codazzi), alors on considère la surface comme le graphe d'une fonction  $z = T(x, y)$ , avec  $t = (x, y, T(x, y))$ , soumis à la condition initiale  $T(x, y) = T(x_0, y_0)$  pour  $x = x_0$  et  $y = y_0$ .

D'après(YUE et WANG, 2010) :

Les coefficients fondamentaux de première espèce  $E, F$  et  $G$  peuvent être formulés comme :

$$\begin{cases} E = 1 + T_x^2 \\ G = 1 + T_y^2 \\ F = T_x \cdot T_y \end{cases} \quad (2.29)$$

Les coefficients fondamentaux de première espèce  $L, M$  et  $N$  peuvent être formulés comme :

$$\begin{cases} L = \frac{T_{xx}}{\sqrt{1+T_x^2+T_y^2}} \\ M = \frac{T_{xy}}{\sqrt{1+T_x^2+T_y^2}} \\ N = \frac{T_{yy}}{\sqrt{1+T_x^2+T_y^2}} \end{cases} \quad (2.30)$$

L'équation de Gauss peut être reformulée comme suit :

$$\begin{cases} T_{xx} = \Gamma_{11}^1 \cdot T_x + \Gamma_{11}^2 \cdot T_y + \frac{L}{\sqrt{E \cdot G - F^2}} \\ T_{yy} = \Gamma_{22}^1 \cdot T_x + \Gamma_{22}^2 \cdot T_y + \frac{N}{\sqrt{E \cdot G - F^2}} \\ T_{xy} = \Gamma_{12}^1 \cdot T_x + \Gamma_{12}^2 \cdot T_y + \frac{M}{\sqrt{E \cdot G - F^2}} \end{cases} \quad (2.31)$$

où

$$\begin{aligned} \Gamma_{11}^1 &= \frac{1}{2} \frac{G \cdot E_x - 2F \cdot F_x + F \cdot E_y}{E \cdot G - F^2} \\ \Gamma_{12}^1 &= \frac{1}{2} \frac{G \cdot E_y - F \cdot G_x}{E \cdot G - F^2} \\ \Gamma_{22}^1 &= \frac{1}{2} \frac{2G \cdot F_y - G \cdot G_x - F \cdot G_y}{E \cdot G - F^2} \\ \Gamma_{11}^2 &= \frac{1}{2} \frac{2E \cdot F_x - E \cdot E_y - F \cdot E_x}{E \cdot G - F^2} \\ \Gamma_{12}^2 &= \frac{1}{2} \frac{E \cdot G_x - F \cdot E_y}{E \cdot G - F^2} \\ \Gamma_{22}^2 &= \frac{1}{2} \frac{E \cdot G_y - 2F \cdot F_y + F \cdot G_x}{E \cdot G - F^2} \end{aligned}$$

Les symboles de Christoffel du second type  $\Gamma_{11}^1$ ,  $\Gamma_{12}^1$ ,  $\Gamma_{22}^1$ ,  $\Gamma_{11}^2$ ,  $\Gamma_{12}^2$  et  $\Gamma_{22}^2$  dépendent uniquement des coefficients fondamentaux de première espèce et de leurs dérivées.

Pour le calcul, on prend le cas d'un maillage d'un domaine orthogonal  $\Omega = [0, N_x] \times [0, N_y]$  avec  $h = \frac{N_x}{I+1} = \frac{N_y}{J+1}$  le pas de discrétisation, où  $(I+1)$  et  $(J+1)$  sont les nombres de points respectifs suivant la direction de  $x$  et la direction de  $y$ . L'ensemble des points sur les mailles est noté  $\{(x_i, y_j)\}$ . Les valeurs de l'attribut des points situés aux quatre bords du domaine sont considérées comme des conditions initiales.

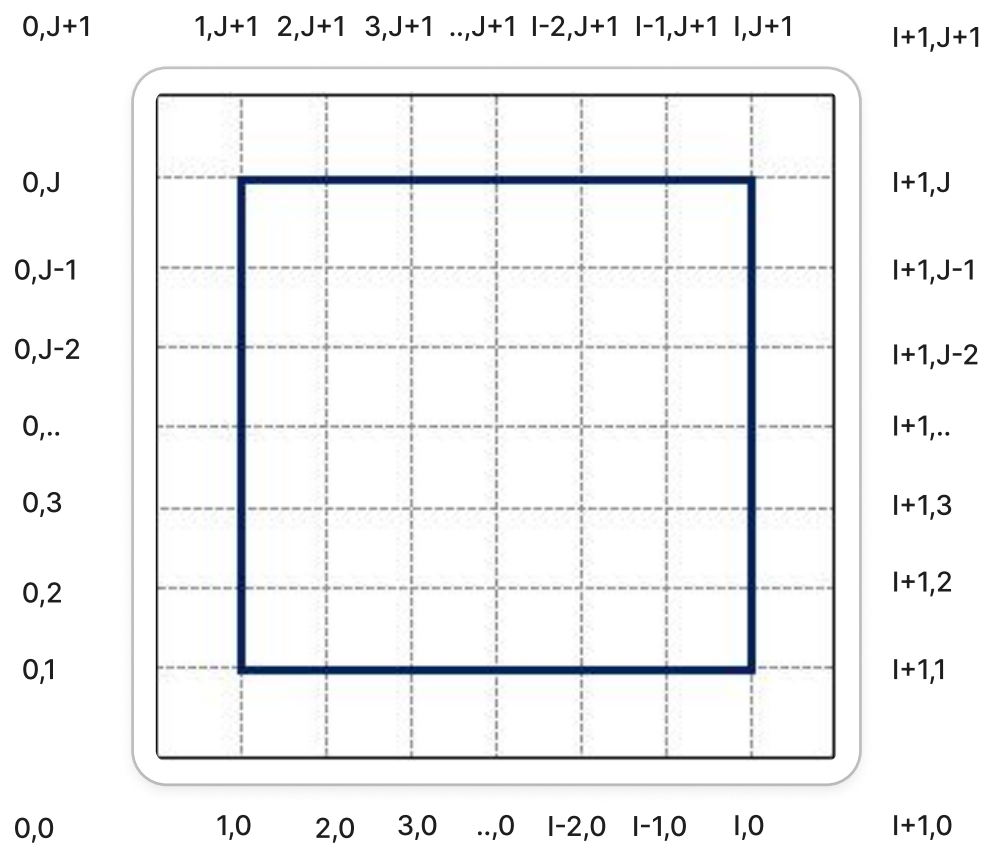


FIGURE 2.2 – Maillage domaine

### 2.3.4 Discrétisation par différences finies

Nous allons procéder aux différences finies centrées pour approximer les dérivées partielles d'ordre 1, d'ordre 2, et la dérivée croisée de  $T$ .

Les fonctions  $T(x+h, y)$  et  $T(x-h, y)$  peuvent être formulées par un développement en série de Taylor autour de  $T(x, y)$  tel que :

$$T(x+h, y) = T(x, y) + h \frac{\partial T(x, y)}{\partial x} + \frac{h^2}{2!} \frac{\partial^2 T(x, y)}{\partial x^2} + \frac{h^3}{3!} \frac{\partial^3 T(x, y)}{\partial x^3} + O(h^4) \quad (2.32)$$

$$T(x-h, y) = T(x, y) - h \frac{\partial T(x, y)}{\partial x} + \frac{h^2}{2!} \frac{\partial^2 T(x, y)}{\partial x^2} - \frac{h^3}{3!} \frac{\partial^3 T(x, y)}{\partial x^3} + O(h^4) \quad (2.33)$$

En faisant la soustraction de l'équation (2.32) par l'équation (2.33), on obtient :

$$T(x+h, y) - T(x-h, y) = 2h \frac{\partial T(x, y)}{\partial x} + \frac{2h^3}{3!} \frac{\partial^3 T(x, y)}{\partial x^3} + O(h^5)$$

Donc :

$$\frac{\partial T(x, y)}{\partial x} = \frac{T(x+h, y) - T(x-h, y)}{2h} - \frac{h^2}{3!} \frac{\partial^3 T(x, y)}{\partial x^3} + O(h^4)$$

Pour un  $h$  suffisamment petit, l'approximation aux différences finies de  $\frac{\partial T(x, y)}{\partial x}$  peut s'exprimer comme :

$$\frac{\partial T(x, y)}{\partial x} \approx \frac{T(x+h, y) - T(x-h, y)}{2h} \quad (2.34)$$

De la même manière, on a :

$$\frac{\partial T(x, y)}{\partial y} \approx \frac{T(x, y+h) - T(x, y-h)}{2h} \quad (2.35)$$

En faisant la somme des équations (2.32) et (2.33) on a :

$$T(x+h, y) + T(x-h, y) = 2T(x, y) + \frac{2h^2}{2!} \frac{\partial^2 T(x, y)}{\partial x^2} + O(h^4)$$

Donc :

$$\frac{\partial^2 T(x, y)}{\partial x^2} = \frac{T(x+h, y) - 2T(x, y) + T(x-h, y)}{h^2} + O(h^2)$$

Pour un  $h$  suffisamment petit, l'approximation aux différences finies de  $\frac{\partial^2 T(x, y)}{\partial x^2}$  peut s'exprimer comme :

$$\frac{\partial^2 T(x, y)}{\partial x^2} \approx \frac{T(x+h, y) - 2T(x, y) + T(x-h, y)}{h^2} \quad (2.36)$$

De la même manière, on aura :

$$\frac{\partial^2 T(x, y)}{\partial y^2} \approx \frac{T(x, y+h) - 2T(x, y) + T(x, y-h)}{h^2} \quad (2.37)$$

Analogiquement, on a :

$$\frac{\partial^2 T(x, y)}{\partial x \partial y} \approx \frac{T(x+h, y+h) - T(x-h, y+h) - T(x-h, y-h) + T(x+h, y-h)}{4h^2} \quad (2.38)$$



### 2.3.5 Forme discrétisée du système d'équations de Gauss(2.31)

La forme discrétisée des équations de Gauss est donnée par :

$$\begin{cases} \frac{T_{i+1,j}-2T_{i,j}+T_{i-1,j}}{h^2} = (\Gamma_{11}^1)_{i,j} \cdot \frac{T_{i+1,j}-T_{i-1,j}}{2h} + (\Gamma_{11}^2)_{i,j} \cdot \frac{T_{i,j+1}-T_{i,j-1}}{2h} + \frac{L_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \\ \frac{T_{i,j+1}-2T_{i,j}+T_{i,j-1}}{h^2} = (\Gamma_{22}^1)_{i,j} \cdot \frac{T_{i+1,j}-T_{i-1,j}}{2} + (\Gamma_{22}^2)_{i,j} \cdot \frac{T_{i,j+1}-T_{i,j-1}}{2} + \frac{N_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \\ \frac{T_{i+1,j+1}-T_{i-1,j+1}+T_{i-1,j-1}-T_{i+1,j-1}}{4h^2} = (\Gamma_{12}^1)_{i,j} \cdot \frac{T_{i+1,j}-T_{i-1,j}}{2h} + (\Gamma_{12}^2)_{i,j} \cdot \frac{T_{i,j+1}-T_{i,j-1}}{2} + \frac{M_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \end{cases} \quad (2.39)$$

En multipliant les deux premières équations du système par  $h^2$  et la troisième par  $4h^2$ , on obtient :

$$\begin{cases} T_{i+1,j}-2T_{i,j}+T_{i-1,j} = (\Gamma_{11}^1)_{i,j} \cdot \frac{T_{i+1,j}-T_{i-1,j}}{2 \cdot h} + (\Gamma_{11}^2)_{i,j} \cdot \frac{T_{i,j+1}-T_{i,j-1}}{2 \cdot h} + \frac{L_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \cdot h^2 \\ T_{i,j+1}-2T_{i,j}+T_{i,j-1} = (\Gamma_{22}^1)_{i,j} \cdot \frac{T_{i+1,j}-T_{i-1,j}}{2 \cdot h} + (\Gamma_{22}^2)_{i,j} \cdot \frac{T_{i,j+1}-T_{i,j-1}}{2 \cdot h} + \frac{N_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \cdot h^2 \\ T_{i+1,j+1}-T_{i-1,j+1}+T_{i-1,j-1}-T_{i+1,j-1} = 2 \cdot h \cdot (\Gamma_{12}^1)_{i,j} \cdot (T_{i+1,j}-T_{i-1,j}) \\ + 2 \cdot h \cdot (\Gamma_{12}^2)_{i,j} \cdot (T_{i,j+1}-T_{i,j-1}) + \frac{4 \cdot h^2 \cdot M_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \end{cases} \quad (2.40)$$

D'après les études de (YUE et WANG, 2010), la troisième équation du système produit une estimation biaisée des résultats. Cette équation a été mise à l'écart du système pour l'efficacité de la méthode. Nous allons donc tenir compte seulement des deux premières équations du système :

$$\begin{cases} T_{i+1,j}-2T_{i,j}+T_{i-1,j} = (\Gamma_{11}^1)_{i,j} \cdot \frac{T_{i+1,j}-T_{i-1,j}}{2 \cdot h} + (\Gamma_{11}^2)_{i,j} \cdot \frac{T_{i,j+1}-T_{i,j-1}}{2 \cdot h} + \frac{L_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \cdot h^2 \\ T_{i,j+1}-2T_{i,j}+T_{i,j-1} = (\Gamma_{22}^1)_{i,j} \cdot \frac{T_{i+1,j}-T_{i-1,j}}{2 \cdot h} + (\Gamma_{22}^2)_{i,j} \cdot \frac{T_{i,j+1}-T_{i,j-1}}{2 \cdot h} + \frac{N_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \cdot h^2 \end{cases} \quad (2.41)$$

Les expressions de Christoffel du second type sont définies comme suit :

$$\Gamma_{11}^1 = \frac{1}{2} \frac{GE_x - 2FF_x + FE_y}{EG - F^2}$$

$$\Gamma_{12}^1 = \frac{1}{2} \frac{GE_y - FG_x}{EG - F^2}$$

$$\Gamma_{22}^1 = \frac{1}{2} \frac{2GF_y - GG_x - FG_y}{EG - F^2}$$

$$\Gamma_{11}^2 = \frac{1}{2} \frac{2EF_x - EE_y - FE_x}{EG - F^2}$$

$$\Gamma_{12}^2 = \frac{1}{2} \frac{EG_x - FE_y}{EG - F^2}$$

$$\Gamma_{22}^2 = \frac{1}{2} \frac{EG_y - 2FF_y + FG_x}{EG - F^2}$$

Les expressions des coefficients fondamentaux de première espèce étant connues, il reste à définir celles de leurs dérivées. En appliquant les différences finies centrées, nous avons :

$$E_x \approx \frac{E_{i+1,j} - E_{i-1,j}}{2h}$$

$$E_y \approx \frac{E_{i,j+1} - E_{i,j-1}}{2h}$$

$$F_x \approx \frac{F_{i+1,j} - F_{i-1,j}}{2h}$$

$$F_y \approx \frac{F_{i,j+1} - F_{i,j-1}}{2h}$$

$$G_x \approx \frac{G_{i+1,j} - G_{i-1,j}}{2h}$$

$$G_y \approx \frac{G_{i,j+1} - G_{i,j-1}}{2h}$$

En remplaçant ces expressions de dérivées des coefficients fondamentaux de première espèce dans les expressions des symboles de Christoffel, nous avons :

$$(\Gamma_{11}^1)_{i,j} = \frac{G_{i,j} \cdot (E_{i+1,j} - E_{i-1,j}) - 2F_{i,j} \cdot (F_{i+1,j} - F_{i-1,j}) + F_{i,j} \cdot (E_{i,j+1} - E_{i,j-1})}{4(E_{i,j} \cdot G_{i,j} - F_{i,j}^2) \cdot h} \quad (2.42)$$

$$(\Gamma_{12}^1)_{i,j} = \frac{G_{i,j} \cdot (E_{i,j+1} - E_{i,j-1}) - F_{i,j} \cdot (G_{i+1,j} - G_{i-1,j})}{4(E_{i,j} \cdot G_{i,j} - F_{i,j}^2) \cdot h} \quad (2.43)$$

$$(\Gamma_{22}^1)_{i,j} = \frac{2G_{i,j} \cdot (F_{i,j+1} - F_{i,j-1}) - G_{i,j} \cdot (G_{i+1,j} - G_{i-1,j}) - F_{i,j} \cdot (G_{i,j+1} - G_{i,j-1})}{4(E_{i,j} \cdot G_{i,j} - F_{i,j}^2) \cdot h} \quad (2.44)$$

$$(\Gamma_{11}^2)_{i,j} = \frac{2E_{i,j} \cdot (F_{i+1,j} - F_{i-1,j}) - E_{i,j} \cdot (E_{i,j+1} - E_{i,j-1}) - F_{i,j} \cdot (E_{i+1,j} - E_{i-1,j})}{4(E_{i,j} \cdot G_{i,j} - F_{i,j}^2) \cdot h} \quad (2.45)$$

$$(\Gamma_{12}^2)_{i,j} = \frac{E_{i,j} \cdot (G_{i+1,j} - G_{i-1,j}) - F_{i,j} \cdot (E_{i,j+1} - E_{i,j-1})}{4(E_{i,j} \cdot G_{i,j} - F_{i,j}^2) \cdot h} \quad (2.46)$$

$$(\Gamma_{22}^2)_{i,j} = \frac{E_{i,j} \cdot (G_{i,j+1} - G_{i,j-1}) - 2F_{i,j} \cdot (F_{i,j+1} - F_{i,j-1}) + F_{i,j} \cdot (G_{i+1,j} - G_{i-1,j})}{4(E_{i,j} \cdot G_{i,j} - F_{i,j}^2) \cdot h} \quad (2.47)$$

Les conditions aux bords sont définies comme suit :

$$\begin{aligned} T_{0,j}^{(n+1)} &= T_{0,j}^{(0)} \quad \text{pour } 0 \leq j \leq J+1 \\ T_{i,0}^{(n+1)} &= T_{i,0}^{(0)} \quad \text{pour } 0 \leq i \leq I+1 \\ T_{I+1,j}^{(n+1)} &= T_{I+1,j}^{(0)} \quad \text{pour } 0 \leq j \leq J+1 \\ T_{i,J+1}^{(n+1)} &= T_{i,J+1}^{(0)} \quad \text{pour } 0 \leq i \leq I+1 \end{aligned}$$

La formulation itérative de l'équation (2.41) est donnée par :

$$\begin{cases} T_{i+1,j}^{(n+1)} - 2T_{i,j}^{(n+1)} + T_{i-1,j}^{(n+1)} = (\Gamma_{11}^1)_{i,j} \cdot \frac{T_{i+1,j}^{(n)} - T_{i-1,j}^{(n)}}{2h} + (\Gamma_{11}^2)_{i,j} \cdot \frac{T_{i,j+1}^{(n)} - T_{i,j-1}^{(n)}}{2h} + \frac{L_{i,j}^{(n)}}{\sqrt{E_{i,j}^{(n)} + G_{i,j}^{(n)} - 1}} \cdot h^2 \\ T_{i,j+1}^{(n+1)} - 2T_{i,j}^{(n+1)} + T_{i,j-1}^{(n+1)} = (\Gamma_{22}^1)_{i,j} \cdot \frac{T_{i+1,j}^{(n)} - T_{i-1,j}^{(n)}}{2h} + (\Gamma_{22}^2)_{i,j} \cdot \frac{T_{i,j+1}^{(n)} - T_{i,j-1}^{(n)}}{2h} + \frac{N_{i,j}^{(n)}}{\sqrt{E_{i,j}^{(n)} + G_{i,j}^{(n)} - 1}} \cdot h^2 \end{cases} \quad (2.48)$$

Avec les expressions :

$$E_{i,j}^{(n)} = 1 + \left( \frac{T_{i+1,j}^{(n)} - T_{i-1,j}^{(n)}}{2h} \right)^2 \quad (2.49)$$

$$G_{i,j}^{(n)} = 1 + \left( \frac{T_{i,j+1}^{(n)} - T_{i,j-1}^{(n)}}{2h} \right)^2 \quad (2.50)$$

$$F_{i,j}^{(n)} = 1 + \left( \frac{T_{i+1,j}^{(n)} - T_{i-1,j}^{(n)}}{2h} \right) \left( \frac{T_{i,j+1}^{(n)} - T_{i,j-1}^{(n)}}{2h} \right) \quad (2.51)$$

$$L_{i,j}^{(n)} = \frac{\frac{T_{i+1,j}^{(n)} - 2T_{i,j}^{(n)} + T_{i-1,j}^{(n)}}{h^2}}{\sqrt{1 + \left( \frac{T_{i+1,j}^{(n)} - T_{i-1,j}^{(n)}}{2h} \right)^2 + \left( \frac{T_{i,j+1}^{(n)} - T_{i,j-1}^{(n)}}{2h} \right)^2}} \quad (2.52)$$

$$N_{i,j}^{(n)} = \frac{\frac{T_{i,j+1}^{(n)} - 2T_{i,j}^{(n)} + T_{i,j-1}^{(n)}}{h^2}}{\sqrt{1 + \left( \frac{T_{i+1,j}^{(n)} - T_{i-1,j}^{(n)}}{2h} \right)^2 + \left( \frac{T_{i,j+1}^{(n)} - T_{i,j-1}^{(n)}}{2h} \right)^2}} \quad (2.53)$$

Les symboles de Christoffel du second type sont également donnés par :

$$(\Gamma_{11}^1)_{i,j} = \frac{G_{i,j}^{(n)} \cdot (E_{i+1,j}^{(n)} - E_{i-1,j}^{(n)}) - 2F_{i,j}^{(n)} \cdot (F_{i+1,j}^{(n)} - F_{i-1,j}^{(n)}) + F_{i,j}^{(n)} \cdot (E_{i,j+1}^{(n)} - E_{i,j-1}^{(n)})}{4(E_{i,j}^{(n)} \cdot G_{i,j}^{(n)} - (F_{i,j}^{(n)})^2) \cdot h} \quad (2.54)$$

$$(\Gamma_{12}^1)_{i,j} = \frac{G_{i,j}^{(n)} \cdot (E_{i,j+1}^{(n)} - E_{i,j-1}^{(n)}) - F_{i,j}^{(n)} \cdot (G_{i+1,j}^{(n)} - G_{i-1,j}^{(n)})}{4(E_{i,j}^{(n)} \cdot G_{i,j}^{(n)} - (F_{i,j}^{(n)})^2) \cdot h} \quad (2.55)$$

$$(\Gamma_{22}^1)_{i,j} = \frac{2G_{i,j}^{(n)} \cdot (F_{i,j+1}^{(n)} - F_{i,j-1}^{(n)}) - G_{i,j}^{(n)} \cdot (G_{i+1,j}^{(n)} - G_{i-1,j}^{(n)}) - F_{i,j}^{(n)} \cdot (G_{i,j+1}^{(n)} - G_{i,j-1}^{(n)})}{4(E_{i,j}^{(n)} \cdot G_{i,j}^{(n)} - (F_{i,j}^{(n)})^2) \cdot h} \quad (2.56)$$

$$(\Gamma_{11}^2)_{i,j} = \frac{2E_{i,j}^{(n)} \cdot (F_{i+1,j}^{(n)} - F_{i-1,j}^{(n)}) - E_{i,j}^{(n)} \cdot (E_{i,j+1}^{(n)} - E_{i,j-1}^{(n)}) - F_{i,j}^{(n)} \cdot (E_{i+1,j+1}^{(n)} - E_{i,j-1}^{(n)})}{4(E_{i,j}^{(n)} \cdot G_{i,j}^{(n)} - (F_{i,j}^{(n)})^2) \cdot h} \quad (2.57)$$

$$(\Gamma_{12}^2)_{i,j} = \frac{E_{i,j}^{(n)} \cdot (G_{i+1,j}^{(n)} - G_{i-1,j}^{(n)}) - F_{i,j}^{(n)} \cdot (E_{i,j+1}^{(n)} - E_{i,j-1}^{(n)})}{4(E_{i,j}^{(n)} \cdot G_{i,j}^{(n)} - (F_{i,j}^{(n)})^2) \cdot h} \quad (2.58)$$

$$(\Gamma_{22}^2)_{i,j} = \frac{E_{i,j}^{(n)} \cdot (G_{i,j+1}^{(n)} - G_{i,j-1}^{(n)}) - 2F_{i,j}^{(n)} \cdot (F_{i,j+1}^{(n)} - F_{i,j-1}^{(n)}) + F_{i,j}^{(n)} \cdot (G_{i+1,j}^{(n)} - G_{i-1,j}^{(n)})}{4(E_{i,j}^{(n)} \cdot G_{i,j}^{(n)} - (F_{i,j}^{(n)})^2) \cdot h} \quad (2.59)$$

$$(\Gamma_{12}^1)_{i,j} = \frac{G_{i,j}^{(n)} \cdot (E_{i,j+1}^{(n)} - E_{i,j-1}^{(n)}) - F_{i,j}^{(n)} \cdot (G_{i+1,j}^{(n)} - G_{i-1,j}^{(n)})}{4(E_{i,j}^{(n)} \cdot G_{i,j}^{(n)} - (F_{i,j}^{(n)})^2) \cdot h} \quad (2.60)$$

$$(\Gamma_{12}^2)_{i,j} = \frac{E_{i,j}^{(n)} \cdot (G_{i+1,j}^{(n)} - G_{i-1,j}^{(n)}) - F_{i,j}^{(n)} \cdot (F_{i,j+1}^{(n)} - F_{i,j-1}^{(n)})}{4(E_{i,j}^{(n)} \cdot G_{i,j}^{(n)} - (F_{i,j}^{(n)})^2) \cdot h} \quad (2.61)$$

Le vecteur des inconnus est :

$$Z^{(n+1)} = \left( T_{1,1}^{(n+1)}, \dots, T_{1,J}^{(n+1)}, \dots, T_{I-1,1}^{(n+1)}, \dots, T_{I-1,J}^{(n+1)}, T_{I,1}^{(n+1)}, \dots, T_{I,J}^{(n+1)} \right)$$

Les méthodes itératives telles que Gauss-Seidel, Gauss-Seidel modifié, et autres peuvent être utilisées pour résoudre le système constitué des inconnus en des points discrétisés du domaine d'étude.

### 2.3.6 Forme matricielle du système d'équations

Considérons le vecteur des inconnus  $Z$  défini comme :

$$Z = (T_{1,1}, T_{1,2}, \dots, T_{1,J}, T_{2,1}, T_{2,2}, \dots, T_{2,J}, \dots, T_{I,1}, T_{I,2}, \dots, T_{I,J})^T \quad (2.62)$$

Les conditions aux bords doivent être fournies pour prédire les valeurs des points à l'intérieur du domaine. Les points à l'intérieur du domaine sont ceux qui ne se trouvent pas sur les bords.

#### Première Équation du Système

La première équation du système (2.41) peut être écrite sous la forme matricielle suivante :

$$A \cdot Z = d \quad (2.63)$$

où  $A$  est la matrice des coefficients et  $d$  est le vecteur du second membre de la première équation du système (2.41).

#### Vecteur $d$

$$d = (d_1, d_2, \dots, d_{I-1}, d_I)^T$$

On avait posé comme inconnues, les valeurs à l'intérieur du domaine, comme l'indique la figure de maille en fixant à chaque fois la variable  $i$  (axe des abscisses) et en variant la variable  $J$  (axe des ordonnées).

Donc

$$Z = (T_{1,1}, \dots, T_{1,J}, \dots, T_{I-1,1}, \dots, T_{I-1,J}, T_{I,1}, \dots, T_{I,J})$$

Pour  $1 < i < I$ , l'équation garde sa forme expressive intacte.

Dans le cas contraire où  $i \in \{1, I\}$ , les conditions initiales interviennent.

Après récurrences des itérations, nous trouvons une formulation générale des paramètres  $A$  et  $d$  de l'équation (2.63).

Le vecteur  $d$  est défini comme suit :

— Pour  $i = 1$  :

$$d_1 = \begin{pmatrix} (\Gamma_{11}^1)_{1,1} \cdot \frac{T_{2,1}-T_{0,1}}{2h} + (\Gamma_{11}^2)_{1,1} \cdot \frac{T_{1,2}-T_{1,0}}{2h} + \frac{L_{1,1}}{\sqrt{E_{1,1}+G_{1,1}-1}} \cdot h^2 - T_{0,1} \\ (\Gamma_{11}^1)_{1,2} \cdot \frac{T_{2,2}-T_{0,2}}{2h} + (\Gamma_{11}^2)_{1,2} \cdot \frac{T_{1,3}-T_{1,1}}{2h} + \frac{L_{1,2}}{\sqrt{E_{1,2}+G_{1,2}-1}} \cdot h^2 - T_{0,2} \\ \vdots \\ (\Gamma_{11}^1)_{1,j} \cdot \frac{T_{2,j}-T_{0,j}}{2h} + (\Gamma_{11}^2)_{1,j} \cdot \frac{T_{1,j+1}-T_{1,j-1}}{2h} + \frac{L_{1,j}}{\sqrt{E_{1,j}+G_{1,j}-1}} \cdot h^2 - T_{0,j} \\ \vdots \\ (\Gamma_{11}^1)_{1,J-1} \cdot \frac{T_{2,J-1}-T_{0,J-1}}{2h} + (\Gamma_{11}^2)_{1,J-1} \cdot \frac{T_{1,J}-T_{1,J-2}}{2h} + \frac{L_{1,J-1}}{\sqrt{E_{1,J-1}+G_{1,J-1}-1}} \cdot h^2 - T_{0,J-1} \\ (\Gamma_{11}^1)_{1,J} \cdot \frac{T_{2,J}-T_{0,J}}{2h} + (\Gamma_{11}^2)_{1,J} \cdot \frac{T_{1,J+1}-T_{1,J-1}}{2h} + \frac{L_{1,J}}{\sqrt{E_{1,J}+G_{1,J}-1}} \cdot h^2 - T_{0,J} \end{pmatrix}$$

— Pour  $2 \leq i \leq I-1$  :

$$d_i = \begin{pmatrix} (\Gamma_{11}^1)_{i,1} \cdot \frac{T_{i+1,1}-T_{i-1,1}}{2h} + (\Gamma_{11}^2)_{i,1} \cdot \frac{T_{i,2}-T_{i,0}}{2h} + \frac{L_{i,1}}{\sqrt{E_{i,1}+G_{i,1}-1}} \cdot h^2 \\ (\Gamma_{11}^1)_{i,2} \cdot \frac{T_{i+1,2}-T_{i-1,2}}{2h} + (\Gamma_{11}^2)_{i,2} \cdot \frac{T_{i,3}-T_{i,1}}{2h} + \frac{L_{i,2}}{\sqrt{E_{i,2}+G_{i,2}-1}} \cdot h^2 \\ \vdots \\ (\Gamma_{11}^1)_{i,j} \cdot \frac{T_{i+1,j}-T_{i-1,j}}{2h} + (\Gamma_{11}^2)_{i,j} \cdot \frac{T_{i,j+1}-T_{i,j-1}}{2h} + \frac{L_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \cdot h^2 \\ \vdots \\ (\Gamma_{11}^1)_{i,J-1} \cdot \frac{T_{i+1,J-1}-T_{i-1,J-1}}{2h} + (\Gamma_{11}^2)_{i,J-1} \cdot \frac{T_{i,J}-T_{i,J-2}}{2h} + \frac{L_{i,J-1}}{\sqrt{E_{i,J}+G_{i,J}-1}} \cdot h^2 \\ (\Gamma_{11}^1)_{i,J} \cdot \frac{T_{i+1,J}-T_{i-1,J}}{2h} + (\Gamma_{11}^2)_{i,J} \cdot \frac{T_{i,J+1}-T_{i,J-1}}{2h} + \frac{L_{i,J}}{\sqrt{E_{i,J}+G_{i,J}-1}} \cdot h^2 \end{pmatrix}$$

— Pour  $i = I$  :

$$d_I = \begin{pmatrix} (\Gamma_{11}^1)_{I,1} \cdot \frac{T_{I+1,1}-T_{I-1,1}}{2h} + (\Gamma_{11}^2)_{I,1} \cdot \frac{T_{I,2}-T_{I,0}}{2h} + \frac{L_{I,1}}{\sqrt{E_{I,1}+G_{I,1}-1}} \cdot h^2 \\ (\Gamma_{11}^1)_{I,2} \cdot \frac{T_{I+1,2}-T_{I-1,2}}{2h} + (\Gamma_{11}^2)_{I,2} \cdot \frac{T_{I,3}-T_{I,1}}{2h} + \frac{L_{I,2}}{\sqrt{E_{I,2}+G_{I,2}-1}} \cdot h^2 \\ \vdots \\ (\Gamma_{11}^1)_{I,j} \cdot \frac{T_{I+1,j}-T_{I-1,j}}{2h} + (\Gamma_{11}^2)_{I,j} \cdot \frac{T_{I,j+1}-T_{I,j-1}}{2h} + \frac{L_{I,j}}{\sqrt{E_{I,j}+G_{I,j}-1}} \cdot h^2 \\ \vdots \\ (\Gamma_{11}^1)_{I,J-1} \cdot \frac{T_{I+1,J-1}-T_{I-1,J-1}}{2h} + (\Gamma_{11}^2)_{I,J-1} \cdot \frac{T_{I,J}-T_{I,J-2}}{2h} + \frac{L_{I,J-1}}{\sqrt{E_{I,J}+G_{I,J}-1}} \cdot h^2 \\ (\Gamma_{11}^1)_{I,J} \cdot \frac{T_{I+1,J}-T_{I-1,J}}{2h} + (\Gamma_{11}^2)_{I,J} \cdot \frac{T_{I,J+1}-T_{I,J-1}}{2h} + \frac{L_{I,J}}{\sqrt{E_{I,J}+G_{I,J}-1}} \cdot h^2 \end{pmatrix}$$

### Matrice A

La matrice A est définie comme :

$$A = \begin{pmatrix} -2I_J & I_J & 0 & \cdots & 0 & \cdots \\ I_J & -2I_J & I_J & \cdots & 0 & \cdots \\ 0 & I_J & -2I_J & \cdots & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & I_J & -2I_J & I_J \\ 0 & \cdots & 0 & 0 & I_J & -2I_J \end{pmatrix}_{(I \cdot J) \times (I \cdot J)}$$

où  $I_J$  est une matrice identité de dimension  $J \times J$ . Les éléments non représentés par des ... désignent les matrices nulles.

### Deuxième Équation du Système

La deuxième équation du système d'équations (2.41) peut s'écrire sous la forme matricielle suivante :

$$B \cdot Z = q \quad (2.64)$$

**vecteur  $q$**

$$q = (q_1, q_2, \dots, q_{I-1}, q_I)^T$$

le vecteur de second membre et B la matrice des coefficients de la deuxième équation du système (2.41)

où  $q_i$  est défini comme suit :

$$q_i = \begin{pmatrix} (\Gamma_{11}^1)_{i,1} \cdot \frac{T_{i+1,1}-T_{i-1,1}}{2h} + (\Gamma_{11}^2)_{i,1} \cdot \frac{T_{i,2}-T_{i,0}}{2h} + \frac{L_{i,1}}{\sqrt{E_{i,1}+G_{i,1}-1}} \cdot h^2 - T_{i,0} \\ (\Gamma_{11}^1)_{i,2} \cdot \frac{T_{i+1,2}-T_{i-1,2}}{2h} + (\Gamma_{11}^2)_{i,2} \cdot \frac{T_{i,3}-T_{i,1}}{2h} + \frac{L_{i,2}}{\sqrt{E_{i,2}+G_{i,2}-1}} \cdot h^2 \\ \vdots \\ (\Gamma_{11}^1)_{i,j} \cdot \frac{T_{i+1,j}-T_{i-1,j}}{2h} + (\Gamma_{11}^2)_{i,j} \cdot \frac{T_{i,j+1}-T_{i,j-1}}{2h} + \frac{L_{i,j}}{\sqrt{E_{i,j}+G_{i,j}-1}} \cdot h^2 \\ \vdots \\ (\Gamma_{11}^1)_{i,J} \cdot \frac{T_{i+1,J}-T_{i-1,J}}{2h} + (\Gamma_{11}^2)_{i,J} \cdot \frac{T_{i,J+1}-T_{i,J-1}}{2h} + \frac{L_{i,J}}{\sqrt{E_{i,J}+G_{i,J}-1}} \cdot h^2 - T_{i,J+1} \end{pmatrix}_{J \times 1}$$

La matrice  $B$  est définie comme :

$$B = \begin{pmatrix} B_J & * & 0 & \cdots & 0 \\ * & B_J & * & \cdots & 0 \\ 0 & * & B_J & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & * & B_J \\ 0 & \cdots & 0 & 0 & B_J \end{pmatrix}_{(I \cdot J) \times (I \cdot J)}$$

où  $B_J$  est défini comme :

$$B_J = \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 & \cdots \\ 1 & -2 & 1 & \cdots & 0 & \cdots \\ 0 & 1 & -2 & \cdots & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \\ 0 & \cdots & 0 & 0 & 1 & -2 \end{pmatrix}_{J \times J}$$

Les éléments non représentés par des ... désignent les matrices nulles.

### 2.3.7 Problème des Moindres Carrés avec Contraintes

Pour garantir que la vraie valeur du point d'échantillonnage soit égale ou proche de la valeur estimée du point d'échantillonnage, l'expression de HASM est transformée en un problème des moindres carrés avec des contraintes d'égalité.

Le modèle de surface basé sur l'Équation (2.63) peut être exprimé comme une approximation par moindres carrés avec contraintes :

$$\begin{cases} \min \|A \cdot Z - d\|_2 \\ \text{s.t. } S \cdot Z = k \end{cases} \quad (2.65)$$

où  $S$  et  $k$  sont respectivement les points d'échantillonnage et les valeurs correspondantes de  $z = f(x, y)$  aux points d'échantillonnage.

Pareillement, le modèle de surface basé sur l'Équation (39) peut être exprimé comme une approximation par moindres carrés avec contraintes :

$$\begin{cases} \min \|B \cdot Z - q\|_2 \\ \text{s.t. } S \cdot Z = k \end{cases} \quad (2.66)$$

La combinaison des systèmes (2.65) et (2.66) donne :

$$\begin{cases} \min \|A \cdot Z - d\|_2 \\ \min \|B \cdot Z - q\|_2 \\ \text{s.t. } S \cdot Z = k \end{cases} \quad (2.67)$$

où  $S$  et  $k$  représentent respectivement la matrice de coefficients des points d'échantillonnage et les valeurs de ces points d'échantillonnage. Ils peuvent être exprimés comme  $S(t, (i-1) \cdot J + j) = 1$  et  $k(t) = T_{i,j}$ , c'est-à-dire, la valeur du  $t$ -ième point d'échantillonnage  $(x_i, y_j)$  est  $T_{i,j}$ .

Pour résoudre le problème des moindres carrés contraints dans l'équation (2.65), un paramètre positif  $\lambda$ , qui est le poids des points d'échantillonnage, est introduit suivant la contribution des points d'échantillonnage à la surface simulée (YUE et WANG, 2010). Pour un  $\lambda$  donné, (2.67) peut être transformé en un problème des moindres carrés non contraint :

$$\min \left\| \begin{pmatrix} A \\ B \\ \lambda \cdot S \end{pmatrix} \cdot Z - \begin{pmatrix} d \\ q \\ \lambda \cdot k \end{pmatrix} \right\|_2 \quad (2.68)$$

En multipliant (2.68) par le vecteur suivant, on obtient :

$$\min \left\| \begin{pmatrix} A^T & B^T & (\lambda S)^T \end{pmatrix} \left( \begin{pmatrix} A \\ B \\ \lambda \cdot S \end{pmatrix} \cdot Z - \begin{pmatrix} d \\ q \\ \lambda \cdot k \end{pmatrix} \right) \right\|_2$$

Ce qui équivaut à :

$$(A^T \cdot A + B^T \cdot B + \lambda^2 \cdot S^T \cdot S) \cdot Z = (A^T \cdot d + B^T \cdot q + \lambda^2 \cdot S^T \cdot k)$$

D'où :

$$Z = (A^T \cdot A + B^T \cdot B + \lambda^2 \cdot S^T \cdot S)^{-1} (A^T \cdot d + B^T \cdot q + \lambda^2 \cdot S^T \cdot k) \quad (2.69)$$

L'équation (2.69) est la solution du vecteur d'inconnus des valeurs de l'attribut en tout point à l'intérieur du domaine, y compris les points échantillonnés sauf les points aux quatre bords du domaine. Les vérifications peuvent être faites en évaluant l'erreur de prédiction des valeurs aux points échantillonnés.

---

## MATÉRIEL ET MÉTHODES

---

### 3.1 Matériel / Outils

#### 3.1.1 Matériels utilisés

- Langage [R](#) pour la simulation des données
- Langage [Python](#) pour la programmation du code HASM
- Langage [Matlab](#) pour l'implémentation du code BME

#### 3.1.2 Outils utilisés

- Package [gstat](#) de R pour la modélisation géostatistique
- Package [BMElib2.0b](#) pour les prédictions avec la méthode BME
- Bibliothèque [distances](#) pour l'évaluation des distances maximales

### 3.2 Méthodes

#### 3.2.1 Mesure d'asymétrie

##### Définition

Les coefficients d'asymétrie (ou *skewness*) sont des mesures statistiques qui quantifient le degré d'asymétrie d'une distribution de données par rapport à une distribution normale. Dans la littérature il existe plusieurs formules pour mesurer le coefficient d'asymétrie :



### Coefficient d'asymétrie de Pearson

Le coefficient d'asymétrie classique (de Pearson) (PEARSON, 1894) est défini comme le rapport entre le troisième moment centré et l'écart-type élevé à la puissance trois. Ce coefficient mesure la symétrie d'une distribution par rapport à sa moyenne :

$$\text{Skew} = \frac{\mu - \text{mode}}{\sigma}$$

où  $\mu$  représente la moyenne et  $\sigma$  l'écart-type. Le coefficient d'asymétrie de Pearson est généralement compris dans l'intervalle  $[-1, 1]$ .

### Coefficient d'asymétrie de Fisher-Pearson

Le coefficient d'asymétrie de Fisher-Pearson (FISHER, 1930) est une version ajustée du coefficient de Pearson. Cette mesure est basée sur les moments :

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s} \right)^3$$

où  $n$  est le nombre d'observations,  $\bar{x}$  la moyenne de l'échantillon, et  $s$  l'écart-type de l'échantillon.

### Coefficient d'asymétrie de Bowley (Asymétrie des quartiles)

(BOWLEY, 1926) Ce coefficient, fondé sur les quartiles de la distribution, est utilisé pour évaluer l'asymétrie d'une distribution tout en étant moins sensible aux valeurs extrêmes :

$$\text{Asymétrie} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

où  $Q_1$ ,  $Q_2$ , et  $Q_3$  sont respectivement le premier, le deuxième (médiane), et le troisième quartile.

### Coefficient d'asymétrie de Kelley (Asymétrie de la moyenne du décile)

(KELLEY, 1947) a introduit une mesure d'asymétrie basée sur les déciles d'une distribution.

$$K = \frac{D_9 + D_1 - D_5}{D_9 - D_1}$$

où  $D_1$ ,  $D_5$ , et  $D_9$  sont respectivement le premier, le cinquième (médiane), et le neuvième décile.

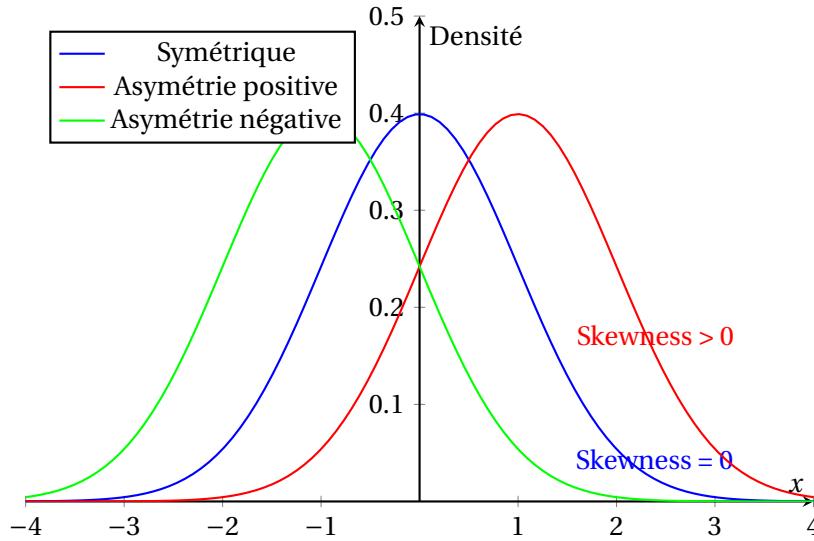


FIGURE 3.1 – Illustration des Coefficients d'Asymétrie ( symétrique, asymétrie positive et asymétrie négative ).

Dans votre étude, nous allons utiliser le coefficient d'asymétrie de Fisher ( $G_1$ ).

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} \frac{m^3}{m^{3/2}}$$

où  $m^3$  est le troisième moment centré et  $m^{3/2}$  est l'écart-type élevé à la puissance 3/2.

### 3.2.2 Mesure de la dépendance spatiale

la dépendance spatiale est généralement décrite par un indice unique qui permet de classer le degré de cette dépendance. Dans cette étude nous allons utiliser l'IDS ( Indice de Dépendance Spaciale) de (SEIDEL et OLIVEIRA, 2016) . L'indice de dépendance spatiale (IDS), proposé par (SEIDEL et OLIVEIRA, 2016), est exprimé par les formules suivantes pour les modèles sphérique, exponentiel et gaussien, respectivement :

$$IDS_{\text{sphérique}}(\%) = 0.375 \times \frac{C_1}{C_0 + C_1} \times 100 \times \frac{a}{0,5 \times MD}$$

$$IDS_{\text{exponentiel}}(\%) = 0.317 \times \frac{C_1}{C_0 + C_1} \times 100 \times \frac{a}{0,5 \times MD}$$

$$IDS_{\text{gaussien}}(\%) = 0.504 \times \frac{C_1}{C_0 + C_1} \times 100 \times \frac{a}{0,5 \times MD}$$

où  $a$  est la portée,  $C_0$  est l'effet de nugget,  $C_1$  est la contribution,  $C_0 + C_1$  est le seuil,  $MD$  est la distance maximale entre les points échantillonnés (la distance de la grille d'échantillonnage),  $(MF) \in \{0.375, 0.317, 0.504\}$  est le facteur du modèle de variogramme.

Ces indices permettent de classer la dépendance spatiale selon les propositions de (SEIDEL et OLIVEIRA, 2016), et ils sont catégorisés comme une dépendance spatiale faible, modérée ou forte .

TABLE 3.1 – Classification des IDS pour les modèles par (SEIDEL et OLIVEIRA, 2016)

Modèle	Faible	Modéré	Forte
<b>Sphérique</b>	$IDS (\%) \leq 7\%$	$7\% < IDS (\%) \leq 15\%$	$IDS (\%) > 15\%$
<b>Exponentiel</b>	$IDS (\%) \leq 6\%$	$6\% < IDS (\%) \leq 13\%$	$IDS (\%) > 13\%$
<b>Gaussien</b>	$IDS (\%) \leq 9\%$	$9\% < IDS (\%) \leq 20\%$	$IDS (\%) > 20\%$

### 3.2.3 Simulation des données

Dans notre étude, nous avons utilisé le modèle de variogramme **sphérique** et défini un domaine de travail carré de  $[1, 10] \times [1, 10]$  pour simuler les coordonnées et les valeurs d'attributs en fonction des tailles d'échantillonnage et du niveau de dépendance spatiale. Cela nous permet d'obtenir une distance maximale fixe,  $MD = 12,72792$ . Ensuite, nous avons utilisé des valeurs de portée (*range*) basées sur cette distance maximale :  $MD/8$  pour un niveau de dépendance *faible*,  $MD/4$  pour un niveau de dépendance *modéré*, et  $MD/2$  pour un niveau de dépendance *fort*. Ces choix correspondent respectivement aux indices de dépendance spatiale (*IDS*) suivants, selon la formule de Seidel :  $IDS = 0,9375$  pour niveau de dépendance faible,  $IDS = 7,5$  pour niveau de dépendance modéré et  $IDS = 33,75$  pour niveau de dépendance fort. Le plan de simulation se présente comme suit :

#### Étape 1 : Taille de l'échantillon

Nous avons généré 1000 attributs à partir des coordonnées d'entrée en utilisant différentes tailles d'échantillons avec la fonction `predict` prédéfinie du langage R. Selon les besoins de l'étape 4, cinq tailles d'échantillons différentes (64, 80, 169, 250 et 400) ont été utilisées pour générer les ensembles de données.

#### Étape 2 : Échantillonnage des points

Pour chaque taille d'échantillon ( $n$ ), un maillage régulier a été effectué sur chaque axe, avec des pas qui ne sont pas nécessairement identiques pour les deux axes, dans un domaine carré  $[1, 10] \times [1, 10]$ , afin d'échantillonner tous les points réguliers du domaine. Cela nous permet d'évaluer la distance maximale ( $MD$ ) entre les points échantillonnés par l'utilisation du package `library(distances)`.

#### Étape 3 : Estimation des paramètres du variogramme

La distance maximale ( $MD$ ) calculée a été utilisée pour générer le niveau de dépendance spatiale. Nous avons utilisé le variogramme théorique (seuil = 1) et supposé pour chaque variogramme, indépendamment de la dépendance spatiale, que la distance au-delà de laquelle les données ne sont plus corrélées était 0.5 du seuil de la distance maximale ( $MD$ ). Nous avons fixé la dépendance spatiale faible à 0,9375%, la dépendance spatiale modérée

à 7.5% et la dépendance spatiale forte à 33,75% en utilisant la classification basée sur l'IDS pour le modèle sphérique proposée par (SEIDEL et OLIVEIRA, 2016). Les tableaux 3.2, 3.3 et 3.4 présentent les paramètres du variogramme en fonction de la taille de l'échantillon et de la dépendance spatiale.

Étant donné que la méthode HASM nécessite des conditions aux bords, considérées comme des données parmi les attributs, nous avons dû simuler l'ensemble des valeurs internes des attributs pour chaque donnée (45 données au total, chacune ayant les coordonnées  $x$ ,  $y$  et 1000 exemplaires pour  $n$  éléments de tailles quelconques) dans le domaine carré. Ensuite, des valeurs fixes pour les attributs ont été générées le long des quatre côtés (bordures) du domaine, considérées comme les conditions aux bords. Ces valeurs fixes ont été ajoutées aux 45 données internes pour assurer la complétude des informations sur toutes les coordonnées du domaine. Après cet ajout, chaque exemplaire contiendra une taille d'éléments parmi {64, 80, 169, 250, 400}. Pour générer ces données, nous avons utilisé les mêmes paramètres que ceux décrits précédemment.

**TABLE 3.2** – Paramètres du variogramme pour la dépendance spatiale faible en fonction de la taille de l'échantillon

Échantillon	MD	Portée	Psill	Nugget	SD%
64	12.72792	1.59099	0.1	0.9	0.9375
80	12.727923	1.59099	0.1	0.9	0.9375
169	12.72792	1.59099	0.1	0.9	0.9375
250	12.72792	1.59099	0.1	0.9	0.9375
400	12.72792	1.59099	0.1	0.9	0.9375

**TABLE 3.3** – Paramètres du variogramme pour la dépendance spatiale modérée en fonction de la taille de l'échantillon

Échantillon	MD	Portée	Psill	Nugget	SD%
64	12.72792	3.18198	0.4	0.6	7.5
80	12.72792	3.18198	0.4	0.6	7.5
169	12.72792	3.18198	0.4	0.6	7.5
250	12.72792	3.18198	0.4	0.6	7.5
400	12.72792	3.18198	0.4	0.6	7.5

**TABLE 3.4** – Paramètres du variogramme pour la dépendance spatiale forte en fonction de la taille de l'échantillon

Échantillon	MD	Portée	Psill	Nugget	SD%
64	12.72792	6.36396	0.9	0.1	33.75
80	12.72792	6.36396	0.9	0.1	33.75
169	12.72792	6.36396	0.9	0.1	33.75
250	12.72792	6.36396	0.9	0.1	33.75
400	12.72792	6.36396	0.9	0.1	33.75

#### Étape 4. Génération des données

Le package `gstat` de `R`, qui permet la modélisation géostatistique multivariable, les prévisions et les simulations, a été utilisé pour simuler les attributs à partir des points d'échantillonnage réguliers et des paramètres de variogramme estimés. Pour chaque taille d'échantillon, nous avons simulé 3 attributs pour chaque niveau de dépendance spatiale (SD), avec 3 degrés de symétrie pour chaque niveau de dépendance. Cela nous a permis de générer 9 attributs différents ( $3 \times 3$ ) pour chaque taille d'échantillon. Étant donné que nous avons travaillé avec 5 tailles d'échantillon distinctes, cela a conduit à un total de 45 jeux de données ( $5 \times 9$ ). Le processus a été répété 1000 fois, produisant pour chaque taille d'échantillon 9 ensembles de données. Chaque ensemble de données correspond à un attribut caractéristique répété 1000 fois, associé aux coordonnées des points d'échantillon générés à l'étape 3.

La formule pour rechercher des données en fonction du skewness imposé est :

$$|\text{skew} - \text{skewness}| < 0.3 \text{ où } \text{skew} \in \{-1, 0, 1\}$$

On attribue  $\text{skew} = -1$  pour une distribution négative,  $\text{skew} = 0$  pour une distribution symétrique et  $\text{skew} = 1$  pour une distribution positive. Cette formule garantit que la valeur du skewness sera proche de la distribution recherchée, avec un écart maximal de 0.3 par rapport au  $\text{skew}$ .

#### Étape 5 : Application des données avec BME et HASM

Pour évaluer les performances des méthodes BME et HASM, nous avons divisé chacune des 45 données en deux parties : une partie pour l'entraînement des modèles et une autre pour le test. Nous avons utilisé 70% des données pour l'entraînement et les 30% restantes pour les tests. De plus, les modèles de ces méthodes peuvent traiter les données de manière aléatoire. Cependant, pour garantir l'inclusion des conditions aux bords, nous avons utilisé les 70% des premières données, incluant ces conditions, pour l'entraînement des modèles.

La compilation du code BME pour la prédiction a été réalisée en Matlab en utilisant le package `BMElib2.0b`, version 2, développé par (SERRE, 2008); (CHRISTAKOS, BOGAERT et SERRE, 2002). La fonction `BMEintervalMode` de `BMElib2.0b` a été utilisée pour prédire les données aux points de test considérés comme non échantillonnés dans le domaine. Cette fonction fait également appel à des dépendances d'autres fonctions du package et

renvoie comme sortie la prédiction des attributs aux coordonnées des points tests (non échantillonnés). Les arguments que prend cette fonction sont détaillés dans le tableau 3.5 :

TABLE 3.5 – Paramètres pour le calcul du BME

Paramètre	Définition
ck	Matrice des coordonnées pour les emplacements d'estimation
ch	Matrice des coordonnées pour les emplacements des <a href="#">hard data</a>
cs	Matrice des coordonnées pour les emplacements des <a href="#">soft data</a>
zh	Vecteur des valeurs pour les <a href="#">hard data</a>
a	Vecteur des valeurs pour la borne inférieure des intervalles
b	Vecteur des valeurs pour la borne supérieure des intervalles
covmodel	Nom du modèle de covariance
covparam	Paramètres du modèle de covariance
nhmax	Nombre maximum de valeurs de <a href="#">hard data</a> par emplacement d'estimation
nsmax	Nombre maximum de valeurs de <a href="#">soft data</a> par emplacement d'estimation
dmax	Distance maximale entre un emplacement d'estimation et les <a href="#">hard/soft data</a> existantes
order	Ordre du décalage polynômial le long des axes spatiaux aux emplacements d'estimation

Étant donné que la méthode [HASM](#) ne prend pas en compte les données molles ([soft data](#)), nous avons défini `nsmax=0` lors de la compilation du code [BME](#) afin d'assurer l'uniformité dans l'application des données.

La méthode [HASM](#) fournit des valeurs de prédiction pour l'ensemble du domaine, à l'exception des quatre bords. Si nous faisons une discrétisation de  $I$  points sur l'axe des abscisses et de  $J$  points sur l'axe des ordonnées, nous aurons  $2I \times 2(J - 2) = 2(I - 2) \times J$  points au bord comme conditions au bord.

Nous avons programmé nous-mêmes le code pour la méthode [HASM](#), qui compte [700 lignes](#) écrites en Python conformément à sa méthodologie. Ce code gère des [entrées](#) et des [sorties](#), dont les paramètres sont consignés dans le tableau 3.6.

TABLE 3.6 – Paramètres pour le calcul de la méthode HASM

Entrées	
Arguments	Définitions
nom_fichier	Adresse du fichier de données Excel (.xlsx)
forma	Format d'enregistrement des valeurs prédites
pourcentage	Pourcentage de données à considérer pour l'entraînement du modèle
i	Position de l'attribut à étudier dans la donnée parmi les 1000
$\lambda$	Poids des points d'échantillonnage (par défaut : 1000)
Sorties	
Les valeurs retournées	Définitions
X	Vecteur coordonnées des points en abscisse
Y	vecteur coordonnées des points en ordonnée
Solu	vecteur valeurs prédites des attributs aux points concernés
donnee_initiale	vecteur valeurs des attributs de la base de données initiale
colonne0	vecteur positions des points en abscisse
colonne1	vecteur positions des points en ordonnée
RMSE	RMSE de prédiction des points à l'intérieur du domaine
MAE	MAE de prédiction des points à l'intérieur du domaine
RMSE_rela	RMSE relative seulement aux points échantionnés
R	Coefficient de détermination après prédiction
real_RMSE	RMSE relative seulement aux valeurs prédites
real_MAE	MAE relative seulement aux valeurs prédites

Les résultats de prédiction obtenus par les deux méthodes ont été enregistrés en fichier excel (xlsx) dans leurs répertoires respectifs.

### Étape 6 : Critères de comparaison des mesures d'erreur : RMSE et MAE

Les critères de comparaison tels que le Root Mean Square Error (RMSE) et le Mean Absolute Error (MAE) sont couramment utilisés pour évaluer la précision des modèles de prédiction, notamment en géostatistique et en modélisation spatiale.

#### Root Mean Square Error (RMSE)

Le RMSE est une mesure statistique qui représente la racine carrée de la moyenne des carrés des écarts entre les valeurs observées et les valeurs prédites par un modèle. Formellement, il est défini comme suit :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

où  $y_i$  représente les valeurs observées,  $\hat{y}_i$  les valeurs prédites, et  $n$  le nombre de points de données. Le RMSE est sensible aux grandes erreurs si les erreurs présentent une variance élevée.

### Mean Absolute Error (MAE)

Le MAE, quant à lui, est la moyenne des valeurs absolues des écarts entre les valeurs observées et prédites. Il est calculé comme suit :

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Le MAE est plus facile à interpréter, donne une mesure directe de l'erreur moyenne en unités de la variable mesurée et n'accentue pas les grandes erreurs, ce qui le rend plus robuste en présence de valeurs aberrantes comparativement au RMSE.

Pour chaque  $i$ , considérons l'écart  $e_i = y_i - \hat{y}_i$ . Le MAE et le RMSE peuvent alors être réécrits comme :

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |e_i|$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2}$$

### Inégalité de Jensen

L'inégalité de Jensen stipule que pour une fonction convexe  $f(x)$ , on a :

$$f\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \leq \frac{1}{n} \sum_{i=1}^n f(x_i)$$

$$f(\mathbb{E}(X)) \leq \mathbb{E}[f(X)]$$

Dans notre cas, la fonction  $f(x) = x^2$  est convexe. Appliquons cette inégalité à notre situation :

$$\left(\frac{1}{n} \sum_{i=1}^n |e_i|\right)^2 \leq \frac{1}{n} \sum_{i=1}^n e_i^2$$

Prendre la racine carrée des deux côtés donne :

$$\frac{1}{n} \sum_{i=1}^n |e_i| \leq \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2}$$

Cela se traduit par :

$$\text{MAE} \leq \text{RMSE}$$

Cette inégalité montre que le MAE est toujours inférieur ou égal au RMSE. Le seul cas où ils sont égaux est lorsque toutes les erreurs  $e_i$  sont identiques, c'est-à-dire lorsque les erreurs sont constantes, et il n'y a pas de variance dans les erreurs.



Le plan de simulation présenté ci-dessous est illustré graphiquement comme suit :

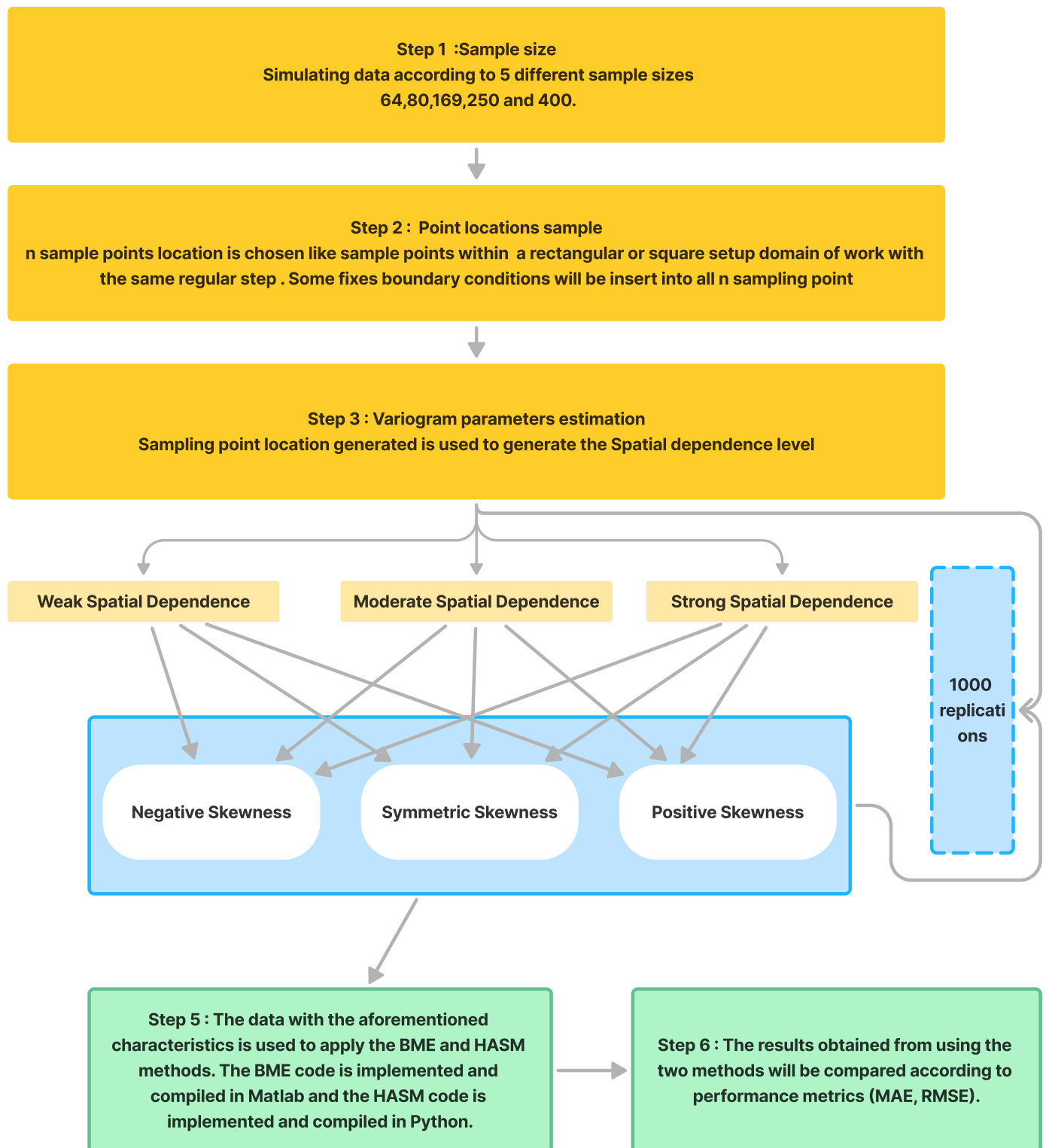


FIGURE 3.2 – Simulation Flowchart

---

## RÉSULTATS ET DISCUSSIONS

---

### 4.1 Résultats des analyses

Les résultats obtenus à partir de l'application des 45 bases de données ont généré 90 ensembles de données comportant deux colonnes, RMSE et MAE, chacune contenant 1000 lignes. Parmi ces ensembles, 45 proviennent des prédictions effectuées avec la méthode BME, tandis que les 45 autres sont issus des prédictions réalisées avec la méthode HASM. Chaque ensemble de données est lié à des catégories spécifiques : la taille de l'échantillon (64, 80, 169, 250, 400), le niveau d'asymétrie (négatif, symétrique, positif) et le niveau de dépendance spatiale (faible, modéré, fort).

#### 4.1.1 Statistique descriptives des résultats

Pour représenter les données concernant les critères d'erreur, nous utiliserons comme paramètre de position la moyenne et comme paramètre de dispersion l'écart-type. C'est ainsi que nous avons regroupé les données par chaque catégorie de paramètres, comme sur les figures 4.1 et 4.2.

TABLE 4.1 – Average value of MAE and RMSE determined from BME

SD	MAE										
	Skew	64		80		169		250		400	
	sk	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
W	Ns	1.28	0.17	1.28	0.15	1.24	0.11	1.25	0.09	1.22	0.08
	Sy	1.21	0.18	1.22	0.17	1.22	0.11	1.25	0.10	1.22	0.08
	Ps	1.18	0.19	1.17	0.17	1.2	0.12	1.17	0.10	1.16	0.09
M	Ns	1.30	0.20	1.36	0.20	1.17	0.17	1.13	0.15	1.06	0.14
	Sy	1.38	0.21	1.40	0.21	1.14	0.16	1.13	0.15	1.11	0.15
	Ps	1.32	0.22	1.34	0.20	1.17	0.18	1.23	0.17	1.05	0.18
S	Ns	1.18	0.32	1.34	0.36	0.78	0.18	1.23	0.37	0.90	0.28
	Sy	1.12	0.27	1.10	0.25	0.95	0.25	1.16	0.30	0.7	0.19
	Ps	1.56	0.33	1.66	0.38	1.14	0.27	0.80	0.18	0.72	0.22
RMSE											
W	Ns	1.46	0.18	1.47	0.16	1.46	0.12	1.48	0.10	1.46	0.08
	Sy	1.45	0.20	1.46	0.18	1.48	0.13	1.50	0.10	1.49	0.09
	Ps	1.46	0.23	1.46	0.20	1.49	0.14	1.46	0.12	1.45	0.11
M	Ns	1.56	0.22	1.63	0.21	1.42	0.19	1.38	0.17	1.31	0.16
	Sy	1.66	0.23	1.69	0.23	1.40	0.19	1.39	0.17	1.37	0.17
	Ps	1.63	0.26	1.65	0.24	1.46	0.22	1.51	0.21	1.32	0.23
S	Ns	1.46	0.39	1.62	0.41	0.98	0.23	1.48	0.41	1.14	0.35
	Sy	1.38	0.32	1.37	0.30	1.17	0.29	1.40	0.33	0.96	0.23
	Ps	1.87	0.34	1.93	0.39	1.36	0.28	0.99	0.21	0.92	0.28

W : Weak Spatial Dependence, M : Moderate Spatial Dependence, S : Strong Spatial Dependence

Sy : Symmetric, Ps : Positive skewed, Ns : Negative skewed

TABLE 4.2 – Average value of MAE and RMSE determined from HASM

SD	Skew	MAE									
		64		80		169		250		400	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
W	Ns	1.06	0.18	0.94	0.14	0.97	0.15	1.03	0.12	0.95	0.13
	Sy	0.96	0.17	0.94	0.15	0.98	0.13	1.00	0.11	0.99	0.14
	Ps	0.98	0.19	0.95	0.16	0.96	0.14	0.94	0.10	0.93	0.12
M	Ns	1.15	0.22	1.11	0.19	1.02	0.18	1.20	0.18	0.91	0.13
	Sy	1.21	0.25	1.10	0.21	0.95	0.15	1.26	0.18	0.93	0.14
	Ps	1.16	0.26	1.09	0.22	0.99	0.19	1.43	0.23	0.92	0.16
S	Ns	1.03	0.23	1.03	0.21	1.12	0.28	2.76	0.47	0.85	0.17
	Sy	1.18	0.32	1.34	0.38	0.91	0.21	2.60	0.45	0.92	0.21
	Ps	1.03	0.23	1.05	0.22	0.88	0.21	2.31	0.50	1.06	0.33
		RMSE									
		64		80		169		250		400	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
W	Ns	1.24	0.19	1.13	0.16	1.19	0.16	1.25	0.13	1.25	0.14
	Sy	1.17	0.20	1.15	0.17	1.21	0.15	1.24	0.13	1.23	0.16
	Ps	1.26	0.24	1.23	0.20	1.23	0.17	1.21	0.13	1.19	0.15
M	Ns	1.35	0.23	1.32	0.20	1.23	0.20	1.46	0.20	1.13	0.16
	Sy	1.45	0.27	1.34	0.23	1.17	0.18	1.55	0.20	1.16	0.16
	Ps	1.44	0.29	1.37	0.26	1.25	0.23	1.74	0.25	1.16	0.19
S	Ns	1.23	0.25	1.23	0.23	1.29	0.28	4.21	0.65	1.03	0.19
	Sy	1.39	0.34	1.54	0.39	1.09	0.23	4.08	0.63	1.10	0.23
	Ps	1.26	0.26	1.28	0.25	1.07	0.24	3.24	0.76	1.26	0.36

W : Weak Spatial Dependence, M : Moderate Spatial Dependence, S : Strong Spatial Dependence

Sy : Symmetric, Ps : Positive skewed, Ns : Negative skewed

Nous avons retenu deux décimales pour enregistrer les résultats dans le tableau.

TABLE 4.3 – PERMANOVA results

Source	MAE			RMSE		
	Df	F value	p-value	Df	F value	p-value
Method	1	5277.5	0.001 ***	1	4112.6	0.001 ***
Size	4	110.6	0.001 ***	4	46.5	0.001 ***
SD	2	14.1	0.001 ***	2	0.6	0.560
Skew	2	1.4	0.202	2	0.2	0.805
Method :Size	4	165.6	0.001 ***	4	267.4	0.001 ***
Method :SD	2	1.1	0.327	2	1.4	0.274
Size :SD	8	14.1	0.001 ***	8	25.7	0.001 ***
Method :Skew	2	3.4	0.038 *	2	1.0	0.369
Size :Skew	8	2.3	0.024 *	8	3.4	0.002 **
SD :Skew	4	5.9	0.001 ***	4	5.0	0.001 ***
Method :Size :SD	8	9.6	0.001 ***	8	24.9	0.001 ***
Method :Size :Skew	8	1.3	0.219	8	3.6	0.001 ***
Method :SD :Skew	4	2.2	0.069 .	4	2.6	0.043 *
Size :SD :Skew	16	2.6	0.002 **	16	2.1	0.005 **
Method :Size :SD :Skew	16	0.9	0.527	16	0.7	0.829

SD : Spatial Dependence

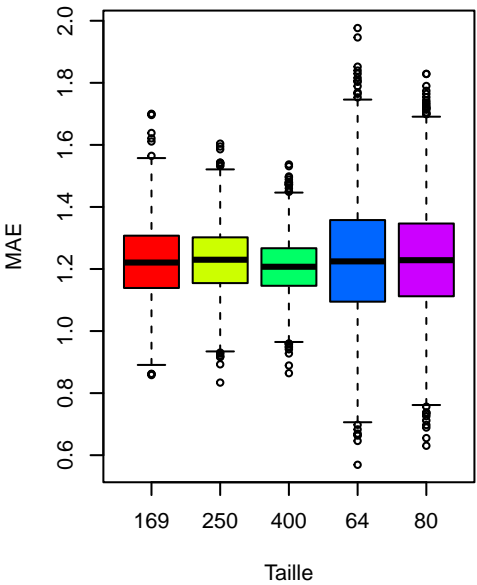
Le tableau 4.3 présente les résultats de l'analyse de PERMANOVA sur l'ensemble des données. Les variables explicatives considérées sont : la méthode avec les modalités ("BME", "HASM"), la taille d'échantillon avec les modalités ("64", "80", "169", "250", "400"), le niveau de dépendance spatiale avec les modalités ("Weak", "Moderate", "Strong") et le degré d'asymétrie ("Négative", "Symétrique", "Positive").

TABLE 4.4 – Effet de la taille et niveau de dépendance sur la prédiction de BME : PERMANOVA results

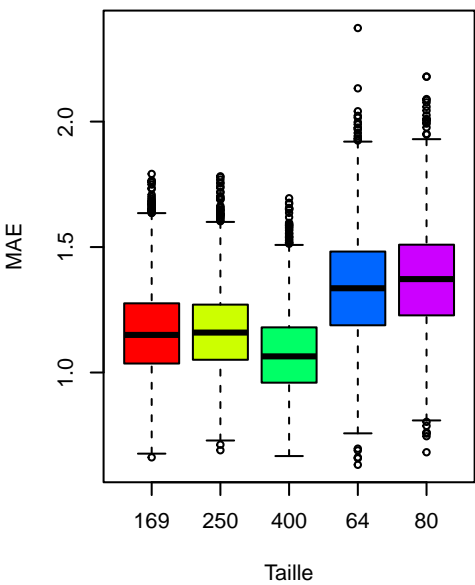
Source	Df	F value	p-value
Size	4	209.565	0.001 ***
SD	2	10.210	0.001 ***
Size :SD	8	12.988	0.001 ***

SD : Spatial Dependence

Dépendance: weak



Dépendance: moderate



Dépendance: strong

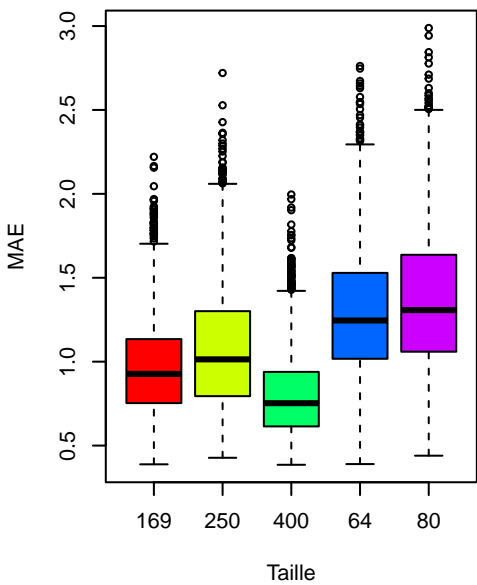


FIGURE 4.1 – Boxplots de MAE\_BME des niveaux de dépendances en fonction des tailles

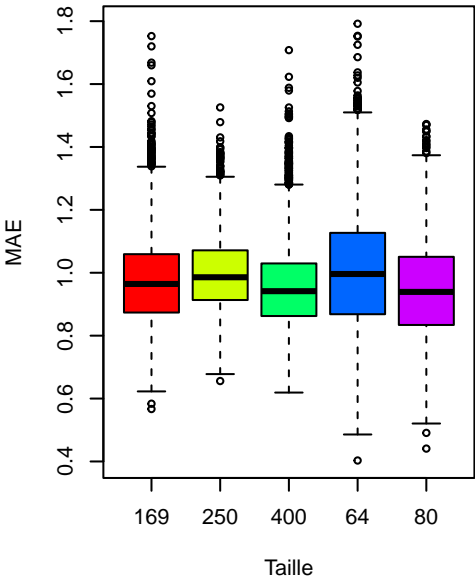
Les tableaux 4.4 et 4.5 analysent les effets des paramètres significatifs du tableau 4.3, de la taille d'échantillon et du niveau de dépendance spatiale sur chacune des méthodes.

TABLE 4.5 – Effet de la taille et niveau de dépendance sur la prédiction de HASM : PERMANOVA results

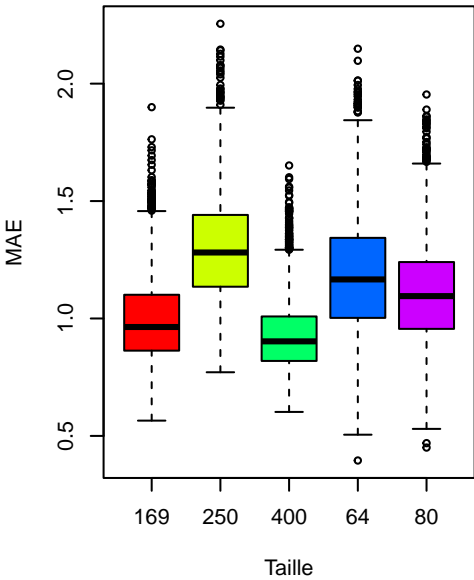
Source	Df	F value	p-value
Size	4	42.192	0.001 ***
SD	2	3.755	0.031 *
Size :SD	8	8.915	0.001 ***

SD : Spatial Dependence

Dépendance: weak



Dépendance: moderate



Dépendance: strong

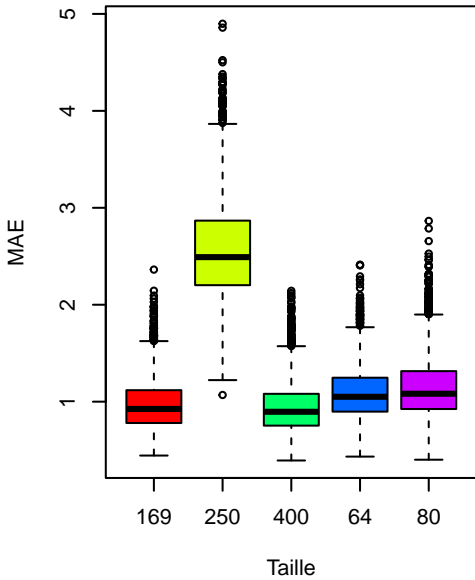


FIGURE 4.2 – Boxplots de MAE\_HASM des niveaux de dépendances en fonction des tailles



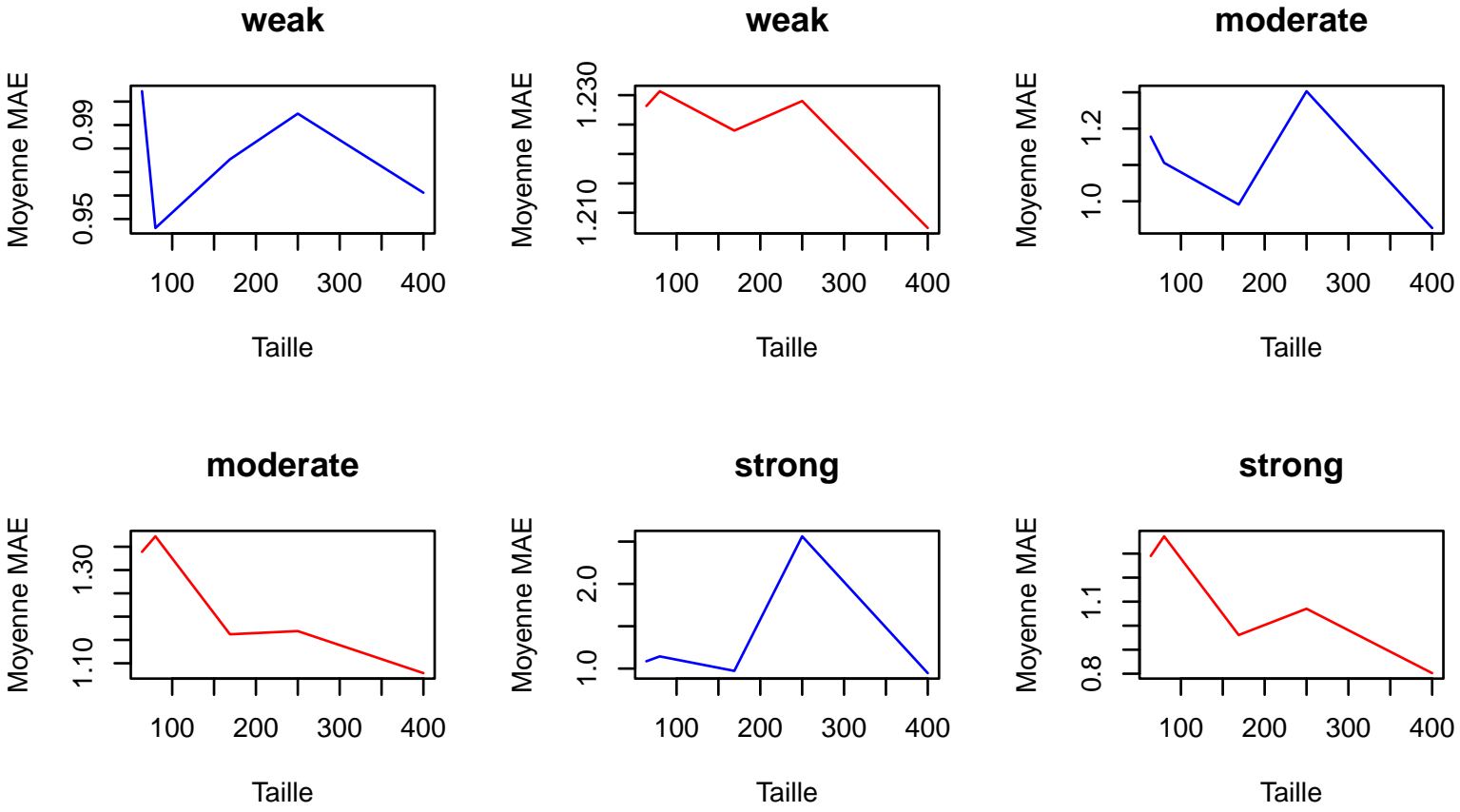


FIGURE 4.3 – Boxplots de MAE\_HASM des niveaux de dépendances en fonction des tailles

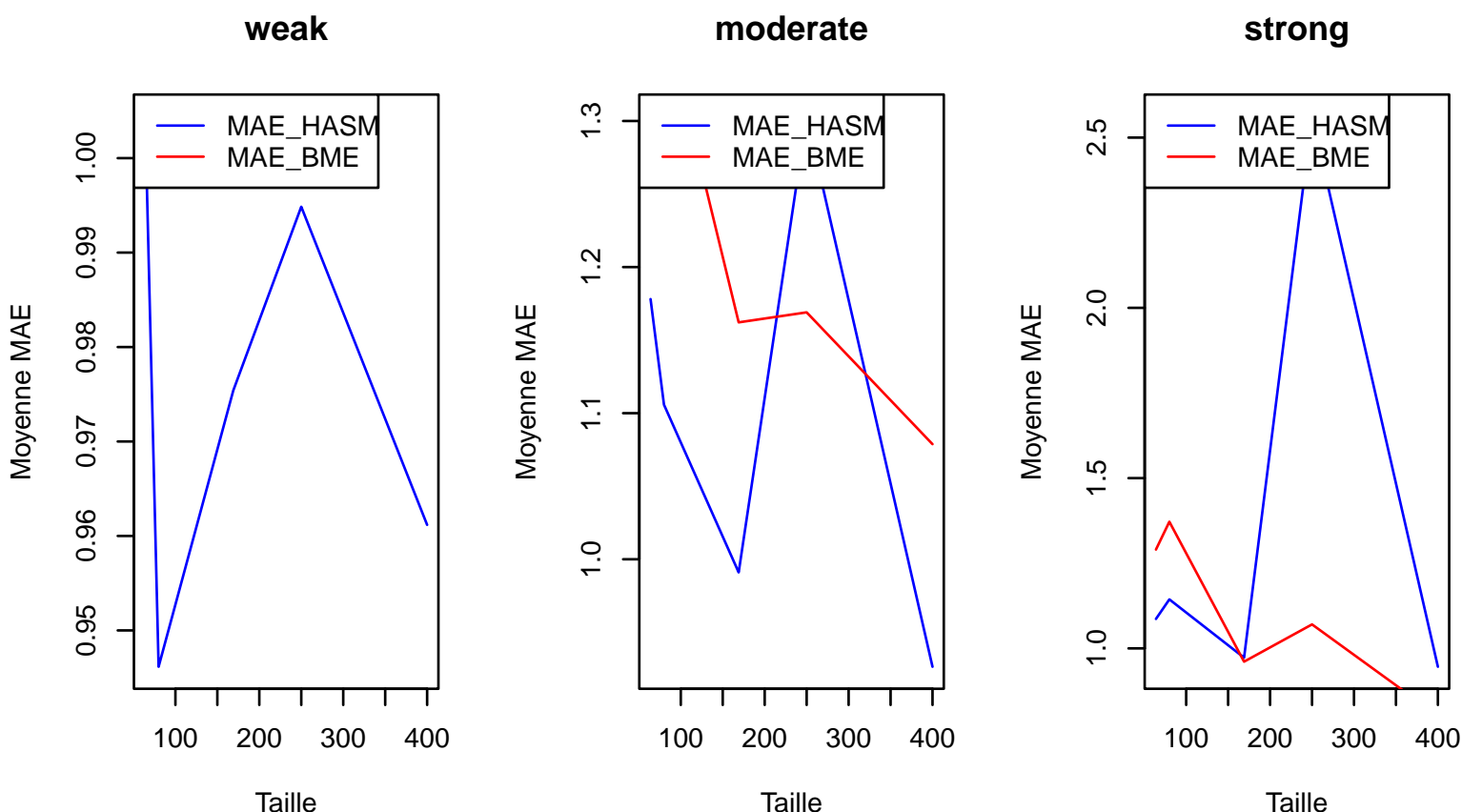


FIGURE 4.4 – Boxplots de MAE\_HASM des niveaux de dépendances en fonction des tailles

## 4.2 Discussion

Ces résultats correspondent aux valeurs moyennes des erreurs de prédiction des méthodes BME et HASM sur les différentes données, selon la **taille d'échantillon**, le **niveau de dépendance spatiale** et le **degré d'asymétrie**. Pour effectuer une analyse approfondie, nous allons utiliser la méthode **PERMANOVA** (Permutation Multivariate Analysis of Variance). Le tableau 4.3 indique que la méthode, la taille d'échantillon et le niveau de dépendance spatiale ont un effet hautement significatif sur l'erreur de prédiction MAE. De même, les **interactions** (Méthode : Asymétrie), (Taille : Dépendance spatiale), (Dépendance spatiale : Asymétrie) et (Méthode : Taille : Dépendance spatiale) sont également significatives. En revanche, les facteurs ayant un impact significatif sur le RMSE sont : la méthode, la taille d'échantillon et les interactions (Taille : Dépendance spatiale), (Dépendance spatiale : Asymétrie), (Méthode : Taille : Dépendance spatiale) et (Méthode : Taille : Asymétrie). Les deux erreurs sont indifférentes à la variation du degré d'asymétrie. Nous allons donc travailler sur les méthodes, les différentes tailles et le niveau de dépendance spatiale, en considérant le critère **MAE** ( $MAE < RMSE$ ) pour l'analyse des résultats. Pour cela, des analyses **PERMANOVA** ont été réalisées individuellement sur chacune des méthodes, comme le montrent les tableaux 4.5 et 4.4, en fonction des autres paramètres importants. Le résumé issu de cette analyse révèle que la taille d'échantillon et le niveau de dépendance spatiale impactent significativement les méthodes de prédiction. Des représentations en

boxplot des deux méthodes en fonction du niveau de dépendance et de la taille sont présentées (voir figures 4.9 et 4.2). Ces représentations montrent généralement une légère supériorité de **HASM** par rapport à **BME**. Pour confirmer ces observations, les courbes moyennes de MAE des deux méthodes sont tracées sur différentes figures (4.3) et sur les mêmes figures (4.4). Nous avons enfin calculé la moyenne totale de **MAE** sur **BME** ( $MAE = 1.182439$ ) et celle sur **HASM** ( $MAE = 1.139815$ ). Cela témoigne de la concurrence rude entre les deux méthodes, puisque **HASM** est légèrement supérieur à **BME** ( $MAE\_HASM < MAE\_BME$ ).

---

## CONCLUSION ET PERSPECTIVES

---

En conclusion, la comparaison entre BME (Bayesian Maximum Entropy) et HASM (High Accuracy Surface Modeling) révèle que ces deux approches sont efficaces pour la modélisation spatiale. BME s'appuie sur une intégration des probabilités bayésiennes et l'entropie maximale pour mesurer l'incertitudes des données, ce qui en fait une méthode robuste face au Hard et soft data. HASM, quant à lui, se distingue par sa précision élevée dans la modélisation de surfaces continues, souvent plus adapté pour des données bien définies et denses. Le choix entre ces deux méthodes dépend du critère MAE. Ce critère montre que le HASM est légèrement supérieur au BME. Les limites de notre étude mettent en lumière plusieurs défis et perspectives. Tout d'abord, il est crucial de trouver des moyens d'intégrer les données dures et souples, car cela pourrait améliorer la pertinence des comparaisons entre les méthodes BME et HASM. Ensuite, adapter la méthode HASM pour inclure la dimension temporelle enrichirait notre compréhension des phénomènes dynamiques. Il serait également intéressant d'utiliser les méthodes des éléments finis (FEM) et des volumes finis (FVM) pour discrétiser des équations différentielles partielles (EDP) sur des domaines complexes avec des maillages irréguliers. Cependant, il est important de noter que la résolution des EDP nécessite des conditions aux bords, ce qui peut limiter la flexibilité des résultats.

---

## BIBLIOGRAPHIE

---

- ABDENNOUR, Mohamed Amine (2021). « Variabilité spatio-temporelle de la salinisation des sols du périmètre irrigué du Ziban (Biskra)–Apport de la géostatistique et de la télédétection ». Thèse de doct. Université Mohamed Khider de Biskra.
- ADAM-POUPART, Ariane et al. (2014). « Spatiotemporal modeling of ozone levels in Quebec (Canada) : a comparison of kriging, land-use regression (LUR), and combined Bayesian maximum entropy–LUR approaches ». In : *Environmental health perspectives* 122.9, p. 970-976.
- BAYAT, Bardia, Mohsen NASSERI et Gholamreza NASER (2014). « Improving Bayesian maximum entropy and ordinary Kriging methods for estimating precipitations in a large watershed : a new cluster-based approach ». In : *Canadian Journal of Earth Sciences* 51.1, p. 43-55.
- BAYES, Thomas (1958). « An essay towards solving a problem in the doctrine of chances ». In : *Biometrika* 45.3-4, p. 296-315.
- BOWLEY, Arthur Lyon (1926). *Elements of statistics*. 8. King.
- BURG, John Parker (1972). « The relationship between maximum entropy spectra and maximum likelihood spectra ». In : *Geophysics* 37.2, p. 375-376.
- CHEN, Chuanfa, Yanyan LI et Tianxiang YUE (2013). « Surface modeling of DEMs based on a sequential adjustment method ». In : *International Journal of Geographical Information Science* 27.7, p. 1272-1291.
- CHRISTAKOS, George, Patrick BOGAERT et Marc SERRE (2002). *Temporal GIS : advanced functions for field-based applications*. Springer Science & Business Media.
- CHRISTAKOS, George et Xinyang LI (1998). « Bayesian maximum entropy analysis and mapping : a farewell to kriging estimators? ». In : *Mathematical Geology* 30, p. 435-462.
- COMAN, Adriana (2008). « Modélisation spatio-temporelle de la pollution atmosphérique urbaine à partir d'un réseau de surveillance de la qualité de l'air ». Thèse de doct. Université Paris-Est.
- DAHLSTRÖM, Bengt (2003). « La conception d'un réseau d'observation climatologique-Un point de vue européen ». In : *La Météorologie* 2003.40, p. 59-63.
- DELANG, Marissa N et al. (2021). « Mapping yearly fine resolution global surface ozone through the Bayesian maximum entropy data fusion of observations and model output for 1990–2017 ». In : *Environmental science & technology* 55.8, p. 4389-4398.
- EVANS, David J (1968). « The use of pre-conditioning in iterative methods for solving linear equations with symmetric positive definite matrices ». In : *IMA Journal of Applied Mathematics* 4.3, p. 295-314.
- FISHER, Ronald Aylmer (1930). « Moments and product moments of sampling distributions ». In : *Proceedings of the London Mathematical Society* 2.1, p. 199-238.
- GAO, Zekun et al. (2022). « Bayesian maximum entropy interpolation of sea surface temperature data : a comparative assessment ». In : *International Journal of Remote Sensing* 43.1, p. 148-166.

- GONGNET, Emmanuel Ehnon et al. (2022). « Spatial prediction of soil organic matter in Adingnigon (Benin) using Bayesian Maximum Entropy (BME) ». In : *African Journal of Applied Statistics* 9.1, p. 1279-1295.
- GONGNET, Emmanuel Ehnon et al. (2023). « Practical considerations on data patterns in Bayesian maximum entropy estimation : a systematic and critical review. » In.
- GUEROUI, Yacine (2014). « Caractérisation Hydrochimique et Bactériologique des Eaux Souterraines de L'aquifère Superficiel de la Plaine de Tamlouka (Nord-Est Algérie) ». Thèse de doct.
- GUIASU, Silviu et Abe SHENITZER (1985). « The principle of maximum entropy ». In : *The mathematical intelligencer* 7, p. 42-48.
- HE, Junyu et George CHRISTAKOS (2023). « Bayesian maximum entropy ». In : *Encyclopedia of Mathematical Geosciences*. Springer, p. 71-79.
- HENINE, H et al. (2015). « Spatialisation de la mesure de la teneur en eau du sol à l'échelle du champ agricole, en utilisant le suivi géoélectrique et la méthode géostatistique ». In.
- JAYNES, Edwin T (1982). « On the rationale of maximum-entropy methods ». In : *Proceedings of the IEEE* 70.9, p. 939-952.
- KANG, Dengjie, Wenkai CHEN et Yijiao JIA (2024). « Bayesian maximum entropy interpolation analysis for rapid assessment of seismic intensity using station and ground motion prediction equations ». In : *Frontiers in Earth Science* 12, p. 1394937.
- KANG, Jiyoung et al. (2021). « Bayesian estimation of maximum entropy model for individualized energy landscape analysis of brain state dynamics ». In : *Human brain mapping* 42.11, p. 3411-3428.
- KARAMI, Hojat et al. (2019). « Prediction of scour pattern around hydraulic structures using geostatistical methods ». In : *Arabian Journal of Geosciences* 12, p. 1-14.
- KELLEY, Truman Lee (1947). *Fundamentals of statistics*. 2. Harvard University Press.
- KULLBACK, Solomon (1968). « Probability densities with given marginals ». In : *The Annals of Mathematical Statistics* 39.4, p. 1236-1243.
- LEBEL, T et al. (1996). « La distribution spatio-temporelle des pluies au Sahel : apports de l'expérience EPSAT-Niger ». In : *IAHS PUBLICATION*, p. 77-98.
- LI, Shuxian (2015). « Modélisation spatio-temporelle pour l'esca de la vigne à l'échelle de la parcelle ». Thèse de doct. Université de Bordeaux.
- LISEIKIN, Vladimir D (2006). *A computational differential geometry approach to grid generation*. Springer Science & Business Media.
- LOUVET, Samuel et al. (2011). « Variabilité spatio-temporelle passée et future de la pluie sur le bassin du Bani en Afrique de l'Ouest ». In : *Hydro-climatology : variability and change* 344, p. 125-130.
- MODIS, K, KI VATALIS et Ch SACHANIDIS (2013). « Spatiotemporal risk assessment of soil pollution in a lignite mining region using a Bayesian maximum entropy (BME) approach ». In : *International Journal of Coal Geology* 112, p. 173-179.
- PASADAS, Miguel et Miguel Luis RODRÍGUEZ (2009). « An approximation method with data selection process ». In : *Mathematics and Computers in Simulation* 79.12, p. 3567-3576.
- PEARSON, Karl (1894). « Contributions to the mathematical theory of evolution ». In : *Philosophical Transactions of the Royal Society of London. A* 185, p. 71-110.
- POUPRY, Sylvain (2023). « Contribution à la conception et à la mise en oeuvre d'un système de surveillance de la qualité de l'air : application à la surveillance de la qualité de l'air dans les vallées des Gaves ». Thèse de doct. Institut National Polytechnique de Toulouse-INPT.
- PROTOPODESCU, V, RT SANTORO et J DOCKERY (1989). « Combat modeling with partial differential equations ». In : *European Journal of Operational Research* 38.2, p. 178-183.

- PUANGTHONGTHUB, Sitthichok (2006). *Bayesian Maximum Entropy space/time analysis of ambient particulate matter and mortality in Thailand*. The University of North Carolina at Chapel Hill.
- RABOULI, Sara et al. (2021). « Spatialization of saturated hydraulic conductivity using the Bayesian Maximum Entropy method : Application to wastewater infiltration areas ». In : *Water Research* 204, p. 117607.
- RAMOS, Yuddy (2017). « Développement et évaluation d'approches géostatistiques à l'échelle urbaine pour l'estimation de l'exposition aux particules fines et à l'ozone troposphérique ». In.
- RATA, Mohamed (2010). « Variabilité spatio-temporelle de la salinité des sols dans la plaine du Bas Chélif-Etablissement d'une Banque de Données ». Thèse de doct. DOUAOUI A.
- SCHMIDT, David M, John S GEORGE et Chris C WOOD (1999). « Bayesian inference applied to the electromagnetic inverse problem ». In : *Human brain mapping* 7.3, p. 195-212.
- SEIDEL, Enio Júnior et Marcelo Silva de OLIVEIRA (2016). « A classification for a geostatistical index of spatial dependence ». In : *Revista Brasileira de Ciência do Solo* 40, e0160007.
- SERRE, M (2008). « A Bayesian Maximum Entropy Approach to Adjust for the Sampling Variability of Space/Time Disease Data ». In : *Epidemiology* 19.6, S20.
- SHI, Wenjiao et al. (2011). « Surface modelling of soil properties based on land use information ». In : *Geoderma* 162.3-4, p. 347-357.
- SHORE, John et Rodney JOHNSON (1980). « Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy ». In : *IEEE Transactions on information theory* 26.1, p. 26-37.
- SOIZE, Christian (2008). « Construction of probability distributions in high dimension using the maximum entropy principle : Applications to stochastic processes, random fields and random matrices ». In : *International Journal for Numerical Methods in Engineering* 76.10, p. 1583-1611.
- SOURIS, Marc (2019). *Épidémiologie et géographie : Principes, méthodes et outils de l'analyse spatiale*. ISTE Group.
- TOPONOGOV, Victor A (2006). *Differential geometry of curves and surfaces*. Springer.
- VIGNAUX, M et al. (1998). « Fine-scale mapping of fish distribution from commercial catch and effort data using maximum entropy tomography ». In : *Canadian Journal of Fisheries and Aquatic Sciences* 55.5, p. 1220-1227.
- VIHOTOGBÉ, R et al. (2022). « On the impacts of hard data patterns on Bayesian maximum entropy performance : Simulation-based analysis ». In.
- WANG, Chenxing et al. (2023). « Bayesian maximum entropy method for stochastic model updating using measurement data and statistical information ». In : *Mechanical Systems and Signal Processing* 188, p. 110012.
- YUE, Tian-Xiang, Chuan-Fa CHEN et Bai-Lian LI (2010). « An adaptive method of high accuracy surface modeling and its application to simulating elevation surfaces ». In : *Transactions in GIS* 14.5, p. 615-630.
- YUE, Tian-Xiang et al. (2010). « High-accuracy surface modelling and its application to DEM generation ». In : *International Journal of Remote Sensing* 31.8, p. 2205-2226.
- YUE, Tian-Xiang et al. (2013). « A multi-grid method of high accuracy surface modeling and its validation ». In : *Transactions in GIS* 17.6, p. 943-952.
- YUE, Tian-Xiang et al. (2015). « A review of recent developments in HASM ». In : *Environmental Earth Sciences* 74, p. 6541-6549.
- YUE, TX et SH WANG (2010). « Adjustment computation of HASM : a high-accuracy and high-speed method ». In : *International Journal of Geographical Information Science* 24.11, p. 1725-1743.
- ZHAO, Na et TianXiang YUE (2021). *High Accuracy Surface Modeling Method : The Robustness*. Springer.

---

ZHAO, Na et al. (2014). « Sensitivity studies of a high accuracy surface modeling method ». In : *Science China Earth Sciences* 57, p. 2386-2396.



## ANNEXES

Nous illustrons ici les résultats d'une des données simulées et la prédiction d'un attribut de cette donnée : (Taille d'échantillon = 64, Dépendance spatiale = "Faible", Asymétrie = "Négative"). Comme mentionné précédemment dans notre développement, nous avons défini le domaine d'étude comme suit :  $D = [1, 10] \times [1, 10]$ . De plus, nous avons pris en compte 70% des valeurs de l'attribut pour l'entraînement du modèle et les 30% restants pour la prédiction.

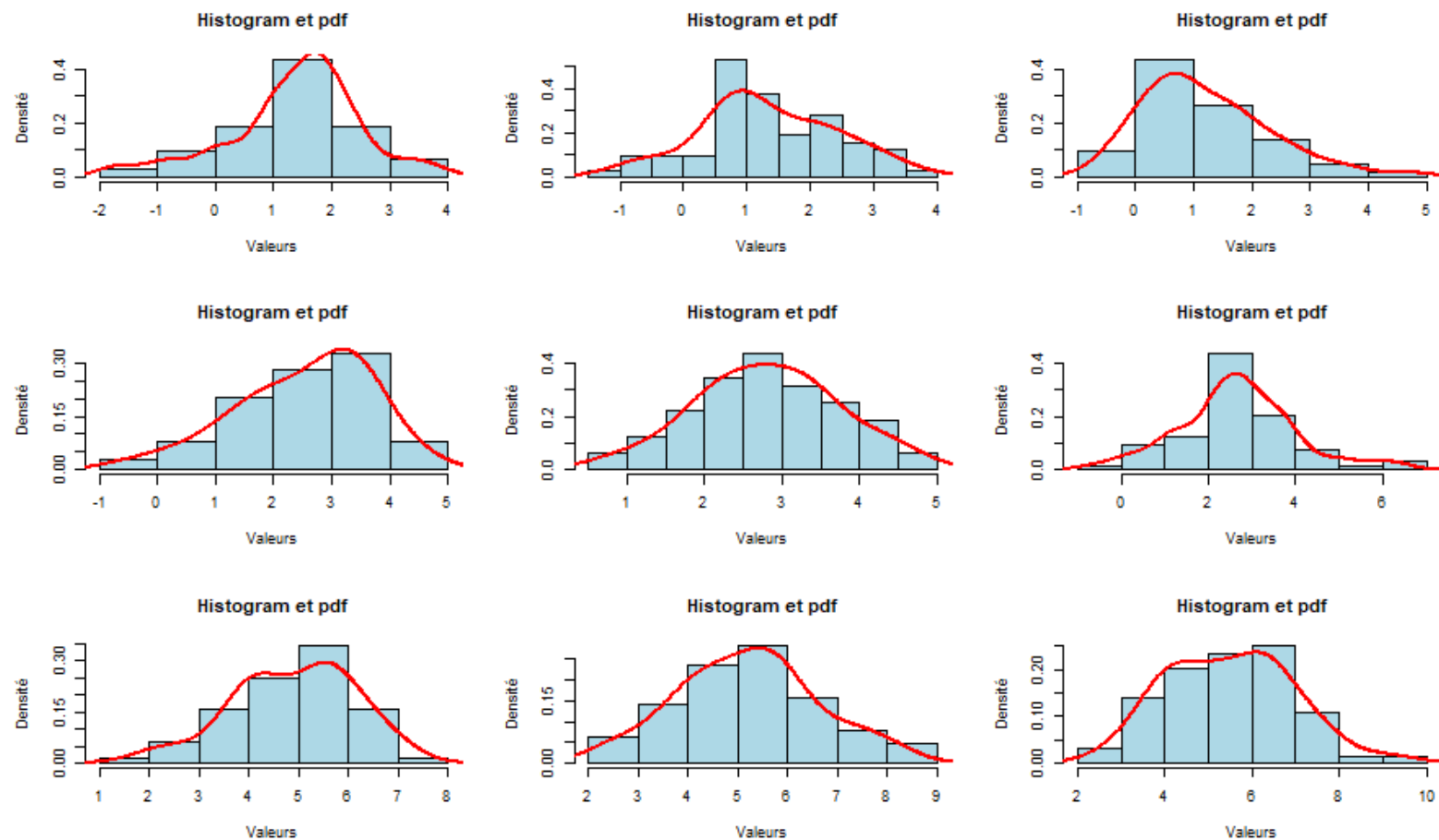


FIGURE 4.5 – Boxplots de la donnée suivant le SD et le skew

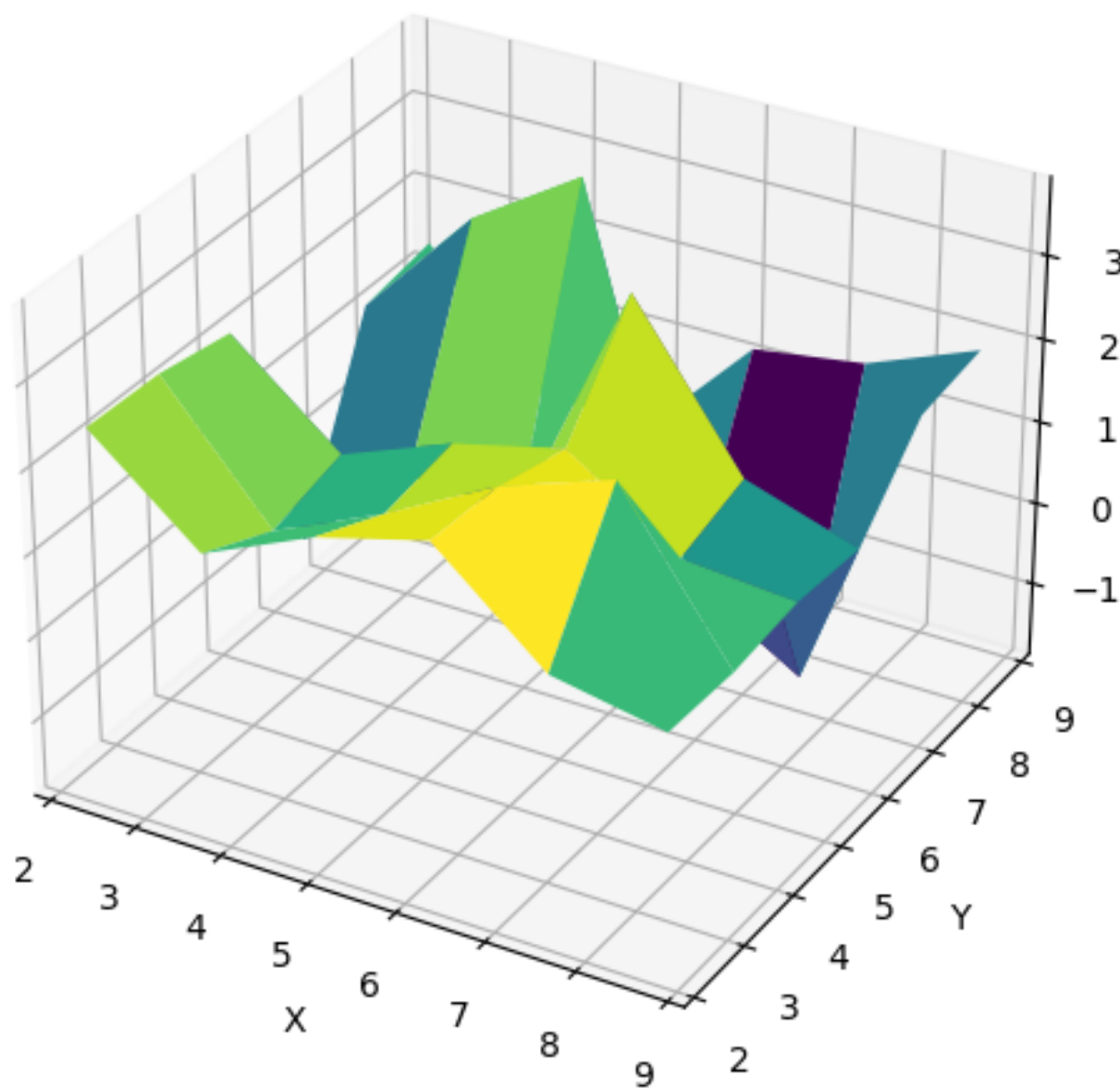
Surface plot of  $f(x, y)$ 

FIGURE 4.6 – Plot 3D d'un attribut de la donnée

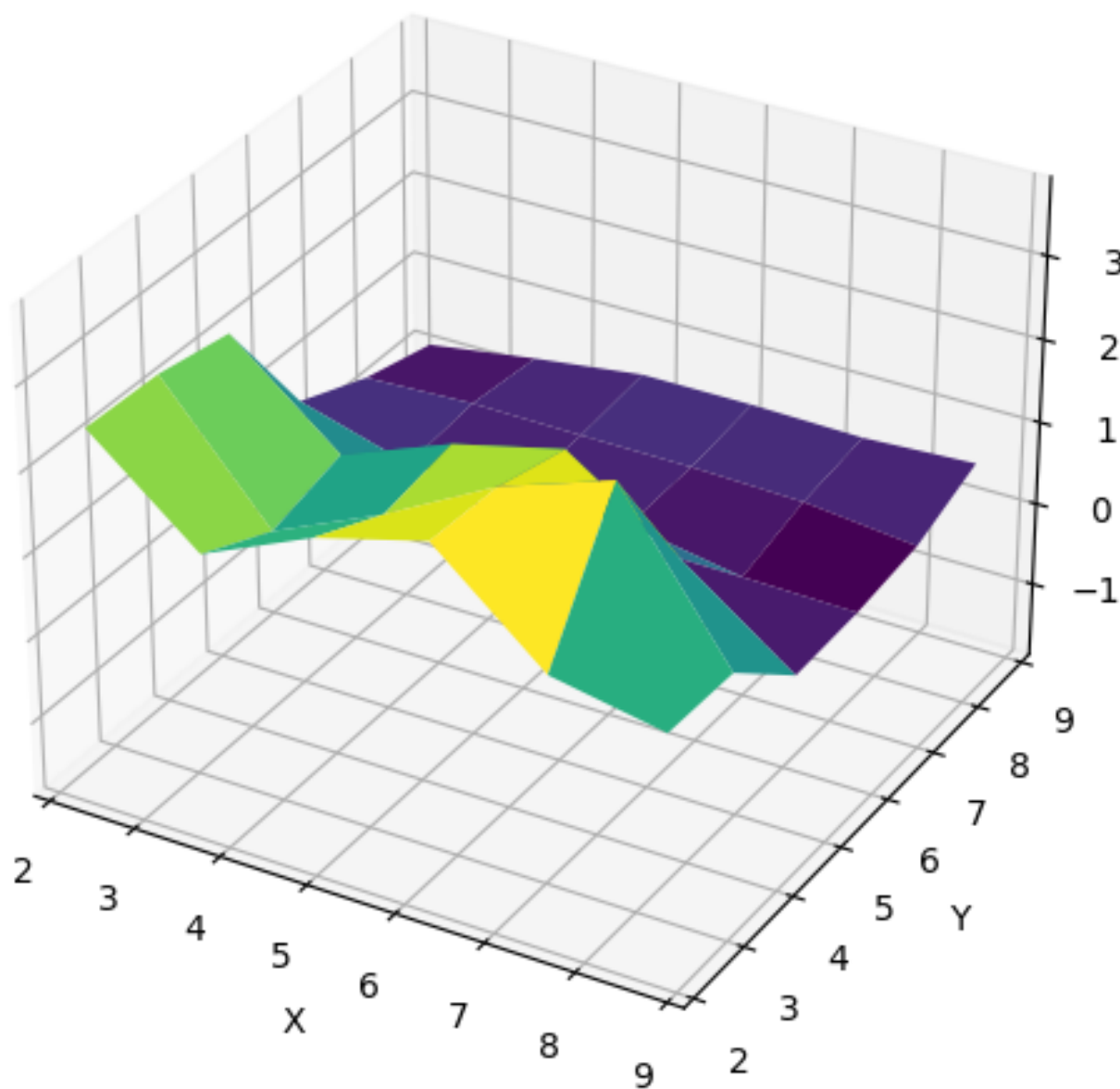
Surface plot of  $f(x, y)$ 

FIGURE 4.7 – Plot 3D de l'attribut prédit

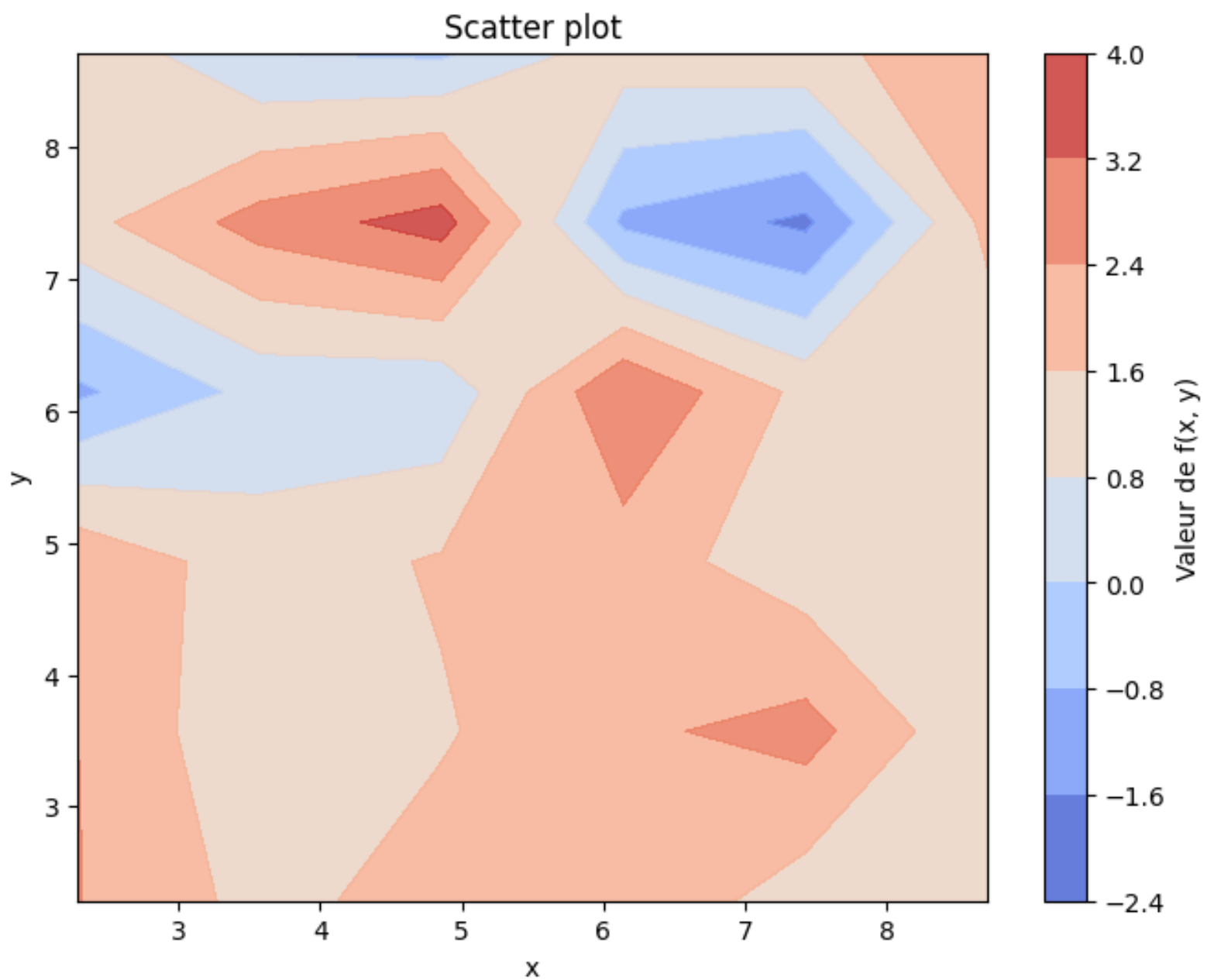


FIGURE 4.8 – Plot scatter d'un attribut de la donnée

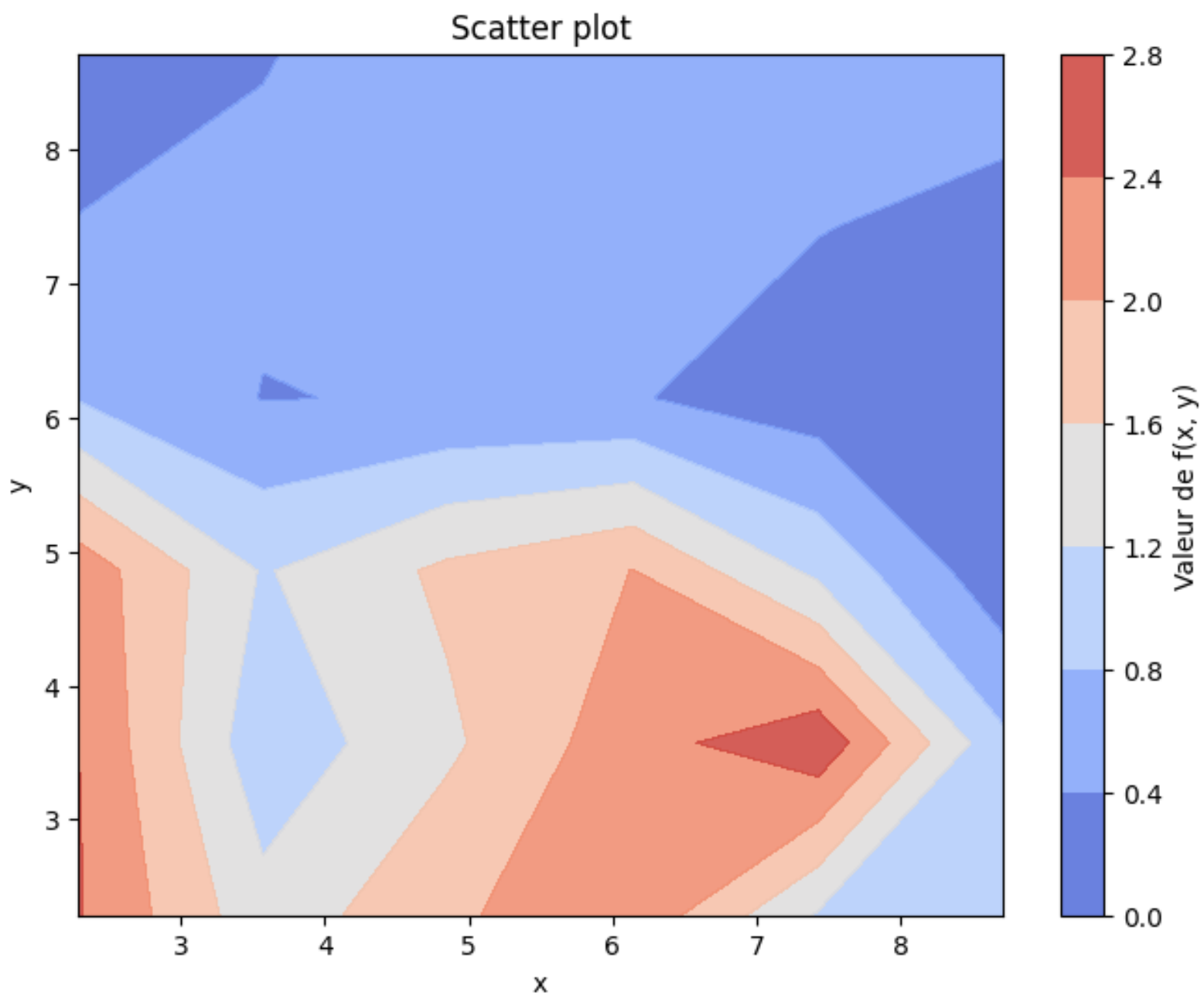


FIGURE 4.9 – Plot scatter de l'attribut prédit

