

République Algérienne Démocratique et Populaire  
Ministère de l'enseignement supérieur et de la recherche scientifique



Université de Saida - Dr Moulay Tahar.  
Faculté de Mathématiques, Informatique et  
Télécommunications  
Département de Mathématiques.



Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Filière : MATHÉMATIQUES

Spécialité : Analyse Stochastique, statistique des processus et  
applications

par

**Bendamou Assia**<sup>1</sup>

Sous la direction de

**Dr. Djerfi Kouider**

Thème :

## Interpolation Spatiale par Krigeage : Méthodologie et Applications

Soutenue le 15/06/2026 devant le jury composé de

<b>Imane Bouazza</b>	Université de Saïda Dr. Moulay Tahar	Président
<b>Kouider Djerfi</b>	Université de Saïda Dr. Moulay Tahar	Encadreur
<b>Chafia Ayhar</b>	Université de Saïda Dr. Moulay Tahar	Examinatrice

Année univ. : 2025/2026

---

1. E-mail : bendamouassia20@gmail.com

# Remerciements

Je tiens tout d'abord à remercier Allah le Tout-Puissant de m'avoir accordé la santé, la patience et la persévérance nécessaires à l'aboutissement de ce travail.

J'exprime ma profonde gratitude à Monsieur **Djerfi Kouider**, mon encadrant, pour sa disponibilité, ses précieux conseils, son soutien constant et l'intérêt qu'il a porté à ce travail tout au long de sa réalisation.

Mes sincères remerciements s'adressent également aux membres du jury :

- Madame **Imane Bouazza**, Présidente du jury ;
- Madame **Chafia Ayhar**, Examinatrice.

Je les remercie chaleureusement d'avoir accepté d'évaluer ce mémoire et pour l'intérêt qu'elles ont porté à ce travail.

Je remercie également l'ensemble des enseignants du Département de Mathématiques pour la qualité de leur enseignement et les connaissances qu'ils m'ont transmises tout au long de mon parcours universitaire.

Enfin, j'adresse mes plus sincères remerciements à ma famille pour son soutien indéfectible, sa confiance et ses encouragements qui m'ont permis d'achever ce travail.

À toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce mémoire, j'exprime ma profonde reconnaissance.

# Dédicace

*À ma très chère mère,*

*Pour son amour infini, ses sacrifices, sa tendresse, ses prières et son soutien inconditionnel. Aucune expression ne saurait traduire toute ma gratitude et mon profond respect envers elle.*

*À mes chères soeurs,*

*Pour leur affection, leur présence constante et leurs encouragements qui m'ont accompagnée tout au long de mon parcours.*

*À mon cher époux,*

*Pour sa patience, sa compréhension, son soutien précieux et sa confiance permanente. Merci d'avoir partagé avec moi les moments de doute et de joie et d'avoir toujours cru en ma réussite.*

*À mes honorables enseignants,*

*Qui ont contribué à ma formation par leur savoir, leurs conseils et leur dévouement. Qu'ils trouvent ici l'expression de ma profonde reconnaissance.*

*À mes chers amis et amies,*

*Pour leur amitié sincère, leur soutien moral et tous les moments inoubliables partagés durant ces années d'études.*

*À tous ceux qui me sont chers et à toutes les personnes qui ont contribué, de près ou de loin, à l'aboutissement de ce travail.*

***Assia***

# Table des matières

<b>Introduction générale</b>	<b>1</b>
<b>1 Données spatiales et méthodes d'interpolation</b>	<b>4</b>
1.1 Introduction . . . . .	4
1.2 Données spatiales et information géographique . . . . .	5
1.3 Formulation du problème d'interpolation spatiale . . . . .	6
1.4 Distance, voisinage et échelle spatiale . . . . .	7
1.5 Méthodes déterministes d'interpolation . . . . .	8
1.6 Critères de comparaison des méthodes d'interpolation . . . . .	11
1.7 Limites des méthodes déterministes . . . . .	12
1.8 Transition vers la géostatistique . . . . .	13
<b>2 Fondements mathématiques de la géostatistique</b>	<b>14</b>
2.1 Introduction . . . . .	14
2.2 Variable régionalisée . . . . .	15
2.3 Moments d'un champ aléatoire . . . . .	16
2.4 Stationnarité . . . . .	17
2.5 Covariance, corrélation, isotropie et anisotropie . . . . .	18
2.6 Variogramme . . . . .	19
2.7 Hypothèse intrinsèque . . . . .	20
2.8 Tendances, dérive et résidus . . . . .	21
2.9 Support spatial et effet d'échelle . . . . .	21
2.10 Conclusion du chapitre . . . . .	22

<b>3</b>	<b>Variogramme : estimation, modélisation et validation</b>	<b>23</b>
3.1	Introduction . . . . .	23
3.2	Nuage variographique . . . . .	23
3.3	Variogramme expérimental . . . . .	24
3.4	Effet de pépite, palier et portée . . . . .	25
3.5	Isotropie et anisotropie dans le variogramme . . . . .	26
3.6	Admissibilité d'un variogramme . . . . .	26
3.7	Modèles théoriques classiques . . . . .	27
3.8	Ajustement d'un modèle de variogramme . . . . .	28
3.9	Validation du modèle variographique . . . . .	28
3.10	Conclusion du chapitre . . . . .	29
<b>4</b>	<b>Théorie du krigeage</b>	<b>30</b>
4.1	Introduction . . . . .	30
4.2	Prédiction linéaire . . . . .	30
4.3	Variance de l'erreur dans le cas stationnaire . . . . .	31
4.4	Krigeage simple . . . . .	31
4.5	Krigeage ordinaire . . . . .	32
4.6	Krigeage universel . . . . .	34
4.7	Interprétation des poids de krigeage . . . . .	35
4.8	Variance de krigeage et incertitude . . . . .	35
4.9	Conclusion du chapitre . . . . .	36
<b>5</b>	<b>Application numérique : interpolation spatiale et comparaison des méthodes</b>	<b>37</b>
5.1	Introduction . . . . .	37
5.2	Organisation générale du protocole numérique . . . . .	38
5.3	Domaine spatial et génération des données . . . . .	39
5.4	Analyse exploratoire des observations . . . . .	40
5.5	Analyse variographique . . . . .	43

5.6	Interpolation par krigeage ordinaire . . . . .	46
5.7	Interpolation par méthodes déterministes . . . . .	49
5.8	Validation croisée . . . . .	53
5.9	Analyse des erreurs du krigeage . . . . .	56
5.10	Discussion critique des résultats . . . . .	58
5.11	Conclusion du chapitre . . . . .	60
	<b>Conclusion générale</b>	<b>61</b>
	<b>A Rappels statistiques utiles à la géostatistique</b>	<b>64</b>
A.1	Variabes aléatoires et lois de probabilité . . . . .	64
A.2	Espérance mathématique . . . . .	65
A.3	Variance et écart-type . . . . .	66
A.4	Covariance et corrélation . . . . .	67
A.5	Indépendance et absence de corrélation . . . . .	68
A.6	Vecteurs aléatoires et matrices de covariance . . . . .	69
A.7	Lois usuelles en statistique . . . . .	70
A.8	Échantillon, moyenne empirique et variance empirique . . . . .	71
A.9	Biais et erreur quadratique moyenne . . . . .	72
A.10	Notions de convergence . . . . .	73
A.11	Moindres carrés et ajustement de modèle . . . . .	73
A.12	Intervalles de confiance et incertitude . . . . .	74
A.13	Lien entre statistiques classiques et géostatistique . . . . .	75
	<b>B Code Python utilisé pour la simulation</b>	<b>76</b>
	<b>Bibliographie</b>	<b>87</b>

# Table des figures

5.1	Répartition spatiale des sites d'observation dans le domaine d'étude. . . . .	41
5.2	Histogramme des valeurs observées. . . . .	42
5.3	Variogramme expérimental calculé à partir des observations. . . . .	44
5.4	Variogramme expérimental et modèle théorique ajusté. . . . .	46
5.5	Carte interpolée par krigeage ordinaire. . . . .	48
5.6	Carte de variance associée au krigeage ordinaire. . . . .	49
5.7	Carte interpolée par la méthode du plus proche voisin. . . . .	50
5.8	Carte interpolée par pondération inverse de la distance. . . . .	51
5.9	Sensibilité de la méthode IDW au choix du paramètre $p$ . . . . .	52
5.10	Comparaison graphique des cartes obtenues par plus proche voisin, IDW et krigeage ordinaire. . . . .	52
5.11	Distribution des erreurs de validation croisée pour les différentes méthodes.	54
5.12	Comparaison des valeurs observées et des valeurs prédites en validation croisée. . . . .	55
5.13	Carte des erreurs de validation du krigeage ordinaire. . . . .	56
5.14	Erreurs standardisées obtenues par validation croisée du krigeage ordinaire.	57

# Liste des tableaux

5.1	Statistiques descriptives des valeurs observées. . . . .	41
5.2	Paramètres estimés pour les modèles de variogramme. . . . .	45
5.3	Sensibilité de la méthode IDW au paramètre de puissance $p$ . . . . .	51
5.4	Comparaison des méthodes par validation croisée leave-one-out. . . . .	53
5.5	Classement des méthodes selon les critères de validation croisée. . . . .	55
5.6	Résumé des erreurs standardisées du krigeage ordinaire. . . . .	57

# Introduction générale

L'analyse des phénomènes spatiaux occupe une place centrale dans de nombreux domaines des sciences appliquées. En hydrologie, en climatologie, en géologie, en agronomie, en environnement ou en cartographie, les observations disponibles sont généralement associées à des localisations géographiques précises. On mesure par exemple une hauteur de précipitation dans certaines stations météorologiques, une teneur minérale dans quelques sondages, une concentration de polluant dans des points de prélèvement, ou encore une propriété du sol dans une zone agricole. Dans toutes ces situations, l'information est disponible seulement en un nombre fini de points, alors que l'objectif pratique consiste souvent à produire une estimation sur tout un domaine spatial.

Cette difficulté conduit naturellement au problème de l'interpolation spatiale. Il s'agit d'estimer la valeur d'une variable en un point non observé à partir des valeurs connues aux points voisins. Les premières réponses à ce problème reposent sur des méthodes déterministes. Parmi elles, on peut citer la méthode du plus proche voisin, la pondération par l'inverse de la distance, les splines ou encore les méthodes fondées sur la triangulation. Ces approches sont simples, faciles à mettre en œuvre et largement utilisées dans les systèmes d'information géographique. Les revues de [8], [13] et [9] montrent la diversité de ces méthodes et leur importance dans les applications environnementales.

Cependant, les méthodes déterministes présentent une limite essentielle : elles ne modélisent pas explicitement la dépendance spatiale du phénomène étudié. Elles utilisent souvent la distance géométrique entre les points, mais elles ne décrivent pas la manière dont la similarité entre deux observations évolue avec l'éloignement. De plus, elles ne fournissent généralement pas une mesure naturelle de l'incertitude associée à la prédiction. Or, dans de nombreuses applications, il ne suffit pas de produire une carte interpolée ; il est également important de savoir dans quelles zones cette carte est fiable et dans quelles zones l'incertitude reste élevée.

C'est dans ce contexte que la géostatistique apporte une réponse plus structurée. Son développement est étroitement lié aux problèmes d'estimation des ressources minières. Dans les années 1950, l'ingénieur sud-africain Daniel G. Krige étudie des méthodes empiriques destinées à améliorer l'estimation des teneurs en minerai à partir d'échan-

tillons localisés. Ces travaux ont mis en évidence l'importance de la structure spatiale des observations dans les problèmes d'estimation [7]. Par la suite, Georges Matheron a donné à ces idées une formulation mathématique rigoureuse à travers la théorie des variables régionalisées [10, 11]. C'est dans ce cadre que la méthode d'interpolation optimale issue de ces travaux a pris le nom de *krigeage*, en hommage à Krige ; une présentation synthétique de cette filiation est également donnée par [6].

L'apport fondamental de la géostatistique est de considérer la variable étudiée non comme une simple fonction inconnue de l'espace, mais comme la réalisation d'un champ aléatoire. Cette modélisation probabiliste permet de définir une moyenne, une covariance et un variogramme. Le variogramme occupe une place centrale : il mesure la dissimilarité moyenne entre deux valeurs du phénomène en fonction de leur séparation spatiale. Il permet ainsi de traduire mathématiquement l'idée intuitive selon laquelle deux observations proches ont tendance à être plus semblables que deux observations éloignées. Les ouvrages de [15], [14] et [4] présentent en détail cette construction théorique et son rôle dans les applications géostatistiques.

Le krigeage peut alors être compris comme une méthode de prédiction linéaire fondée sur la structure spatiale du phénomène. Pour estimer la valeur en un point non observé, on construit une combinaison linéaire des observations disponibles. Les poids de cette combinaison ne sont pas choisis de façon arbitraire ni uniquement à partir de la distance. Ils sont déterminés par un problème d'optimisation : minimiser la variance de l'erreur de prédiction sous une contrainte de non-biais. Dans un cadre approprié, le krigeage apparaît ainsi comme le meilleur prédicteur linéaire sans biais.

Cette propriété distingue profondément le krigeage des méthodes déterministes classiques. Dans la pondération par l'inverse de la distance, les poids dépendent directement de la distance entre le point à prédire et les observations. Dans le krigeage, les poids dépendent aussi des relations entre les observations elles-mêmes, ce qui permet de prendre en compte la redondance entre points proches. De plus, le krigeage fournit une variance de prédiction, qui permet de construire des cartes d'incertitude. Cette information est particulièrement importante dans les domaines où les décisions doivent tenir compte du risque d'erreur.

Le présent mémoire a pour objectif de présenter les fondements de l'interpolation spatiale par krigeage, en allant progressivement des méthodes déterministes vers le cadre géostatistique. L'accent est mis sur la compréhension mathématique des notions essentielles : champ aléatoire, stationnarité, covariance, variogramme, hypothèse intrinsèque, krigeage simple, krigeage ordinaire et krigeage universel. Une attention particulière est également accordée à l'analyse variographique, car la qualité d'un krigeage dépend fortement du modèle de variogramme retenu.

Le mémoire comporte cinq chapitres. Le premier chapitre introduit les données spatiales et le problème général de l'interpolation. Il présente plusieurs méthodes déterministes classiques, notamment le plus proche voisin, l'IDW, les splines et la triangulation. Le deuxième chapitre présente les fondements mathématiques de la géostatistique. On y introduit la notion de variable régionalisée, son interprétation comme réalisation d'un champ aléatoire, ainsi que les notions de moyenne, covariance, stationnarité et variogramme. Le troisième chapitre est consacré à l'analyse variographique. Il explique la construction du variogramme expérimental, l'interprétation de l'effet de pépite, du palier et de la portée, ainsi que l'ajustement de modèles théoriques. Le quatrième chapitre développe la théorie du krigeage. Enfin, le cinquième chapitre propose une application numérique et une comparaison entre le krigeage ordinaire, l'IDW et le plus proche voisin.

Ainsi, ce mémoire vise à donner une présentation cohérente et progressive du krigeage, depuis le problème élémentaire d'interpolation spatiale jusqu'à la construction d'une méthode statistique fondée sur la dépendance spatiale.

# Chapitre 1

## Données spatiales et méthodes d'interpolation

### 1.1 Introduction

L'étude des phénomènes spatiaux occupe une place importante dans les sciences de l'environnement, la géographie, l'hydrologie, la climatologie, la géologie, l'agronomie et la cartographie. Dans ces domaines, l'information observée n'est pas seulement une valeur numérique : elle est aussi liée à une localisation. Une hauteur de précipitation mesurée dans une station météorologique, une concentration de polluant mesurée en un point de prélèvement, une teneur minérale obtenue par sondage ou une altitude relevée sur le terrain n'ont de sens que si l'on connaît la position où la mesure a été effectuée.

Dans la plupart des situations réelles, les mesures disponibles sont ponctuelles et limitées. On dispose d'un nombre fini d'observations, alors que l'on souhaite décrire un phénomène sur l'ensemble d'une région. Cette différence entre l'information observée et l'information recherchée conduit au problème fondamental de l'interpolation spatiale : estimer la valeur d'une variable en un point non observé à partir des valeurs mesurées en des points voisins.

Les méthodes d'interpolation spatiale sont nombreuses. Certaines sont déterministes : elles reposent sur des règles géométriques ou analytiques. C'est le cas de la méthode du plus proche voisin, de la pondération inverse de la distance, des méthodes polynomiales locales, des splines ou de l'interpolation par triangulation. D'autres méthodes, dites géostatistiques, reposent sur un modèle probabiliste de la dépendance spatiale. Le krigeage appartient à cette seconde famille. Les revues de [8], [13], [9] et [5] soulignent la richesse de ces approches et l'importance du choix de la méthode selon la nature des données et l'objectif de l'étude.

Le but de ce chapitre est de poser les bases du problème d'interpolation spatiale. Nous introduisons d'abord la notion de donnée spatiale et les principales formes qu'elle peut prendre. Nous formulons ensuite mathématiquement le problème d'interpolation. Nous présentons enfin plusieurs méthodes déterministes classiques, leurs propriétés, leurs avantages et leurs limites. Cette discussion prépare naturellement l'introduction de la géostatistique dans les chapitres suivants.

## 1.2 Données spatiales et information géographique

### Définition générale

Une donnée spatiale est une information associée à une position dans l'espace. Dans un cadre plan, cette position peut être représentée par un point

$$x = (x_1, x_2) \in D \subset \mathbb{R}^2,$$

où  $D$  désigne le domaine d'étude. La valeur observée au point  $x$  est notée  $z(x)$ . Un ensemble de données spatiales ponctuelles peut donc s'écrire sous la forme

$$\mathcal{D}_n = \{(x_i, z_i) : i = 1, \dots, n\},$$

où  $x_i \in D$  est le site d'observation et  $z_i = z(x_i)$  la valeur mesurée.

**Définition 1.1.** *On appelle donnée spatiale ponctuelle un couple  $(x, z)$  où  $x \in D \subset \mathbb{R}^d$  représente une position et  $z \in \mathbb{R}$  représente la valeur observée d'une variable quantitative en cette position.*

La localisation peut être donnée en coordonnées cartésiennes, en coordonnées géographiques ou dans un système de projection. Dans les applications cartographiques, le choix du système de coordonnées est important. Une distance calculée directement sur des latitudes et longitudes n'a pas toujours la même signification qu'une distance calculée dans une projection plane adaptée. Dans ce mémoire, afin de concentrer l'étude sur les méthodes d'interpolation, nous considérons principalement un domaine plan  $D \subset \mathbb{R}^2$ .

### Exemples de phénomènes spatiaux

Les phénomènes spatiaux apparaissent dans de nombreuses situations :

- en hydrologie, on cherche à estimer les précipitations sur un bassin versant à partir de stations pluviométriques ;

- en environnement, on veut cartographier une concentration de polluant à partir de quelques points de prélèvement ;
- en géologie, on estime une teneur minérale entre plusieurs sondages ;
- en agronomie, on décrit l’humidité, la salinité ou la fertilité du sol ;
- en climatologie, on interpole la température, la pression ou l’humidité relative.

Ces exemples ont en commun une propriété essentielle : la valeur observée dépend de la localisation, et les observations proches sont souvent plus semblables que les observations éloignées. Cette idée intuitive est au coeur de la plupart des méthodes d’interpolation spatiale.

### Données ponctuelles, surfaciques et maillées

Les données spatiales ne sont pas toujours de même nature. On distingue généralement plusieurs types.

Les *données ponctuelles* sont mesurées en des points précis. Elles correspondent par exemple à des stations météorologiques, des capteurs, des forages ou des sites de prélèvement. Le krigeage classique est principalement construit pour ce type de données.

Les *données surfaciques* sont associées à des zones. Une densité de population par commune, un rendement agricole par parcelle ou une moyenne de pollution par secteur administratif sont des exemples de données surfaciques. Dans ce cas, la valeur observée représente souvent une moyenne ou un total sur une unité spatiale.

Les *données maillées* sont observées sur une grille régulière. Elles apparaissent par exemple dans les images satellitaires, les modèles climatiques ou les modèles numériques de terrain. Même lorsque les données sont déjà sur une grille, il peut être nécessaire de les interpoler sur une autre grille ou à une autre résolution.

**Remarque 1.1.** *Dans ce mémoire, l’accent est mis sur les données ponctuelles quantitatives. Ce choix permet de présenter clairement le variogramme et le krigeage ordinaire. Les problèmes liés au changement de support spatial ou à l’agrégation des données sont importants, mais ils dépassent le cadre principal de cette étude.*

## 1.3 Formulation du problème d’interpolation spatiale

Soit  $D \subset \mathbb{R}^2$  un domaine spatial. On suppose que l’on connaît les valeurs d’une variable  $z$  en  $n$  sites distincts :

$$z_i = z(x_i), \quad i = 1, \dots, n.$$

Le problème d'interpolation consiste à estimer la valeur  $z(x_0)$  en un point  $x_0 \in D$  où aucune observation n'est disponible.

Une méthode d'interpolation construit une fonction estimée

$$\widehat{z} : D \longrightarrow \mathbb{R}$$

telle que  $\widehat{z}(x_0)$  soit une approximation de  $z(x_0)$ . Dans de nombreuses méthodes, l'estimation est linéaire par rapport aux observations :

$$\widehat{z}(x_0) = \sum_{i=1}^n w_i(x_0) z(x_i),$$

où  $w_i(x_0)$  sont des poids dépendant du point à estimer et de la configuration des observations.

La différence fondamentale entre les méthodes réside dans la manière de choisir ces poids. Dans les méthodes déterministes, les poids sont souvent construits à partir de la distance géométrique. Dans les méthodes géostatistiques, les poids sont déterminés à partir d'un modèle de dépendance spatiale, exprimé par une covariance ou un variogramme.

**Définition 1.2.** *Une méthode d'interpolation est dite exacte si elle reproduit les observations aux sites mesurés, c'est-à-dire si*

$$\widehat{z}(x_i) = z(x_i), \quad i = 1, \dots, n.$$

*Elle est dite lissante si cette propriété n'est pas imposée.*

Les méthodes exactes sont adaptées lorsque les observations sont considérées comme fiables. Les méthodes lissantes sont utiles lorsque les données contiennent du bruit ou lorsque l'on souhaite représenter une tendance générale plutôt que reproduire toutes les fluctuations locales.

## 1.4 Distance, voisinage et échelle spatiale

La distance est une notion centrale en interpolation spatiale. Dans un domaine plan, la distance euclidienne entre deux points  $x = (x_1, x_2)$  et  $y = (y_1, y_2)$  est donnée par

$$d(x, y) = \|x - y\| = \left( (x_1 - y_1)^2 + (x_2 - y_2)^2 \right)^{1/2}.$$

Cette distance est appropriée lorsque les coordonnées sont exprimées dans un système cartésien ou dans une projection plane adaptée.

Le voisinage d'un point  $x_0$  désigne l'ensemble des observations utilisées pour prédire  $z(x_0)$ . On peut utiliser toutes les observations disponibles ou seulement un sous-ensemble local, par exemple les  $k$  plus proches voisins. Le choix du voisinage est important : un voisinage trop petit peut produire des prédictions instables, tandis qu'un voisinage trop large peut masquer des variations locales.

L'échelle spatiale joue également un rôle fondamental. Un phénomène peut apparaître régulier à grande échelle, mais très irrégulier à petite échelle. Par exemple, la température varie généralement de façon progressive à l'échelle régionale, tandis que certaines concentrations chimiques du sol peuvent changer fortement sur de petites distances. Une méthode d'interpolation doit donc être choisie en fonction de l'échelle du phénomène et de la densité des observations.

**Exemple 1.1.** *Supposons que deux stations météorologiques soient distantes de deux kilomètres et enregistrent des températures très proches. Cela suggère une forte continuité locale. En revanche, si deux prélèvements de sol distants de quelques mètres donnent des concentrations très différentes, la variable étudiée présente une variabilité à petite échelle. Dans ce second cas, une interpolation très lisse peut être trompeuse.*

## 1.5 Méthodes déterministes d'interpolation

Les méthodes déterministes construisent une estimation à partir de règles explicites, sans introduire de modèle probabiliste du phénomène spatial. Elles sont largement utilisées en pratique, notamment en raison de leur simplicité et de leur disponibilité dans les logiciels de cartographie et de systèmes d'information géographique. Les synthèses de [8], [13] et [9] montrent que ces méthodes restent des outils de référence dans de nombreux contextes appliqués.

### Méthode du plus proche voisin

La méthode du plus proche voisin consiste à attribuer au point non observé  $x_0$  la valeur de l'observation la plus proche. Si

$$i^* = \arg \min_{1 \leq i \leq n} \|x_0 - x_i\|,$$

alors

$$\widehat{z}(x_0) = z(x_{i^*}).$$

Cette méthode est exacte et très simple à mettre en oeuvre. Elle ne nécessite aucun paramètre. Elle peut être utile pour certaines variables qualitatives ou lorsque l'on souhaite associer chaque point du domaine à la station la plus proche.

Sa principale limite est la discontinuité de la surface obtenue. Le domaine est découpé en zones d'influence, et la valeur prédite est constante dans chaque zone. Lorsque l'on traverse la frontière entre deux zones, la prédiction change brutalement. Cette propriété est peu adaptée aux phénomènes continus tels que la température ou la précipitation.

### Pondération inverse de la distance

La méthode de pondération inverse de la distance, ou IDW, repose sur l'idée que les observations proches de  $x_0$  doivent avoir plus d'influence que les observations éloignées. Pour un paramètre  $p > 0$ , on définit

$$\widehat{z}(x_0) = \sum_{i=1}^n w_i(x_0) z(x_i),$$

où

$$w_i(x_0) = \frac{\|x_0 - x_i\|^{-p}}{\sum_{j=1}^n \|x_0 - x_j\|^{-p}}.$$

Les poids vérifient

$$w_i(x_0) \geq 0, \quad \sum_{i=1}^n w_i(x_0) = 1.$$

Le paramètre  $p$  contrôle la décroissance de l'influence avec la distance. Lorsque  $p$  est petit, plusieurs observations contribuent à la prédiction. Lorsque  $p$  est grand, les observations les plus proches dominent fortement.

**Proposition 1.1.** *Si  $x_0$  ne coïncide avec aucun site d'observation, l'estimateur IDW est une combinaison convexe des valeurs observées. En particulier,*

$$\min_{1 \leq i \leq n} z(x_i) \leq \widehat{z}(x_0) \leq \max_{1 \leq i \leq n} z(x_i).$$

*Démonstration.* Les poids  $w_i(x_0)$  sont positifs et leur somme est égale à 1. La prédiction est donc une moyenne pondérée des valeurs observées. Elle appartient nécessairement à l'intervalle compris entre la plus petite et la plus grande observation.  $\square$

Cette propriété évite des extrapolations extrêmes, mais elle limite aussi la méthode : l>IDW ne peut pas prédire une valeur supérieure au maximum observé ni inférieure au minimum observé, même si le phénomène réel peut présenter un maximum ou un minimum entre les sites d'observation.

## Méthodes polynomiales globales et locales

Les méthodes polynomiales cherchent à représenter la variable par une fonction polynomiale des coordonnées. Par exemple, un modèle linéaire s'écrit

$$P(x_1, x_2) = a_0 + a_1x_1 + a_2x_2,$$

et un modèle quadratique peut s'écrire

$$P(x_1, x_2) = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_1x_2 + a_5x_2^2.$$

Les coefficients sont estimés par moindres carrés à partir des observations.

Un polynôme global peut représenter une tendance spatiale générale, mais il est souvent incapable de décrire des variations locales complexes. Un polynôme de degré trop élevé peut au contraire produire des oscillations artificielles. Les méthodes polynomiales locales tentent de résoudre cette difficulté en ajustant un polynôme dans un voisinage du point à estimer.

Dans la théorie géostatistique, la notion de tendance polynomiale réapparaît dans le krigeage universel, où la moyenne du champ est modélisée comme une combinaison de fonctions connues de l'espace.

## Splines et surfaces régulières

Les méthodes de splines construisent une surface régulière passant près des observations ou les interpolant exactement. Une formulation générale consiste à chercher une fonction  $s$  minimisant un critère du type

$$\sum_{i=1}^n (z(x_i) - s(x_i))^2 + \lambda J(s),$$

où  $J(s)$  mesure la rugosité de la surface et  $\lambda > 0$  contrôle le degré de lissage.

Lorsque  $\lambda$  est faible, la surface suit fortement les observations. Lorsque  $\lambda$  est grand, la surface devient plus lisse. Les splines sont efficaces pour produire des cartes régulières, mais leur interprétation statistique dépend du cadre utilisé. Dans une présentation introductive, elles restent principalement des méthodes de lissage géométrique.

## Triangulation et interpolation linéaire par morceaux

Une autre approche consiste à construire une triangulation des sites d'observation, puis à interpoler linéairement dans chaque triangle. Si  $x_0$  appartient au triangle de

sommets  $x_i, x_j, x_k$ , on peut écrire

$$x_0 = \lambda_i x_i + \lambda_j x_j + \lambda_k x_k,$$

avec

$$\lambda_i + \lambda_j + \lambda_k = 1, \quad \lambda_i, \lambda_j, \lambda_k \geq 0.$$

La valeur interpolée est alors

$$\widehat{z}(x_0) = \lambda_i z(x_i) + \lambda_j z(x_j) + \lambda_k z(x_k).$$

Cette méthode produit une surface continue par morceaux. Elle est simple et géométriquement naturelle, mais la surface n'est pas nécessairement lisse aux frontières des triangles. De plus, la méthode dépend fortement de la géométrie du maillage.

## 1.6 Critères de comparaison des méthodes d'interpolation

Le choix d'une méthode d'interpolation ne doit pas reposer uniquement sur l'apparence visuelle de la carte produite. Plusieurs critères doivent être examinés.

### Exactitude et effet de lissage

Une méthode exacte reproduit les observations. Cette propriété est souhaitable lorsque les mesures sont fiables. En présence d'erreurs de mesure, une méthode lissante peut être préférable. Le choix entre interpolation exacte et lissage dépend donc de l'objectif : représenter les données observées ou estimer un phénomène sous-jacent.

### Régularité de la surface

Une carte peut être discontinue, continue par morceaux ou lisse. Le plus proche voisin donne une surface discontinue. L'interpolation par triangulation donne une surface continue mais non lisse. Les splines et le krigeage produisent souvent des surfaces plus régulières. La régularité attendue doit être cohérente avec la nature du phénomène.

### Sensibilité aux paramètres

Certaines méthodes dépendent fortement de paramètres :  $p$  dans l>IDW,  $\lambda$  dans les splines, degré du polynôme, taille du voisinage. Une méthode peut donner des résultats très différents selon ces choix. Une analyse sérieuse doit donc discuter la sensibilité des résultats.

## Capacité à mesurer l'incertitude

Les méthodes déterministes produisent généralement une valeur prédite, mais pas une variance de prédiction. Cette limite est importante. Une carte interpolée peut donner une impression de précision, même dans les zones éloignées des observations. Le krigeage se distingue précisément par sa capacité à produire une carte d'incertitude.

## Validation croisée

La validation croisée permet de comparer les méthodes sur les données disponibles. Dans la validation leave-one-out, on retire successivement chaque observation  $z(x_i)$ , on prédit sa valeur à partir des autres observations, puis on calcule l'erreur

$$e_i = z(x_i) - \hat{z}_{-i}(x_i).$$

On peut ensuite calculer l'erreur moyenne

$$ME = \frac{1}{n} \sum_{i=1}^n e_i,$$

l'erreur absolue moyenne

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

et la racine de l'erreur quadratique moyenne

$$RMSE = \left( \frac{1}{n} \sum_{i=1}^n e_i^2 \right)^{1/2}.$$

Ces critères seront utilisés dans l'application numérique.

## 1.7 Limites des méthodes déterministes

Les méthodes déterministes ont l'avantage d'être simples, rapides et faciles à expliquer. Elles sont utiles pour une première exploration des données et pour produire rapidement des cartes. Cependant, elles présentent plusieurs limites.

Premièrement, elles ne modélisent pas explicitement la dépendance spatiale. La distance intervient souvent directement, mais la manière dont la similarité diminue avec l'éloignement n'est pas estimée à partir des données de façon probabiliste.

Deuxièmement, elles ne tiennent pas toujours compte de la redondance entre observations proches. Dans l>IDW, deux observations très proches l'une de l'autre peuvent

exercer une influence excessive, alors qu'elles apportent une information en grande partie similaire.

Troisièmement, elles ne fournissent pas naturellement une variance de prédiction. Or l'incertitude est essentielle dans les applications environnementales et géologiques.

Quatrièmement, elles ne permettent pas toujours de séparer la tendance globale, la dépendance spatiale locale et le bruit de mesure. Cette séparation est précisément l'un des objectifs de la géostatistique.

## 1.8 Transition vers la géostatistique

La géostatistique propose une réponse à ces limites en introduisant un modèle probabiliste. La variable étudiée est considérée comme la réalisation d'un champ aléatoire. La dépendance spatiale est décrite par une covariance ou par un variogramme. Les poids de prédiction ne sont plus fixés seulement par une règle géométrique, mais déterminés par un problème d'optimisation.

Cette différence est fondamentale. Dans l>IDW, les poids dépendent de la distance entre le point à prédire et les observations. Dans le krigeage, les poids dépendent également de la structure de dépendance entre les observations elles-mêmes. Le krigeage peut donc réduire l'effet de redondance et fournir une variance de prédiction.

Le chapitre suivant introduit les notions mathématiques nécessaires à cette approche : variable régionalisée, champ aléatoire, moyenne, covariance, stationnarité, variogramme et hypothèse intrinsèque.

# Chapitre 2

## Fondements mathématiques de la géostatistique

### 2.1 Introduction

Le chapitre précédent a montré que les méthodes déterministes d'interpolation spatiale reposent principalement sur la distance et sur des règles géométriques. Pour aller plus loin, il est nécessaire de modéliser la structure de dépendance du phénomène étudié. C'est l'objet de la géostatistique.

La géostatistique est née dans le contexte de l'estimation minière. Les travaux de Krige ont mis en évidence l'importance de la continuité spatiale dans l'estimation des teneurs en minerai, et ceux de Matheron ont fourni une théorie mathématique des variables régionalisées [7, 10, 11]. Cette théorie est aujourd'hui utilisée dans de nombreux domaines : sciences de la Terre, hydrologie, environnement, agronomie, climatologie et cartographie. Des exposés détaillés sont donnés dans [15], [14], [4], [3] et [1].

L'objectif de ce chapitre est de présenter les notions mathématiques qui fondent le krigeage. Nous introduisons d'abord la notion de variable régionalisée et son interprétation probabiliste comme réalisation d'un champ aléatoire. Nous définissons ensuite les moments d'un champ aléatoire, la stationnarité, la covariance et le variogramme. Enfin, nous discutons l'hypothèse intrinsèque et le rôle de la tendance spatiale.

## 2.2 Variable régionalisée

### Point de vue déterministe

Une variable régionalisée est une grandeur définie sur un domaine spatial. Si  $D \subset \mathbb{R}^d$ , on peut la représenter par une fonction

$$z : D \longrightarrow \mathbb{R}.$$

La valeur  $z(x)$  dépend de la position  $x$ . Cette dépendance spatiale distingue une variable régionalisée d'une simple série de valeurs numériques.

**Définition 2.1.** *On appelle variable régionalisée une grandeur numérique définie sur un domaine spatial  $D \subset \mathbb{R}^d$ , dont la valeur dépend de la position. Elle est représentée par une fonction  $z : D \rightarrow \mathbb{R}$ .*

Dans les applications, la fonction  $z$  n'est pas connue partout. On observe seulement

$$z(x_1), \dots, z(x_n),$$

et l'on cherche à prédire  $z(x_0)$  pour un point non observé  $x_0$ .

### Point de vue probabiliste

La géostatistique introduit une interprétation probabiliste. La fonction observée est considérée comme une réalisation d'un champ aléatoire

$$Z = \{Z(x) : x \in D\}.$$

Pour chaque position  $x$ ,  $Z(x)$  est une variable aléatoire. La valeur observée  $z(x)$  correspond à une réalisation particulière de cette variable.

**Définition 2.2.** *Un champ aléatoire réel sur  $D \subset \mathbb{R}^d$  est une famille de variables aléatoires*

$$Z = \{Z(x) : x \in D\},$$

*définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ .*

Pour chaque  $\omega \in \Omega$ , l'application

$$x \longmapsto Z(x, \omega)$$

est une réalisation possible du champ. Dans une étude réelle, on observe généralement

une seule réalisation. Cette particularité rend les données spatiales différentes des données issues de répétitions indépendantes.

**Remarque 2.1.** *La modélisation par champ aléatoire ne signifie pas que le phénomène réel est aléatoire au sens physique. Elle fournit un cadre mathématique permettant de parler de moyenne, de covariance, de variogramme et d'incertitude de prédiction.*

## 2.3 Moments d'un champ aléatoire

### Moyenne

La fonction moyenne du champ est définie par

$$m(x) = \mathbb{E}[Z(x)],$$

lorsque cette espérance existe. Elle représente le niveau moyen du phénomène au point  $x$ .

Si  $m(x)$  est constante, on écrit

$$m(x) = m.$$

Si  $m(x)$  varie avec  $x$ , on parle de tendance ou de dérive spatiale.

**Définition 2.3.** *Soit  $Z = \{Z(x) : x \in D\}$  un champ aléatoire tel que  $\mathbb{E}|Z(x)| < \infty$  pour tout  $x \in D$ . La fonction  $m(x) = \mathbb{E}[Z(x)]$  est appelée fonction moyenne du champ.*

### Variance et covariance

Lorsque  $Z(x)$  est de carré intégrable, sa variance est

$$\text{Var}(Z(x)) = \mathbb{E}[(Z(x) - m(x))^2].$$

La covariance entre deux points  $x$  et  $y$  est

$$C(x, y) = \text{Cov}(Z(x), Z(y)) = \mathbb{E}[(Z(x) - m(x))(Z(y) - m(y))].$$

La covariance mesure la dépendance linéaire entre les valeurs du champ en deux positions. Si  $C(x, y)$  est grande et positive, les valeurs ont tendance à varier dans le même sens. Si elle est proche de zéro, la dépendance linéaire est faible.

**Proposition 2.1.** *Pour tout ensemble fini de points  $x_1, \dots, x_n \in D$ , la matrice*

$$\Sigma = (C(x_i, x_j))_{1 \leq i, j \leq n}$$

*est symétrique et semi-définie positive.*

*Démonstration.* La symétrie provient de la symétrie de la covariance. Pour tout vecteur  $a = (a_1, \dots, a_n)^T \in \mathbb{R}^n$ ,

$$a^T \Sigma a = \sum_{i=1}^n \sum_{j=1}^n a_i a_j C(x_i, x_j) = \text{Var} \left( \sum_{i=1}^n a_i Z(x_i) \right) \geq 0.$$

Donc  $\Sigma$  est semi-définie positive. □

Cette propriété est essentielle : toute fonction ne peut pas être utilisée comme covariance. Une covariance admissible doit produire des matrices semi-définies positives pour tout choix fini de points.

## 2.4 Stationnarité

La stationnarité exprime l'idée que certaines propriétés probabilistes du champ sont invariantes par translation. Elle permet de simplifier fortement la modélisation.

### Stationnarité stricte

**Définition 2.4.** *Un champ  $Z = \{Z(x) : x \in D\}$  est dit strictement stationnaire si, pour tout entier  $k \geq 1$ , pour tous points  $x_1, \dots, x_k$  et tout vecteur  $h$  tel que les points  $x_i + h$  appartiennent à  $D$ , les vecteurs*

$$(Z(x_1), \dots, Z(x_k))$$

*et*

$$(Z(x_1 + h), \dots, Z(x_k + h))$$

*ont la même loi.*

Cette propriété est forte : elle concerne toutes les lois de dimension finie. Elle est difficile à vérifier directement en pratique.

### Stationnarité d'ordre deux

**Définition 2.5.** *Un champ de carré intégrable est dit stationnaire d'ordre deux si*

$$\mathbb{E}[Z(x)] = m$$

*est constante et si la covariance dépend seulement du vecteur de séparation :*

$$\text{Cov}(Z(x), Z(x+h)) = C(h).$$

Dans ce cas,

$$C(0) = \text{Var}(Z(x)),$$

et  $C(h)$  décrit la dépendance spatiale en fonction du décalage  $h$ .

**Remarque 2.2.** *La stationnarité d'ordre deux ne signifie pas que la réalisation observée est constante. Elle signifie seulement que les moments de premier et de second ordre sont invariants par translation.*

### Stationnarité locale

Dans les applications, la stationnarité globale est rarement parfaitement vraie. On l'utilise souvent localement, dans une zone où le phénomène peut être considéré comme suffisamment homogène. Cette idée justifie l'utilisation de voisinages locaux dans le krigeage.

Si une tendance globale est présente, on peut écrire

$$Z(x) = m(x) + \varepsilon(x),$$

où  $m(x)$  représente la tendance et  $\varepsilon(x)$  le résidu spatial. On cherche alors à modéliser la dépendance spatiale des résidus plutôt que celle de la variable brute.

## 2.5 Covariance, corrélation, isotropie et anisotropie

Sous stationnarité d'ordre deux, la covariance s'écrit  $C(h)$ . La corrélation spatiale est

$$\rho(h) = \frac{C(h)}{C(0)},$$

lorsque  $C(0) > 0$ . Elle vérifie  $\rho(0) = 1$ .

Un champ est dit isotrope si sa covariance dépend seulement de la distance :

$$C(h) = C_0(\|h\|).$$

Dans le cas contraire, on parle d'anisotropie. L'anisotropie apparaît lorsque la continuité spatiale dépend de la direction : par exemple dans une vallée, le long d'un cours d'eau, sous l'effet d'un vent dominant ou dans une structure géologique orientée.

**Exemple 2.1.** *Dans une région montagneuse, la température peut être plus fortement corrélée le long d'une même altitude que dans la direction verticale. Dans ce cas, une modélisation isotrope peut être insuffisante.*

L'isotropie simplifie la modélisation, mais elle doit être considérée comme une hypothèse. Dans un travail appliqué complet, elle doit être examinée à l'aide de variogrammes directionnels.

## 2.6 Variogramme

### Définition

Le variogramme mesure la variabilité des différences spatiales. Sous des conditions appropriées, on définit le semi-variogramme par

$$\gamma(h) = \frac{1}{2} \text{Var}(Z(x+h) - Z(x)).$$

Il ne dépend pas de  $x$  lorsque les accroissements sont stationnaires.

**Définition 2.6.** *La fonction  $\gamma$  définie par*

$$\gamma(h) = \frac{1}{2} \text{Var}(Z(x+h) - Z(x))$$

*est appelée semi-variogramme. Par abus de langage, on l'appelle souvent simplement variogramme.*

Le variogramme est une mesure de dissimilarité. Si  $\gamma(h)$  est faible, les valeurs séparées par  $h$  sont proches en moyenne quadratique. Si  $\gamma(h)$  est grande, elles peuvent différer fortement.

### Relation avec la covariance

Si le champ est stationnaire d'ordre deux, alors

$$\begin{aligned} \text{Var}(Z(x+h) - Z(x)) &= \text{Var}(Z(x+h)) + \text{Var}(Z(x)) - 2\text{Cov}(Z(x+h), Z(x)) \\ &= 2C(0) - 2C(h). \end{aligned}$$

Par conséquent,

$$\gamma(h) = C(0) - C(h).$$

**Proposition 2.2.** *Sous stationnarité d'ordre deux, le variogramme et la covariance sont liés par*

$$\gamma(h) = C(0) - C(h).$$

Cette relation montre que covariance et variogramme contiennent la même information lorsque la stationnarité d'ordre deux est valable. Une forte covariance correspond à une faible dissimilarité.

### Propriétés élémentaires

Le variogramme vérifie

$$\gamma(0) = 0, \quad \gamma(h) \geq 0, \quad \gamma(h) = \gamma(-h).$$

La première propriété provient du fait que  $Z(x) - Z(x) = 0$ . La positivité provient du fait que  $\gamma$  est une demi-variance. La symétrie provient du fait que changer le signe d'une différence ne change pas sa variance.

Ces propriétés sont nécessaires, mais non suffisantes. Un variogramme admissible doit satisfaire une condition de négativité conditionnelle, qui sera rappelée au chapitre suivant.

## 2.7 Hypothèse intrinsèque

La stationnarité d'ordre deux peut être trop restrictive. L'hypothèse intrinsèque est plus faible : elle impose des conditions seulement sur les accroissements.

**Définition 2.7.** *Un champ  $Z$  satisfait l'hypothèse intrinsèque si, pour tout vecteur  $h$ ,*

$$\mathbb{E}[Z(x+h) - Z(x)] = 0$$

et

$$\text{Var}(Z(x+h) - Z(x)) = 2\gamma(h),$$

où  $\gamma(h)$  ne dépend pas de  $x$ .

**Proposition 2.3.** *Tout champ stationnaire d'ordre deux satisfait l'hypothèse intrinsèque.*

*Démonstration.* Si  $Z$  est stationnaire d'ordre deux, sa moyenne est constante. Donc

$$\mathbb{E}[Z(x+h) - Z(x)] = m - m = 0.$$

De plus,

$$\text{Var}(Z(x+h) - Z(x)) = 2(C(0) - C(h)),$$

qui dépend seulement de  $h$ . En posant  $\gamma(h) = C(0) - C(h)$ , on obtient l'hypothèse intrinsèque.  $\square$

L'hypothèse intrinsèque justifie l'utilisation directe du variogramme, même lorsque la covariance n'est pas l'objet le plus naturel. Elle est donc fondamentale dans la théorie des variables régionalisées.

## 2.8 Tendances, dérive et résidus

Dans de nombreuses applications, la moyenne n'est pas constante. On écrit alors

$$Z(x) = m(x) + \varepsilon(x),$$

où  $m(x)$  est la tendance et  $\varepsilon(x)$  le résidu spatial. La tendance peut être modélisée par

$$m(x) = \sum_{k=0}^p \beta_k f_k(x),$$

où les fonctions  $f_k$  sont connues et les coefficients  $\beta_k$  inconnus.

Cette décomposition est importante pour distinguer la variation à grande échelle de la dépendance locale. Le krigeage ordinaire suppose une moyenne constante inconnue, tandis que le krigeage universel autorise une tendance dépendant de la position.

## 2.9 Support spatial et effet d'échelle

Une observation n'est pas toujours strictement ponctuelle. Elle peut représenter une moyenne sur un pixel, une parcelle, un volume ou une période de temps. Le support spatial influence la variabilité observée. En général, plus le support est grand, plus les variations locales sont lissées.

Dans ce mémoire, les observations sont considérées comme ponctuelles. Cette hypothèse est classique dans une introduction au krigeage, mais elle doit être gardée à l'esprit lors de l'interprétation de données réelles.

## 2.10 Conclusion du chapitre

Ce chapitre a présenté les notions mathématiques de base de la géostatistique. La variable régionalisée a été interprétée comme la réalisation d'un champ aléatoire. Les notions de moyenne, covariance, stationnarité, isotropie, anisotropie, variogramme et hypothèse intrinsèque ont été introduites. Ces outils permettent de construire le cadre du krigeage. Le chapitre suivant sera consacré à l'estimation et à la modélisation pratique du variogramme.

# Chapitre 3

## Variogramme : estimation, modélisation et validation

### 3.1 Introduction

Le chapitre précédent a montré que le variogramme est un outil central de la géostatistique. Il mesure la dissimilarité moyenne entre deux valeurs d'un champ en fonction de leur séparation spatiale. Dans le krigeage, il intervient directement dans le calcul des poids et de la variance de prédiction. Un modèle variographique mal choisi peut donc conduire à des prédictions peu fiables, même si les équations de krigeage sont correctement résolues.

L'objectif de ce chapitre est de présenter l'analyse variographique du point de vue pratique. Nous partons des données observées et nous expliquons comment construire un variogramme expérimental, comment interpréter ses paramètres, comment ajuster un modèle théorique admissible et comment valider ce modèle. Cette démarche constitue une étape indispensable avant l'application du krigeage. Les références classiques sont [15], [14], [4], [3], [1] et [2].

### 3.2 Nuage variographique

Avant de regrouper les distances en classes, il est utile de considérer le nuage variographique. Pour chaque paire de points  $(x_i, x_j)$ , on calcule

$$d_{ij} = \|x_i - x_j\|$$

et

$$v_{ij} = \frac{1}{2}(z(x_i) - z(x_j))^2.$$

Le nuage variographique est l'ensemble des points

$$(d_{ij}, v_{ij}), \quad 1 \leq i < j \leq n.$$

Ce nuage donne une première vision de la relation entre distance et dissimilarité. S'il existe une continuité spatiale, les valeurs  $v_{ij}$  tendent à être faibles pour les petites distances. Cependant, le nuage est souvent dispersé. Il devient donc nécessaire de regrouper les paires par classes de distance.

### 3.3 Variogramme expérimental

#### Définition

Soit  $N(h)$  l'ensemble des paires de points dont la distance appartient à une classe centrée en  $h$ . Le variogramme expérimental est défini par

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} (z(x_i) - z(x_j))^2.$$

**Définition 3.1.** *Le variogramme expérimental est la fonction empirique qui associe à chaque classe de distance  $h$  la moyenne des demi-différences quadratiques des observations séparées par cette distance.*

Cette définition explique pourquoi le variogramme est bien adapté à l'hypothèse intrinsèque : il repose sur les différences entre observations, et non directement sur la moyenne du champ.

#### Choix des classes de distance

La construction du variogramme expérimental dépend du choix des classes de distance. Il faut fixer une largeur de classe, un nombre de classes et une distance maximale. Ces choix influencent fortement l'allure du variogramme.

Si les classes sont trop étroites, elles contiennent peu de paires et les estimations sont instables. Si elles sont trop larges, le variogramme est trop lissé et peut masquer des structures importantes. Une règle pratique consiste à limiter l'analyse aux distances pour lesquelles le nombre de paires reste suffisant.

## Nombre de paires et fiabilité

La fiabilité d'une valeur  $\hat{\gamma}(h)$  dépend du nombre de paires  $|N(h)|$ . Une valeur calculée à partir de nombreuses paires est généralement plus stable qu'une valeur calculée à partir de quelques paires. Il est donc utile de représenter ou de conserver le nombre de paires par classe.

**Remarque 3.1.** *Le variogramme expérimental n'est pas un modèle. C'est une statistique descriptive. Il doit être ajusté par une fonction théorique admissible avant d'être utilisé dans un système de krigeage.*

## 3.4 Effet de pépité, palier et portée

### Effet de pépité

L'effet de pépité correspond à une discontinuité éventuelle du variogramme à l'origine. Théoriquement, si le champ est continu en moyenne quadratique, on a

$$\gamma(h) \rightarrow 0 \quad \text{lorsque} \quad \|h\| \rightarrow 0.$$

Dans les données réelles, le variogramme peut tendre vers une valeur positive. Cette valeur, notée  $c_0$ , peut traduire une erreur de mesure, une variabilité à très petite échelle ou une imprécision de localisation.

### Palier

Le palier est la valeur vers laquelle le variogramme se stabilise lorsque la distance augmente. Dans le cas stationnaire d'ordre deux, il correspond à la variance du champ. Si le variogramme atteint un palier, cela signifie qu'au-delà d'une certaine distance les observations ne sont plus fortement dépendantes.

### Portée

La portée est la distance à partir de laquelle le variogramme atteint son palier ou devient proche de celui-ci. Elle représente l'échelle de continuité spatiale. Une grande portée indique une dépendance spatiale longue. Une faible portée indique que la dépendance disparaît rapidement.

### 3.5 Isotropie et anisotropie dans le variogramme

Dans le cas isotrope, le variogramme dépend seulement de la distance :

$$\gamma(h) = \gamma_0(\|h\|).$$

Dans le cas anisotrope, il dépend également de la direction. Pour détecter l'anisotropie, on calcule des variogrammes directionnels. On regroupe alors les paires selon leur distance, mais aussi selon l'orientation du vecteur  $x_j - x_i$ .

L'anisotropie peut être géométrique ou zonale. Dans l'anisotropie géométrique, le palier reste le même mais la portée varie selon la direction. Dans l'anisotropie zonale, le palier lui-même peut dépendre de la direction. Ces situations demandent des modèles plus complexes.

**Remarque 3.2.** *Dans ce mémoire, l'application numérique est principalement isotrope. Cependant, la prise en compte de l'anisotropie constitue une extension importante dans les applications réelles.*

### 3.6 Admissibilité d'un variogramme

Un modèle de variogramme doit être mathématiquement admissible. Il ne suffit pas qu'une fonction soit croissante et positive. Elle doit pouvoir correspondre à une structure de dépendance cohérente.

**Définition 3.2.** *Une fonction  $\gamma : \mathbb{R}^d \rightarrow \mathbb{R}$  est dite conditionnellement négative si, pour tout  $n \geq 1$ , tous points  $x_1, \dots, x_n$  et tous réels  $a_1, \dots, a_n$  vérifiant*

$$\sum_{i=1}^n a_i = 0,$$

on a

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(x_i - x_j) \leq 0.$$

Cette condition garantit la positivité des variances des combinaisons linéaires d'accroissements. En pratique, on utilise souvent des familles de modèles déjà connues pour être admissibles.

## 3.7 Modèles théoriques classiques

### Modèle pépité pur

Le modèle pépité pur est défini par

$$\gamma(h) = \begin{cases} 0, & h = 0, \\ c_0, & h \neq 0. \end{cases}$$

Il correspond à une absence de continuité spatiale à l'échelle observée.

### Modèle sphérique

Le modèle sphérique est défini par

$$\gamma(h) = \begin{cases} c_0 + c \left[ \frac{3\|h\|}{2a} - \frac{1}{2} \left( \frac{\|h\|}{a} \right)^3 \right], & 0 < \|h\| \leq a, \\ c_0 + c, & \|h\| > a, \\ 0, & h = 0. \end{cases}$$

Il atteint exactement le palier à la distance  $a$ .

### Modèle exponentiel

Le modèle exponentiel est

$$\gamma(h) = c_0 + c \left( 1 - \exp\left(-\frac{\|h\|}{a}\right) \right), \quad h \neq 0,$$

avec  $\gamma(0) = 0$ . Il atteint son palier asymptotiquement. Une portée pratique souvent utilisée est environ  $3a$ .

### Modèle gaussien

Le modèle gaussien est

$$\gamma(h) = c_0 + c \left( 1 - \exp\left(-\frac{\|h\|^2}{a^2}\right) \right), \quad h \neq 0,$$

avec  $\gamma(0) = 0$ . Il est très régulier près de l'origine et convient aux phénomènes lisses.

### Choix entre les modèles

Le modèle sphérique décrit une dépendance qui disparaît à distance finie. Le modèle exponentiel décrit une dépendance qui décroît rapidement au voisinage de l'origine. Le

modèle gaussien correspond à une continuité plus forte. Le choix doit tenir compte de l'allure du variogramme expérimental, des critères d'ajustement et de l'interprétation du phénomène.

### 3.8 Ajustement d'un modèle de variogramme

Soient  $h_1, \dots, h_K$  les centres des classes de distance. On ajuste un modèle  $\gamma_\theta$  en minimisant un critère de moindres carrés :

$$Q(\theta) = \sum_{k=1}^K (\hat{\gamma}(h_k) - \gamma_\theta(h_k))^2.$$

Une version pondérée est

$$Q_w(\theta) = \sum_{k=1}^K w_k (\hat{\gamma}(h_k) - \gamma_\theta(h_k))^2,$$

où  $w_k$  peut dépendre du nombre de paires dans la classe.

L'ajustement peut être visuel, numérique ou mixte. Dans une analyse sérieuse, il est préférable de compléter l'ajustement visuel par une validation croisée.

### 3.9 Validation du modèle variographique

La validation croisée "leave-one-out" consiste à retirer une observation, à la prédire à partir des autres, puis à comparer la prédiction avec la valeur observée. L'erreur est

$$e_i = z(x_i) - \hat{z}_{-i}(x_i).$$

On calcule ensuite des critères tels que *ME*, *MAE* et *RMSE*.

Dans le cas du krigeage, on peut aussi analyser les erreurs standardisées :

$$e_i^* = \frac{z(x_i) - \hat{z}_{-i}(x_i)}{\sigma_{K,-i}(x_i)}.$$

Si le modèle est bien calibré, ces erreurs doivent être centrées autour de zéro et avoir une dispersion proche de un.

#### Difficultés pratiques

L'analyse variographique est sensible à plusieurs facteurs. Un nombre faible d'obser-

vations peut rendre le variogramme instable. Des valeurs atypiques peuvent fortement influencer les demi-différences quadratiques. Une tendance spatiale non prise en compte peut gonfler artificiellement le variogramme. L'anisotropie peut être masquée par un variogramme isotrope. Enfin, le choix des classes de distance peut modifier l'interprétation.

Ces difficultés montrent que l'analyse variographique n'est pas une étape mécanique. Elle demande une lecture statistique et géographique des données.

### Lien avec le krigeage

Une fois le variogramme théorique choisi, il est utilisé dans les systèmes de krigeage. Dans le krigeage ordinaire, les poids  $\lambda_i$  vérifient

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \mu = \gamma(x_i - x_0), & i = 1, \dots, n, \\ \sum_{j=1}^n \lambda_j = 1. \end{cases}$$

Cette formule montre que le variogramme intervient à deux niveaux : entre observations et entre observations et point à prédire.

## 3.10 Conclusion du chapitre

Ce chapitre a présenté l'analyse variographique : construction du nuage variographique, calcul du variogramme expérimental, interprétation de l'effet de pépite, du palier et de la portée, choix des modèles théoriques, ajustement et validation. Le chapitre suivant utilise ces éléments pour construire les systèmes de krigeage.

# Chapitre 4

## Théorie du krigeage

### 4.1 Introduction

Le krigeage est une méthode d'interpolation statistique qui utilise la structure de dépendance spatiale du phénomène. Contrairement aux méthodes déterministes, il ne choisit pas les poids seulement à partir de la distance. Il construit un prédicteur linéaire sans biais dont la variance de l'erreur est minimale. Dans un cadre approprié, il est donc le meilleur prédicteur linéaire sans biais, ou *Best Linear Unbiased Predictor*.

Cette méthode est issue des travaux de D. Krige sur l'estimation minière et de la formalisation mathématique de Matheron [7, 10, 11]. Elle est aujourd'hui l'une des méthodes centrales de la géostatistique [15, 4, 2].

Ce chapitre présente la théorie du krigeage. Nous commençons par la prédiction linéaire et la variance de prédiction. Nous étudions ensuite le krigeage simple, le krigeage ordinaire et le krigeage universel. Enfin, nous discutons l'interprétation des poids et de la variance de krigeage.

### 4.2 Prédiction linéaire

Soit  $Z = \{Z(x) : x \in D\}$  un champ aléatoire observé aux sites  $x_1, \dots, x_n$ . On veut prédire  $Z(x_0)$  en un point non observé. Le krigeage cherche un prédicteur de la forme

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i).$$

Les coefficients  $\lambda_i$  sont les poids de krigeage.

L'erreur de prédiction est

$$e(x_0) = Z(x_0) - \widehat{Z}(x_0).$$

La variance de cette erreur est

$$\sigma_K^2(x_0) = \text{Var}(e(x_0)).$$

Le principe du krigeage est de minimiser cette variance sous une contrainte de non-biais.

**Définition 4.1.** *Un prédicteur linéaire  $\widehat{Z}(x_0)$  est dit sans biais si*

$$\mathbb{E}[\widehat{Z}(x_0) - Z(x_0)] = 0.$$

La forme de la contrainte de non-biais dépend de l'hypothèse faite sur la moyenne du champ.

### 4.3 Variance de l'erreur dans le cas stationnaire

Supposons que  $Z$  soit stationnaire d'ordre deux, de covariance  $C(h)$ . Pour un prédicteur linéaire,

$$\begin{aligned} \sigma_K^2(x_0) &= \text{Var} \left( Z(x_0) - \sum_{i=1}^n \lambda_i Z(x_i) \right) \\ &= C(0) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i - x_j) - 2 \sum_{i=1}^n \lambda_i C(x_i - x_0). \end{aligned}$$

Cette expression dépend du modèle de covariance et de la configuration géométrique des points. Elle ne dépend pas directement des valeurs observées.

### 4.4 Krigeage simple

#### Hypothèses

Le krigeage simple suppose que la moyenne est connue et constante :

$$\mathbb{E}[Z(x)] = m,$$

où  $m$  est connu. On suppose aussi que la covariance  $C(h)$  est connue.

## Prédicteur

Le prédicteur du krigeage simple est

$$\widehat{Z}_{SK}(x_0) = m + \sum_{i=1}^n \lambda_i (Z(x_i) - m).$$

Ce prédicteur est automatiquement sans biais, car les écarts  $Z(x_i) - m$  ont une moyenne nulle.

## Système

La minimisation de la variance conduit au système

$$\sum_{j=1}^n \lambda_j C(x_i - x_j) = C(x_i - x_0), \quad i = 1, \dots, n.$$

En notation matricielle,

$$\mathbf{C}\boldsymbol{\lambda} = \mathbf{c}_0,$$

où

$$\mathbf{C} = (C(x_i - x_j))_{i,j}, \quad \mathbf{c}_0 = (C(x_1 - x_0), \dots, C(x_n - x_0))^T.$$

Si  $\mathbf{C}$  est inversible,

$$\boldsymbol{\lambda} = \mathbf{C}^{-1}\mathbf{c}_0.$$

## Variance

La variance du krigeage simple est

$$\sigma_{SK}^2(x_0) = C(0) - \mathbf{c}_0^T \mathbf{C}^{-1} \mathbf{c}_0.$$

Elle diminue lorsque le point à prédire est fortement corrélé aux observations.

## 4.5 Krigeage ordinaire

### Motivation

Dans les applications, la moyenne est rarement connue. Le krigeage ordinaire suppose qu'elle est constante mais inconnue :

$$\mathbb{E}[Z(x)] = m.$$

Pour garantir le non-biais quelle que soit la valeur de  $m$ , on impose une contrainte sur les poids.

### Contrainte de non-biais

Le prédicteur est

$$\widehat{Z}_{OK}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i).$$

Son espérance vaut

$$\mathbb{E}[\widehat{Z}_{OK}(x_0)] = m \sum_{i=1}^n \lambda_i.$$

Pour que cette espérance soit égale à  $m$ , il faut et il suffit que

$$\sum_{i=1}^n \lambda_i = 1.$$

**Proposition 4.1.** *Dans le krigage ordinaire, la condition*

$$\sum_{i=1}^n \lambda_i = 1$$

*est équivalente au non-biais pour toute moyenne constante inconnue.*

*Démonstration.* Le résultat découle directement de

$$\mathbb{E}[\widehat{Z}_{OK}(x_0)] = m \sum_{i=1}^n \lambda_i.$$

Cette quantité est égale à  $m$  pour toute valeur de  $m$  si et seulement si la somme des poids vaut 1. □

### Système avec covariance

On minimise la variance de l'erreur sous la contrainte  $\sum_i \lambda_i = 1$ . En utilisant un multiplicateur de Lagrange  $\mu$ , on obtient

$$\begin{cases} \sum_{j=1}^n \lambda_j C(x_i - x_j) + \mu = C(x_i - x_0), & i = 1, \dots, n, \\ \sum_{j=1}^n \lambda_j = 1. \end{cases}$$

En notation matricielle,

$$\begin{pmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \mu \end{pmatrix} = \begin{pmatrix} \mathbf{c}_0 \\ 1 \end{pmatrix}.$$

### Système avec variogramme

Sous l'hypothèse intrinsèque, on utilise souvent le variogramme. Le système devient

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \mu = \gamma(x_i - x_0), & i = 1, \dots, n, \\ \sum_{j=1}^n \lambda_j = 1. \end{cases}$$

Cette formulation est la plus fréquente en pratique, car le variogramme est directement estimé à partir des données.

### Variance du krigeage ordinaire

La variance de krigeage ordinaire est

$$\sigma_{OK}^2(x_0) = \sum_{i=1}^n \lambda_i \gamma(x_i - x_0) + \mu.$$

Elle dépend des positions des points et du modèle de variogramme. Elle ne dépend pas directement des valeurs observées.

## 4.6 Krigeage universel

### Tendance spatiale

Lorsque la moyenne varie dans l'espace, on écrit

$$m(x) = \mathbb{E}[Z(x)] = \sum_{k=0}^p \beta_k f_k(x),$$

où les fonctions  $f_k$  sont connues et les coefficients  $\beta_k$  inconnus. Le krigeage universel est adapté à ce cadre.

## Contraintes de non-biais

Le prédicteur garde la forme

$$\widehat{Z}_{UK}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i).$$

Pour être sans biais pour toute valeur des coefficients  $\beta_k$ , il faut imposer

$$\sum_{i=1}^n \lambda_i f_k(x_i) = f_k(x_0), \quad k = 0, \dots, p.$$

## Système

Avec le variogramme, le système du krigeage universel est

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \sum_{k=0}^p \mu_k f_k(x_i) = \gamma(x_i - x_0), & i = 1, \dots, n, \\ \sum_{i=1}^n \lambda_i f_k(x_i) = f_k(x_0), & k = 0, \dots, p. \end{cases}$$

Le krigeage ordinaire est le cas particulier  $p = 0$  avec  $f_0(x) = 1$ .

## 4.7 Interprétation des poids de krigeage

Les poids de krigeage dépendent de la distance au point cible, mais aussi de la dépendance entre observations. Deux observations proches l'une de l'autre peuvent être redondantes. Le krigeage réduit alors leur influence combinée. C'est une différence essentielle avec l'IDW.

Les poids peuvent être négatifs. Cela n'est pas nécessairement une erreur : ils résultent d'une minimisation sous contrainte. Toutefois, des poids très grands en valeur absolue peuvent indiquer un problème de stabilité numérique, de modèle variographique ou de configuration des points.

## 4.8 Variance de krigeage et incertitude

Un avantage majeur du krigeage est de fournir une variance de prédiction. On peut donc produire deux cartes : une carte des valeurs prédites et une carte de variance. Cette dernière indique les zones où la prédiction est moins fiable.

Il faut cependant rappeler que la variance de krigeage est conditionnelle au modèle choisi. Si le variogramme est mal ajusté, la carte d'incertitude peut être trompeuse. La validation du modèle reste donc indispensable.

### **Comparaison avec les méthodes déterministes**

Dans l'IDW, les poids sont donnés par une formule directe fondée sur la distance. Dans le krigeage, ils sont issus d'un système linéaire utilisant le variogramme. Le krigeage est donc plus exigeant, mais aussi plus riche : il tient compte de la dépendance spatiale et fournit une variance de prédiction.

Cependant, le krigeage n'est pas automatiquement supérieur dans toutes les situations. Si le variogramme est mal estimé ou si le nombre de données est trop faible, une méthode déterministe simple peut donner des résultats compétitifs. C'est pourquoi les comparaisons par validation croisée sont importantes.

### **Difficultés numériques**

Les systèmes de krigeage peuvent être mal conditionnés lorsque des points sont très proches ou lorsque le modèle variographique produit des dépendances fortes. En pratique, on peut utiliser un voisinage local, supprimer les doublons, ajouter un effet de pépite ou vérifier la stabilité de la matrice.

Le coût de calcul augmente avec le nombre de points. Pour de très grands jeux de données, il faut recourir à des approximations, au krigeage local ou à des méthodes numériques adaptées.

## **4.9 Conclusion du chapitre**

Ce chapitre a présenté la théorie du krigeage. Le krigeage simple suppose une moyenne connue. Le krigeage ordinaire suppose une moyenne inconnue mais constante. Le krigeage universel prend en compte une tendance spatiale. Dans tous les cas, le principe est de construire un prédicteur linéaire sans biais et de variance minimale. Le chapitre suivant illustre ces notions par une application numérique.

# Chapitre 5

## Application numérique : interpolation spatiale et comparaison des méthodes

### 5.1 Introduction

Les chapitres précédents ont présenté les bases conceptuelles et mathématiques de l'interpolation spatiale par krigeage. Nous avons d'abord distingué les méthodes déterministes, qui reposent essentiellement sur des règles géométriques ou analytiques, des méthodes géostatistiques, qui introduisent un modèle explicite de dépendance spatiale. Nous avons ensuite étudié le variogramme, son estimation expérimentale, son ajustement par des modèles admissibles et son rôle dans la construction des systèmes de krigeage. Le présent chapitre a pour objectif d'illustrer l'ensemble de cette démarche à travers une application numérique reproductible.

L'exemple étudié concerne une variable environnementale interprétée comme une hauteur de précipitation observée en un nombre fini de stations. Le choix d'un tel exemple est naturel dans un mémoire consacré au krigeage. Les précipitations, les températures, les concentrations de polluants ou les teneurs géologiques sont des grandeurs qui varient dans l'espace et que l'on ne peut mesurer qu'en certains points. Le problème consiste donc à produire une estimation sur une grille ou sur une carte continue à partir d'un nombre limité d'observations ponctuelles.

Les données utilisées ici sont simulées. Ce choix ne diminue pas l'intérêt de l'étude : il permet au contraire de contrôler le protocole, de fixer une structure spatiale connue, de rendre les calculs reproductibles et de comparer les méthodes dans des conditions identiques. Une telle simulation ne remplace évidemment pas une application à des données réelles, mais elle constitue une étape pédagogique importante pour comprendre le rôle du variogramme, l'influence des paramètres et la signification de la validation

croisée.

Trois méthodes sont comparées :

- la méthode du plus proche voisin ;
- la pondération par l'inverse de la distance, notée IDW ;
- le krigeage ordinaire.

Ces trois méthodes ont été choisies car elles représentent trois niveaux de complexité. Le plus proche voisin est une méthode locale très simple. L'IDW est une méthode déterministe plus régulière, reposant sur une pondération décroissante avec la distance. Le krigeage ordinaire est une méthode géostatistique qui utilise un variogramme ajusté et fournit une variance de prédiction.

L'objectif du chapitre n'est donc pas seulement de produire des cartes. Il s'agit aussi de comprendre ce que chaque méthode apporte, ce qu'elle ne peut pas fournir, et comment les résultats numériques doivent être interprétés. En particulier, nous insistons sur trois aspects : la construction du variogramme, la comparaison des erreurs de validation et l'analyse de l'incertitude de prédiction.

## 5.2 Organisation générale du protocole numérique

La démarche suivie dans cette application peut être résumée en plusieurs étapes :

1. génération d'un ensemble de sites d'observation dans un domaine spatial ;
2. simulation d'une variable spatiale présentant à la fois une tendance régulière et une composante aléatoire corrélée ;
3. analyse exploratoire des observations ;
4. calcul du variogramme expérimental ;
5. ajustement de plusieurs modèles théoriques de variogramme ;
6. sélection d'un modèle variographique ;
7. interpolation par krigeage ordinaire ;
8. interpolation par plus proche voisin et par IDW ;
9. comparaison des méthodes par validation croisée ;
10. discussion des cartes, des erreurs et de la variance de krigeage.

Cette organisation reprend la logique générale d'une étude géostatistique appliquée, telle qu'elle est présentée dans les références classiques de la géostatistique [15, 14, 4, 3]. Elle montre aussi que le krigeage ne commence pas par la résolution d'un système linéaire. Avant de calculer les poids de krigeage, il faut examiner les données, estimer

une structure spatiale, choisir un modèle de variogramme et vérifier la qualité de ce modèle.

Tous les calculs sont effectués avec un code Python fourni en annexe. Le code génère les données, construit les figures, exporte les tableaux et permet de reproduire les résultats. Cette reproductibilité est importante, car elle permet de vérifier les calculs numériques et de modifier facilement certains paramètres, par exemple le nombre de stations, le modèle de covariance, la taille de la grille ou le paramètre de puissance de l'IDW.

## 5.3 Domaine spatial et génération des données

### Domaine d'étude

On considère le domaine spatial carré

$$D = [0, 100] \times [0, 100].$$

Ce domaine représente une région fictive dans laquelle une variable environnementale est observée. Les coordonnées sont notées

$$x = (x_1, x_2),$$

où  $x_1$  et  $x_2$  peuvent être interprétées comme deux coordonnées planes. Dans une application réelle, ces coordonnées pourraient provenir d'une projection cartographique.

On génère

$$n = 80$$

sites d'observation

$$x_i = (x_{i1}, x_{i2}), \quad i = 1, \dots, n,$$

répartis aléatoirement dans le domaine. Les observations sont notées

$$z_i = z(x_i), \quad i = 1, \dots, n.$$

Cette configuration irrégulière est volontaire. Dans la pratique, les stations météorologiques, les points de prélèvement ou les forages ne sont presque jamais disposés sur une grille parfaite. Certaines zones peuvent être mieux couvertes que d'autres, ce qui influence directement la qualité de l'interpolation.

## Modèle utilisé pour simuler la variable

La variable simulée est construite selon la décomposition

$$z(x) = m(x) + \varepsilon(x),$$

où  $m(x)$  est une tendance spatiale régulière et  $\varepsilon(x)$  est une composante aléatoire spatialement corrélée. Cette écriture est classique dans les modèles géostatistiques : elle sépare une composante à grande échelle, liée à la tendance, d'une composante résiduelle décrivant les fluctuations locales.

La tendance choisie est

$$m(x, y) = 55 + 0.12x - 0.08y + 7 \sin\left(\frac{x}{18}\right) \cos\left(\frac{y}{22}\right).$$

Elle contient une partie linéaire et une partie oscillante lisse. La partie linéaire introduit une variation globale selon les deux directions du plan, tandis que la partie trigonométrique produit des variations spatiales plus locales.

La composante résiduelle est simulée avec une covariance exponentielle :

$$\text{Cov}\{\varepsilon(x_i), \varepsilon(x_j)\} = \sigma^2 \exp\left(-\frac{\|x_i - x_j\|}{a}\right),$$

où les paramètres utilisés sont

$$\sigma^2 = 35, \quad a = 28.$$

Un terme de pépite est ajouté pour représenter une variabilité de très petite échelle ou une erreur de mesure. Cette construction permet d'obtenir une variable dont les valeurs proches sont corrélées, mais dont les observations ne sont pas parfaitement lisses.

**Remarque 5.1.** *Le modèle utilisé pour simuler les données est connu dans cette expérience, mais il est traité comme inconnu lors de l'analyse. Cela reproduit la situation d'une étude réelle : l'analyste observe les valeurs aux stations, mais il ne connaît pas le mécanisme exact qui a produit ces valeurs.*

## 5.4 Analyse exploratoire des observations

### Répartition spatiale des stations

La première étape consiste à visualiser les sites d'observation. La figure 5.1 présente la position des stations dans le domaine, avec une couleur associée à la valeur observée.

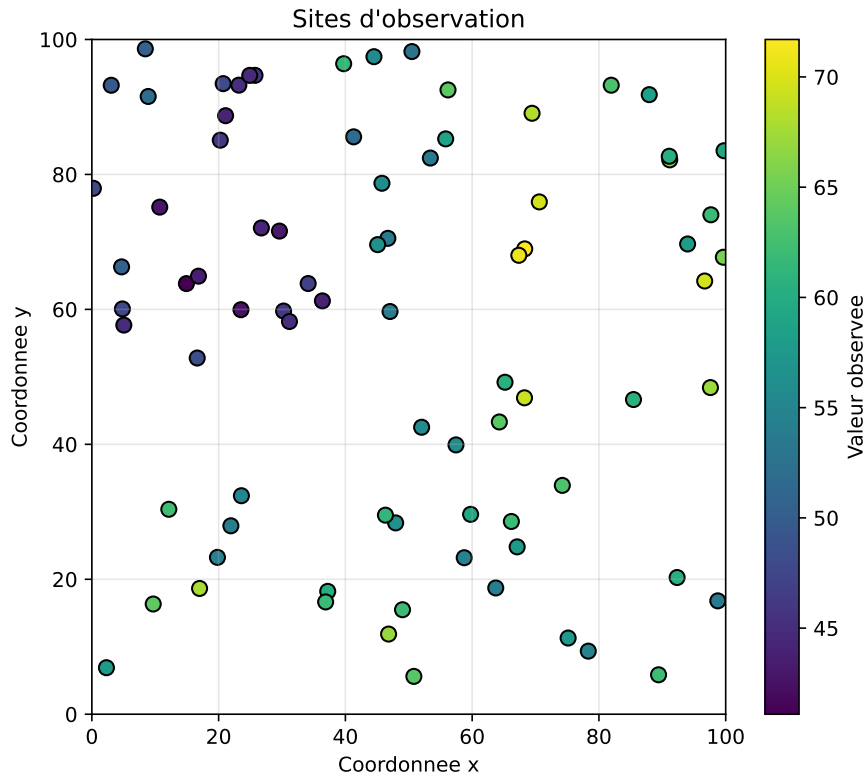


FIGURE 5.1 – Répartition spatiale des sites d’observation dans le domaine d’étude.

La figure montre que les stations couvrent globalement le domaine, mais de manière irrégulière. Certaines régions contiennent plusieurs points proches, tandis que d’autres sont moins échantillonnées. Cette propriété aura deux conséquences. D’une part, les méthodes locales comme le plus proche voisin et l’IDW seront fortement influencées par la densité locale des points. D’autre part, la variance de krigeage sera plus faible dans les zones bien couvertes et plus élevée dans les zones éloignées des observations.

### Statistiques descriptives

Le tableau 5.1 donne quelques statistiques descriptives des observations.

TABLE 5.1 – Statistiques descriptives des valeurs observées.

Indicateur	Valeur
Nombre d’observations	80.0000
Minimum	41.0948
Premier quartile	49.1823
Médiane	56.3066
Moyenne	56.0617
Troisième quartile	61.9321
Maximum	71.6993
Écart-type	8.0515

Les valeurs observées varient entre environ 41.09 et 71.70. La moyenne est proche de 56.06 et l'écart-type est proche de 8.05. Ces indicateurs montrent que la variable présente une variabilité notable, mais sans valeurs extrêmes très éloignées du reste de l'échantillon.

La médiane est proche de la moyenne, ce qui suggère une distribution relativement équilibrée. Toutefois, l'analyse de la distribution ne suffit pas pour comprendre le phénomène, car elle ignore la localisation des observations. Deux jeux de données peuvent avoir les mêmes statistiques descriptives mais des structures spatiales très différentes.

### Histogramme des observations

La figure 5.2 présente l'histogramme des valeurs observées.

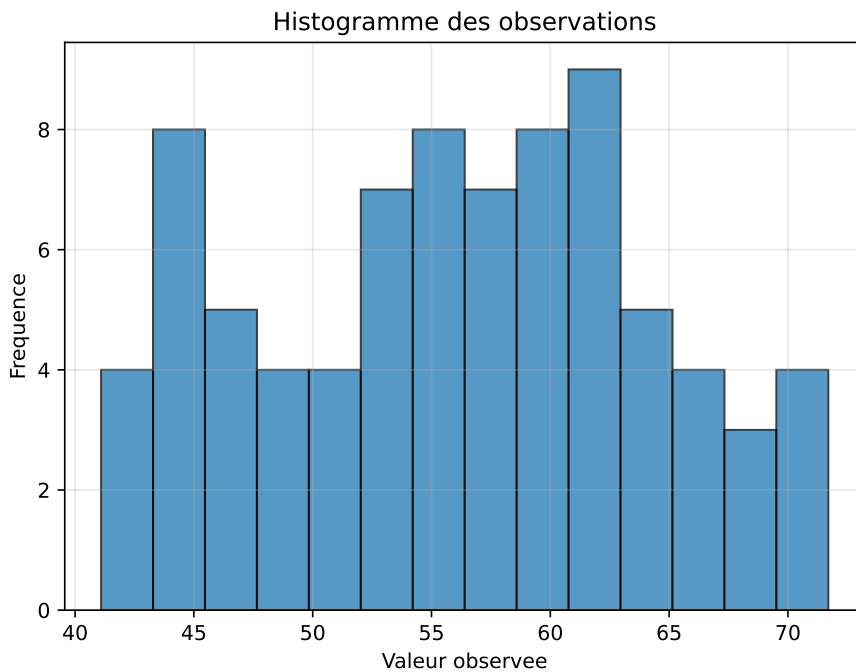


FIGURE 5.2 – Histogramme des valeurs observées.

L'histogramme donne une vision globale de la distribution marginale. Il permet de repérer une éventuelle asymétrie ou la présence de valeurs atypiques. Dans cette simulation, la distribution ne présente pas d'anomalie majeure. On peut donc poursuivre l'analyse variographique sans transformation préalable des données.

Dans une application réelle, cette étape serait plus délicate. Si les données présentaient une forte asymétrie, une transformation logarithmique ou une transformation de type Box-Cox pourrait être envisagée. Si des valeurs atypiques apparaissaient, il faudrait déterminer si elles correspondent à des erreurs de mesure ou à des phénomènes réels

localisés.

## 5.5 Analyse variographique

### Calcul du variogramme expérimental

Le variogramme expérimental est calculé à partir des différences entre paires d'observations. Pour deux stations  $x_i$  et  $x_j$ , on calcule la distance

$$d_{ij} = \|x_i - x_j\|$$

et la demi-différence quadratique

$$v_{ij} = \frac{1}{2}(z_i - z_j)^2.$$

Les distances sont ensuite regroupées en classes. Dans cette application, on utilise 12 classes de distance. Pour une classe centrée en  $h_k$ , le variogramme expérimental est

$$\hat{\gamma}(h_k) = \frac{1}{2|N(h_k)|} \sum_{(i,j) \in N(h_k)} (z_i - z_j)^2,$$

où  $N(h_k)$  désigne l'ensemble des paires dont la distance appartient à la classe considérée.

La figure 5.3 présente le variogramme expérimental obtenu.

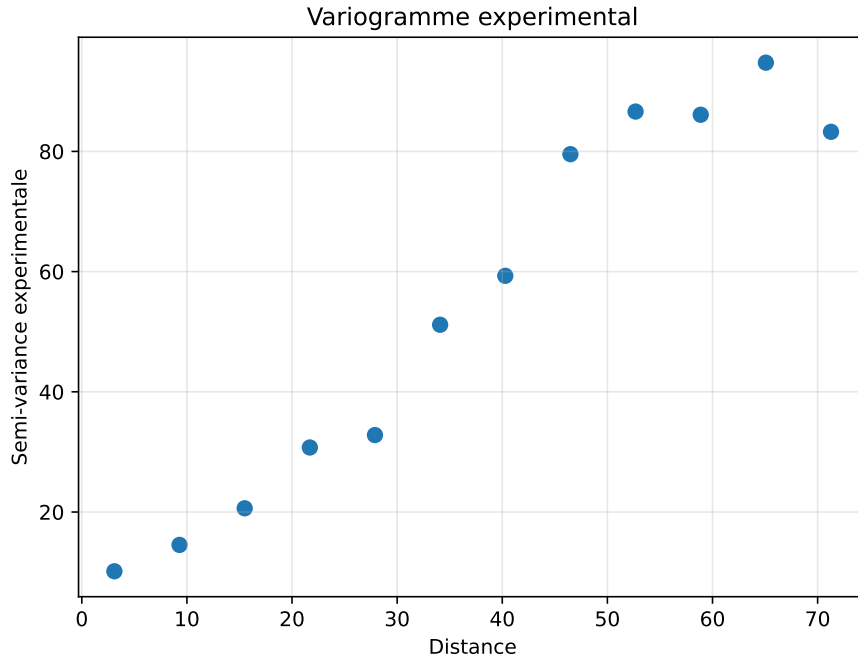


FIGURE 5.3 – Variogramme expérimental calculé à partir des observations.

Le variogramme expérimental augmente globalement avec la distance. Cela confirme que les observations proches sont, en moyenne, plus semblables que les observations éloignées. Cette tendance est l'indice principal d'une dépendance spatiale exploitable par le krigeage.

Il faut cependant rester prudent. Le variogramme expérimental n'est pas une courbe théorique lisse, mais une estimation obtenue à partir d'un nombre fini de paires. Certaines classes peuvent contenir plus de paires que d'autres. Les valeurs correspondant aux grandes distances sont souvent moins stables, car elles reposent parfois sur moins de paires et sont plus sensibles à la géométrie du domaine.

### Choix du nombre de classes

Le choix du nombre de classes est un compromis. Un nombre trop faible de classes produit un variogramme très lissé, qui peut masquer des variations importantes aux petites distances. Un nombre trop élevé de classes produit au contraire une courbe très instable, car chaque classe contient peu de paires. Le choix de 12 classes constitue ici un compromis raisonnable pour 80 observations.

Dans une étude plus approfondie, il serait possible de comparer plusieurs choix de classes. Une analyse robuste devrait vérifier que les conclusions principales ne changent pas radicalement lorsque l'on modifie modérément la largeur des classes ou la distance maximale utilisée.

## Ajustement de modèles théoriques

Un variogramme expérimental ne peut pas être utilisé directement dans le système de krigeage. Il faut lui ajuster un modèle théorique admissible. Trois modèles classiques sont considérés :

- le modèle sphérique ;
- le modèle exponentiel ;
- le modèle gaussien.

Ces modèles ont été présentés au chapitre 3. Ils dépendent d'un effet de pépité, d'une contribution structurale et d'une portée.

L'ajustement est réalisé par moindres carrés pondérés. Pour un modèle  $\gamma_\theta$  dépendant du paramètre  $\theta$ , on minimise un critère de la forme

$$SSE(\theta) = \sum_{k=1}^K w_k \{\widehat{\gamma}(h_k) - \gamma_\theta(h_k)\}^2,$$

où  $w_k$  est un poids associé à la classe  $k$ . Les classes contenant plus de paires reçoivent ainsi une influence plus importante dans l'ajustement.

Les paramètres estimés sont donnés dans le tableau 5.2.

TABLE 5.2 – Paramètres estimés pour les modèles de variogramme.

Modèle	Effet pépité	Contribution	Portée	Palier total	SSE pondérée
Sphérique	0.0000	99.9111	89.7590	99.9111	430.6813
Exponentiel	0.0000	243.8134	139.4823	243.8134	506.3311
Gaussien	5.7266	90.8957	39.9388	96.6223	243.8232

Le modèle gaussien donne la plus faible somme pondérée des carrés. Il est donc retenu pour la suite de l'application. Les paramètres retenus sont approximativement

$$c_0 = 5.7266, \quad c = 90.8957, \quad a = 39.9388.$$

Le palier total est

$$c_0 + c = 96.6223.$$

La figure 5.4 montre le variogramme expérimental et le modèle gaussien ajusté.

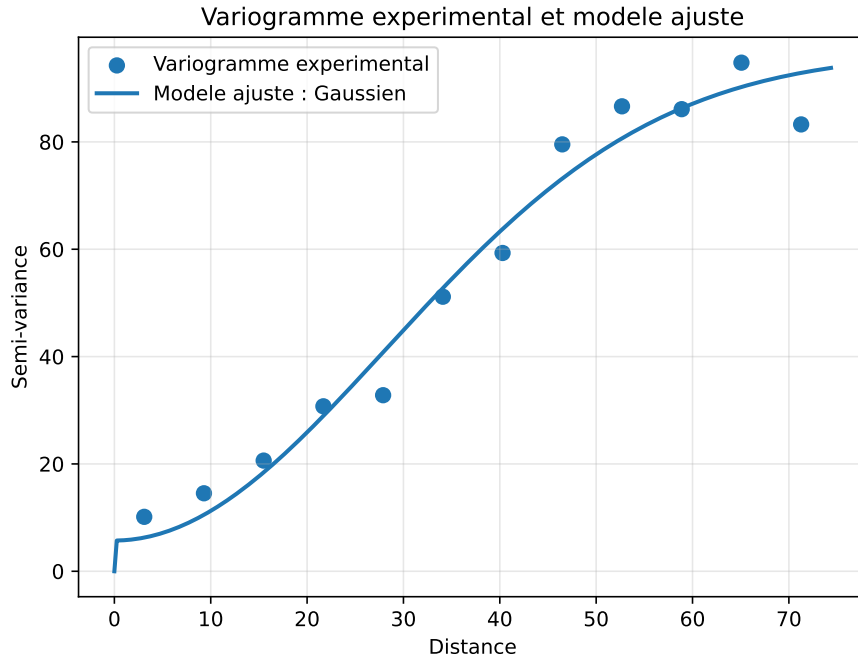


FIGURE 5.4 – Variogramme expérimental et modèle théorique ajusté.

Le modèle gaussien traduit une continuité spatiale relativement régulière. La présence d'un effet de pépite non nul indique une variabilité locale ou une composante de bruit. Ce point est important, car une pépite positive augmente la variance de prédiction près des observations et empêche d'interpréter le phénomène comme parfaitement régulier à très petite échelle.

**Remarque 5.2.** *La sélection du modèle par le seul critère SSE est utile, mais elle ne suffit pas à elle seule. Un bon modèle variographique doit être cohérent avec les données, produire des cartes plausibles, donner des erreurs de validation raisonnables et fournir une variance de krigeage correctement calibrée.*

## 5.6 Interpolation par krigeage ordinaire

### Système utilisé

Après sélection du variogramme gaussien, on applique le krigeage ordinaire. Pour un point  $x_0$  non observé, le prédicteur est

$$\hat{Z}_{OK}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i),$$

avec la contrainte de non-biais

$$\sum_{i=1}^n \lambda_i = 1.$$

Les poids sont obtenus en résolvant le système

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \mu = \gamma(x_i - x_0), & i = 1, \dots, n, \\ \sum_{j=1}^n \lambda_j = 1. \end{cases}$$

La variance de krigeage est ensuite calculée par

$$\sigma_{OK}^2(x_0) = \sum_{i=1}^n \lambda_i \gamma(x_i - x_0) + \mu.$$

Cette variance est l'une des différences fondamentales entre le krigeage et les méthodes déterministes. Elle ne donne pas l'erreur réelle, qui reste inconnue aux points non observés, mais une mesure de l'incertitude sous le modèle variographique adopté.

### Carte de prédiction

Les prédictions sont calculées sur une grille régulière de taille  $80 \times 80$ . La figure 5.5 présente la carte obtenue par krigeage ordinaire.

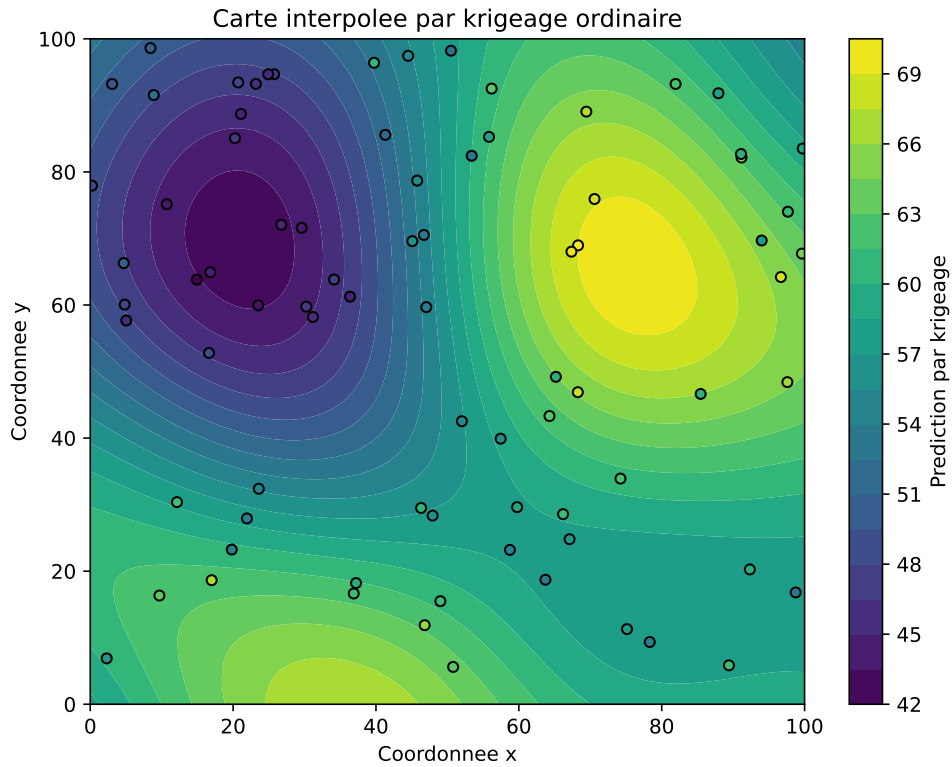


FIGURE 5.5 – Carte interpolée par krigeage ordinaire.

La carte est globalement lisse. Les transitions entre les zones de faibles et de fortes valeurs sont progressives. Cette régularité provient du modèle variographique retenu, mais aussi de la nature linéaire du krigeage. Contrairement au plus proche voisin, le krigeage ne découpe pas le domaine en zones constantes. Contrairement à l'IDW, il ne se contente pas d'une règle directe de décroissance avec la distance : il tient compte des dépendances entre toutes les observations utilisées dans le système.

### Carte de variance

La figure 5.6 présente la carte de variance du krigeage ordinaire.

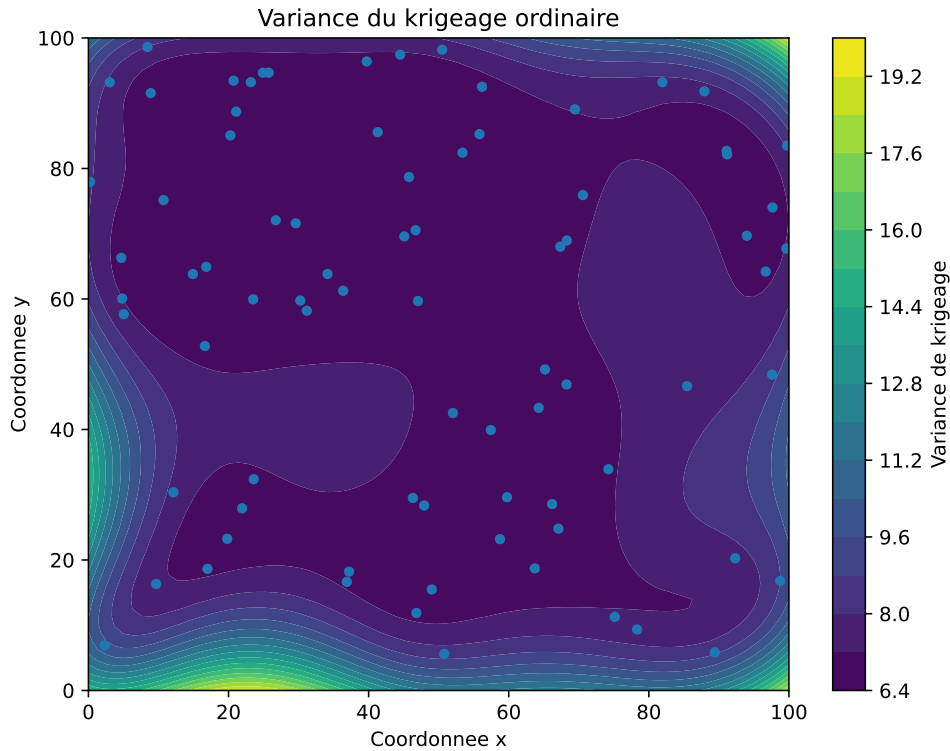


FIGURE 5.6 – Carte de variance associée au krigeage ordinaire.

L'incertitude est plus faible dans les zones proches des stations et plus élevée dans les zones moins couvertes par les observations. Cette carte est importante du point de vue pratique. Deux régions peuvent avoir des valeurs prédites semblables, mais des incertitudes très différentes. Une décision fondée sur la carte d'interpolation devrait donc tenir compte de la carte de variance.

Il faut toutefois rappeler que cette variance est conditionnelle au modèle. Si le variogramme est mal ajusté, la carte de variance peut donner une impression de précision excessive ou insuffisante. La validation croisée reste donc indispensable.

## 5.7 Interpolation par méthodes déterministes

### Méthode du plus proche voisin

La méthode du plus proche voisin attribue au point  $x_0$  la valeur de l'observation la plus proche :

$$\hat{Z}_{NN}(x_0) = Z(x_{i^*}), \quad i^* = \arg \min_{1 \leq i \leq n} \|x_0 - x_i\|.$$

La carte obtenue est donnée par la figure 5.7.

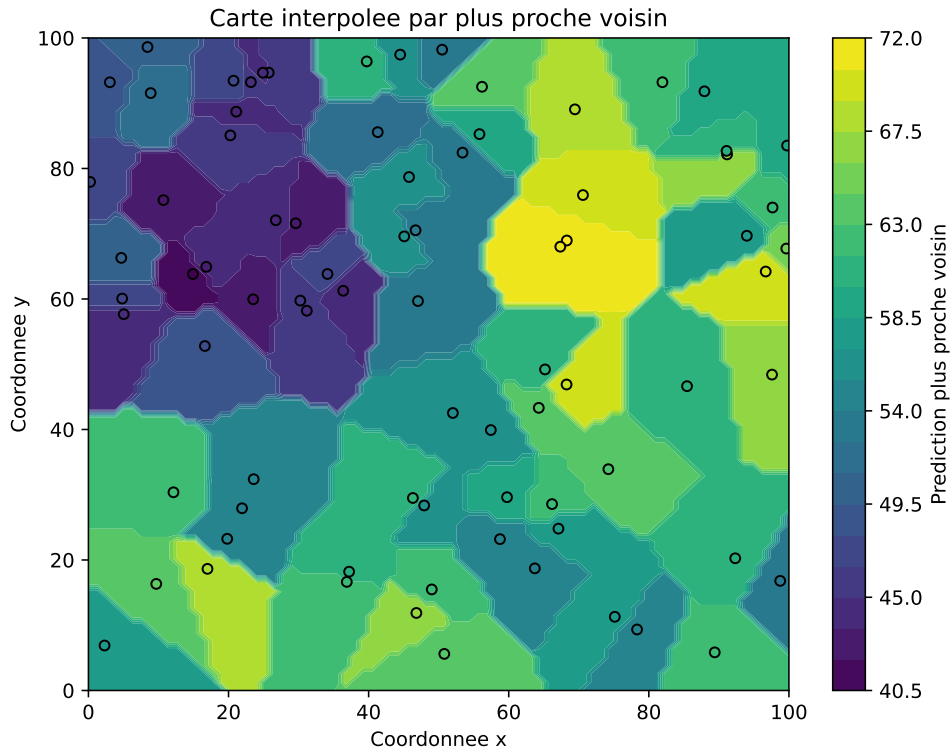


FIGURE 5.7 – Carte interpolée par la méthode du plus proche voisin.

La carte présente des ruptures nettes. Elle est constante dans les zones d'influence associées aux stations. Cette méthode est très simple, mais elle est peu adaptée lorsque le phénomène étudié est supposé varier continûment dans l'espace.

### Méthode IDW

La méthode IDW prédit la valeur en  $x_0$  par

$$\widehat{Z}_{IDW}(x_0) = \sum_{i=1}^n w_i(x_0)Z(x_i),$$

où

$$w_i(x_0) = \frac{\|x_0 - x_i\|^{-p}}{\sum_{j=1}^n \|x_0 - x_j\|^{-p}}.$$

Dans une première comparaison, on prend  $p = 2$ , choix couramment utilisé dans les applications.

La carte obtenue est représentée dans la figure 5.8.

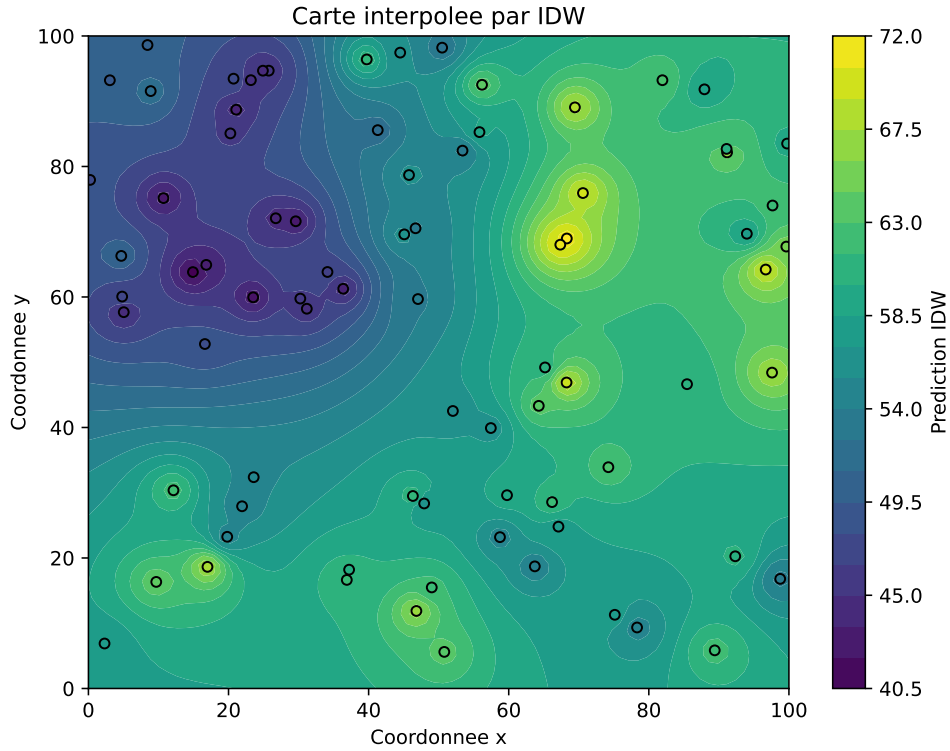


FIGURE 5.8 – Carte interpolée par pondération inverse de la distance.

La carte IDW est plus régulière que celle du plus proche voisin. Elle reste cependant entièrement déterminée par la distance et par le paramètre  $p$ . Elle ne fournit pas de variance de prédiction et ne modélise pas la dépendance entre observations.

### Sensibilité de l’IDW au paramètre de puissance

Le choix du paramètre  $p$  influence fortement la méthode IDW. Lorsque  $p$  est faible, plusieurs observations contribuent à la prédiction. Lorsque  $p$  est élevé, les observations les plus proches dominant fortement. Il est donc utile d’étudier la sensibilité des résultats à ce paramètre.

Le tableau 5.3 donne les erreurs de validation pour plusieurs valeurs de  $p$ .

TABLE 5.3 – Sensibilité de la méthode IDW au paramètre de puissance  $p$ .

$p$	ME	MAE	RMSE
0.5000	0.2216	5.8143	6.9982
1.0000	0.3855	4.6510	5.6391
1.5000	0.4308	3.6759	4.5160
2.0000	0.4013	3.2184	4.0082
2.5000	0.3398	3.1832	3.8764
3.0000	0.2742	3.2276	3.8832
4.0000	0.1841	3.3177	3.9887

La figure 5.9 représente l'évolution de la MAE et de la RMSE en fonction de  $p$ .

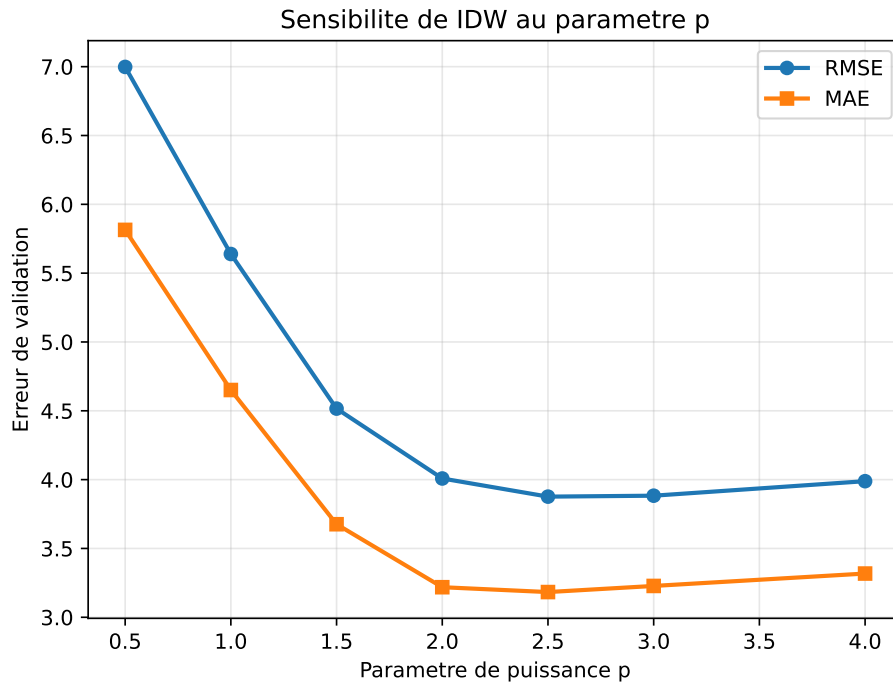


FIGURE 5.9 – Sensibilité de la méthode IDW au choix du paramètre  $p$ .

Dans cette simulation, la valeur  $p = 2.5$  donne la meilleure RMSE parmi les valeurs testées. Le choix initial  $p = 2$  reste toutefois raisonnable. Cette analyse montre surtout que l'IDW dépend d'un paramètre qui doit être choisi avec soin. Une comparaison équitable entre IDW et krigeage devrait donc préciser si  $p$  est fixé a priori ou sélectionné par validation croisée.

### Comparaison visuelle des cartes

La figure 5.10 compare les cartes obtenues par les trois méthodes.

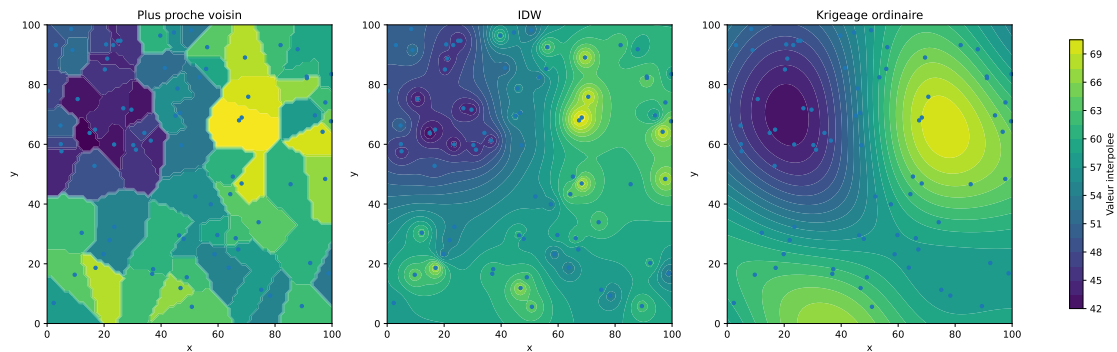


FIGURE 5.10 – Comparaison graphique des cartes obtenues par plus proche voisin, IDW et krigeage ordinaire.

La comparaison visuelle confirme les différences théoriques. Le plus proche voisin produit une carte discontinue. L'IDW produit une carte plus régulière, mais sa régularité dépend directement du choix de  $p$ . Le krigeage produit une carte construite à partir du modèle variographique et s'accompagne d'une carte de variance.

## 5.8 Validation croisée

### Principe de la validation leave-one-out

Pour comparer les méthodes, on utilise une validation croisée de type leave-one-out. Pour chaque observation  $z_i$ , on retire temporairement le point  $x_i$  de l'échantillon. On prédit ensuite la valeur en  $x_i$  à partir des  $n - 1$  observations restantes. L'erreur de validation est

$$e_i = z_i - \hat{z}_{-i}(x_i).$$

Cette procédure est répétée pour tous les points. Elle permet de comparer les méthodes sur les mêmes observations et dans des conditions identiques.

Les critères utilisés sont :

$$ME = \frac{1}{n} \sum_{i=1}^n e_i,$$

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

$$RMSE = \left( \frac{1}{n} \sum_{i=1}^n e_i^2 \right)^{1/2}.$$

Le critère  $ME$  mesure le biais moyen. Le critère  $MAE$  mesure l'erreur absolue moyenne. Le critère  $RMSE$  pénalise davantage les grandes erreurs.

### Résultats de validation

Le tableau 5.4 présente les résultats obtenus.

TABLE 5.4 – Comparaison des méthodes par validation croisée leave-one-out.

Méthode	ME	MAE	RMSE
Plus proche voisin	0.1307	3.8417	4.5319
IDW	0.4013	3.2184	4.0082
Krigeage ordinaire	-0.0483	3.2252	4.1206

La méthode du plus proche voisin est la moins performante selon la MAE et la

RMSE. Cela est cohérent avec la discontinuité de la carte obtenue. L'IDW avec  $p = 2$  donne la plus faible RMSE parmi les trois méthodes du tableau. Le krigeage ordinaire donne une RMSE légèrement plus élevée, mais son biais moyen est le plus faible :

$$ME = -0.0483.$$

Cette valeur est très proche de zéro.

La figure 5.11 montre la distribution des erreurs de validation.

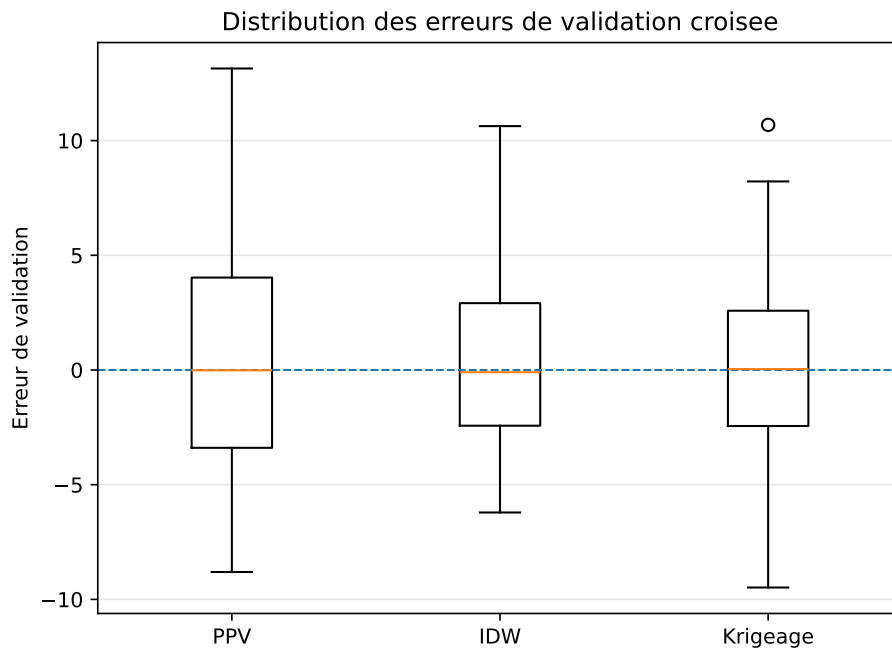


FIGURE 5.11 – Distribution des erreurs de validation croisée pour les différentes méthodes.

La lecture du tableau doit être complétée par l'examen graphique. Deux méthodes peuvent avoir des RMSE proches mais des distributions d'erreurs différentes. Les boîtes à moustaches permettent de visualiser la dispersion, la médiane et les erreurs extrêmes.

### Valeurs observées et valeurs prédites

La figure 5.12 compare, pour chaque méthode, les valeurs observées et les valeurs prédites en validation croisée.

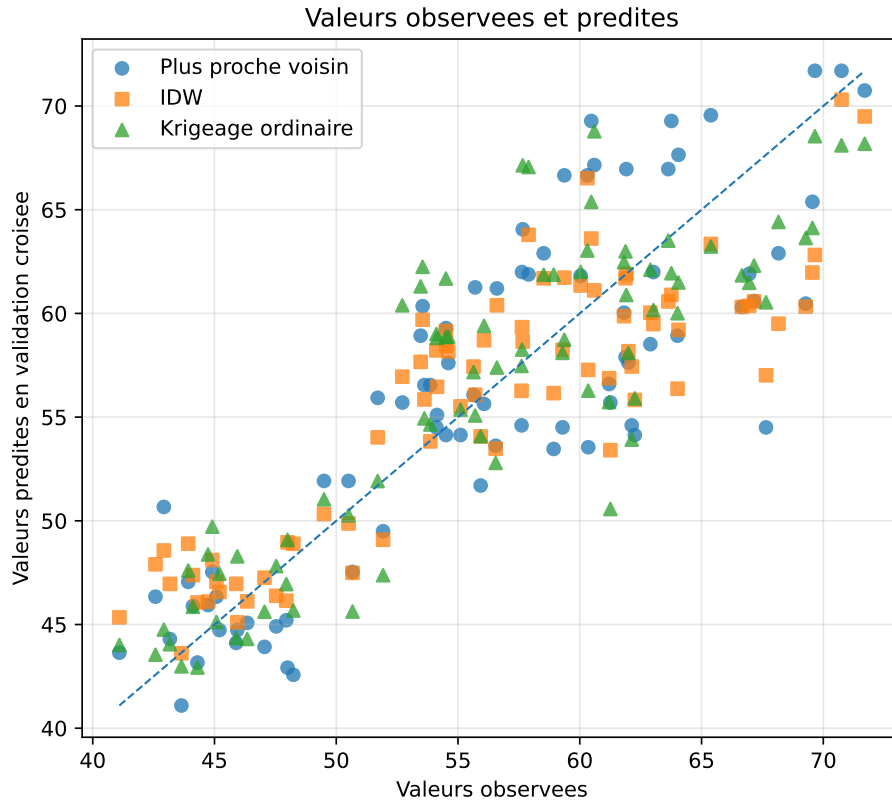


FIGURE 5.12 – Comparaison des valeurs observées et des valeurs prédites en validation croisée.

La droite diagonale correspond à une prédiction parfaite. Plus les points sont proches de cette droite, plus la méthode reproduit correctement les observations retirées. Les écarts à la diagonale montrent les erreurs de prédiction. Ce graphique complète les indicateurs numériques en donnant une vision directe de la qualité des prédictions.

### Classement selon les critères

Le tableau 5.5 résume le classement des méthodes selon trois critères : la valeur absolue de  $ME$ , la MAE et la RMSE.

TABLE 5.5 – Classement des méthodes selon les critères de validation croisée.

Méthode	ME	MAE	RMSE	Rang $ ME $	Rang MAE	Rang RMSE
Plus proche voisin	0.1307	3.8417	4.5319	2	3	3
IDW	0.4013	3.2184	4.0082	3	1	1
Krigeage ordinaire	-0.0483	3.2252	4.1206	1	2	2

Ce tableau évite une conclusion trop rapide. Si l'on regarde uniquement la RMSE, l'IDW est le meilleur dans cette expérience. Si l'on regarde le biais moyen, le krigeage ordinaire est meilleur. Le choix de la méthode dépend donc du critère privilégié. Dans

une application où l'incertitude doit être cartographiée, le krigeage conserve un avantage majeur, même lorsque la RMSE n'est pas strictement minimale.

## 5.9 Analyse des erreurs du krigeage

### Carte des erreurs de validation

La figure 5.13 présente les erreurs de validation du krigeage ordinaire aux sites d'observation.

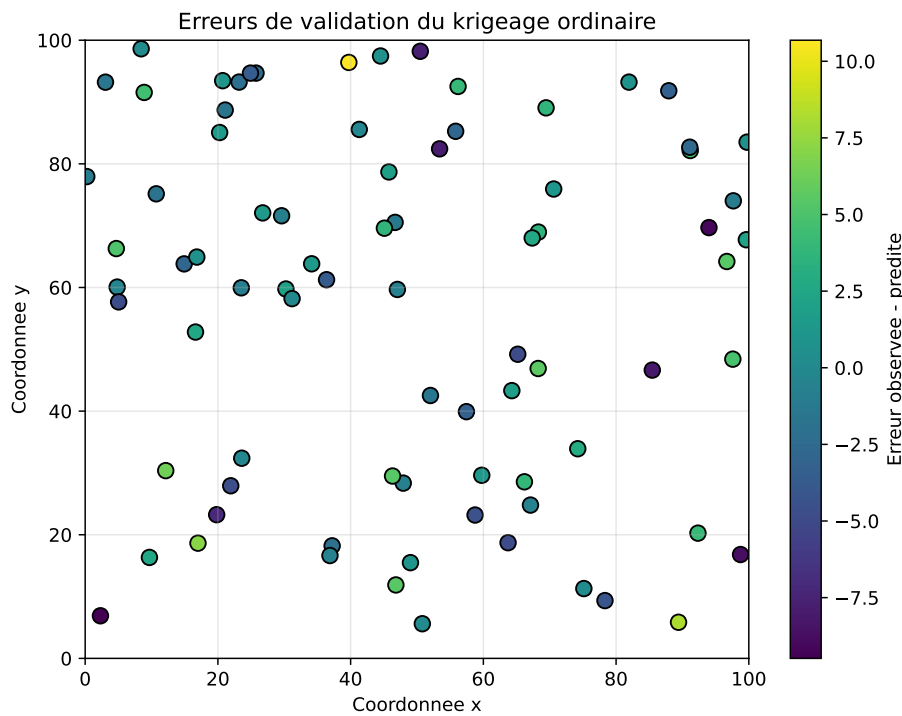


FIGURE 5.13 – Carte des erreurs de validation du krigeage ordinaire.

Cette représentation permet de vérifier si les erreurs sont spatialement organisées. Si de grandes erreurs positives ou négatives étaient concentrées dans une même région, cela pourrait indiquer une tendance mal modélisée, une anisotropie non prise en compte ou un variogramme insuffisant. Dans cette simulation, les erreurs ne montrent pas une structure spatiale dominante évidente, mais certaines zones présentent des écarts plus importants.

### Erreurs standardisées

L'avantage du krigeage est de fournir une variance de prédiction. On peut donc

définir les erreurs standardisées :

$$e_i^* = \frac{z_i - \hat{z}_{-i}(x_i)}{\sigma_{K,-i}(x_i)}.$$

Si le modèle variographique est bien calibré, ces erreurs doivent être centrées autour de zéro et avoir une dispersion proche de un.

Le tableau 5.6 donne un résumé des erreurs standardisées.

TABLE 5.6 – Résumé des erreurs standardisées du krigeage ordinaire.

Indicateur	Valeur
Moyenne	-0.0054
Écart-type	1.4564
Minimum	-3.4339
Maximum	3.9594

La moyenne est très proche de zéro, ce qui confirme l'absence de biais moyen important. En revanche, l'écart-type est supérieur à un. Cela suggère que les erreurs réelles sont plus dispersées que ce que la variance de krigeage prévoit. Autrement dit, l'incertitude fournie par le modèle est ici légèrement sous-estimée.

La figure 5.14 présente l'histogramme des erreurs standardisées.

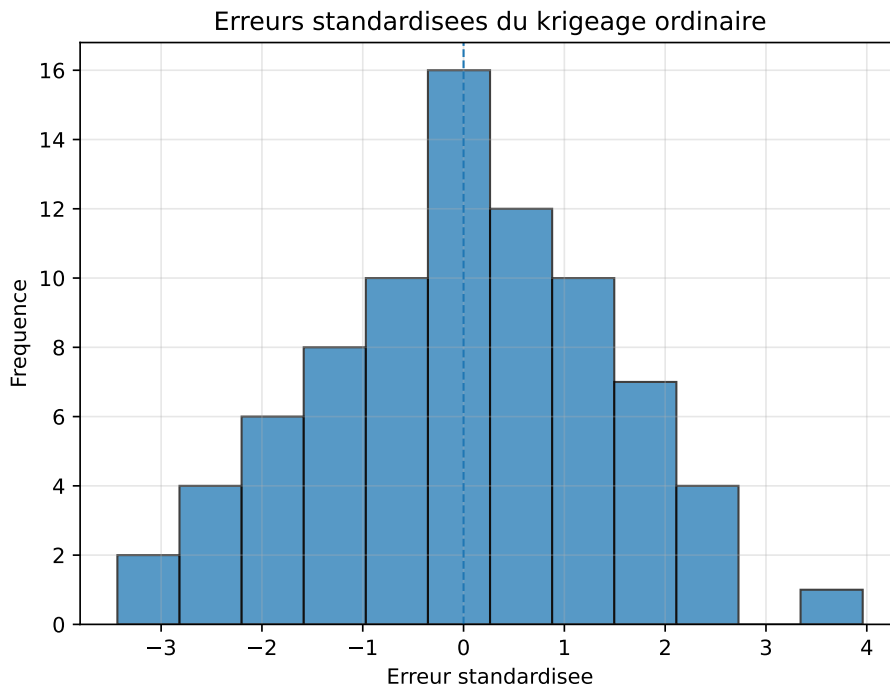


FIGURE 5.14 – Erreurs standardisées obtenues par validation croisée du krigeage ordinaire.

Cette analyse est essentielle. Un krigeage peut produire une carte de prédiction satisfaisante tout en donnant une variance mal calibrée. La validation des erreurs standardisées permet donc d'évaluer non seulement la précision des prédictions, mais aussi la qualité de l'incertitude annoncée.

## 5.10 Discussion critique des résultats

Les résultats obtenus appellent plusieurs commentaires.

Premièrement, le plus proche voisin donne les résultats les plus faibles. Cela est attendu pour un phénomène simulé avec une certaine continuité spatiale. La méthode ignore toutes les observations sauf la plus proche et produit une surface discontinue. Elle peut être utile pour des variables qualitatives, mais elle est généralement limitée pour des variables environnementales continues.

Deuxièmement, l'IDW donne de très bonnes performances numériques dans cette simulation. Avec  $p = 2$ , il obtient la meilleure RMSE dans la comparaison principale. L'analyse de sensibilité montre même que  $p = 2.5$  améliore encore légèrement la RMSE. Ce résultat rappelle qu'une méthode déterministe simple peut être compétitive lorsque la structure spatiale est régulière et lorsque ses paramètres sont bien choisis.

Troisièmement, le krigeage ordinaire ne donne pas ici la plus petite RMSE, mais il donne le biais moyen le plus faible et fournit une variance de prédiction. Il est donc plus riche du point de vue statistique. Dans une application où l'on doit produire une carte d'incertitude ou évaluer le risque associé à une prédiction, cet avantage est déterminant.

Quatrièmement, l'écart-type des erreurs standardisées du krigeage est supérieur à un. Cela montre que le modèle variographique retenu, bien que raisonnable, n'est pas parfaitement calibré. Plusieurs explications sont possibles : présence d'une tendance dans les données simulées, choix d'un modèle isotrope, nombre limité de stations, ou estimation imparfaite des paramètres du variogramme.

Cinquièmement, cette simulation montre que la comparaison des méthodes ne doit pas être réduite à un seul indicateur. Une méthode peut être meilleure en RMSE, une autre en biais, et une autre encore en interprétabilité ou en quantification de l'incertitude. Dans une étude appliquée, il faut donc définir clairement le critère de qualité avant de choisir une méthode.

### Limites de l'application numérique

Cette application présente plusieurs limites qu'il convient de préciser.

La première limite est l'utilisation de données simulées. La simulation permet de

contrôler le protocole et de reproduire les résultats, mais elle ne représente pas toute la complexité des données réelles. Des données de précipitations réelles peuvent dépendre du relief, du vent, de la saison, de l'altitude ou de la distance à la mer. Ces facteurs ne sont pas intégrés dans le modèle simple utilisé ici.

La deuxième limite est l'hypothèse d'isotropie. Le variogramme ajusté dépend seulement de la distance. Or, dans de nombreux phénomènes environnementaux, la continuité varie selon la direction. Une analyse plus complète devrait construire des variogrammes directionnels et tester la présence d'anisotropie.

La troisième limite concerne la tendance spatiale. Les données simulées contiennent une tendance  $m(x, y)$ , mais le krigeage appliqué est un krigeage ordinaire, qui suppose une moyenne localement constante. Une extension naturelle serait donc d'appliquer un krigeage universel ou un krigeage des résidus après estimation de la tendance.

La quatrième limite concerne le choix du modèle de variogramme. Seuls trois modèles classiques ont été comparés. Le modèle de Matérn, par exemple, pourrait être utilisé pour contrôler plus finement la régularité du champ. On pourrait aussi comparer les modèles non seulement par ajustement au variogramme expérimental, mais directement par validation croisée.

La cinquième limite concerne l>IDW. Nous avons montré que ses performances dépendent du paramètre  $p$ . Dans une comparaison plus complète, il faudrait sélectionner  $p$  par validation croisée et comparer l>IDW optimisé au krigeage dont le variogramme est également choisi par validation.

## **Reproductibilité des calculs**

Le code Python fourni en annexe permet de reproduire l'ensemble des figures et tableaux du chapitre. Il fixe une graine aléatoire afin que les mêmes stations, les mêmes observations et les mêmes résultats puissent être obtenus à chaque exécution. Le code exporte également les données simulées, les tableaux au format CSV et les figures au format PDF et PNG.

Cette reproductibilité est importante pour trois raisons. Premièrement, elle permet de vérifier les résultats numériques. Deuxièmement, elle permet de modifier certains paramètres et d'observer leur effet sur les conclusions. Troisièmement, elle facilite l'intégration des figures et des tableaux dans le mémoire.

Dans un travail ultérieur, il serait possible d'ajouter au code :

- une sélection automatique du modèle de variogramme par validation croisée ;
- une comparaison avec le krigeage universel ;
- une étude de l'anisotropie ;

- une application à des données réelles ;
- des intervalles de prédiction fondés sur la variance de krigeage.

## 5.11 Conclusion du chapitre

Ce chapitre a présenté une application numérique complète consacrée à l'interpolation spatiale. À partir de données simulées, nous avons réalisé une analyse exploratoire, construit un variogramme expérimental, ajusté plusieurs modèles théoriques, appliqué le krigeage ordinaire et comparé les résultats à ceux de deux méthodes déterministes.

L'analyse montre que le plus proche voisin est simple mais peu adapté à un phénomène continu. L'IDW est performant dans cette simulation, notamment lorsque le paramètre de puissance est bien choisi. Le krigeage ordinaire, quant à lui, présente un biais moyen très faible et fournit une variance de prédiction, ce qui lui donne une interprétation statistique plus complète.

Le résultat principal n'est donc pas que le krigeage domine toujours les méthodes déterministes. Le résultat important est plus nuancé : le krigeage apporte une structure probabiliste, une prise en compte de la dépendance spatiale et une mesure de l'incertitude. Sa qualité dépend cependant du modèle variographique, de la validation et des hypothèses adoptées.

# Conclusion générale

Ce mémoire a été consacré à l'étude de l'interpolation spatiale par krigeage. L'objectif principal était de présenter une démarche progressive permettant de passer des méthodes déterministes d'interpolation à une approche géostatistique fondée sur une modélisation probabiliste du phénomène spatial.

Nous avons d'abord rappelé que les données spatiales apparaissent dans de nombreux domaines appliqués : hydrologie, climatologie, géologie, environnement, agronomie et cartographie. Dans ces contextes, les observations sont disponibles en un nombre fini de sites, alors que l'on souhaite souvent prédire la valeur d'une variable en des points non observés ou produire une carte continue sur un domaine. Les méthodes déterministes, telles que le plus proche voisin, l>IDW, les splines ou la triangulation, constituent des outils simples et utiles. Toutefois, elles restent limitées lorsqu'il s'agit de modéliser la dépendance spatiale et d'évaluer l'incertitude associée aux prédictions.

La géostatistique apporte une réponse plus complète à ces limites. En interprétant la variable spatiale comme une variable régionalisée ou comme la réalisation d'un champ aléatoire, elle permet de définir des outils mathématiques adaptés : moyenne, covariance, stationnarité, variogramme et hypothèse intrinsèque. Le variogramme joue un rôle central, car il décrit la manière dont la dissimilarité entre deux observations évolue avec leur séparation spatiale. Il constitue le lien essentiel entre les données observées et le système de krigeage.

La théorie du krigeage a ensuite été présentée à travers ses trois formes principales. Le krigeage simple suppose une moyenne connue. Le krigeage ordinaire, plus utilisé en pratique, suppose une moyenne constante mais inconnue et impose une contrainte de somme des poids égale à un. Le krigeage universel permet de prendre en compte une tendance spatiale représentée par une combinaison de fonctions connues. Dans tous les cas, le principe général reste le même : construire un prédicteur linéaire sans biais dont la variance de l'erreur de prédiction est minimale.

L'application numérique a permis d'illustrer concrètement cette démarche. À partir de données simulées, nous avons construit un variogramme expérimental, ajusté un

modèle théorique, puis appliqué le krigeage ordinaire. Les résultats ont été comparés à ceux obtenus par la méthode du plus proche voisin et par l'IDW. Cette comparaison a montré que les méthodes déterministes peuvent parfois donner de bonnes performances numériques. Cependant, le krigeage présente un avantage fondamental : il fournit non seulement une prédiction, mais aussi une variance de prédiction. Cette variance permet de produire une carte d'incertitude, information essentielle dans les applications où les décisions doivent tenir compte du risque d'erreur.

Les résultats obtenus confirment donc que le krigeage ne doit pas être considéré comme une simple technique de lissage ou de cartographie. Il s'agit d'une méthode statistique complète, dont la qualité dépend de la modélisation variographique, des hypothèses de stationnarité ou de validité intrinsèque, de la distribution spatiale des observations et du choix du voisinage utilisé pour la prédiction.

Toutefois, ce travail présente plusieurs limites. La première est liée au fait que l'application numérique repose sur des données simulées. Une étude plus complète devrait inclure des données réelles, par exemple des données de précipitations, de pollution, de température ou de propriétés du sol. Les données réelles introduisent des difficultés supplémentaires : valeurs manquantes, erreurs de mesure, hétérogénéité du domaine, effet du relief, unités spatiales différentes et présence possible de tendances complexes.

La deuxième limite concerne l'hypothèse d'isotropie. Dans ce mémoire, l'analyse est essentiellement menée dans un cadre isotrope. Or, dans de nombreuses situations réelles, la continuité spatiale dépend aussi de la direction. L'étude de l'anisotropie constitue donc une perspective naturelle. Elle nécessite la construction de variogrammes directionnels et l'ajustement de modèles capables de représenter des portées différentes selon les directions.

Plusieurs perspectives et problèmes ouverts peuvent être dégagés à partir de ce travail. Une première perspective concerne le krigeage universel. Lorsque la moyenne du phénomène varie dans l'espace, le krigeage ordinaire peut être insuffisant. Il serait intéressant d'étudier plus en détail les méthodes combinant régression spatiale et krigeage des résidus, en intégrant des variables explicatives comme l'altitude, la distance à une source, la nature du sol ou des variables climatiques.

Une deuxième perspective concerne le cokrigeage. Dans de nombreuses applications, plusieurs variables sont observées simultanément ou partiellement. Le cokrigeage permet d'exploiter les dépendances croisées entre plusieurs variables régionalisées. Son étude soulève des difficultés supplémentaires, notamment la modélisation des variogrammes croisés et la garantie de la validité du modèle multivarié.

Une troisième perspective concerne les modèles spatio-temporels. De nombreux phénomènes ne varient pas seulement dans l'espace, mais aussi dans le temps : précipitations, températures, pollution atmosphérique, humidité du sol ou propagation d'un phénomène environnemental. Le krigeage spatio-temporel cherche à prédire une variable en tenant compte simultanément de la proximité spatiale et de la proximité temporelle. La construction de covariance ou de variogrammes spatio-temporels admissibles reste un sujet riche et complexe.

Une quatrième perspective concerne la quantification de l'incertitude. La variance de krigeage est un outil important, mais elle dépend fortement du modèle choisi. Lorsque les paramètres du variogramme sont estimés à partir des données, une incertitude supplémentaire apparaît. Une question importante est donc de mieux prendre en compte l'incertitude liée à l'estimation du variogramme lui-même. Les approches bayésiennes ou les méthodes de bootstrap spatial peuvent être envisagées dans ce cadre.

Une cinquième perspective concerne les données massives. Avec les capteurs modernes, les satellites et les bases géographiques, les jeux de données spatiales peuvent contenir un très grand nombre d'observations. Le krigeage classique nécessite la résolution de systèmes linéaires dont la taille augmente avec le nombre de points, ce qui peut devenir coûteux. Le développement de méthodes numériques rapides, de krigeage local, d'approximations par voisinage ou de méthodes adaptées aux grandes matrices de covariance constitue un problème important.

Enfin, une dernière perspective concerne la comparaison systématique entre méthodes géostatistiques et méthodes modernes d'apprentissage statistique. Les méthodes comme les forêts aléatoires, les réseaux de neurones ou les modèles hybrides peuvent produire de bonnes prédictions, mais elles n'offrent pas toujours la même interprétabilité spatiale que le krigeage. Un problème actuel consiste donc à combiner la force prédictive de ces méthodes avec la rigueur géostatistique du variogramme et de la variance de prédiction.

En conclusion, le krigeage demeure une méthode fondamentale de l'interpolation spatiale. Sa force réside dans sa capacité à relier la prédiction à un modèle explicite de dépendance spatiale et à fournir une mesure d'incertitude. Sa mise en œuvre exige cependant une attention particulière à la qualité du variogramme, aux hypothèses du modèle et à la validation des résultats.

# Annexe A

## Rappels statistiques utiles à la géostatistique

Cette annexe rassemble les rappels statistiques nécessaires à la lecture du présent mémoire. Elle a pour but de préciser les notions qui interviennent directement dans la géostatistique : variable aléatoire, loi de probabilité, espérance, variance, covariance, corrélation, échantillon, estimation, erreur quadratique moyenne et modèle gaussien. Ces rappels sont utiles car le krigeage repose sur une idée simple mais exigeante : prédire une valeur inconnue en utilisant la structure probabiliste du phénomène spatial.

Dans les chapitres précédents, une variable régionalisée a été modélisée par un champ aléatoire

$$Z = \{Z(x) : x \in D\}.$$

Pour chaque position spatiale  $x$ ,  $Z(x)$  est une variable aléatoire. La covariance entre  $Z(x)$  et  $Z(y)$ , le variogramme de  $Z(x) - Z(y)$  et la variance de l'erreur de krigeage sont donc des notions statistiques. Il est alors important de rappeler leur signification avant d'interpréter les systèmes de krigeage. Cette annexe s'inspire de l'esprit des rappels statistiques proposés dans des supports d'introduction à la géostatistique, notamment le cours de [12], mais elle est rédigée ici de manière autonome et adaptée au contenu du présent mémoire.

### A.1 Variables aléatoires et lois de probabilité

Une variable aléatoire réelle est une grandeur numérique dont la valeur dépend du hasard. On la note généralement  $X$ . Dans un contexte spatial, la valeur observée en un site  $x$  peut être considérée comme une réalisation d'une variable aléatoire  $Z(x)$ .

La notation majuscule  $X$  ou  $Z(x)$  désigne la variable aléatoire, tandis que la notation minuscule  $x$  ou  $z(x)$  désigne une valeur observée.

**Définition A.1.** *Une variable aléatoire réelle  $X$  est une application mesurable définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  et à valeurs dans  $\mathbb{R}$ .*

La loi de probabilité de  $X$  décrit la distribution possible de ses valeurs. Lorsque  $X$  est discrète, elle prend un nombre fini ou dénombrable de valeurs  $x_k$  avec des probabilités  $p_k$ . On a alors

$$\mathbb{P}(X = x_k) = p_k, \quad \sum_k p_k = 1.$$

Lorsque  $X$  est continue, sa loi peut être décrite par une densité  $f$  telle que

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(t) dt.$$

La fonction de répartition de  $X$  est définie par

$$F_X(t) = \mathbb{P}(X \leq t).$$

Elle permet de décrire la loi de  $X$  sans distinguer explicitement le cas discret et le cas continu.

**Exemple A.1.** *Si  $X$  représente une mesure de précipitation journalière en une station donnée, alors  $X$  est une variable aléatoire positive. Si  $Z(x)$  représente la précipitation au point spatial  $x$ , alors la famille  $\{Z(x) : x \in D\}$  forme un champ aléatoire.*

En géostatistique, on ne s'intéresse pas seulement à la loi d'une variable en un point. On s'intéresse aussi aux relations entre les variables associées à plusieurs points. Pour deux sites  $x$  et  $y$ , il faut comprendre la loi conjointe de  $Z(x)$  et  $Z(y)$ , car c'est cette relation qui permet de mesurer la dépendance spatiale.

## A.2 Espérance mathématique

L'espérance d'une variable aléatoire est une mesure de tendance centrale. Elle représente la moyenne théorique de la variable lorsque l'expérience aléatoire est répétée un grand nombre de fois.

Dans le cas discret, si  $X$  prend les valeurs  $x_k$  avec les probabilités  $p_k$ , on définit

$$\mathbb{E}(X) = \sum_k x_k p_k,$$

à condition que cette somme soit bien définie. Dans le cas continu de densité  $f$ , on écrit

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} tf(t) dt,$$

lorsque cette intégrale existe.

**Définition A.2.** *L'espérance de  $X$ , notée  $\mathbb{E}(X)$ , est la moyenne théorique de la variable aléatoire  $X$  lorsqu'elle existe.*

L'espérance possède des propriétés de linéarité fondamentales. Pour deux variables aléatoires intégrables  $X$  et  $Y$ , et pour deux constantes réelles  $a$  et  $b$ , on a

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Cette propriété ne nécessite pas l'indépendance de  $X$  et  $Y$ .

Dans un champ aléatoire  $Z = \{Z(x) : x \in D\}$ , l'espérance dépend éventuellement du point  $x$  :

$$m(x) = \mathbb{E}[Z(x)].$$

La fonction  $m(x)$  est appelée fonction moyenne du champ. Lorsque  $m(x) = m$  est constante, on dit que le champ possède une moyenne constante. Cette hypothèse est importante dans le krigeage simple et le krigeage ordinaire.

**Remarque A.1.** *Dans une application réelle, l'espérance théorique n'est pas directement observable. On observe une seule valeur  $z(x)$  en chaque site. La moyenne théorique est donc une notion de modèle. Elle sert à définir le biais d'un estimateur et à formuler les contraintes du krigeage.*

## A.3 Variance et écart-type

L'espérance décrit la tendance centrale d'une variable, mais elle ne mesure pas sa dispersion. Deux variables peuvent avoir la même moyenne et des variabilités très différentes. La variance mesure la dispersion autour de l'espérance.

**Définition A.3.** *Soit  $X$  une variable aléatoire de carré intégrable. Sa variance est définie par*

$$\text{Var}(X) = \mathbb{E} [(X - \mathbb{E}(X))^2].$$

L'écart-type est

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

On utilise souvent la formule équivalente

$$\text{Var}(X) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2.$$

Cette identité est obtenue en développant le carré

$$(X - \mathbb{E}(X))^2 = X^2 - 2X\mathbb{E}(X) + \{\mathbb{E}(X)\}^2.$$

La variance est toujours positive ou nulle. Elle est nulle si et seulement si la variable est constante presque sûrement.

Si  $a$  et  $b$  sont deux constantes réelles, alors

$$\text{Var}(aX + b) = a^2\text{Var}(X).$$

L'ajout d'une constante ne change pas la dispersion, tandis que la multiplication par  $a$  multiplie la variance par  $a^2$ .

Dans un champ aléatoire, la variance au point  $x$  est

$$\text{Var}(Z(x)) = \mathbb{E} [(Z(x) - m(x))^2].$$

Sous stationnarité d'ordre deux, cette variance est constante :

$$\text{Var}(Z(x)) = C(0) = \sigma^2.$$

## A.4 Covariance et corrélation

La géostatistique repose sur la dépendance entre des valeurs observées en différents points de l'espace. La covariance est l'outil statistique de base permettant de mesurer la dépendance linéaire entre deux variables aléatoires.

**Définition A.4.** Soient  $X$  et  $Y$  deux variables aléatoires de carré intégrable. Leur covariance est définie par

$$\text{Cov}(X, Y) = \mathbb{E} [(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))].$$

Si la covariance est positive, les deux variables ont tendance à varier dans le même sens. Si elle est négative, elles ont tendance à varier en sens opposé. Si elle est nulle, cela signifie seulement l'absence de dépendance linéaire ; cela n'implique pas nécessairement l'indépendance, sauf dans certains cas particuliers comme le cas gaussien.

La covariance vérifie

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Elle est bilinéaire :

$$\text{Cov}(aX + bY, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z).$$

Elle est symétrique :

$$\text{Cov}(X, Y) = \text{Cov}(Y, X).$$

La corrélation normalise la covariance afin d'obtenir une quantité sans unité :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Lorsque les variances sont non nulles, on a

$$-1 \leq \rho(X, Y) \leq 1.$$

Dans un champ aléatoire stationnaire d'ordre deux, la covariance entre  $Z(x)$  et  $Z(y)$  dépend seulement du vecteur  $h = y - x$  :

$$C(h) = \text{Cov}(Z(x), Z(x+h)).$$

Cette fonction de covariance est l'un des objets fondamentaux de la géostatistique. Elle exprime la manière dont la dépendance statistique décroît lorsque la séparation spatiale augmente.

## A.5 Indépendance et absence de corrélation

Deux variables aléatoires  $X$  et  $Y$  sont indépendantes si la connaissance de l'une ne modifie pas la loi de l'autre. Formellement, pour tous ensembles mesurables  $A$  et  $B$ , on a

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

L'indépendance est une propriété forte. Elle implique que

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y),$$

lorsque les espérances existent. Par conséquent,

$$\text{Cov}(X, Y) = 0.$$

Ainsi, l'indépendance implique l'absence de corrélation.

La réciproque est fautive en général. Deux variables peuvent avoir une covariance nulle sans être indépendantes. En géostatistique, cette distinction est importante : une covariance faible ou nulle à grande distance ne signifie pas toujours que les variables sont totalement indépendantes. Elle signifie que la dépendance linéaire mesurée par le modèle est faible.

Dans le cas particulier d'un vecteur gaussien, l'absence de covariance implique l'indépendance. C'est l'une des raisons pour lesquelles les modèles gaussiens sont très utilisés en statistique spatiale. Ils permettent de décrire une structure de dépendance complète à partir de la moyenne et de la covariance.

## A.6 Vecteurs aléatoires et matrices de covariance

En géostatistique, on observe simultanément plusieurs variables aléatoires :

$$Z(x_1), \dots, Z(x_n).$$

Il est donc naturel de les regrouper dans un vecteur aléatoire

$$\mathbf{Z} = (Z(x_1), \dots, Z(x_n))^T.$$

La moyenne de ce vecteur est

$$\mathbb{E}(\mathbf{Z}) = (\mathbb{E}[Z(x_1)], \dots, \mathbb{E}[Z(x_n)])^T.$$

Sa matrice de covariance est

$$\Sigma = (\text{Cov}(Z(x_i), Z(x_j)))_{1 \leq i, j \leq n}.$$

Cette matrice est symétrique et semi-définie positive. En effet, pour tout vecteur  $\mathbf{a} = (a_1, \dots, a_n)^T$ , on a

$$\mathbf{a}^T \Sigma \mathbf{a} = \text{Var} \left( \sum_{i=1}^n a_i Z(x_i) \right) \geq 0.$$

Cette propriété explique pourquoi toutes les fonctions ne peuvent pas être utilisées comme covariance spatiale. Une fonction de covariance doit produire une matrice semi-définie positive pour tout ensemble fini de sites.

**Remarque A.2.** *Les systèmes de krigage reposent directement sur des matrices de covariance ou de variogramme. Si la matrice est mal conditionnée ou non admissible, les poids obtenus peuvent être instables. Les conditions de positivité et d'admissibilité ne sont donc pas seulement théoriques ; elles sont nécessaires au calcul numérique.*

## A.7 Lois usuelles en statistique

Plusieurs lois de probabilité apparaissent fréquemment dans les rappels statistiques et dans les modèles de simulation. Les plus utiles pour ce mémoire sont la loi uniforme, la loi normale et, de manière indirecte, les lois liées aux erreurs quadratiques.

### Loi uniforme

Une variable aléatoire  $X$  suit la loi uniforme sur l'intervalle  $[a, b]$ , notée

$$X \sim \mathcal{U}([a, b]),$$

si sa densité est constante sur cet intervalle :

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

Son espérance et sa variance sont

$$\mathbb{E}(X) = \frac{a+b}{2}, \quad \text{Var}(X) = \frac{(b-a)^2}{12}.$$

Dans le chapitre numérique, les positions des stations sont générées de manière aléatoire dans un domaine carré. Une manière simple de le faire consiste à tirer indépendamment les coordonnées selon une loi uniforme.

### Loi normale

La loi normale, ou loi gaussienne, est centrale en statistique. Une variable  $X$  suit la loi normale de moyenne  $m$  et de variance  $\sigma^2$ , notée

$$X \sim \mathcal{N}(m, \sigma^2),$$

si sa densité est

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

La loi normale est symétrique autour de sa moyenne. Environ 95% de sa masse se trouve dans l'intervalle

$$[m - 1.96\sigma, m + 1.96\sigma].$$

La loi normale multivariée est particulièrement importante en géostatistique. Un vecteur aléatoire  $\mathbf{Z}$  suit une loi normale multivariée de moyenne  $\mathbf{m}$  et de covariance  $\Sigma$  si toute combinaison linéaire de ses composantes est normale. On note

$$\mathbf{Z} \sim \mathcal{N}_n(\mathbf{m}, \Sigma).$$

Dans ce cadre, la moyenne et la matrice de covariance déterminent entièrement la loi.

**Remarque A.3.** *Le krigeage linéaire ne nécessite pas toujours une hypothèse de normalité. En revanche, lorsque le champ est gaussien, le krigeage possède une interprétation probabiliste plus forte : la prédiction linéaire coïncide avec l'espérance conditionnelle et la variance de krigeage avec la variance conditionnelle.*

## A.8 Échantillon, moyenne empirique et variance empirique

Dans une étude statistique, on dispose d'un échantillon

$$X_1, \dots, X_n.$$

Lorsque les variables sont indépendantes et de même loi, on parle d'échantillon aléatoire simple. En géostatistique, les observations spatiales ne sont généralement pas indépendantes, car deux sites proches peuvent être corrélés. Cependant, les notions de moyenne empirique et de variance empirique restent utiles pour l'analyse exploratoire.

La moyenne empirique est

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

La variance empirique corrigée est

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Le facteur  $n - 1$  permet d'obtenir un estimateur sans biais de la variance dans le cas indépendant et identiquement distribué.

Dans le chapitre numérique, les statistiques descriptives des observations jouent un rôle exploratoire. Elles permettent d'identifier l'ordre de grandeur de la variable, sa dispersion et la présence éventuelle de valeurs extrêmes. Elles ne remplacent pas l'analyse spatiale, mais elles en constituent une étape préalable.

## A.9 Biais et erreur quadratique moyenne

Un estimateur est une quantité calculée à partir des données afin d'approcher un paramètre inconnu. Par exemple, la moyenne empirique  $\bar{X}$  estime l'espérance  $\mathbb{E}(X)$ .

**Définition A.5.** Soit  $\theta$  un paramètre inconnu et  $\hat{\theta}$  un estimateur de  $\theta$ . Le biais de  $\hat{\theta}$  est

$$\text{Biais}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta.$$

L'estimateur est sans biais si

$$\mathbb{E}(\hat{\theta}) = \theta.$$

L'erreur quadratique moyenne est définie par

$$\text{MSE}(\hat{\theta}) = \mathbb{E} \left[ (\hat{\theta} - \theta)^2 \right].$$

Elle se décompose en

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left\{ \text{Biais}(\hat{\theta}) \right\}^2.$$

Cette décomposition montre qu'un bon estimateur doit avoir à la fois une variance faible et un biais faible.

En krigeage, le paramètre à estimer est la valeur inconnue  $Z(x_0)$ . Le prédicteur est

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i).$$

La condition de non-biais impose

$$\mathbb{E}\{\hat{Z}(x_0) - Z(x_0)\} = 0.$$

Par conséquent, minimiser l'erreur quadratique moyenne revient à minimiser la variance

de l'erreur de prédiction :

$$\text{MSE}\{\widehat{Z}(x_0)\} = \text{Var}\{\widehat{Z}(x_0) - Z(x_0)\}.$$

Cette remarque explique la définition du krigeage comme prédicteur linéaire sans biais de variance minimale.

## A.10 Notions de convergence

Les notions de convergence permettent de comprendre le comportement d'un estimateur lorsque le nombre d'observations augmente. Même si la géostatistique travaille souvent avec une seule réalisation spatiale, ces notions restent importantes pour interpréter les méthodes statistiques.

Une suite d'estimateurs  $\widehat{\theta}_n$  converge en probabilité vers  $\theta$  si, pour tout  $\varepsilon > 0$ ,

$$\mathbb{P}(|\widehat{\theta}_n - \theta| > \varepsilon) \longrightarrow 0 \quad \text{lorsque } n \rightarrow \infty.$$

On dit alors que  $\widehat{\theta}_n$  est consistant.

La loi des grands nombres affirme que, sous des hypothèses appropriées, la moyenne empirique converge vers l'espérance :

$$\overline{X}_n \longrightarrow \mathbb{E}(X).$$

Le théorème central limite affirme que, sous des hypothèses classiques,

$$\sqrt{n} \frac{\overline{X}_n - \mathbb{E}(X)}{\sigma} \Longrightarrow \mathcal{N}(0, 1).$$

En statistique spatiale, l'application directe de ces résultats est plus délicate, car les observations ne sont pas indépendantes. La dépendance spatiale modifie la variance effective des estimateurs. C'est pourquoi la modélisation de la covariance ou du variogramme est essentielle.

## A.11 Moindres carrés et ajustement de modèle

L'ajustement d'un modèle consiste à choisir des paramètres de manière à rapprocher les valeurs prédites des valeurs observées. La méthode des moindres carrés est l'une des méthodes les plus utilisées.

Supposons que l'on observe des couples  $(u_k, y_k)$  et que l'on souhaite ajuster un modèle  $g_\theta(u)$  dépendant d'un paramètre  $\theta$ . Le critère des moindres carrés ordinaires est

$$Q(\theta) = \sum_{k=1}^K \{y_k - g_\theta(u_k)\}^2.$$

On choisit  $\hat{\theta}$  qui minimise  $Q(\theta)$ .

Dans l'analyse variographique, les valeurs observées sont les valeurs du variogramme expérimental

$$\hat{\gamma}(h_k), \quad k = 1, \dots, K,$$

et le modèle ajusté est un variogramme théorique  $\gamma_\theta(h)$ . Le critère peut donc s'écrire

$$Q(\theta) = \sum_{k=1}^K \{\hat{\gamma}(h_k) - \gamma_\theta(h_k)\}^2.$$

Comme toutes les classes de distance ne contiennent pas le même nombre de paires, il est souvent préférable d'utiliser un critère pondéré :

$$Q_w(\theta) = \sum_{k=1}^K w_k \{\hat{\gamma}(h_k) - \gamma_\theta(h_k)\}^2.$$

Les poids  $w_k$  peuvent être choisis en fonction du nombre de paires dans chaque classe. Cela donne plus d'importance aux points expérimentaux estimés avec davantage d'information.

## A.12 Intervalles de confiance et incertitude

Un intervalle de confiance fournit une plage de valeurs plausibles pour un paramètre inconnu. Dans un cadre simple, si un estimateur  $\hat{\theta}$  est approximativement normal de moyenne  $\theta$  et d'écart-type  $s$ , alors un intervalle approximatif à 95% est

$$\hat{\theta} \pm 1.96s.$$

Cette formule repose sur la loi normale standard.

En krigeage, une idée analogue apparaît avec la variance de krigeage. Si l'on suppose que l'erreur de prédiction est approximativement normale, alors on peut écrire un intervalle de prédiction approximatif :

$$\hat{Z}(x_0) \pm 1.96\sigma_K(x_0),$$

où  $\sigma_K(x_0)$  est l'écart-type de krigeage. Cette interprétation est naturelle dans le cas d'un champ gaussien.

Il faut toutefois être prudent. La variance de krigeage dépend du modèle variographique choisi. Si le variogramme est mal ajusté, l'intervalle de prédiction peut être trop large ou trop étroit. De plus, l'incertitude liée à l'estimation des paramètres du variogramme n'est pas toujours prise en compte dans la variance de krigeage classique.

## A.13 Lien entre statistiques classiques et géostatistique

Les rappels précédents permettent de mieux comprendre les objets utilisés dans le mémoire. La moyenne d'un champ aléatoire est une espérance. La covariance spatiale est une covariance entre deux variables aléatoires localisées. Le variogramme est une demi-variance d'un accroissement spatial :

$$\gamma(h) = \frac{1}{2} \text{Var}\{Z(x+h) - Z(x)\}.$$

La matrice de covariance est une matrice de covariances entre observations. La variance de krigeage est la variance de l'erreur de prédiction. La condition de non-biais est une condition portant sur l'espérance de l'erreur.

Ainsi, le krigeage combine trois idées statistiques fondamentales :

1. imposer une condition de non-biais ;
2. minimiser une variance ou une erreur quadratique moyenne ;
3. utiliser la covariance ou le variogramme pour décrire la dépendance entre observations.

Par suite, le krigeage utilise ces notions pour construire une prédiction optimale dans une classe de prédicteurs linéaires. La contrainte de non-biais repose sur l'espérance, tandis que le critère d'optimalité repose sur la variance de l'erreur. La matrice du système de krigeage est construite à partir de la covariance ou du variogramme. La variance de krigeage fournit ensuite une mesure de l'incertitude de prédiction.

# Annexe B

## Code Python utilisé pour la simulation

Cette annexe contient le code Python utilisé pour générer les données simulées, les figures et les tableaux du chapitre numérique. Le code est conçu pour être exécuté dans un environnement Python standard ou dans Google Colab.

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib
4 matplotlib.use('Agg')
5 import matplotlib.pyplot as plt
6 from pathlib import Path
7 from scipy.spatial.distance import pdist, squareform, cdist
8 from scipy.spatial import cKDTree
9 from scipy.optimize import minimize
10 import shutil, warnings
11 warnings.filterwarnings('ignore')
12
13 SEED=2026
14 np.random.seed(SEED)
15 BASE=Path('kriging_outputs')
16 FIG_DIR=BASE/'figures'
17 TAB_DIR=BASE/'tables'
18 DATA_DIR=BASE/'data'
19 for d in [FIG_DIR,TAB_DIR,DATA_DIR]: d.mkdir(parents=True,
20     exist_ok=True)
21 X_MIN,X_MAX=0.,100.; Y_MIN,Y_MAX=0.,100.; N_STATIONS=80;
22     GRID_SIZE=80; IDW_POWER=2.0; N_LAGS=12; MAX_DIST_FRAC=0.60
23
24 def savefig(name):
```

```

23     plt.savefig(FIG_DIR/f'{name}.pdf', bbox_inches='tight')
24     plt.savefig(FIG_DIR/f'{name}.png', dpi=220, bbox_inches='
        tight')
25     plt.close()
26
27 def export_latex_table(df, filename, caption=None, label=None,
        index=False, float_format='% .4f'):
28     tex = df.to_latex(index=index, escape=False, float_format=
        float_format, caption=caption, label=label)
29     (TAB_DIR/filename).write_text(tex, encoding='utf-8')
30
31 def make_grid(grid_size=80):
32     gx=np.linspace(X_MIN,X_MAX,grid_size); gy=np.linspace(Y_MIN,
        Y_MAX,grid_size); XX,YY=np.meshgrid(gx,gy); grid_points=np
        .column_stack([XX.ravel(),YY.ravel()]); return gx,gy,XX,YY
        ,grid_points
33 coords=np.column_stack([np.random.uniform(X_MIN,X_MAX,N_STATIONS)
        , np.random.uniform(Y_MIN,Y_MAX,N_STATIONS)])
34 x=coords[:,0]; y=coords[:,1]
35 trend=55+0.12*x-0.08*y+7.0*np.sin(x/18.0)*np.cos(y/22.0)
36 sigma2_true=35.0; range_true=28.0; nugget_true=4.0
37 Dmat=squareform(pdist(coords)); Cov=sigma2_true*np.exp(-Dmat/
        range_true)+nugget_true*np.eye(N_STATIONS)
38 residual=np.random.multivariate_normal(np.zeros(N_STATIONS), Cov)
39 z=trend+residual
40 if z.min()<0: z=z-z.min()+1.0
41 pd.DataFrame({'x':coords[:,0], 'y':coords[:,1], 'z':z}).to_csv(
        DATA_DIR/'observations_simulees.csv', index=False)
42 # figures exploratory
43 plt.figure(figsize=(7,6)); sc=plt.scatter(coords[:,0],coords
       [:,1],c=z,s=55,edgecolor='black'); plt.colorbar(sc,label='
        Valeur observee'); plt.xlabel('Coordonnee x'); plt.ylabel('
        Coordonnee y'); plt.title("Sites d'observation"); plt.xlim(
        X_MIN,X_MAX); plt.ylim(Y_MIN,Y_MAX); plt.grid(True, alpha=.3);
        savefig('stations_observation')
44 stats=pd.DataFrame({'Indicateur':['Nombre d\'observations', '
        Minimum', 'Premier quartile', 'Mediane', 'Moyenne', 'Troisieme
        quartile', 'Maximum', 'Ecart-type'], 'Valeur':[len(z),np.min(z),
        np.quantile(z,.25),np.median(z),np.mean(z),np.quantile(z,.75),
        np.max(z),np.std(z,ddof=1)]})
45 stats.to_csv(TAB_DIR/'stats_descriptives.csv', index=False);

```

```

    export_latex_table(stats, 'stats_descriptives.tex', '
    Statistiques descriptives des valeurs observees.', 'tab:stats-
    descriptives')
46 plt.figure(figsize=(7,5)); plt.hist(z,bins=14,edgecolor='black',
    alpha=.75); plt.xlabel('Valeur observee'); plt.ylabel('
    Frequence'); plt.title('Histogramme des observations'); plt.
    grid(True,alpha=.3); savefig('histogramme_observations')
47 # variogram
48
49 def empirical_variogram(coords,values,n_lags=12,max_dist_frac=.6)
    :
50     distances=pdist(coords); diffs=pdist(values.reshape(-1,1));
        semivariances=.5*diffs**2; max_dist=max_dist_frac*
        distances.max(); bins=np.linspace(0,max_dist,n_lags+1);
        centers=.5*(bins[:-1]+bins[1:]); gamma_hat=np.full(n_lags,
        np.nan); counts=np.zeros(n_lags,dtype=int)
51     for k in range(n_lags):
52         mask=(distances>=bins[k])&(distances<bins[k+1]); counts[k
            ]=np.sum(mask)
53         if counts[k]>0: gamma_hat[k]=np.mean(semivariances[mask])
54     valid=~np.isnan(gamma_hat); return centers[valid],gamma_hat[
        valid],counts[valid],max_dist
55 lags,gamma_exp,pair_counts,max_dist_used=empirical_variogram(
    coords,z,N_LAGS,MAX_DIST_FRAC)
56 pd.DataFrame({'distance':lags,'gamma_exp':gamma_exp,'
    nombre_paires':pair_counts}).to_csv(TAB_DIR/'
    variogramme_experimental.csv', index=False)
57 plt.figure(figsize=(7,5)); plt.scatter(lags,gamma_exp,s=50); plt.
    xlabel('Distance'); plt.ylabel('Semi-variance experimentale');
    plt.title('Variogramme experimental'); plt.grid(True,alpha
    =.3); savefig('variogramme_experimental')
58 # models
59 def spherical_model(h,nugget,sill_contrib,rang):
60     h=np.asarray(h,dtype=float); out=np.zeros_like(h); positive=h
        >0; hr=h[positive]/rang
61     out[positive]=np.where(h[positive]<=rang,nugget+sill_contrib
        *(1.5*hr-.5*hr**3),nugget+sill_contrib); return out
62
63 def exponential_model(h,nugget,sill_contrib,rang):
64     h=np.asarray(h,dtype=float); out=np.zeros_like(h); positive=h
        >0; out[positive]=nugget+sill_contrib*(1-np.exp(-h[

```

```

        positive]/rang)); return out
65
66 def gaussian_model(h,nugget,sill_contrib,rang):
67     h=np.asarray(h,dtype=float); out=np.zeros_like(h); positive=h
        >0; out[positive]=nugget+sill_contrib*(1-np.exp(-(h[
        positive]/rang)**2)); return out
68 MODELS={'Spherique':spherical_model,'Exponentiel':
        exponential_model,'Gaussien':gaussian_model}
69 def fit_variogram_model(model_func,lags,gamma_exp,counts):
70     sample_var=np.var(z,ddof=1); x0=np.array([max(1e-6,.05*
        sample_var), max(1e-6,.95*sample_var), max(1e-6,.35*
        max_dist_used)]); weights=counts/np.max(counts)
71     def obj(theta):
72         return np.sum(weights*(gamma_exp-model_func(lags,*theta))
        **2)
73     bounds=[(0,2*sample_var),(1e-8,4*sample_var),(1e-6,2*
        max_dist_used)]
74     res=minimize(obj,x0=x0,bounds=bounds,method='L-BFGS-B')
75     return {'nugget':res.x[0],'sill_contrib':res.x[1],'range':res
        .x[2],'weighted_sse':res.fun,'success':res.success}
76 fit_results=[]
77 for name,func in MODELS.items():
78     f=fit_variogram_model(func,lags,gamma_exp,pair_counts); f['
        modele']=name; fit_results.append(f)
79 fit_df=pd.DataFrame(fit_results)[['modele','nugget','sill_contrib
        ','range','weighted_sse','success']]; fit_df['palier_total']=
        fit_df['nugget']+fit_df['sill_contrib']
80 best_idx=fit_df['weighted_sse'].idxmin(); best_model_name=fit_df.
        loc[best_idx,'modele']; best_model_func=MODELS[best_model_name
        ]; best_params=fit_df.loc[best_idx,['nugget','sill_contrib','
        range']].values.astype(float)
81 param_table=fit_df.rename(columns={'modele':'Modele','nugget':'
        Effet pepite','sill_contrib':'Contribution','range':'Portee','
        palier_total':'Palier total','weighted_sse':'SSE ponderee','
        success':'Convergence'})
82 param_table.to_csv(TAB_DIR/'parametres_variogramme.csv', index=
        False); export_latex_table(param_table[['Modele','Effet pepite
        ','Contribution','Portee','Palier total','SSE ponderee']],
        parametres_variogramme.tex,'Parametres estimates pour les
        modeles de variogramme.','tab:parametres-variogramme')
83 h_line=np.linspace(0,max_dist_used,300); gamma_line=

```

```

    best_model_func(h_line,*best_params)
84 plt.figure(figsize=(7,5)); plt.scatter(lags,gamma_exp,s=50,label=
    'Variogramme experimental'); plt.plot(h_line,gamma_line,
    linewidth=2,label=f'Modele ajuste : {best_model_name}'); plt.
    xlabel('Distance'); plt.ylabel('Semi-variance'); plt.title('
    Variogramme experimental et modele ajuste'); plt.legend(); plt
    .grid(True,alpha=.3); savefig('variogramme_ajuste')
85 # kriging
86 def variogram_values(h): return best_model_func(h,*best_params)
87 def build_ok_matrix(train_coords):
88     n=train_coords.shape[0]; D=squareform(pdist(train_coords));
        Gamma=variogram_values(D); np.fill_diagonal(Gamma,0.0); K=
        np.ones((n+1,n+1)); K[:n,:n]=Gamma; K[n,n]=0.0; return K
89 def ordinary_kriging(train_coords,train_values,pred_points):
90     train_coords=np.asarray(train_coords,float); train_values=np.
        asarray(train_values,float); pred_points=np.asarray(
        pred_points,float); n=train_coords.shape[0]; m=pred_points
        .shape[0]; K=build_ok_matrix(train_coords); D0=cdist(
        train_coords,pred_points); Gamma0=variogram_values(D0); B=
        np.ones((n+1,m)); B[:n,:]=Gamma0; B[n,:]=1.0
91     try: sol=np.linalg.solve(K,B)
92     except np.linalg.LinAlgError:
93         K_reg=K.copy(); K_reg[:n,:n]+=1e-10*np.eye(n); sol=np.
            linalg.solve(K_reg,B)
94     weights=sol[:n,:]; mu=sol[n,:]; predictions=
        train_values@weights; variances=np.sum(weights*Gamma0,axis
        =0)+mu; return predictions,np.maximum(variances,0.0)
95 gx,gy,XX,YY,grid_points=make_grid(GRID_SIZE); ok_pred,ok_var=
    ordinary_kriging(coords,z,grid_points); OK_MAP=ok_pred.reshape
    (GRID_SIZE,GRID_SIZE); OK_VAR_MAP=ok_var.reshape(GRID_SIZE,
    GRID_SIZE)
96 plt.figure(figsize=(8,6)); cf=plt.contourf(XX,YY,OK_MAP,levels
    =20); plt.scatter(coords[:,0],coords[:,1],c=z,edgecolor='black
    ',s=25); plt.colorbar(cf,label='Prediction par krigeage'); plt
    .xlabel('Coordonnee x'); plt.ylabel('Coordonnee y'); plt.title
    ('Carte interpolee par krigeage ordinaire'); plt.xlim(X_MIN,
    X_MAX); plt.ylim(Y_MIN,Y_MAX); savefig('
    carte_krigeage_ordinaire')
97 plt.figure(figsize=(8,6)); cf=plt.contourf(XX,YY,OK_VAR_MAP,
    levels=20); plt.scatter(coords[:,0],coords[:,1],s=18); plt.
    colorbar(cf,label='Variance de krigeage'); plt.xlabel('

```

```

    Coordonnee x'); plt.ylabel('Coordonnee y'); plt.title('
    Variance du krigage ordinaire'); plt.xlim(X_MIN,X_MAX); plt.
    ylim(Y_MIN,Y_MAX); savefig('variance_krigeage_ordinaire')
98 # deterministic
99 def idw_predict(train_coords ,train_values ,pred_points ,power=2.0,
    eps=1e-12):
100     D=cdist(pred_points ,train_coords); pred=np.empty(pred_points.
        shape[0])
101     for i in range(pred_points.shape[0]):
102         d=D[i]; zero=np.where(d<eps)[0]
103         if len(zero)>0: pred[i]=train_values [zero [0]]
104         else:
105             w=1.0/(d**power); pred[i]=np.sum(w*train_values)/np.
                sum(w)
106     return pred
107 def nearest_neighbor_predict(train_coords ,train_values ,
    pred_points):
108     tree=cKDTree(train_coords); _,idx=tree.query(pred_points ,k=1)
        ; return train_values [idx]
109 idw_pred=idw_predict(coords ,z ,grid_points ,IDW_POWER); nn_pred=
    nearest_neighbor_predict(coords ,z ,grid_points); IDW_MAP=
    idw_pred.reshape(GRID_SIZE ,GRID_SIZE); NN_MAP=nn_pred.reshape(
    GRID_SIZE ,GRID_SIZE)
110 plt.figure(figsize=(8,6)); cf=plt.contourf(XX,YY,IDW_MAP ,levels
    =20); plt.scatter(coords[:,0],coords[:,1],c=z,edgecolor='black
    ',s=25); plt.colorbar(cf,label='Prediction IDW'); plt.xlabel('
    Coordonnee x'); plt.ylabel('Coordonnee y'); plt.title('Carte
    interpolee par IDW'); plt.xlim(X_MIN,X_MAX); plt.ylim(Y_MIN,
    Y_MAX); savefig('carte_idw')
111 plt.figure(figsize=(8,6)); cf=plt.contourf(XX,YY,NN_MAP ,levels
    =20); plt.scatter(coords[:,0],coords[:,1],c=z,edgecolor='black
    ',s=25); plt.colorbar(cf,label='Prediction plus proche voisin'
    ); plt.xlabel('Coordonnee x'); plt.ylabel('Coordonnee y'); plt
    .title('Carte interpolee par plus proche voisin'); plt.xlim(
    X_MIN,X_MAX); plt.ylim(Y_MIN,Y_MAX); savefig('
    carte_plus_proche_voisin')
112 fig,axes=plt.subplots(1,3,figsize=(16,5),constrained_layout=True)
    ; maps=[NN_MAP,IDW_MAP,OK_MAP]; titles=['Plus proche voisin','
    IDW','Krigage ordinaire']; vmin=min(map(np.min,maps)); vmax=
    max(map(np.max,maps))
113 for ax,M,title in zip(axes ,maps ,titles):

```

```

114     im=ax.contourf(XX,YY,M,levels=20,vmin=vmin,vmax=vmax); ax.
        scatter(coords[:,0],coords[:,1],s=12); ax.set_title(title)
        ; ax.set_xlabel('x'); ax.set_ylabel('y'); ax.set_xlim(
            X_MIN,X_MAX); ax.set_ylim(Y_MIN,Y_MAX)
115 fig.colorbar(im, ax=axes.ravel().tolist(), shrink=.9, label='
        Valeur interpolée'); plt.savefig(FIG_DIR/'comparaison_cartes.
        pdf',bbox_inches='tight'); plt.savefig(FIG_DIR/'
        comparaison_cartes.png',dpi=220,bbox_inches='tight'); plt.
        close()
116 # LOO validation
117 def loo_nearest_neighbor(coords,values):
118     preds=np.empty(len(values))
119     for i in range(len(values)):
120         mask=np.ones(len(values),dtype=bool); mask[i]=False;
            preds[i]=nearest_neighbor_predict(coords[mask],values[
            mask],coords[i:i+1])[0]
121     return preds
122 def loo_idw(coords,values,power=2.0):
123     preds=np.empty(len(values))
124     for i in range(len(values)):
125         mask=np.ones(len(values),dtype=bool); mask[i]=False;
            preds[i]=idw_predict(coords[mask],values[mask],coords[
            i:i+1],power)[0]
126     return preds
127 def loo_ordinary_kriging(coords,values):
128     preds=np.empty(len(values)); vars_=np.empty(len(values))
129     for i in range(len(values)):
130         mask=np.ones(len(values),dtype=bool); mask[i]=False; p,v=
            ordinary_kriging(coords[mask],values[mask],coords[i:i
            +1]); preds[i]=p[0]; vars_[i]=v[0]
131     return preds,vars_
132 pred_nn_loo=loo_nearest_neighbor(coords,z); pred_idw_loo=loo_idw(
        coords,z,IDW_POWER); pred_ok_loo,var_ok_loo=
        loo_ordinary_kriging(coords,z); errors_nn=z-pred_nn_loo;
        errors_idw=z-pred_idw_loo; errors_ok=z-pred_ok_loo
133 def metrics(errors): return {'ME':np.mean(errors),'MAE':np.mean(
        np.abs(errors)),'RMSE':np.sqrt(np.mean(errors**2))}
134 validation=pd.DataFrame([{'Methode':'Plus proche voisin',**
        metrics(errors_nn)},{ 'Methode':'IDW',**metrics(errors_idw)},{ '
        Methode':'Krigage ordinaire',**metrics(errors_ok)}])
135 validation.to_csv(TAB_DIR/'validation_croisee.csv', index=False);

```

```

    export_latex_table(validation, 'validation_croisee.tex', '
    Comparaison des methodes par validation croisee leave-one-out.
    ', 'tab:validation-croisee')
136 pd.DataFrame({'x': coords[:,0], 'y': coords[:,1], 'z_observe': z, '
    pred_nn': pred_nn_loo, 'pred_idw': pred_idw_loo, 'pred_ok':
    pred_ok_loo, 'erreur_nn': errors_nn, 'erreur_idw': errors_idw, '
    erreur_ok': errors_ok, 'variance_ok': var_ok_loo}).to_csv(
    DATA_DIR/'erreurs_validation_croisee.csv', index=False)
137 plt.figure(figsize=(7,5)); plt.boxplot([errors_nn, errors_idw,
    errors_ok], labels=['PPV', 'IDW', 'Krigage']); plt.axhline(0.0,
    linestyle='--', linewidth=1); plt.ylabel('Erreur de validation'
    ); plt.title('Distribution des erreurs de validation croisee')
    ; plt.grid(True, axis='y', alpha=.3); savefig('
    boxplot_erreurs_validation')
138 std_ok=np.sqrt(np.maximum(var_ok_loo, 1e-12)); std_errors_ok=
    errors_ok/std_ok; std_summary=pd.DataFrame({'Indicateur': ['
    Moyenne', 'Ecart-type', 'Minimum', 'Maximum'], 'Valeur': [np.mean(
    std_errors_ok), np.std(std_errors_ok, ddof=1), np.min(
    std_errors_ok), np.max(std_errors_ok)]}); std_summary.to_csv(
    TAB_DIR/'erreurs_standardisees_krigeage.csv', index=False);
    export_latex_table(std_summary, 'erreurs_standardisees_krigeage
    .tex', 'Resume des erreurs standardisees du krigeage ordinaire.
    ', 'tab:erreurs-standardisees-krigeage')
139 plt.figure(figsize=(7,5)); plt.hist(std_errors_ok, bins=12,
    edgecolor='black', alpha=.75); plt.axvline(0.0, linestyle='--',
    linewidth=1); plt.xlabel('Erreur standardisee'); plt.ylabel('
    Frequence'); plt.title('Erreurs standardisees du krigeage
    ordinaire'); plt.grid(True, alpha=.3); savefig('
    erreurs_standardisees_krigeage')
140 # summary json-ish
141 (BASE/'resultats_resume.txt').write_text(f'Modele retenu: {
    best_model_name}\nParametres: {best_params}\n\nValidation:\n{
    validation.to_string(index=False)}\n', encoding='utf-8')
142 zip_path = shutil.make_archive('kriging_results', 'zip', BASE)
143 print('Fichier ZIP cree :', zip_path)
144 try:
145     from google.colab import files
146     files.download(zip_path)
147 except Exception:
148     print('Telechargement automatique non disponible hors Colab.'
    )

```

```

149 print('done', best_model_name, best_params)
150 print(validation)
151
152 # -----
153 # Complements pour le chapitre 5 enrichi
154 # -----
155 # Ces sorties supplementaires servent a approfondir la discussion
156 # :
157 # sensibilite de l'IDW au parametre p, diagrammes observe/predit
158 # et carte des erreurs de validation du krigeage.
159
160 powers = [0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0]
161 idw_sensitivity_rows = []
162 for pwr in powers:
163     pred_p = loo_idw(coords, z, power=pwr)
164     err_p = z - pred_p
165     idw_sensitivity_rows.append({'p': pwr, **metrics(err_p)})
166 idw_sensitivity = pd.DataFrame(idw_sensitivity_rows)
167 idw_sensitivity.to_csv(TAB_DIR/'sensibilite_idw_p.csv', index=
168     False)
169 export_latex_table(idw_sensitivity, 'sensibilite_idw_p.tex',
170     'Sensibilite de la methode IDW au parametre de
171     puissance p.',
172     'tab:sensibilite-idw-p')
173
174 plt.figure(figsize=(7,5))
175 plt.plot(idw_sensitivity['p'], idw_sensitivity['RMSE'], marker='o',
176     linewidth=2, label='RMSE')
177 plt.plot(idw_sensitivity['p'], idw_sensitivity['MAE'], marker='s',
178     linewidth=2, label='MAE')
179 plt.xlabel('Parametre de puissance p')
180 plt.ylabel('Erreur de validation')
181 plt.title('Sensibilite de IDW au parametre p')
182 plt.grid(True, alpha=.3)
183 plt.legend()
184 savefig('sensibilite_idw_p')
185
186 # Diagrammes valeurs observees / valeurs predites en validation
187 # croisee
188 plt.figure(figsize=(7,6))
189 plt.scatter(z, pred_nn_loo, alpha=.75, marker='o', label='Plus

```

```

    proche voisin')
184 plt.scatter(z, pred_idw_loo, alpha=.75, marker='s', label='IDW')
185 plt.scatter(z, pred_ok_loo, alpha=.75, marker='^', label='
    Krigeage ordinaire')
186 minv = min(np.min(z), np.min(pred_nn_loo), np.min(pred_idw_loo),
    np.min(pred_ok_loo))
187 maxv = max(np.max(z), np.max(pred_nn_loo), np.max(pred_idw_loo),
    np.max(pred_ok_loo))
188 plt.plot([minv, maxv], [minv, maxv], linestyle='--', linewidth=1)
189 plt.xlabel('Valeurs observees')
190 plt.ylabel('Valeurs predites en validation croisee')
191 plt.title('Valeurs observees et predites')
192 plt.grid(True, alpha=.3)
193 plt.legend()
194 savefig('predictions_vs_observees_validation')
195
196 # Carte des erreurs de validation du krigeage ordinaire
197 plt.figure(figsize=(8,6))
198 sc = plt.scatter(coords[:,0], coords[:,1], c=errors_ok, s=70,
    edgecolor='black')
199 plt.colorbar(sc, label='Erreur observee - predite')
200 plt.axhline(Y_MIN, linewidth=0)
201 plt.xlabel('Coordonnee x')
202 plt.ylabel('Coordonnee y')
203 plt.title('Erreurs de validation du krigeage ordinaire')
204 plt.xlim(X_MIN, X_MAX)
205 plt.ylim(Y_MIN, Y_MAX)
206 plt.grid(True, alpha=.3)
207 savefig('carte_erreurs_validation_krigeage')
208
209 # Comparaison synthetique : meilleur indicateur par critere
210 rank_table = validation.copy()
211 rank_table['Rang_RMSE'] = rank_table['RMSE'].rank(method='min')
212 rank_table['Rang_MAE'] = rank_table['MAE'].rank(method='min')
213 rank_table['Abs_ME'] = rank_table['ME'].abs()
214 rank_table['Rang_abs_ME'] = rank_table['Abs_ME'].rank(method='min
    ')
215 rank_table.to_csv(TAB_DIR/'classement_methodes_validation.csv',
    index=False)
216 export_latex_table(rank_table[['Methode', 'ME', 'MAE', 'RMSE', '
    Rang_abs_ME', 'Rang_MAE', 'Rang_RMSE']],

```

```
217         'classement_methodes_validation.tex',
218         'Classement des methodes selon les criteres de
           validation croisee.',
219         'tab:classement-methodes-validation')
```

# Bibliographie

- [1] Allard, D. (2012). *Statistiques spatiales : introduction à la géostatistique*. Notes de cours, Université de Montpellier.
- [2] Arnaud, M. et Emery, X. (2000). *Estimation et interpolation spatiale : méthodes déterministes et méthodes géostatistiques*. Hermès.
- [3] Baillargeon, S. (2005). *Le krigeage : revue de la théorie et application à l'interpolation spatiale de données de précipitations*.
- [4] Chilès, J.-P. et Delfiner, P. (2012). *Geostatistics : Modeling Spatial Uncertainty*. John Wiley & Sons.
- [5] Dubois, G. (1998). Spatial interpolation comparison 97 : foreword and introduction. *Journal of Geographic Information and Decision Analysis*, 2(2), 1–10.
- [6] Gratton, Y. (2002). *Le krigeage : la méthode optimale d'interpolation spatiale*. Les articles de l'Institut d'Analyse Géographique, 1(4).
- [7] Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6), 119–139.
- [8] Lam, N. S. N. (1983). Spatial interpolation methods : a review. *The American Cartographer*, 10(2), 129–150.
- [9] Li, J. and Heap, A. D. (2008). *A Review of Spatial Interpolation Methods for Environmental Scientists*.
- [10] Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8), 1246–1266.
- [11] Matheron, G. (1971). *The Theory of Regionalized Variables and its Applications*. Les Cahiers du Centre de Morphologie Mathématique, École des Mines de Paris.
- [12] Méneroux, Y. (2019). *Introduction à la géostatistique : variographie, krigeage, interpolation et simulation*. École Nationale des Sciences Géographiques, [https://cours-fad-public.ensg.eu/pluginfile.php/2965/mod\\_resource/content/2/geostatistique.pdf](https://cours-fad-public.ensg.eu/pluginfile.php/2965/mod_resource/content/2/geostatistique.pdf).

- [13] Mitas, L. et Mitasova, H. (1999). Spatial interpolation. In *Geographical Information Systems : Principles, Techniques, Management and Applications*, 1(2), 481–492.
- [14] Wackernagel, H. (2004). *Géostatistique et assimilation séquentielle de données*.
- [15] Wackernagel, H. (1993). *Cours de géostatistique multivariable*. ENSMP, Fontainebleau, France.

## ملخص

تتناول هذه المذكرة دراسة الاستيفاء المكاني باستعمال طريقة الكريغينغ. نعرض في البداية الإشكالية العامة المتمثلة في تقدير قيمة متغير مكاني في مواقع غير مرصودة انطلاقاً من قياسات متوفرة في عدد محدود من المواقع. وبعد التطرق إلى بعض طرق الاستيفاء الحتمية، مثل طريقة أقرب جار، وطريقة الترجيح بعكس المسافة، والشرائح، والتثليث، يتم تقديم الإطار الجيوبوصائي الذي يعتمد على نمذجة الظاهرة المكانية باعتبارها تحقيقاً لحقل عشوائي. وقد تم التركيز على مفاهيم أساسية مثل السكون، ودالة التباين، والفارايو غرام، والفرضية الجوهرية. كما تعرض المذكرة مبادئ الكريغينغ البسيط، والكريغينغ العادي، والكريغينغ الشامل، مع إبراز مبدأ تقليل تباين خطأ التنبؤ تحت شرط عدم التحيز. وفي الجانب التطبيقي، تمت مقارنة الكريغينغ العادي مع طريقة الترجيح بعكس المسافة وطريقة أقرب جار. وتبين النتائج أهمية الكريغينغ، لأنه يوفر تنبؤاً مكانياً مصحوباً بقياس كمي لعدم اليقين.

الكلمات المفتاحية: الاستيفاء المكاني، الجيوبوصاء، الفارايو غرام، الكريغينغ، عدم اليقين.

## Abstract

This dissertation is devoted to the study of spatial interpolation by kriging. It first presents the general problem of predicting a spatial variable at unobserved locations from measurements available at a finite number of sites. After reviewing classical deterministic methods, such as nearest neighbor interpolation, inverse distance weighting, splines and triangulation, the work introduces the geostatistical framework, where the spatial phenomenon is modeled as a realization of a random field. Particular attention is given to stationarity, covariance, the variogram and the intrinsic hypothesis. The dissertation then develops the principles of simple, ordinary and universal kriging, with emphasis on the minimization of the prediction error variance under unbiasedness constraints. A numerical application compares ordinary kriging with IDW and nearest neighbor interpolation. The results highlight the main advantage of kriging: it provides not only spatial predictions, but also a quantitative measure of prediction uncertainty.

Keywords: spatial interpolation, geostatistics, variogram, kriging, uncertainty.

## Résumé

Ce mémoire est consacré à l'étude de l'interpolation spatiale par krigeage. Il présente d'abord le problème général de la prédiction d'une variable spatiale en des points non observés à partir de mesures disponibles en un nombre fini de sites. Après un rappel des principales méthodes déterministes, telles que le plus proche voisin, l'IDW, les splines et la triangulation, le travail introduit le cadre géostatistique fondé sur la modélisation du phénomène spatial comme réalisation d'un champ aléatoire. Une attention particulière est accordée aux notions de stationnarité, de covariance, de variogramme et d'hypothèse intrinsèque. Le mémoire développe ensuite les principes du krigeage simple, ordinaire et universel, en insistant sur la minimisation de la variance de prédiction sous contrainte de non-biais. Une application numérique compare le krigeage ordinaire à l'IDW et au plus proche voisin. Les résultats soulignent l'intérêt du krigeage, qui fournit à la fois une prédiction et une mesure de l'incertitude.

Mots-clés: interpolation spatiale, géostatistique, variogramme, krigeage, incertitude.