



الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي  
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

جامعة سعد دحلب البلدية-1-  
Université Saad Dahlab Blida -1-



**Mémoire de fin D'études**

En vue de l'obtention du diplôme Master

**Faculté des Sciences**  
**Département : Mathématiques**

**Spécialité : Modélisation stochastique et statistique**

**THEME**

**Interpolation spatiale par Krigeage et  
CoKrigeage**

**Présenté par :**  
**CHABANE HENADI**  
**SELLES SARA**  
**Devant le Jury :**

**Soutenu le :**

**Président : A. RASSOUL** Prof. **ENSH de Blida**  
**Promoteur: O.TAMI** MCB **Université de Blida 1**  
**Examineur : R.FRIHI** MAA **Université de Blida 1**

**Promotion : 2020/2021**



# Remerciement

Nous tenons avant toute chose, à remercier DIEU et à lui rendre grâce pour la force et le courage qu'il nous a donné.

Ainsi nous tenons également à exprimer nos vifs remerciements à notre encadreur monsieur Rassoul.

Pour avoir d'abord proposé ce thème, pour le suivi continué tout long de la réalisation de ce mémoire et qui n'a pas cessé de nous donner ses conseils.

Nous voudrions aussi remercier nos familles, pour le soutien et l'aide inconditionnel qu'elle nous apporte chaque jour.

Et je remercie mon binôme pour son aide continue et son souci de travailler parfaitement.

Nous adressons nos sincères remerciements et gratitude au membre de Jury pour nous avoir consacré son précieux temps.

Nous tenons à remercier vivement toutes les personnes qui nous ont aidés et réaliser ce mémoire, ainsi à tous ceux qui nous ont aidés de près ou de loin à accomplir ce travail.

Merci à tous



# DEDICACE

*Je dédie ce modeste travail*

*A mes parents, mes estimes pour eux sont immenses,  
je vous remercie pour tout ce*

*que vous avez fait pour moi. Que dieu vous préserve  
une longue vie heureuse, Avec toute ma tendresse.*

*A tous ceux qui m'ont enseigné et surtout à  
monsieur **A.Rasoul***

*Pour vous, mon chère fiancé : Boubaker*

*Pour vous, mes chères frères : Sallah, Amine*

*Pour vous, mes chères seurs: Fatima*

*Aux personnes que je n'oublierais jamais mes amies :  
Nesrine, Sara, Fatima, Fatiha*

*pour son soutien moral et sa présence dans bons et les  
mauvais moments.*

*A tous les membres de ma promotion.*

*A tous mes professeurs*

*Sarah*



# DEDICACE

*Je dédie ce modeste travail à ma mère, ma raison de vivre pour ces encouragements et ses efforts de m'avoir permis de continuer le chemin dans de bonne condition, A mon père.*

*A tous ceux qui m'ont enseigné et surtout à monsieur  
A. Rassoul*

*A mes adorables sœurs, Hadjer, Roumaïssa, Safaa*

*Et mon beau-frère, Ismaïl*

*A ma grande famille, mes amis qui ont contribué de près ou de loin à moi et mes cousines, Cousins.*

---

*HENADI*

---

# الملخص

تستخدم أساليب الاستيفاء المكاني لتقدير قيم المتغير في النقاط التي لا يعرف فيها من قيم ذلك المتغير في النقاط الملاحظة. الهدف الرئيسي من هذا العمل هو دراسة أساليب الاستيفاء العشوائي ، ونهتم بالطريقتين المعتادتين كريجياج و كوكريجياج ، لفهم مبادئهم الأساسية والمقارنة بينهم .

تتوقع كريجياج قيمة ظاهرة طبيعية في المواقع التي لم يتم أخذ عينات من قبل مزيج خطي غير متحيز من الملاحظات على هذه الظاهرة في المواقع القريبة. كوكريجياج يأتي من الجيوستاتيكيات متعددة المتغيرات ، فإنه يجعل من الممكن لبناء تقريب وظيفية درست من قيمة عدة متغيرات في نقاط عينات. تطبيق طريقتين على البيانات الحقيقية يعطي نتائج هامة جدا، وأخيرا قارنا بين طريقتي الاستيفاء.

## Résumé

Les méthodes d'interpolations spatiale permettent d'estimer les valeurs d'une variable en des points où elle n'est pas connue à partir des valeurs de cette variable aux points observés. L'objectif principal de ce travail est l'étude des méthodes stochastique d'interpolation spatiales, on s'intéresse par les deux méthodes usuelles, krigeage et cokrigeage de comprendre leurs principes de base et de faire une comparaison entres eux. Le krigeage prévoit la valeur d'un phénomène naturel en des sites non échantillonnés par une combinaison linéaire sans biais et à variance minimale des observations du phénomène en des sites voisins. Le cokrigeage provient de la géostatique multivariable, il permet de construire une approximation d'une fonction étudiée à partir de la valeur de plusieurs variables en des points échantillonnés. L'application des deux méthodes sur des données réelles donne des résultats assez important, enfin nous avons comparé entre les deux méthodes d'interpolations.

## Abstract

Spatial interpolation methods are used to estimate the values of a variable at points where it is not known from the values of that variable at the observed points.

The main objective of this work is the study of stochastic methods of spatial interpolation, we are interested in the two usual methods, kriging and cokriging to understand their basic principles and to make a comparison between them. Kriging predicts the value of a natural phenomenon at sites not sampled by an unbiased linear combination of observations of the phenomenon at nearby sites.

Cokriging comes from multivariate geostatic, it makes it possible to construct an approximation of a function studied from the value of several variables at sampled points

The application of the two methods on real data gives quite important results, finally we compared between the two interpolation methods.

# TABLE DES MATIÈRES

<b>Introduction Générale</b>	<b>2</b>
<b>1 Rappel sur statistique et géostatique</b>	<b>4</b>
1.1 Historique de la géostatistique . . . . .	4
1.2 Définitions et propriétés de base . . . . .	5
1.2.1 Espace probabilisé . . . . .	5
1.2.2 Variables aléatoires . . . . .	5
1.2.3 Fonction aléatoire . . . . .	5
1.2.4 Fonction de répartition . . . . .	5
1.2.5 Densité de probabilité . . . . .	5
1.2.6 Espérance mathématique . . . . .	6
1.2.7 Variance . . . . .	6
1.2.8 Covariance . . . . .	6
1.2.9 Coefficient de Corrélation . . . . .	7
1.2.10 Moment statistique . . . . .	7
1.2.11 Champ aléatoire . . . . .	7
1.2.12 Distribution spatiale . . . . .	7
1.2.13 Vecteur aléatoire gaussien . . . . .	7
1.2.14 Fonction aléatoire gaussienne . . . . .	8
1.3 Élément de la géostatistique . . . . .	8
1.3.1 Variance d'extension . . . . .	8
1.3.2 Erreur d'estimation . . . . .	9
1.3.3 Variance de dispersion . . . . .	9
1.3.4 Variables régionalisées . . . . .	9
1.4 <b>La géostatistique</b> . . . . .	<b>9</b>
1.4.1 Données spatiales . . . . .	10
1.4.2 Données géostatique . . . . .	10

1.4.2.1	Processus stochastique . . . . .	10
1.5	Interpolation spatiale . . . . .	11
1.5.1	Interpolation numérique . . . . .	12
1.5.2	Interpolation spatiale . . . . .	12
1.6	Analyse variographique . . . . .	12
1.6.1	Les hypothèses fondamentales . . . . .	12
1.6.1.1	La Stationnarité . . . . .	13
1.6.2	Stationnarité stricte . . . . .	13
1.6.3	Stationnarité faible . . . . .	13
1.6.3.1	Ergodicité . . . . .	14
1.7	Variogrammes . . . . .	15
1.7.1	Propriétés concernant le variogramme . . . . .	15
1.7.1.1	Propriétés du variogramme . . . . .	15
1.7.1.2	Relation entre le variogramme et la covariance . . . . .	16
1.7.1.3	Comparaison entre variogramme et covariance . . . . .	16
1.7.2	Variogramme empirique . . . . .	16
1.7.3	Comportement du variogramme . . . . .	17
1.7.4	Caractéristique du variogramme . . . . .	17
1.7.5	Les modèles de variogramme . . . . .	17
1.7.5.1	Le modèle linéaire . . . . .	17
1.7.5.2	Le modèle linéaire avec palier . . . . .	18
1.7.5.3	Le modèle exponentiel . . . . .	18
1.7.5.4	Le modèle gaussien . . . . .	18
1.7.5.5	<b>Le modèle puissance</b> . . . . .	18
1.7.5.6	Le modèle cubique . . . . .	18
1.7.5.7	Le modèle sphérique . . . . .	18
1.7.5.8	Le modèle à effet de pépite . . . . .	19
1.7.5.9	Le modèle à effet de trou . . . . .	19
1.7.5.10	Le modèle pépitique pur . . . . .	19
1.7.5.11	Les modèle gigognes . . . . .	19
<b>2</b>	<b>Méthodes d'interpolation spatiales</b> . . . . .	<b>20</b>
2.1	Introduction . . . . .	20
2.2	La méthode barycentrique . . . . .	20
2.3	La méthode krigeage . . . . .	21
2.3.1	Définitions et propriétés du krigeage . . . . .	21
2.3.1.1	Formule de Krige . . . . .	21
2.3.1.2	Intérêt du krigeage . . . . .	21
2.3.1.3	Définition du krigeage . . . . .	21

2.3.1.4	Définition de la variance de krigeage . . . . .	22
2.3.1.5	Propriétés du Krigeage . . . . .	22
2.3.1.6	Caractéristiques du Krigeage . . . . .	22
2.3.2	Démarche de résolution . . . . .	23
2.3.2.1	Contrainte de linéarité . . . . .	23
2.3.2.2	Contrainte d'autorisation . . . . .	23
2.3.2.3	Contrainte de non-biais (l'universalité) . . . . .	23
2.3.2.4	Contrainte d'optimalité . . . . .	24
2.3.3	Géostatistique linéaire . . . . .	24
2.3.3.1	Krigeage simple . . . . .	24
2.3.3.2	krigeage ordinaire . . . . .	26
2.3.3.3	Contrainte d'universalité . . . . .	27
2.3.4	Géostatistique non-stationnaire . . . . .	31
2.3.4.1	Krigeage universel . . . . .	31
2.3.4.2	Contrainte de linéarité . . . . .	32
2.3.4.3	Contrainte d'autorisation . . . . .	32
2.3.4.4	Contrainte d'universalité . . . . .	33
2.3.4.5	Contrainte d'optimalité . . . . .	33
2.3.4.6	Krigeage avec modèle de tendance . . . . .	35
2.3.4.7	Krigeage universel . . . . .	35
2.3.4.8	Krigeage avec dérive externe . . . . .	35
2.3.5	D'autres types de krigeage . . . . .	37
2.3.6	Mise en œuvre du krigeage . . . . .	38
2.3.6.1	L'analyse exploratoire . . . . .	38
2.3.6.2	La modélisation . . . . .	38
2.3.6.3	Le krigeage . . . . .	39
2.3.7	Validation croisée . . . . .	39
2.4	La méthode cokrigeage . . . . .	39
2.4.1	Généralités sur le cokrigeage d'une ou plusieurs variables . . . . .	39
2.4.2	Cokrigeage Isotopie et hétérotopie . . . . .	40
2.4.3	Cokrigeage ordinaire . . . . .	41
2.4.4	Cokrigeage simple . . . . .	43
2.4.5	Cokrigeage intrinsèque . . . . .	43
2.4.6	Cokrigeage de facteurs . . . . .	44
<b>3</b>	<b>Applications sur les données Meuse</b>	<b>46</b>
3.1	Introduction . . . . .	46
3.2	Packages géostatistiques R . . . . .	47
3.3	Variable cible et covariables . . . . .	47

3.3.1	Variable cible . . . . .	48
3.3.1.1	Résumer des paramètres descriptives . . . . .	48
3.3.2	Sélection des covariables . . . . .	48
3.3.2.1	Résumer des paramètres descriptives des deux covariables . . . . .	49
3.3.3	Modélisation de la structure spatiale de la variable cible . . . . .	50
3.3.4	Le variogramme . . . . .	51
3.4	Interpolation par krigeage ordinaire de la variable cible . . . . .	52
3.4.1	Prédiction de la variable cible . . . . .	52
3.4.2	Evaluation des erreurs . . . . .	53
3.4.3	Validation . . . . .	53
3.4.4	Evaluation de l'erreur de KO . . . . .	54
3.5	Modélisation des covariables . . . . .	56
3.5.1	Construction des données pour modéliser la co-régionalisation . . . . .	57
3.5.2	Comparaison des modèles régionale et co-régionale . . . . .	58
3.6	Co-krigeage avec une seule co-variable . . . . .	58
3.7	Comparaison du co-krigeage et du krigeage ordinaire . . . . .	59
3.7.1	Prédictions . . . . .	59
3.7.2	Prédiction de l'erreur . . . . .	60
3.7.3	Validation . . . . .	61
3.8	CoKrigeage avec une autre covariable . . . . .	62
3.8.1	Modélisation de la covariable . . . . .	62
3.8.2	Ajustement linéaire de la covariable . . . . .	64
3.8.3	CoKrigeage avec une covariable . . . . .	64
3.8.4	Validation et validation croisée . . . . .	64
3.9	Comparaison des trois interpolations . . . . .	66
3.9.1	Prédiction et erreur de prédiction . . . . .	66
3.10	Validation . . . . .	68
3.10.1	Validation croisée . . . . .	68

## TABLE DES FIGURES

3.1	Histogramme de la variable cible Pb et $\log_{10}(Pb)$ . . . . .	48
3.2	Histogramme des deux covariables et leurs $\log$ decimal . . . . .	50
3.3	La concentration de la matière dans la rivière . . . . .	51
3.4	Variogramme expérimentale des données de la variable cible . . . . .	51
3.5	Ajustement du variogramme . . . . .	52
3.6	Prédictions et erreur OK, $\log_{10}(Pb)$ ; points de sous-échantillon rouges, points supplémentaires verts . . . . .	53
3.7	Erreur de validation de krigeage ordinaire . . . . .	54
3.8	Histogramme de l'erreur de validation du KO . . . . .	55
3.9	Histogramme de l'erreur de la validation croisée du KO . . . . .	55
3.10	Nuage des points, Pb et OM . . . . .	56
3.11	Le variogramme expérimentale de la covariable OM . . . . .	57
3.12	Le graphe d'ajustement du variogramme de la covariable OM . . . . .	58
3.13	Expérimental et ajustement du Variogrammes directs et croisés, $\log_{10}(Pb)$ et $\log_{10}(OM)$ . . . . .	59
3.14	Projection des erreurs de CK . . . . .	60
3.15	Différence entre CK et OK, $\log_{10}(Pb)$ en points de validation, co-variable OM . . . . .	61
3.16	Différence entre CK et OK des résidus du validation croisée $\log_{10}(Pb)$ . . . . .	62
3.17	Nuage de point de la variable cible Pb avec la covariable Zn . . . . .	63
3.18	Variogramme expérimentale de la covariable Zn . . . . .	63
3.19	Ajustement du variogramme de la covariable Zn . . . . .	64
3.20	CK et prédiction, erreur, $\log_{10}(Pb)$ et covariable Zn . . . . .	65
3.21	Différence entre CK et OK prédictions, $\log_{10}(Pb)$ au points de valida- tion, co-variable Zn . . . . .	65
3.22	Différence entre CK and OK validation croisé résiduels, $\log_{10}(Pb)$ , co- variable Zn . . . . .	66

3.23 Prédications de krigeage et leurs erreurs comparées,  $\log_{10}(\text{Pb})$  . . . . . 67

## LISTE DES TABLEAUX

3.1	Paramètres descriptive de la variable cible Pb . . . . .	48
3.2	Paramètres descriptive des covariables : OM et Zn . . . . .	49
3.3	Caractéristiques des prédictions par le krigeage . . . . .	52
3.4	Évaluation de l'erreur de prédiction par krigeage . . . . .	53
3.5	Paramètres d'erreurs de la prédiction . . . . .	53
3.6	Résultats de la validation d'interpolation par le krigeage . . . . .	54
3.7	Résultats de prédiction et erreurs de prédiction . . . . .	58
3.8	Résultats des coefficients statistiques des erreurs . . . . .	59
3.9	Résultats de comparaison entre KO et CK . . . . .	59
3.10	Comparaison des erreurs des résultats entre KO et CK . . . . .	60
3.11	Comparaison entre CK et KO . . . . .	62
3.12	Résultats de prédiction du CK avec la covariable Zn . . . . .	64
3.13	Comparaison entre KO et CK avec la covariable Zn . . . . .	66
3.14	Comparaison entre les trois interpolations de la variable cible Pb . . . . .	68
3.15	Comparaison entre KO et CK avec les deux covariables . . . . .	68

12 juillet 2021

## INTRODUCTION GÉNÉRALE

La géostatistique désigne l'étude statistique des phénomènes naturels. ([Matheron, G. (1965)]) fut le premier à utiliser de façon extensive ce terme et nous retiendrons sa définition :« La géostatistique est l'application du formalisme des fonctions aléatoires à la reconnaissance et l'estimation des phénomènes naturels».

La géostatistique est une application des méthodes d'analyse probabiliste à l'étude de phénomènes corrélés dans l'espace appelés phénomènes régionalisés. On suppose que le phénomène régionalisé peut être décrit par la donnée d'une fonction numérique  $Z$  définie dans un domaine circonscrit de l'espace (le champ) et désignée sous le terme de variable régionalisée. Cette fonction n'est connue que partiellement, par l'intermédiaire d'un échantillonnage.

A partir des mesures disponibles et d'une information relative à la position géographique des données, les techniques de la géostatistique permettent de représenter la structure spatiale du phénomène et d'estimer la répartition de la variable régionalisée dans la zone d'étude. Elles introduisent la notion de fonction aléatoire afin de traduire d'une part l'aspect erratique de la variable régionalisée étudiée, qui empêche de prédire avec certitude les valeurs prises en différents points, et d'autre part, l'existence d'une certaine structure spatiale.

Les méthodes géostatistiques supposent que les valeurs de concentration mesurées sont la réalisation d'un processus aléatoire dont est modélisée la fonction de covariance ou le variogramme.

Les données de référence spatiales sont de plus en plus utilisées, et ce de différentes manières. Pour rechercher dans les surfaces. Que ce soit la densité du minerai dans les échantillons de sol à certains endroits, ces données contiennent toutes des points communs : ils sont situés dans l'espace géographique. Traitement statistique Ce type de données requiert une attention particulière car l'hypothèse classique que les observations soient indépendantes et distribuées symétriquement est rarement vérifiée. Des méthodes statistiques adaptées à l'analyse de données à référence spatiale

ont été développé ([Ripley, B. D. (1981).] [Cressie, N. A. C. (1993).]). Ce mémoire porte sur des méthodes d'interpolation spatiale, (**bary centrique le krigeage et le cokrigeage**). L'application pratique des ces méthodes géostatistiques a pour but caractériser la dépendance spatiale entres différentes observations à l'aide des outils statistiques spatiales .et de faire une comparaison entre la meilleure méthode à l'utiliser .

les sous-objectifs du mémoire sont d'approfondir les fondements mathématiques du krigeage et cokrigeage et d'examiner la méthodologie générale de Ce sujet consiste à mettre en lumière la notion de statistique spatiale ou modélisation de données spatiales .

Le problème qui se pose régulièrement dans le cadre spatial est celui de l'interpolation en tous points de mesure en quelques points de données. Lorsque l'étude se fait sur un champ aléatoire, les methodes d'interpolation (**bary centrique,krigeage,cokrigeage**) permet de prédire la valeur du champ en un point arbitraire en utilisant les valeurs prises par les points de données.

L'objectif de ce mémoire est d'étudier l'interpolation spatiale. Ce travail est organisé en trois chapitres :

Dans le premier chapitre nous ferons un rappel de quelques notions préliminaires en statistique et en géostatistique et les fondements théoriques de l'analyse variographique sont plus approfondis.

Dans le second chapitre, nous nous intéressons aux différentes méthodes d'interpolation spatiales après illustration des théorèmes de bases de l'interpolation numérique. Ces méthodes sont décrites brièvement y compris le krigeage et cokrigeage et bary centre, la théorie du krigeage comparées sont techniques de krigeage, soit le krigeage ordinaire, le krigeage universel, le krigeage optimale le krigeage avec dérive externe, la théorie du cokrigeage en liaison avec les notions d'isotopie/hétérotopie, de corrélation intrinsèque.cokrigeage ordinaire.

Dans le troisième chapitre, nous étudierons un exemple de la Meuse, nous allons définir les différentes formes de krigeage et cokrigeage et leurs propriétés associées, ainsi qu'une validation de la méthode et validation croisée et prédiction des erreurs et évaluation. Ensuite, nous comparons la méthode de krigeage et la méthode de cokrigeage.

# CHAPITRE 1

## RAPPEL SUR STATISTIQUE ET GÉOSTATIQUE

### 1.1 Historique de la géostatistique

L'histoire de la géostatistique remonte aux travaux faits par Kolmogorov et Wiener sur la théorie des fonctions aléatoires entre 1930 et 1950.

En 1955, Daniel Krige, un professeur de l'université du Witwatersrand en Afrique du sud s'est mis à une approche empirique pour corriger les problèmes de biais conditionnel observés dans les mines.

Entre 1960 et 1970, Georges Matheron (ingénieur français du corps des mines), Matern (en foresterie) et Gandin (en météorologie) ont développé les théories et les outils utilisés précédemment par Krige; d'où la naissance de la géostatistique linéaire stationnaire.

Matheron développa une méthode d'estimation qu'il appela krigeage. Il fonda le centre de géostatistique de l'Ecole des Mines de Paris à Fontaine bleau.

En 1970, polytechnique fut la première université en France à enseigner la géostatistique par M. David.

C'est en 1973 que la géostatistique linéaire non-stationnaire fait son apparition et 1975 pour la géostatistique non-linéaire. En 1980 apparut la géostatistique linéaire.

C'est en 1985 que commencèrent les premières simulations (un domaine très actif en recherche aujourd'hui).

Parmi les domaines où la géostatistique est le plus répand : les mines, la foresterie, l'agriculture, le pétrole, l'hydrogéologie, l'environnement, la pêche, la biologie, la géotechnique, la biomédicale et bien d'autres domaines.

## 1.2 Définitions et propriétés de base

### 1.2.1 Espace probabilisé

Un espace probabilisé est un triplé  $(\Omega, A, P)$  qui vérifié :

- $\Omega$  est l'ensemble des états possibles du système étudié.
- $A$  l'ensemble des états du système ayant un sens.
- $P$  est une loi de probabilité.

### 1.2.2 Variables aléatoires

$(\Omega, A, P)$  un espace probabilisé. On appelle une variable aléatoire de  $\Omega$  dans l'ensemble  $E$ , toutes applications mesurable  $X$  de  $\Omega$  vers  $E$ . L'image correspondante est appelée loi de la variable aléatoire. Ce type de fonction permet de modéliser un phénomène aléatoire, Comme par exemple le résultat d'un jet de dés.

### 1.2.3 Fonction aléatoire

Soient  $(\Omega, A, P)$  un espace probabilisé et  $D \subseteq \mathbb{R}^n$

$$\begin{aligned} Z(D, \Omega) &\rightarrow \mathbb{R} \\ (s, w) &\mapsto Z(s, w) \end{aligned}$$

Est une fonction aléatoire (f.a).

### 1.2.4 Fonction de répartition

Soit  $(\Omega, A, P)$  un espace probabilisé, et  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire réelle. On appelle fonction de répartition de  $X$ , la fonction  $F_X$  définie par :

$$F_X(x) = P(X \in ]-\infty, x]) = P(X \leq x). \quad (1.1)$$

### 1.2.5 Densité de probabilité

Soit  $X$  une variable aléatoire réelle et  $F_X$  sa fonction de répartition. Si  $F_X$  est dérivable sur  $\mathbb{R}$ , alors on peut définir la densité de probabilité  $\pi$  par :

$$\pi(x) = \frac{dF_X}{dx}(x) \quad (1.2)$$

### 1.2.6 Espérance mathématique

L'espérance mathématique d'une variable aléatoire réelle est définie comme étant la valeur de cette variable pondérée par sa probabilité. Cette quantité est connue sous le nom de moyenne. Sa formule est donnée par :

$$m(x) = E(Z(x)) \quad (1.3)$$

- Dans le cas discret,  $E[X]$  est la moyenne des valeurs  $x_i$  potentiellement prises par  $X$  pondérées par les probabilités d'apparition associées  $p_i$  :

$$E[Z] = \sum_{i=0}^{+\infty} x_i p_i \quad (1.4)$$

sous réserve que la suite de terme général  $x_i p_i$  soit sommable.

- Dans le cas continu, si  $X$  admet  $\pi$  pour densité et si la fonction  $x\pi(x)$  est intégrable :

$$E[Z] = \int_{-\infty}^{+\infty} x\pi(x)dx \quad (1.5)$$

### 1.2.7 Variance

L'espérance d'une variable aléatoire permet de localiser le centre d'une distribution. Pour caractériser la dispersion de la distribution autour de cette valeur centrale, on a besoin d'un second indicateur. C'est précisément le rôle de la variance.

Soit  $X$  une variable aléatoire (discrète ou continue). On définit la variance de  $X$  comme la valeur moyenne (lorsqu'elle existe) du carré des écarts de  $X$  à son espérance :

$$var(Z(x)) = E[(Z(x) - m(x))^2]. \quad (1.6)$$

### 1.2.8 Covariance

Pour traiter le cas de la somme de variables dépendantes, on doit introduire la notion de covariance.

Sa formule est :

$$cov[Z(x), Z(y)] = E[(Z(x) - m(x))(Z(y) - m(y))]. \quad (1.7)$$

### 1.2.9 Coefficient de Corrélation

Soit  $(X, Y)$  un couple de variables aléatoires réelles. On définit la corrélation par la quantité :

$$\rho_{XY} = \frac{\text{cov}[Z(x), Z(y)]}{\delta_X \delta_Y} \quad (1.8)$$

### 1.2.10 Moment statistique

Soit  $Z$  une variable aléatoire réelle. Alors, pour  $k \in \mathbb{N}$ , on définit le moment d'ordre  $k$  par :

$$m_k = E(Z(x)^k) \quad (1.9)$$

### 1.2.11 Champ aléatoire

Soient  $D$  un sous ensemble de  $\mathbb{R}^d$  et  $(\Omega, A, P)$  un espace de probabilité. Un champ aléatoire est une fonction de deux variables  $Z(x, \omega)$ , telles que pour tout  $x \in D$ ,  $Z(x, \bullet)$  est une variable aléatoire sur  $(\Omega, A, P)$ . Chaque fonction  $Z(x, \omega)$  est une réalisation de la variable aléatoire.

### 1.2.12 Distribution spatiale

La distribution spatiale associée à une fonction aléatoire  $Z$  est définie par l'ensemble :

$$\{Z(s_1), \dots, Z(s_k) \forall k \in \mathbb{N}^*, \forall (s_1, \dots, s_k) \in D^k\} \quad (1.10)$$

En théorie, la distribution spatiale est insuffisante pour calculer la probabilité d'un évènement impliquant un nombre infini non dénombrable de points. C'est pourquoi on introduit la notion de séparation. Une fonction aléatoire est dite séparable si toutes les probabilités impliquant un ensemble de points non dénombrables peuvent être uniquement déterminées par des probabilités sur des ensembles de points dénombrables.

### 1.2.13 Vecteur aléatoire gaussien

Un vecteur aléatoire  $X$  de taille  $n \in \mathbb{N}$  suit une loi normale multidimensionnelle s'il possède une densité de probabilité de type :

$$f(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sqrt{\det \Sigma}}\right) \exp\left(-\frac{1}{2} Y' Y\right)$$

### 1.2.14 Fonction aléatoire gaussienne

Une fonction aléatoire est dite gaussienne si toutes ses distributions spatiales suivent des lois normales multidimensionnelles. Une distribution gaussienne est entièrement déterminée par ses deux premiers moments. Ainsi, la connaissance de la moyenne et de la fonction de covariance suffit pour déterminer la distribution spatiale d'une fonction aléatoire gaussienne. Une fonction aléatoire intrinsèque est gaussienne si ses incréments sont gaussiens multivariés.

Les processus gaussiens sont importants car la loi normale est facile à manipuler et beaucoup de lois de distributions convergent en loi 1 vers la loi gaussienne d'après le théorème de la limite centrale.

## 1.3 Élément de la géostatistique

### 1.3.1 Variance d'extension

Soit  $\bar{Z}(v)$ , la moyenne aléatoire de la moyenne spatiale  $Z(x)$  notée :

$$\bar{Z}(v) = \frac{1}{[v]} \int_D Z(x) dx$$

Où  $[v]$  est la mesure du domaine  $v$ .

$\bar{Z}(v)$ , admet une variance notée :

$$\bar{C}(v, v) = \text{var}[\bar{Z}(v)] = \frac{1}{[v^2]} \int_v \int_v C(x, y) dx dy$$

On appelle variance d'extension du domaine  $v$  au domaine  $v'$  notée  $\delta_E^2(v, v')$  la variance de la différence  $\bar{Z}(v) - \bar{Z}(v')$ .

$$\delta_E^2(v, v') = \text{var}[\bar{Z}(v) - \bar{Z}(v')]$$

La variance d'extension se développe en :

$$\delta_E^2(v, v') = \frac{1}{[v^2]} \int_v \int_v C(x, y) dx dy + \frac{1}{[v'^2]} \int_{v'} \int_{v'} C(x, y) dx dy - \frac{2}{[v][v']} \int_v \int_{v'} C(x, y) dx dy$$

où

$$\delta_E^2(v, v') = \bar{C}(v, v) + \bar{C}(v', v') - 2\bar{C}(v, v')$$

### 1.3.2 Erreur d'estimation

Dans le cas où on considère  $\bar{Z}(v)$  comme une moyenne à estimer et  $\bar{Z}(\hat{v})$  comme son estimateur, l'écart  $\bar{Z}(v) - \bar{Z}(\hat{v})$  sera alors leur erreur d'estimation.

### 1.3.3 Variance de dispersion

Soit  $V$  un espace géographique et  $v$ , un sous-domaine de  $V$  tel que  $N$  est la taille de l'échantillon. On appelle dispersion statistique de  $v$  dans  $V$  et on la note  $S^2 = (v|V)$  la quantité :

$$S^2(v|V) = \frac{1}{N} \sum_i (\bar{Z}_i - \bar{Z})^2$$

La variance de dispersion de  $v$  dans  $V$  est définie comme étant l'espérance mathématique de  $S^2(v|V)$  et est notée :

$$\delta^2(v|V) = E[S^2(v|V)]$$

Ou

$$\delta^2(v|V) = \bar{C}(v, v) - \bar{C}(V, V)$$

### 1.3.4 Variables régionalisées

Le terme variable régionalisé a été choisi par Matheron pour souligner les deux aspects contradictoires de cette variable :

- 1 Un aspect aléatoire (explique les irrégularités locales)
- 2 Un aspect structuré (reflète les tendances du phénomène à grande échelle)

Le phénomène à étudier prend des valeurs  $x$  dans l'espace. Si  $Z(x)$  désigne la valeur du point  $x$  d'une caractéristique  $Z$  de ce phénomène nous dirons que  $Z(x)$  est une variable régionalisée qui dépend de la position de  $x$  dans l'espace.

## 1.4 La géostatistique

**Définition 1.1** *La géostatistique est l'application de la théorie des fonctions aléatoires à des données localisées dans un espace géographique.*

La géostatistique est un ensemble de méthodes statistiques dédiées à l'analyse des phénomènes répartis dans le temps ou dans l'espace. En mathématique, ce phénomène est une fonction qui dépend du temps  $Z(t)$  que l'on considère comme une réalisation d'une fonction aléatoire  $Z(t)$ .

La fonction aléatoire  $Z(t)$  n'est connue qu'en un certain nombre de points donnés  $t_n$  et les valeurs  $Z(t_n)$ , qui sont la réalisation des variables aléatoires  $Z(t)$ .

**Définition 1.2** *La géostatistique est basée sur la théorie des variables régionalisées. C'est à dire une fonction  $z(x)$  dans un domaine  $D$  de l'espace géographique (qui inclut éventuellement un intervalle de l'axe temporel), où  $X$  est un vecteur de coordonnées spatiales et  $z$  est une valeur de la quantité d'intérêt.*

La géostatistique est définie comme étant une application de méthodes probabilistes à des variables régionalisées.

La géostatistique permet une analyse cohérente des données, des incertitudes et des erreurs qui les entachent, ainsi que de la structure spatiale de la teneur

.La géostatistique est classiquement subdivisée en géostatistique linéaire, multivariable, non-stationnaire et non-linéaire.

### 1.4.1 Données spatiales

Les données spatiales sont des informations (ou observations) qui nous donnent à la fois les valeurs mais aussi les localisations geo-référencées. Ces objets geolocalisés définissent un ensemble des coordonnées spatiales et constituent des informations riches lors de l'analyse.

### 1.4.2 Données géostatistique

Les données de type géostatistique, appelées également données continues sont modélisées à partir des processus aléatoires  $\{Y(s), s \in \Omega\}$ .

A chaque mesure de position  $s$  obtenue on l'associe à une variable aléatoire discret  $Y_s$ . Ainsi il existe une valeur pour la variable d'intérêt en tout point  $s$  du territoire étudié. Ils sont utiles lorsqu'on analyse la composition chimique du sol, la qualité de l'eau ou encore la précipitation des pluies. L'analyse de ces types de données spatiales consiste à prédire la valeur d'une variable en un point où elle n'a pas été échantillonnée. La méthode d'interpolations la plus utilisée pour analyser spatialement est le krigeage. Elle s'agit de prédire  $Y$  sur l'ensemble des sites  $S$  en site  $s \in S$  inconnu.

#### 1.4.2.1 Processus stochastique

Reprenons notre problème modèle : nous disposons d'un semis d'observations ponctuelles d'altitudes  $z_1, z_2, \dots, z_n$  mesurées par GPS sur un domaine  $D$  (par exemple un massif montagneux délimité). L'objectif du problème de production d'un Modèle Numérique de Terrain (MNT) consiste à estimer une fonction  $z$  retournant l'altitude en chaque point de  $D$  et telle qu'en chaque point  $z_i = z(x_i)$  soit respectée. Il s'agit d'un problème classique d'interpolation spatiale il existe un certain nombre de méthodes déterministes permettant de traiter ce problème : la méthode du plus proche voisin,

l'interpolation linéaire, bilinéaire, pondérée par l'inverse des distances aux sites d'observations, les splines de lissages et de régression. [Mitas, L. et Mitasova, H. (1999).], [Arnaud, M. et Emery, X. (2000).].

Toutes ces méthodes présentent l'avantage de ne nécessiter aucune hypothèse forte sur le phénomène à estimer (ici le relief du terrain). Nommément, la seule sup position effectuée est que l'altitude sur le domaine  $D$  puisse être décrite par une fonction  $z : D \rightarrow \mathbb{R}$ . En un couple de coordonnées cartographiques  $x$  donné, il ne peut donc y avoir qu'une seule altitude correspondante  $z(x)$ , ce qui interdit donc la modélisation des dévers, grottes, sous-terrains, tunnels... Cette fonction  $z$  est appelée une variable régionalisée.

En Géostatistique, nous devons ajouter une seconde couche d'abstraction à ce modèle primaire, afin de prendre en compte la nature stochastique du phénomène étudié. Nous supposons à présent que la variable régionalisée  $z$  est une réalisation  $Z(\omega)$  d'une variable aléatoire fonctionnelle  $Z$ . Nous allons voir par la suite que ce niveau supplémentaire d'abstraction nécessite de poser des hypothèses un peu plus lourdes sur le phénomène.

**Définition 1.3 (processus gaussien)** *Un processus  $\{Z(s) : s \in X\}$  est un processus gaussien si ses loi fini-dimensionnelles sont Gaussiennes.*

Un processus  $\{Z(s) : s \in X\}$  est dit d'ordre 2 si pour tout  $s \in X$

$$\text{Var}\{Z(s)\} < \infty$$

(et donc a fortiori  $E\{Z(s)\} < \infty$ )

Ceci nous permet alors de considérer la tendance (trend // drift)

$$\begin{aligned} \mu & : X \rightarrow E \\ s & \rightarrow \mu(s) = E\{Z(s)\} \end{aligned}$$

et la fonction de covariance

$$\begin{aligned} K & : X \times X \rightarrow R \\ (s_1, s_2) & \mapsto K(s_1, s_2) = \text{cov}\{Z(s_1), Z(s_2)\} \end{aligned}$$

## 1.5 Interpolation spatiale

L'interpolation est une procédure d'estimation de la valeur d'une variable dans un endroit déterminé à partir de la connaissance des valeurs d'un certain nombre de

points pour permettre de les calculer en chaque point de la zone. Si le point que nous voulons estimer se trouve à l'extérieur du champ des points connus on parle d'extrapolation.

### 1.5.1 Interpolation numérique

Considérons une fonction  $f$  supposée définie positive sur un intervalle  $I$  de  $\mathbb{R}$  connue en un nombre fini de points de  $I$  notés  $x_0, x_1, \dots, x_n, n \in \mathbb{N}$

Pouvons-nous et comment déterminer une fonction polynôme  $P$  de degré au plus  $n$  prenant les mêmes valeurs que la fonction donnée  $F$  en les points  $x_0, x_1, \dots, x_n$ ? Dans le cas de l'existence de cette fonction  $P$ , on dira que  $P$  interpole  $f$  aux points  $\{x_0, x_1, \dots, x_n\}$ .

Cet outil est crucial pour les mathématiciens. Il peut fournir des valeurs approchées de  $f$  en des points autres que les nœuds et ainsi de faire de la prévision.

### 1.5.2 Interpolation spatiale

La statistique spatiale concerne l'étude de phénomènes observés dans un domaine spatial. On note  $s \in D \subset \mathbb{R}^2$  la localisation d'un site de mesure et :

$$Z = \{Z(s) : s \in D\}$$

le phénomène étudié, ou  $Z$  est une variable aléatoire indexée par l'ensemble  $D$ .

## 1.6 Analyse variographique

En tant qu'outil central de la Géostatistique, le variogramme constitue une signature probabiliste (partielle mais suffisante) de l'autocorrélation des phénomènes. Nous définissons en premier lieu plus rigoureusement la notion de phénomène à l'aide des processus stochastiques. Nous établirons alors les conditions sous lesquelles il est possible d'extraire un variogramme, dont nous étudierons les propriétés théoriques, avant de présenter la méthode standard d'estimation du variogramme (inconnu par nature) à partir d'un semis d'observation ponctuelles.

### 1.6.1 Les hypothèses fondamentales

Un processus  $Z$  est entièrement déterminé par ses lois jointes ni-dimensionnelles

$$F_x(z_1, \dots, z_k) = [Z(x_1) \leq z_1, \dots, Z(x_k) \leq z_k]$$

où  $X = [x_1, x_2, \dots, x_n]$ .

Est un vecteur (de dimension finie) répertoriant les  $n$  sites concernées par la loi ni-dimensionnelle. Remarquons que pour un domaine  $D$  de taille finie  $p$ , il y a autant de lois  $F_x$  que de vecteurs  $X$  possibles, à savoir  $2^p$ , le nombre de partitions de  $D$ . Pour un domaine continu (e.g.  $\mathbb{R}$ ), la connaissance de la famille (infinie) de ces lois est impossible en pratique, et on choisit de caractériser  $Z$  uniquement à l'aide de ses 2 premiers moments : moyenne et covariance.

### 1.6.1.1 La Stationnarité

Informellement, la propriété de stationnarité d'un processus stochastique décrit l'uniformité de ses propriétés statistiques sur son domaine de définition analytique. Cela ne signifie en aucun cas qu'il est constant, mais seulement que ses réalisations ont les mêmes chances de produire les mêmes motifs en tout lieu du domaine.

### 1.6.2 Stationnarité stricte

Un processus  $\{Z(s) : s \in X\}$  est dit strictement stationnaire si pour tout  $n \geq 1$  et  $(s_1, \dots, s_n) \in X^n$

$$Pr\{Z(s_1) \in B_1, \dots, Z(s_n) \in B_n\} = Pr\{Z(s_1 + h) \in B_1, \dots, Z(s_n + h) \in B_n\}$$

$h \in X$  et  $B_1, \dots, B_n$  ont des Boréliens de  $\mathbb{R}^n$ .

C'est généralement une propriété bien trop forte (et impossible à vérifier en pratique) qui fait que l'on considère souvent une version assouplie.

### 1.6.3 Stationnarité faible

Un processus d'ordre 2  $\{Z(s) : s \in X\}$  est dit faiblement stationnaire si pour tout  $s, s_1, s_2, h \in X$

$$E\{Z(s)\} = E\{Z(s+h)\}$$

$$cov\{Z(s_1), Z(s_2)\} = cov\{Z(s_1+h), Z(s_2+h)\}$$

i.e,

$$\mu(s) = \mu, \quad K(s_1, s_2) = K(s_1+h, s_2+h)$$

**Proposition 1.1** Si  $\{Z(s) : s \in X\}$  est faiblement stationnaire alors

$$\mu(s) = \mu, \quad K(s_1, s_2) = K(0, s_2 - s_1) = K(s_2 - s_1)$$

ou 0 est une origine arbitraire de  $X$ .

**Définition 1.4 (La stationnarité au sens fort (sss))** *Un processus stochastique est dit stationnaire au sens fort, ou au sens strict si, et seulement si, toutes ses lois jointes ni-dimensionnelles sont invariantes par translation  $h \in D$  :*

$$F_X(z_1, \dots, z_n) = F_{x+h}(z_1, \dots, z_n)$$

**Définition 1.5 (Stationnarité a l'ordre 2 (WSS))** *Un processus stochastique est dit stationnaire au sens faible, ou au sens large, si et seulement si, ses deux premiers moments existent et sont invariants par translation  $h \in D$  :*

1.

$$E[Z(x)] = E[Z(x+h)]$$

2.

$$\text{Cov}[Z(x_1), Z(x_2)] = \text{Cov}[Z(x_1+h), Z(x_2+h)].$$

**Définition 1.6 (isotroph)** *Soit  $Z$  un processus stationnaire au sens faible. On dit que  $Z$  est isotrope si et seulement si, sa covariance ne dépend que de la distance séparant les sites :*

$$\text{Cov}[Z(x_1); Z(x_2)] = C(\|(x_1 - x_2)\|).$$

### 1.6.3.1 Ergodicité

Pour estimer le covariogramme en pratique, nous devons partir des observations  $z_1, z_2, \dots, z_n$ , qui ne sont que des mesures ponctuelles prélevées sur une unique réalisation. Il est important ici de bien saisir l'importance de cette limitation. Dans le cadre de la Géostatistique, les données collectées sur la variable régionalisée (altitudes, températures...) sont le résultat d'un double échantillonnage, d'abord dans l'espace probabiliste  $\Omega$ , puisque seule une réalisation est disponible, puis dans l'espace analytique  $D$  (i.e. l'espace physique réel) puisqu'on ne peut observer qu'un nombre fini de sites.

**Définition 1.7 (Ergodicité a l'ordre 1)** *Soit  $Z$  un processus stochastique. On dit que  $Z$  est ergodique d'ordre 1, ou encore ergodique pour la moyenne, si et seulement si, pour toute réalisation  $z$  du processus*

$$\frac{1}{D} \int_D Z(x) dx \xrightarrow{|D| \rightarrow +\infty} \int_{\Omega} Z(x, \omega) dP(\omega) = E(x)$$

avec  $|D|$  qui désigne la taille de  $D$  (longueur, surface, volume...).

**Définition 1.8 (Ergodicité a l'ordre 2)** *Soit  $Z$  un processus stochastique ergodique a l'ordre 1. On dit que  $Z$  est ergodique a l'ordre 2, ou encore ergodique pour la covariance, si et seulement si, pour toute réalisation  $z$  du processus (de moyenne spatiale  $m_z$  sur  $D$ ), et pour tout*

vecteur de translation  $h$  :

$$\frac{1}{D} \int_D (Z(x+h) - m_z)(Z(x) - m_z) \xrightarrow{|D| \rightarrow +\infty} \text{Cov}[Z(x+h), Z(x)]$$

avec  $|D|$  qui désigne la taille de  $D$  (longueur, surface, volume...).

## 1.7 Variogrammes

C'est l'outil fondamental en géostatistique pour analyser et modéliser la structure spatiale de la variable régionalisée. Il est défini aussi bien dans le cadre stationnaire d'ordre 2 que dans le cadre strictement intrinsèque. Il représente la variabilité moyenne des concentrations entre deux points en fonction de la distance  $h$  qui les sépare.

**Définition 1.9 (Hypothèse intrinsèque et Variogramme)** *Un processus stochastique  $Z$  est dit intrinsèquement stationnaire (ou intrinsèque) s'il est stationnaire pour la moyenne et si la variance de la différence entre deux sites  $x$  et  $x+h$  du domaine  $D$  distants de  $h$ , existe et ne dépend que de  $h$  :*

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

La fonction  $\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  est appelée variogramme (ou parfois semi-variogramme) de  $Z$ .

### 1.7.1 Propriétés concernant le variogramme

#### 1.7.1.1 Propriétés du variogramme

La définition même du variogramme comme variance des accroissements

$$2\gamma(h) = E\{[Z(x+h) - Z(x)]^2\}$$

Entraîne les propriétés suivantes :

$$\begin{cases} \gamma(0) = 0 \\ \gamma(h) = \gamma(-h) \geq 0 \end{cases}$$

En général le variogramme croît à partir de l'origine ; à mesure que  $h$  grandit l'écart quadratique moyen entre les 2 variables  $Z(x)$  et  $Z(x+h)$  tend à croître.

Cette croissance, plus au moins rapide, caractérise ainsi d'une manière statistique la façon dont se détériore l'information apportée par une mesure ponctuelle quand on s'éloigne de ce point connu.

### 1.7.1.2 Relation entre le variogramme et la covariance

La covariance est liée au variogramme par la relation suivante :

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

Comme

$$C(0) = \text{var}[Z(x)] = \delta^2$$

Alors la relation devient :

$$\gamma(h) = \delta^2 - C(h)$$

### 1.7.1.3 Comparaison entre variogramme et covariance

**Espérance :** L'espérance est constante pour les deux cas. Par contre elle est connue pour la covariance.

**Variance :** La variance n'existe forcément pas pour le variogramme ; mais si elle existe, on l'appelle « palier ». Pour la covariance, elle doit exister et est égale à :

$$\text{cov}(h) = 0$$

**Le vecteur  $h$  :** Il croît pour le variogramme et décroît pour la variance.

## 1.7.2 Variogramme empirique

Sous l'hypothèse d'ergodicité au second ordre, on peut estimer  $\gamma$  à partir d'un ensemble d'observations ponctuelles, prélevées sur une unique réalisation du processus. Son estimateur (que l'on note  $\hat{\gamma}$ , prononcé gamma chapeau) est appelé variogramme expérimental.

Pour construire le variogramme expérimental, on doit donc estimer les valeurs de  $\gamma(h)$  pour toutes les distances  $h$  possibles, ce qui est bien évidemment impossible en pratique. On discrétise alors le problème en une séquence de nombres réels  $h_1, h_2, \dots, h_n$  régulièrement échantillonnés (i.e.  $\forall k \in \mathbb{N}^* h_k = h_0 + k\Delta h$ , avec  $\Delta h$  et  $h_0$  des paramètres à calibrer en fonction du nombre de données disponibles).

**Théorème 1.1 (Biais du variogramme empirique)** *Le variogramme expérimental est un estimateur non-biaisé du variogramme théorique :*

$$\forall h \in \mathbb{R}^+ : E[\hat{\gamma}(h)] = \gamma(h).$$

**Théorème 1.2 (Variance du variogramme empirique)** *La dispersion du variogramme expérimental autour du variogramme théorique, est inversement proportionnelle au nombre*

$N(h)$  de couples de sites utilisés dans l'estimation :

$$\text{var}[\hat{\gamma}(h)] = \frac{\gamma(h)^2}{N(h)}.$$

### 1.7.3 Comportement du variogramme

L'examen du comportement du graphe de  $\gamma(h)$  est donc particulièrement important; car le variogramme donne une description synthétique de la structure du phénomène étudié et permet en outre la liaison entre cette structure et la précision avec laquelle pourront être résolus les différents problèmes d'interpolation et d'estimation posés, on étudie en particulier avec soin le comportement au voisinage de l'origine (c'est-à-dire aux faibles distances) et à l'infini (c'est-à-dire aux grandes distances).

### 1.7.4 Caractéristique du variogramme

**Définition 1.10 (le palier)**  $C$  d'un variogramme  $\gamma$  (lorsqu'il existe) est la limite de  $\gamma$  à l'infini :

$$C = \lim_{n \rightarrow +\infty} \gamma(h).$$

**Définition 1.11 (porté)** Pour un seuil de tolérance  $\varepsilon \geq 0$  fixé, la portée d'un variogramme  $\gamma$  de palier  $C$  est la plus petite valeur  $h_{\max}$  telle que :

$$\forall h \geq h_{\max} : 1 - \varepsilon \leq \frac{\gamma(h)}{C} \leq 1 + \varepsilon.$$

**Définition 1.12 (Pépite)** Soit  $\gamma$  un variogramme quelconque (non-nécessairement borné). On appelle pépite du variogramme, la quantité (éventuellement nulle) :

$$\eta = \lim_{h \rightarrow 0^+} \gamma(h).$$

La pente en 0 du variogramme reflète le degré de régularité de la variable régionalisée :

- Parabolique : ( $\gamma'(0) = 0$ ), la variable régionalisée est dérivable.
- Linéaire : ( $\gamma'(0) = a$ ), la variable régionalisée est continue mais non-dérivable.
- Discontinu : ( $\gamma'(0) = +\infty$ ), la variable régionalisée est discontinue.

### 1.7.5 Les modèles de variogramme

#### 1.7.5.1 Le modèle linéaire

$$\gamma(h) = kh$$

Où  $k$  est une constante réelle positive.

### 1.7.5.2 Le modèle linéaire avec palier

$$\gamma(h) = \begin{cases} \frac{C}{a}h & \text{si } h \leq a \\ C & \text{sinon} \end{cases}$$

Où  $C$  et  $a$  sont deux constantes réelles positives.

### 1.7.5.3 Le modèle exponentiel

$$\gamma(h) = C \left(1 - e^{-\frac{h}{a}}\right)$$

Où  $C$  et  $a$  sont deux constantes réelles positives.

### 1.7.5.4 Le modèle gaussien

$$\gamma(h) = C \left(1 - e^{-\frac{h^2}{a^2}}\right)$$

Où  $C$  et  $a$  sont deux constantes réelles positives.

### 1.7.5.5 Le modèle puissance

$$\gamma(h) = kh^\theta \text{ avec } 0 \leq \theta \leq 2$$

### 1.7.5.6 Le modèle cubique

$$\gamma(h) = \begin{cases} C \left(7\frac{h^2}{a^2} - \frac{35}{4}\frac{h^3}{a^3} + \frac{7}{2}\frac{h^5}{a^5} - \frac{3}{4}\frac{h^7}{a^7}\right) & \text{si } h \leq a \\ C & \text{sinon} \end{cases}$$

Où  $C$  et  $a$  sont deux constantes réelles positives.

### 1.7.5.7 Le modèle sphérique

$$\gamma(h) = \begin{cases} C \left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a^3}\right) & \text{si } h \leq a \\ C & \text{sinon} \end{cases}$$

Où  $C$  et  $a$  sont deux constantes réelles positives.

#### 1.7.5.8 Le modèle à effet de pépite

$$\gamma(h) = \begin{cases} C \left( 7\frac{h^2}{a^2} - \frac{35}{4}\frac{h^3}{a^3} + \frac{7}{2}\frac{h^5}{a^5} - \frac{3}{4}\frac{h^7}{a^7} \right) & \text{si } h \leq 0 \\ C & \text{sinon} \end{cases}$$

Où  $C$ ,  $C_0$  et  $a$  sont trois constantes réelles positives.

#### 1.7.5.9 Le modèle à effet de trou

$$\gamma(h) = C \left( 1 - \frac{\sin(h/a)}{h/a} \right)$$

Où  $C$  et  $a$  sont deux constantes réelles positives.

#### 1.7.5.10 Le modèle pépitique pur

$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C & \text{sinon} \end{cases}$$

où  $C$  une constante réelle positive.

où  $\gamma$  a un comportement discontinu à l'origine.

#### 1.7.5.11 Les modèle gigognes

Les modèles de variogrammes sont additifs. Autrement dit, si  $Y_1$  et  $Y_2$  sont des variogrammes valides, alors  $Y_1 + Y_2$  en est un aussi Avec ce procédé, on peut facilement construire des variogrammes étagés, dits variogrammes en gigogne.

$$\gamma(h) = \gamma_1(h) + \gamma_2(h) + \dots + \gamma_n(h)$$

où  $\gamma_1, \gamma_2, \dots, \gamma_n$  sont des modèles de variogramme licites.

Il est possible de regrouper ces modelés en trois classes (Baillargeon, 2005) : les modèles sans palier (linéaire, puissance) avec palier exact (linéaire avec palier, pépitique pur, sphérique et cubique) et enfin avec palier asymptotique (gaussien, exponentiel et à effet de trou). Il existe par ailleurs de nombreux autres modèles moins classiques : Matérn, Legendre, Bessel.

## 2.1 Introduction

Ce chapitre présente des principales méthodes d'interpolation spatiale de comprendre leurs principes de base et de comparer le krigeage aux autres méthodes.

Il ressort alors que malgré les nombreux points communs entre les méthodes, le krigeage se distingue par sa prise en compte de la structure de dépendance spatiale des données.

## 2.2 La méthode barycentrique

La méthode d'interpolation barycentrique prévoit la valeur d'une variable régionalisée en un point non échantillonné  $E$  par une moyenne pondérée des valeurs régionalisées observés :

$$\hat{Z}(s_0) = \sum_{i \in V(s_0)} \lambda_i Z(s_i) \text{ et } \sum_{i \in V(s_0)} \lambda_i = 1$$

Avec

$$\sum_{i \in V(s_0)} \lambda_i = 1.$$

Les poids  $\lambda_i$  sont contraints de sommer à la valeur 1 afin que la prévision ne présente pas de distorsion par rapport à la valeur réelle. Ces poids sont fonction de la distance euclidienne  $|s_i - s_j|$  entre le site d'observation  $s_i$  et le site de prévision  $s_0$  de façon à ce que les sites les plus proches aient plus d'influence dans l'interpolation. Souvent, un poids nul est accordé aux observations les plus éloignées. Ainsi, seules les observations localisées dans un certain voisinage de  $s_0$ , noté  $V(s_0)$  sont prises en compte.

La façon la plus simple de déterminer un voisinage est de prendre les  $n_0$  sites d'observation les plus proches ou les sites tom à l'intérieur d'un cercle centré en  $s_0$  de rayon prédéterminé. Le champ peut aussi être divisé en quadrants ou en octants dont l'origine est  $s_0$ . Le voisinage peut alors comprendre les  $n_0^*$  sites d'observation le plus proche de  $s_0$  pour chaque secteur .

Un exemple populaire de méthode barycentrique est la méthode de l'inverse de la distance à une certaine puissance  $d$ . Par cette méthode, la prévision en un point  $s_0$  prend la forme :

$$\hat{Z}(s_0) = \sum_{i \in V(s_0)} \frac{\frac{1}{|s_i - s_0|^d}}{\sum_{i \in V(s_0)} \frac{1}{|s_i - s_0|^d}} Z(s_i), \quad d > 0$$

## 2.3 La méthode krigage

### 2.3.1 Définitions et propriétés du krigage

#### 2.3.1.1 Formule de Krige

Elle découle de l'expression générale de la variance de dispersion. Pour trois ensembles quelconques  $v, V, W$  elle s'écrit :

$$\sigma^2(V|W) = \sigma^2(v/V) + \sigma^2(V/W)$$

#### 2.3.1.2 Intérêt du krigage

Comme l'échantillonnage fournit des informations exactes sur le domaine d'où proviennent les données mais ne précise pas ce qui se passe en ces points échantillonnées ; alors on a besoin d'une méthode précise pour estimer les valeurs des points intermédiaires ou la moyenne des blocs.

Le krigage doit son nom au professeur Krige-un ingénieur des mines Sud-Africain- qui a développé une technique simple de régression permettant d'estimer les teneurs en or sans biais. Cette technique fut améliorée par le professeur Matheron qui lui a donné le nom de krigage.

#### 2.3.1.3 Définition du krigage

Le krigage est un estimateur linéaire sans biais optimal. Il fournit la moyenne pondérée des valeurs des échantillons qui conduit à l'erreur d'estimation de variance minimale.

Le voisinage de krigage est lui définit comme étant le Domaine du champ qui contient le site à estimer et les données utilisées dans l'estimation. Le voisinage de krigage se résume à un cercle ou à une ellipse autour du point à estimer.

#### 2.3.1.4 Définition de la variance de krigeage

La variance de krigeage est la valeur minimale de la variance de l'erreur de prévision. La variance de krigeage mesure la variance de la différence entre  $Z$ , variable à estimer, inconnue en dehors des points expérimentaux, et son estimateur par krigeage :

$$\delta^2 = \text{var}(Z^* - Z)$$

#### 2.3.1.5 Propriétés du Krigeage

- Le krigeage est un estimateur linéaire sans biais.
- Le krigeage est une interpolation exacte. Chaque point connu estimé à une valeur connue.
- Le krigeage tient en compte la taille du champ à estimer et la position des points entre eux.
- Le krigeage présente un effet local.
- Le krigeage est un estimateur presque sans-biais conditionnel. Ce qui implique que l'estimateur utilisé (le krigeage) est plus lisse que la valeur à estimer.

En général, le krigeage effectue un lissage ; ce qui nous donne des estimations moins variables que les points qu'on cherche à estimer.

- Le krigeage est transitif.
- Le krigeage tient en compte la continuité du phénomène étudié en cas d'utilisation du variogramme tels que l'effet de pépité ou anisotropie

#### 2.3.1.6 Caractéristiques du Krigeage

Cette méthode présente trois caractéristiques :

- Elle est linéaire, c'est-à-dire que l'estimation  $Z$  en  $x_0$  est une combinaison linéaire des valeurs expérimentales  $Z_i$  aux points  $x_i$  ;
- Elle est non biaisée, c'est un choix selon lequel, l'erreur moyenne doit être nulle :

$$E|z^* - z| = 0$$

- Elle prend en compte une condition d'optimalité, minimisant la variance de l'écart entre  $Z$  et son estimation  $Z^*$

$$\text{var}|Z^* - Z| \text{ est minimale.}$$

Ces caractéristiques permettent de calculer les poids et une variance d'estima-

tion à l'aide du variogramme. Le système s'écrit :

$$\sum_{i=1}^n \lambda_i \gamma(x_i - x_i) + \mu = \gamma(x_j - x_0); j = 1, \dots, n$$

$$\sum_{i=1}^n \lambda_i = 1; \mu : \text{Coefficient de Lagrange}$$

La variance de l'estimation vaut alors

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

- L'estimation est exacte, c'est-à-dire que les valeurs estimées aux points de mesures sont égales aux valeurs mesurées.

### 2.3.2 Démarche de résolution

En géostatistique linéaire, la démarche générale de résolution des équations de krigeage proposée par Chauvet(1994) se fait selon la démarche appelée *LAUO* qui comporte les étapes de linéarité, d'autorisation, d'universalité et d'optimalité qui doivent être suivies selon cet ordre.

#### 2.3.2.1 Contrainte de linéarité

En krigeage, le modèle de base considéré est sous forme de combinaison linéaire des données. Il s'écrit :

$$\hat{Z}(s_i) = a + \sum_{i \in V(S_0)} \lambda_i \hat{Z}(s_i),$$

où  $\lambda_i$  et  $a$  sont des constantes.

#### 2.3.2.2 Contrainte d'autorisation

L'erreur d'estimation  $\hat{Z}(s_0) - Z(s_0)$  doit être d'espérance et de variance finies.

Cette contrainte est vérifiée pour le cas des fonctions stationnaires d'ordre 2; donc elle n'intervient que pour les fonctions de stationnarité intrinsèque.

#### 2.3.2.3 Contrainte de non-biais (l'universalité)

L'absence de biais en géostatistique est appelé universalité. Cette contrainte doit être vérifiée pour un estimateur tel que le krigeage. C'est pour cela qu'elle impose que

l'espérance de la variance de l'erreur de prévision soit nulle :

$$E\left(\hat{Z}(s_0) - Z(s_0)\right) = 0$$

### 2.3.2.4 Contrainte d'optimalité

les poids  $\lambda_i$  et la constante  $a$  sont déterminés de façon à minimiser  $Var\left(\hat{Z}(s_0) - Z(s_0)\right)$  sous les contraintes précédentes Cette démarche mènera à la résolution d'équations. Pour faciliter l'écriture de ces équations qui s'avèrent parfois longues, une notation matricielle sera employée. Voici donc tout d'abord quelques remarques sur cette notation :

$Z$  : le vecteur  $n_0 \times 1$  des variables aléatoires  $\hat{Z}(s_1)$  à  $Z(s_{n_0})$  intervenant dans la prévision de  $Z(s_0)$

$\lambda$  : le vecteur  $n_0 \times 1$  des poids associées aux variables aléatoires  $\hat{Z}(s_1)$  à  $Z(s_{n_0})$

$\delta$  : le vecteur  $n_0 \times 1$  des erreurs associées aux variables aléatoires  $\hat{Z}(s_1)$  à  $Z(s_{n_0})$ .

Dans le cadre stationnaire de second ordre :

$\Sigma$  est la matrice  $n_0 \times n_0$  de variances-covariances de  $\delta$ , dont la diagonale est composée uniquement de  $\delta^2$ , la variance commune  $\mu$  a toutes les erreurs  $\delta(s)$  pour  $s \in D$ .

$c_0$  est le vecteur  $n_0 \times 1$  des covariances entre  $\delta$  et  $\delta(s_0)$

Dans le cadre stationnaire intrinsèque :

$\Gamma$  est la matrice  $n_0 \times n_0$  dont l'élément  $(i, j)$  est  $\gamma(s_{(i)} - s_{(j)})$ , soit le semi-variogramme entre  $\delta(s_{(i)})$  et  $\delta(s_{(j)})$ , les éléments  $i$  et  $j$  de  $\delta$

$\gamma_0$  est le vecteur  $n_0 \times 1$  dont l'élément  $i$  est  $\gamma(s_{(i)} - s_0)$ , le semi-variogramme entre  $\delta(s_{(i)})$  et  $\delta(s_0)$

Notons que la prévision  $\hat{Z}(s_0)$  obtenue est une variable aléatoire. Sa valeur numérique est calculée en remplaçant les variables aléatoires  $\hat{Z}(s_i)$  par les valeurs régionalisées observées  $z(s_i)$ .

## 2.3.3 Géostatistique linéaire

### 2.3.3.1 Krigeage simple

La stationnarité qu'utilise ce modèle est la stationnarité de second ordre où l'espérance de la fonction aléatoire étudiée  $\delta(\cdot)$  est constante et connue sur tout le domaine. C'est ce qui rend cette technique facile à exécuter comparé aux autres. L'étude de ce modèle commence par la modélisation de la fonction aléatoire  $Z(\cdot)$  suivante :

$$Z(s) = \delta(s) + m, \forall s \in D,$$

Où :  $m$  est une constante connue.

$\delta(\cdot)$  est fonction aléatoire

on note  $Z = (Z(s_1), \dots, Z(s_i))^t = (Z(s_0), Z^t)$  et  $\delta^* = (\delta(s_0), \delta^t)$

$\Sigma$  : la matrice de  $n_0 \times n_0$  de variances -covariances de  $\delta$

$C_0$  : le vecteur  $n_0 \times 1$  des covariances entre  $\delta$  et  $\delta(s_0)$

le modèle s'écrit alors

$$Z^* = m1_{(n_0+1)} + \delta^* \text{ tel que } \begin{cases} m = \text{cste connue} \\ E(\delta^*) = 0 \\ \text{var}(\delta^*) = \begin{pmatrix} \delta^2 & C_0^t \\ C_0 & \Sigma \end{pmatrix} \end{cases}$$

**2.3.3.1.1 Contrainte de linéarité**  $\hat{Z}(s_0)$  la prévision de  $Z(s_0)$  doit être de la forme

$$\hat{Z}(s_0) = a + \sum_{i \in V(s_0)} \lambda_i Z(s_i) = a + \lambda^t Z$$

**2.3.3.1.2 Contrainte d'autorisation** Comme l'hypothèse de stationnarité du krigeage simple est d'ordre 2, alors l'espérance et la variance de toute fonction ou variable aléatoire existent de l'espérance et de la variance de tout variable aléatoire  $Z(s)$  pour  $s \in D$ .

**2.3.3.1.3 Contrainte d'universalité** La condition d'absence de biais doit être vérifiée. Donc :

soit :

$$E[\hat{Z}(s_0) - Z(s_0)] = E[a + \lambda^t Z - Z(s_0)] = a + \lambda^t m1_{n_0} - m = 0$$

La contrainte de non-biais implique que

$$a = \lambda^t m1_{n_0} - m = m(1 - \lambda^t 1_{n_0})$$

Donc  $\hat{Z}(s_0)$  sera réécrite comme suit :

$$\hat{Z}(s_0) = m(1 - \lambda^t 1_{n_0} + \lambda^t Z) = m + (Z - m1_{n_0})$$

**2.3.3.1.4 Contrainte d'optimalité** On cherche à minimiser la variance de l'erreur de prévision.

C'est-à-dire :

$$\begin{aligned}
 \text{Var}[\hat{Z}(s_0) - Z(s_0)] &= \text{Var}[m + (Z - m\mathbf{1}_{n_0}) - Z(s_0)] \\
 &= \text{Var}[m + \lambda^t \delta - m - \delta(s_0)] \\
 &= \text{Var}(\lambda^t \delta) + \text{Var}(\delta(s_0)) - 2\lambda^t \text{Cov}[\delta, \delta(s_0)] \\
 &= \lambda^t \Sigma \lambda + \delta^2 - 2\lambda^t c_0 = f(\lambda)
 \end{aligned}$$

tels que :  $f(\lambda) = f(\lambda_1, \dots, \lambda_{n_0})$ . Le gradient de cette fonction est :

$$\begin{aligned}
 \frac{\partial}{\partial \lambda} f(\lambda) &= \frac{\partial}{\partial \lambda} (\lambda^t \Sigma \lambda + \delta^2 - 2\lambda^t c_0) \\
 &= 2\Sigma \lambda - 2c_0
 \end{aligned}$$

La prévision de  $Z$  par le krigeage simple est :

$$Z(s_0) = m + c_0^t \Sigma^{-1} (Z - m\mathbf{1}_{n_0})$$

La variance de krigeage pour krigeage simple vaut :

$$\begin{aligned}
 \delta^2(s_0) &= \hat{Z}(s_0) - Z(s_0) = c_0^t \Sigma^{-1} \Sigma \Sigma^{-1} c_0 + \delta^2 - 2c_0^t \Sigma^{-1} c_0 \\
 &= \delta^2 - c_0^t \Sigma^{-1} c_0
 \end{aligned}$$

### 2.3.3.2 krigeage ordinaire

Contrairement au krigeage simple, la stationnarité utilisée ne requière pas forcément que l'espérance de la fonction aléatoire étudiée soit connue mais qu'elle soit constante où au moins localement. Elle ne requière pas aussi une stationnarité d'ordre 2; elle est généralement développer sous l'hypothèse de stationnarité intrinsèque. Donc le krigeage simple est un cas particulier du krigeage ordinaire. Cette technique de krigeage est la plus fréquemment utilisée. La modélisation de la fonction d'intérêt  $z(\cdot)$  de cette technique est :

$$Z(s) = \mu + \delta(s), s \in D$$

Où :  $\mu$  est une constante inconnue

$\delta(\cdot)$  est la fonction aléatoire  $SI - 0$  d'espérance nulle et de structure de dépendance connue.

On note :

$\Gamma$  : La matrice de  $n_0 \times n_0$  dont l'élément  $(i, j)$  de  $\delta$  est  $\gamma(s_i, s_j)$ .

$\gamma_0$  : le vecteur  $n_0 \times 1$  dont l'élément  $i$  est  $\gamma(s_i, s_j)$ .

Le modèle s'écrit alors :

$$Z^* = \mu 1_{(n_0+1)} + \delta^* \text{ tel que } \begin{cases} m = \text{cste inconnue} \\ E(\delta^*) = 0 \\ \Gamma \text{ et } \gamma_0 \text{ connus} \end{cases}$$

**2.3.3.2.1 Contrainte de linéarité** la prévision de  $\hat{Z}(s_0)$  doit être une combinaison linéaire des variables aléatoires  $Z(s_1)$  à  $Z(s_2)$ . elles s'expriment donc sous la forme :

$$\hat{Z}(s_0) = a + \sum_{i \in V(s_0)} \lambda_i Z(s_i) = a + \lambda^t Z$$

**2.3.3.2.2 Contrainte d'autorisation** La stationnarité intrinsèque de  $\delta(\cdot)$  nous oblige à nous assurer que l'erreur de prévision

$$\begin{aligned} \delta^2(s_0) &= \hat{Z}(s_0) - Z(s_0) \\ &= a + \sum_{i \in V(s_0)} \lambda_i Z(s_i) - Z(s_0) \\ &= a + \sum_{i \in V(s_0)} \lambda_i (\mu + \delta(s_i)) - \mu - \delta(s_0) \\ &= a + \mu \sum_{i \in V(s_0)} \lambda_i - \mu + \sum_{i \in V(s_0)} \lambda_i \delta(s_i) - \delta(s_0) \\ &= a + \mu \left( \sum_{i \in V(s_0)} \lambda_i - 1 \right) + \sum_{i \in V(s_0)} \lambda_i \delta(s_i) - \delta(s_0). \end{aligned}$$

Cette erreur de prévision est une combinaison linéaire des  $\delta(\cdot)$  que si  $\sum_{i \in V(s_0)} \lambda_i = 1$ .

Donc la condition de linéarité sera  $\sum_{i \in V(s_0)} \lambda_i = 1$ .

### 2.3.3.3 Contrainte d'universalité

Comme pour le krigeage simple; la prévision doit vérifier la condition de non biais :

$$\begin{aligned}
 E(\hat{Z}(s_0) - Z(s_0)) &= E\left(a + \sum_{i \in V(s_0)} \lambda_i Z(s_i) - Z(s_0)\right) \\
 &= a + \sum_{i \in V(s_0)} \lambda_i \mu - \mu \\
 &= a + \mu \left( \sum_{i \in V(s_0)} \lambda_i - 1 \right) \\
 &= 0
 \end{aligned}$$

Donc la contrainte de non biais est vérifiée que si  $\sum_{i \in V(s_0)} \lambda_i = 1$  Ce qui suppose que  $a = 0$ .

La prévision  $\hat{Z}(S_0)$  sera de la forme :

$$\hat{Z}(s_0) = \sum_{i \in V(s_0)} \lambda_i Z(s_i) \text{ avec } \sum_{i \in V(s_0)} \lambda_i = 1.$$

**2.3.3.3.1 Contrainte d'optimale** On cherche à minimiser la variance de l'erreur de prévision. Pour cela on doit trouver les poids  $\lambda_i$  :

$$\begin{aligned}
 \text{var}(S_0) &= \hat{Z}(s_0) - Z(s_0) \\
 &= E \left[ \left( \hat{Z}(s_0) - Z(s_0) \right)^2 \right] \\
 &= E \left[ \left\{ \sum_{i \in V(S_0)} \lambda_i Z(s_i) - Z(s_0) \right\}^2 \right] \\
 &= E \left\{ \mu \sum_{i \in V(S_0)} \lambda_i + \sum_{i \in V(S_0)} \lambda_i \delta(s_i) - \mu - \sum_{i \in V(S_0)} \lambda_i \delta(s_i) \right\}^2 \\
 &= E \left[ \left\{ \sum_{i \in V(S_0)} \lambda_i \delta(s_i) - \delta(s_0) \right\} \right] \\
 &= E \left[ \left\{ \sum_{i \in V(S_0)} \lambda_i \delta(s_i) \right\}^2 + \{\delta(s_0)\}^2 - 2 \times \delta(s_0) \times \sum_{i \in V(S_0)} \lambda_i \delta(s_i) \right] \\
 &= E \left\{ \sum_{i \in V(S_0)} \sum_{j \in V(S_0)} \lambda_i \lambda_j \delta(s_i) \delta(s_j) - \sum_{i \in V(S_0)} \delta(s_i)^2 \right\} \\
 &\quad + \left\{ \sum_{i \in V(S_0)} \delta(s_i)^2 - 2\delta(s_0) \sum_{i \in V(S_0)} \delta(s_i) + \delta(s_0)^2 \right\}
 \end{aligned}$$

Le premier terme de l'équation s'écrit de la manière suivante :

$$\begin{aligned}
 &\sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j \delta(s_i) \delta(s_j) - \sum_{i \in V(s_0)} \lambda_i \delta(s_i)^2 \\
 = &\sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j \delta(s_i) \delta(s_j) - \sum_{j \in V(s_0)} \lambda_j \sum_{i \in V(s_0)} \lambda_i \delta(s_i)^2 \\
 = &\sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j \delta(s_i) \delta(s_j) - \frac{1}{2} \sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j \delta(s_i)^2 - \frac{1}{2} \sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j \delta(s_j)^2 \\
 = &-\frac{1}{2} \sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j \left( -\delta(s_i) \delta(s_j) + \delta(s_i)^2 + \delta(s_j)^2 \right) \\
 = &-\frac{1}{2} \sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j \left( \delta(s_i) - \delta(s_j) \right)^2
 \end{aligned}$$

Et le deuxième terme s'écrit comme suit :

$$\begin{aligned}
 & \sum_{i \in V(s_0)} \lambda_i \delta(s_i)^2 - 2\delta(s_0) \sum_{i \in V(s_0)} \lambda_i \delta(s_i) + \delta(s_0)^2 \\
 = & \sum_{i \in V(s_0)} \lambda_i \delta(s_i)^2 - 2\delta(s_0) \sum_{i \in V(s_0)} \lambda_i \delta(s_i) + \sum_{i \in V(s_0)} \lambda_i \delta(s_0)^2 \\
 = & \sum_{i \in V(s_0)} \lambda_i (\delta(s_i)^2 - 2\delta(s_0) \delta(s_i) + \delta(s_0)^2) \\
 = & \sum_{i \in V(s_0)} \lambda_i (\delta(s_0) - \delta(s_i))^2
 \end{aligned}$$

D'où :

$$\begin{aligned}
 & \text{var}(\hat{Z}(s_0) - Z(s_0)) \\
 = & E \left[ -\frac{1}{2} \sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j (\delta(s_i) - \delta(s_j))^2 \right] \\
 & + E \left[ \sum_{i \in V(s_0)} \lambda_i (\delta(s_0) - \delta(s_i))^2 \right] \\
 = & -\frac{1}{2} \sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j E \left[ (\delta(s_i) - \delta(s_j))^2 \right] \\
 & + \sum_{i \in V(s_0)} \lambda_i E \left[ (\delta(s_0) - \delta(s_i))^2 \right] \\
 = & - \sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \lambda_i \lambda_j \frac{1}{2} \text{Var}[(\delta(s_i) - \delta(s_j))] \\
 & + \sum_{i \in V(s_0)} \lambda_i \text{Var}[(\delta(s_0) - \delta(s_i))] \\
 = & - \sum_{i \in V(s_0)} \sum_{j \in V(s_0)} \gamma(s_i - s_j) + 2 \sum_{i \in V(s_0)} \gamma(s_0 - s_i) \\
 = & -\lambda^t \Gamma \lambda + 2\lambda^t \gamma_0
 \end{aligned}$$

L'expression de la variance d'erreur s'écrit donc :

$$\delta(s_0)^2 = \hat{Z}(s_0) - Z(s_0) = -\lambda^t \Gamma \lambda + 2\lambda^t \gamma_0 = f(\lambda)$$

On minimise cette équation sous la contrainte

$$\lambda^t \Gamma \mathbf{1}_{n_0} = 1$$

La fonction à minimiser est :

$$f(\lambda, l) = -\lambda^t \Gamma \lambda + 2\lambda^t y_0 + 2(\lambda^t \mathbf{1}_{n_0} - 1)l$$

Où  $l$  est lagrangien.

Le gradient de cette fonction est :

$$\begin{aligned} \frac{\partial}{\partial \lambda} f(\lambda, l) &= \frac{\partial}{\partial \lambda} [-\lambda^t \Gamma \lambda + 2\lambda^t y_0 + 2(\lambda^t \mathbf{1}_{n_0} - 1)l] \\ &= -\Gamma \lambda + 2y_0 + 2l \mathbf{1}_{n_0} \end{aligned}$$

Cette fonction s'annule au point

$$\widehat{\lambda} = \Gamma^{-1}(\gamma_0 + l \mathbf{1}_{n_0})$$

L'estimateur de  $l$  est :

$$\widehat{l} = \frac{1 - y_0 \mathbf{1}_{n_0}^t \Gamma^{-1}}{\mathbf{1}_{n_0}^t \Gamma^{-1} \mathbf{1}_{n_0}}$$

En remplaçant  $l$  par  $\widehat{l}$  dans  $\lambda^t$  on obtient :

$$\widehat{\lambda} = \Gamma^{-1} \left[ y_0 + \frac{(1 - y_0 \mathbf{1}_{n_0}^t \Gamma^{-1})}{\mathbf{1}_{n_0}^t \Gamma^{-1} \mathbf{1}_{n_0}} \mathbf{1}_{n_0} \right]$$

Donc l'expression de la prévision de  $Z(s_0)$  est

$$\widehat{Z}(s_0) = \left[ y_0 + \frac{(1 - y_0 \mathbf{1}_{n_0}^t \Gamma^{-1})}{\mathbf{1}_{n_0}^t \Gamma^{-1} \mathbf{1}_{n_0}} \mathbf{1}_{n_0} \right]^t \Gamma^{-1} Z$$

Et de celle de la variance de krigeage :

$$\begin{aligned} \delta^2(S_0) &= \widehat{Z}(s_0) - Z(s_0) \\ &= y_0^t \Gamma^{-1} y_0 - \frac{(1 - y_0 \mathbf{1}_{n_0}^t \Gamma^{-1})^2}{\mathbf{1}_{n_0}^t \Gamma^{-1} \mathbf{1}_{n_0}} \end{aligned}$$

## 2.3.4 Géostatistique non-stationnaire

### 2.3.4.1 Krigeage universel

Le modèle de base de ce krigeage est :

$$Z(S) = \sum_{j=0} f_j(S) \beta_j + \delta(S) \quad \forall S \in D$$

Où :  $f_j(S)$  sont des fonctions de la position  $S(x, y)$  déterminées.

$\beta_j$  sont des paramètres inconnus.

$\delta(\cdot)$  la fonction aléatoire d'espérance nulle et de structure de dépendance connue.

La tendance linéaire choisit est  $f_0(S) = 1, f_i(S) = W_i \quad i = 1, \dots, p$ .

On note :

$$Z^* = (Z(s_0), Z), \delta^* \delta((s_0), \delta), \beta = (\beta_0, \dots, \beta_p), x_0 = (f_0(s_0), \dots, f_p(s_0))$$

et  $X$  est la matrice  $n_0 \times (p + 1)$  dont l'élément  $(i, j)$  est  $f_j(S_i)$

Le modèle s'écrit alors :

$$Z^* = \begin{pmatrix} \beta x_0 \\ X\beta \end{pmatrix} + \delta^*$$

tel que

$$\begin{cases} \delta(\delta^*) = 0 \\ \delta \text{ et } y_0 \text{ connu} \end{cases}$$

#### 2.3.4.2 Contrainte de linéarité

La prévision  $\hat{Z}(s_0)$  et  $Z(s_0)$  qui doit être aussi une combinaison linéaire de  $Z(S_0)$  est de la forme :

$$\hat{Z}(S_0) = a + \sum_{i \in V(S_0)} \lambda_i Z(S_i) = a + \lambda^t Z$$

#### 2.3.4.3 Contrainte d'autorisation

Puisque la stationnarité ici est intrinsèque alors comme pour le cas ordinaire on doit s'assurer que l'erreur de prévision est une combinaison linéaire d'accroissements de la fonction aléatoire  $\delta(\cdot)$ .

$$\begin{aligned} \delta^2(s_0) &= \hat{Z}(s_0) - Z(s_0) \\ &= a + \sum_{i \in V(S_0)} \lambda_i Z(s_i) - Z(s_0) \\ &= a + \sum_{i \in V(S_0)} \lambda_i (\beta x_i + \delta(s_i)) - \beta x_0 - \delta(s_0) \\ &= a + \sum_{i \in V(S_0)} \lambda_i \beta - \beta x_0 + \sum_{i \in V(S_0)} \lambda_i \delta(s_i - s_0) \end{aligned}$$

La condition de linéarité est d'existence de l'espérance de la variance d'erreur est

$$\sum_{i \in V(S_0)} \lambda_i = 1$$

### 2.3.4.4 Contrainte d'universalité

Pour que la contrainte de non-biais soit vérifiée, l'espérance de la variance d'erreur doit être nulle

$$\begin{aligned}
 & E[\hat{Z}(S_0) - Z(S_0)] \\
 &= E\left[a + \sum_{i \in V(S_0)} \lambda_i Z(S_i) - Z(S_0)\right] \\
 &= a + \sum_{i \in V(S_0)} \lambda_i x_i \beta - x_0 \beta \\
 &= a + \sum_{i \in V(S_0)} \lambda_i \sum_{j=0}^p f_j(S_i) \beta_j - \sum_{j=0}^p f_j(S_0) \beta_j \\
 &= a + \sum_{j=0}^p \left( \sum_{i \in V(S_0)} \lambda_i f_j(S_i) \beta_j - f_j(S_0) \beta_j \right)
 \end{aligned}$$

La condition est vérifiée que si  $\sum_{i \in V(S_0)} \lambda_i = 1$  et  $a = 0$  et  $\sum_{i \in V(S_0)} \lambda_i f_j(S_i) - f_j(S_0) = 0$  pour tout  $\beta_j$  pour  $j = 0, \dots, p$

En posant  $f_i(0) = 1$  on obtient pour  $j = 0$   $\sum_{i \in V(S_0)} \lambda_i f_j(S_i) = f_j(S_0)$  qui devient  $\sum_{i \in V(S_0)} \lambda_i = 1$

La forme matricielle de ces contraintes est :

$$\lambda^t X = x_0^t$$

L'estimateur devient :

$$\hat{Z}(s_0) = \sum_{i \in V(S_0)} \lambda_i Z(s_i)$$

avec

$$\sum_{i \in V(S_0)} \lambda_i f_j(s_i) = f_j(s_0) \text{ pour } j = 1, \dots, p.$$

### 2.3.4.5 Contrainte d'optimalité

L'expression de la variance d'erreur est obtenue en suivant la même étape que le krigeage ordinaire

$$\begin{aligned}
 \delta^2(s_0) &= \hat{Z}(s_0) - Z(s_0) \\
 &= -\lambda^t \Gamma \lambda + 2\lambda^t \gamma_0 \\
 &= f(\lambda)
 \end{aligned}$$

Pour minimiser cette expression on utilise les contraintes d'autorisations vues précédemment ainsi que la contrainte de non-biais :

$$\lambda^t X = x_0^t$$

La fonction à minimiser est :

$$f(\lambda, l) = -\lambda^t \Gamma \lambda + 2\lambda^t y_0 + 2(\lambda^t X - x_0^t)l$$

Où  $Z$  est un vecteur de  $p + 1$  Lagrangiens.

Son gradient est :

$$\frac{\partial}{\partial \lambda} f(\lambda, l) = \frac{\partial}{\partial \lambda} [-\lambda^t \Gamma \lambda + 2\lambda^t y_0 + 2(\lambda^t X - x_0^t)l] = -2\Gamma \lambda + 2y_0 + 2Xl$$

Cette expression s'annule au point

$$\hat{\lambda} = \Gamma^{-1}(y_0 + Xl)$$

En utilisant la contrainte de non-biais :

$$x_0 = X^t \hat{\lambda} + X^t \Gamma^{-1}(y_0 + Xl)$$

l'estimateur du vecteur  $\hat{l}$  est :

$$\hat{l} = (X^t \Gamma^{-1} X)^{-1} (x_0 - X^t \Gamma^{-1} y_0)$$

En remplaçant  $Z$  par sa valeur estimée  $\hat{l}$  dans l'équation de  $\hat{\lambda}$  on obtient :

$$\hat{\lambda} = \Gamma^{-1} \left[ y_0 + X (X^t \Gamma^{-1} X)^{-1} (x_0 - X^t \Gamma^{-1} y_0) \right]$$

D'où résulte l'expression finale de la prévision :

$$\hat{Z}(s_0) = \left[ y_0 + X (X^t \Gamma^{-1} X)^{-1} (x_0 - X^t \Gamma^{-1} y_0) \right] \Gamma^{-1} Z$$

Et de la variance de krigeage :

$$\begin{aligned} \delta^2(s_0) &= Z(s_0) - \hat{Z}(s_0) \\ &= y_0^t \Gamma^{-1} y_0 - (x_0 - X^t \Gamma^{-1} y_0) (X^t \Gamma^{-1} X)^{-1} (x_0 - X^t \Gamma^{-1} y_0) \end{aligned}$$

### 2.3.4.6 Krigeage avec modèle de tendance

L'hypothèse de stationnarité sur laquelle repose les deux types de krigeage présentés précédemment peut souvent être mise en doute. En particulier, il semble souvent erroné de postuler que l'espérance de la fonction aléatoire étudiée reste constante ou quasi-constante sur le champ  $D$ . En krigeage avec un modèle de tendance, aussi appelé krigeage en présence d'une dérive, l'espérance est une fonction des coordonnées spatiales ou de variables régionalisées auxiliaires connues exhaustivement. La majorité des auteurs en géostatistique parlent alors de krigeage universel [Matheron, G. (1969)]. [Cressie, N. A. (1975)] de krigeage avec dérive externe [Goovaerts, P. (1997)]. p.194 [Hans WACKERNAGEL] p.283) respectivement.

### 2.3.4.7 Krigeage universel

Le modèle de base du krigeage universel est :

$$Z(s) = \sum_{j=0}^p f_j(s) \beta_j + \delta(s), s \in D$$

avec  $f_j(s)$  fonctions de la position  $s = (x, y)$ ,  $\beta_j$  paramètres inconnus et  $\delta(\cdot)$  fonction aléatoire stationnaire intrinsèque d'espérance nulle et de structure de dépendance connue. Les  $f_j(s)$  sont déterminés par l'utilisateur. Les graphiques des  $z(s_i)$  en fonction des coordonnées  $x_i$  et  $y_i$  sont utilisés pour guider le choix de ces fonctions. Souvent, une tendance linéaire ou quadratique est choisie. Par exemple, dans le cas d'une tendance linéaire, les  $f_j(s)$  sont :  $f_0(s) = 1, f_1(s) = x, f_2(s) = y$ .

### 2.3.4.8 Krigeage avec dérive externe

En krigeage avec dérive externe, le modèle s'écrit :

$$Z(s) = \sum_{j=0}^p f_j(w) \beta_j + \delta(s), s \in D$$

où  $w = (w_{1(s)}, \dots, w_{q(s)})$  est le vecteur des valeurs prises par les  $q$  variables régionalisées auxiliaires au point  $s$ . Si une tendance linéaire est choisie, les  $f_j(w)$  seront  $f_0(w) = 1, f_1(w) = w_1, f_2(w) = w_2, f_q(w) = w_q$ .

La résolution des équations du krigeage est la même pour ces deux types de krigeage. Dans les paragraphes qui suivent, nous utilisons la notation du krigeage universel, mais la démarche est exactement la même en krigeage avec dérive externe en changeant les  $f_j(s)$  pour des  $f_j(w)$ . En fait, le modèle de tendance peut aussi être composé de fonctions des coordonnées spatiales et de variables régionalisées auxiliaires simultanément.

Sous forme matricielle, le Modèle utilisé pour prévoir  $Z(s_0)$  s'énonce comme suit :

$$Z^* = \begin{pmatrix} x_0\beta \\ X\beta \end{pmatrix} + \delta^* \text{ avec } \begin{cases} E[\delta^*] = 0 \\ \Gamma, \gamma_0 \text{ connus} \end{cases}$$

**2.3.4.8.1 Contrainte linéarité**  $\hat{Z}(S_0)$  la prévision  $Z(S_i)$  doit être de la forme :

$$\hat{Z}(S_0) = a + \sum_{i \in V(S_0)} \lambda_i Z(S_i) = a + \lambda^t Z$$

**2.3.4.8.2 Contrainte d'autorisation** Comme en krigeage ordinaire,  $\mu$  a cause de l'hypothèse de stationnarité intrinsèque, il faut s'assurer que l'erreur  $\hat{Z}(S_0) - Z(S_0)$  soit une combinaison linéaire d'accroissements

de  $\delta(\cdot)$ . Pour en être une, on a vu modèle la section  $\mu$  la section précédente que la somme des poids des termes aléatoires doit valoir zéro. L'erreur de prévision s'écrit :

$$\begin{aligned} \hat{Z}(S_0) - Z(S_0) &= a + \sum_{i \in V(S_0)} \lambda_i Z(S_i) - Z(S_0) \\ &= a + \sum_{i \in V(S_0)} \lambda_i (x_i\beta + \delta(s_i)) - x_0\beta - \delta(s_0) \\ &= \underbrace{a + \sum_{i \in V(S_0)} \lambda_i x_i\beta - x_0\beta}_{\text{termes non aléatoires}} + \sum_{i \in V(S_0)} \lambda_i \delta(s_i) - \delta(s_0) \end{aligned}$$

Il faut donc encore que  $\sum_{i \in V(S_0)} \lambda_i = 1$  pour assurer l'existence de l'espérance et de la variance de  $\hat{Z}(S_0) - Z(S_0)$ .

**2.3.4.8.3 Contrainte d'optimalité** L'expression de la variance d'erreur est obtenue en suivant la même étape que le krigeage ordinaire

$$\begin{aligned} \delta^2(s_0) &= \hat{Z}(S_0) - Z(S_0) \\ &= -\lambda^t \Gamma \lambda + 2\lambda^t \gamma_0 \\ &= f(\lambda) \end{aligned}$$

Cette expression doit être minimisée en respectant les contraintes d'autorisation et de non-biais  $-\lambda^t \Omega^t = w_0^t$ . Cette étape sera effectuée à l'aide d'un vecteur, noté  $l$ , de  $p + 1$  lagrangiens. La fonction à minimiser est donc :

$$f(\lambda, l) = -\lambda^t \Gamma \lambda + 2\lambda^t \gamma_0 + 2(\lambda^t \Omega - w_0^t) l$$

Son gradient est

$$\frac{\partial}{\partial \lambda} f(\lambda, l) = \frac{\partial}{\partial \lambda} \left[ -\lambda^t \Gamma \lambda + 2\lambda^t y_0 + 2(\lambda^t \Omega - w_0^t) l \right] = -2\Gamma \lambda + 2y_0 + 2\Omega l$$

Il vaut zéro au point

$$\hat{\lambda} = \Gamma^{-1} (y_0 + \Omega l)$$

En utilisant la contrainte

$$w_0 = \Omega^t \hat{\lambda} + X^t \Gamma^{-1} (y_0 + \Omega l)$$

le vecteur de Lagrangiens  $l$  est estimé par : l'estimateur du vecteur  $\hat{l}$  est :

$$\hat{l} = (\Omega^t \Gamma^{-1} \Omega)^{-1} (w_0 - \Omega^t \Gamma^{-1} y_0)$$

En remplaçant  $Z$  par sa valeur estimée  $\hat{l}$  dans l'équation de  $\hat{\lambda}$  on obtient :

$$\hat{\lambda} = \Gamma^{-1} \left[ y_0 + \Omega (\Omega^t \Gamma^{-1} \Omega)^{-1} (w_0 - \Omega^t \Gamma^{-1} y_0) \right]$$

D'où résulte l'expression finale de la prévision :

$$\hat{Z}(s_0) = \left[ y_0 + \Omega (\Omega^t \Gamma^{-1} \Omega)^{-1} (w_0 - \Omega^t \Gamma^{-1} y_0) \right] \Gamma^{-1} Z$$

Et de la variance de krigeage :

$$\begin{aligned} \delta^2(s_0) &= Z(s_0) - \hat{Z}(s_0) \\ &= y_0^t \Gamma^{-1} y_0 - (w_0 - \Omega^t \Gamma^{-1} y_0) (\Omega^t \Gamma^{-1} \Omega)^{-1} (w_0 - \Omega^t \Gamma^{-1} y_0) \end{aligned}$$

## 2.3.5 D'autres types de krigeage

**2.3.5.0.1 Krigeage lognormal** Le krigeage lognormal est le krigeage qui utilise la transformation logarithmique pour transformer les données obtenues

$Y(s) = \log[z(s)]$ ,  $\forall s \in D$  est la nouvelle fonction aléatoire associée à la variable régionalisée transformée

La prévision de  $z(s_0)$  par krigeage lognormal est de la forme :

$$\widehat{Z}(s_0) = \exp \left\{ \widehat{Y}(s_0) + \frac{\delta_y^2}{2 - I_Y} \right\}$$

Où :  $\widehat{Y}(s_0)$  est la prévision de  $Y(s_0)$  par krigeage ordinaire.

$\delta_y^2(s_0)$  est la variance de  $\widehat{Y}(s_0)$

$l_Y$  est le lagrangien de la résolution des équations de krigeage ordinaire sur les

$$Y(s_0) = \phi^{-1}[z(s_i)], i \in V(s_0).$$

**2.3.5.0.2 Krigeage trans-gaussien** Cette technique suppose que la fonction aléatoire transformée est :

$$Y(s) = \Phi^{-1}(Z(s)), \forall s \in D$$

La prévision de  $Z(s_0)$  par krigeage trans-gaussien est donc de la forme :

$$\hat{Z}(s_0) = \Phi(\hat{Y}(s_0)) + \Phi''(\mu_Y) \left\{ \frac{\delta_Y^2}{2 - l_Y} \right\}$$

où :  $\Phi(\cdot)$  est une fonction deux fois différentiable tel que  $Y(\cdot)$  est gaussienne .

$\hat{Y}(s_0)$  est la prévision de  $Y(s_0)$  par krigeage ordinaire

$\delta_Y^2(s_0)$  est la variance de  $\hat{Y}(s_0)$ .

$l_Y$  est le lagrangien de la résolution des équations de krigeage ordinaire sur les  $Y(s_i) = \Phi^{-1}(Z(s_i)), i \in V(s_0)$

$$\hat{\mu}_Y = \frac{1_{n_0}^t \sum_Y^{-1} Y}{1_{n_0}^t \sum_Y^{-1} 1_{n_0}}$$

## 2.3.6 Mise en œuvre du krigeage

La mise en œuvre du krigeage se fait en trois étapes :

### 2.3.6.1 L'analyse exploratoire

Dans cette étape, une analyse exploratoire de données multivariées est faite. Cette exploration spatiale des données permet de déterminer la stationnarité de la variable, d'examiner le comportement directionnel de la variable (anisotropie)

et d'identifier les valeurs aberrantes.

### 2.3.6.2 La modélisation

Cette étape se caractérise par la formulation du modèle qui nécessite le choix de la forme de  $\mu$  et l'estimation du variogramme. Le modèle de variogramme peut ensuite être validé par la méthode de bootstrap paramétrique (approche de type Monte Carlo) ou par la méthode de validation croisée. L'erreur quadratique normalisée moyenne, est alors proche de 1 lorsque le modèle est bien estimé.

### 2.3.6.3 Le krigeage

Quand le modèle est sélectionné, l'interpolation peut être effectuée en n'importe quels points. Le krigeage sera réalisé sur une grille régulière pour obtenir une cartographie de la variable d'étude ainsi qu'à l'incertitude associée aux valeurs interpolées.

### 2.3.7 Validation croisée

A fin d'évaluer le modèle de variogramme et le voisinage utilisée pour le krigeage, on utilise une pratique importante nommée validation croisée. Le principe de cette méthode est d'éliminer temporairement et à tour de rôle une observation et de l'estimer à l'aide de ses voisins. Plus précisément, éliminer un point de l'ensemble des données puis estimer sa valeur par krigeage à l'aide des données restantes et du modèle de variogramme qui a été ajusté. En répétant l'opération pour chaque point noté  $s_i$  on obtient une valeur mesurée  $Z_i$  et une valeur estimée  $Z_i^*$  et un écart type de krigeage  $\delta_{ki}^2$ . On définit  $e_i = Z_i - Z_i^*$  comme étant l'erreur d'estimation (ou résidu),  $n_i = \frac{e_i}{\delta_{ki}^2}$  erreur réduite (ou résidu normalisé) et  $\rho = corr(Z, Z^*)$  la corrélation entre la valeur vraie et la valeur estimée. Ainsi le modèle est meilleur si :

- La moyenne des erreurs d'estimation et des erreurs réduites est plus proche de 0 ; i.e.  $\sum_i e_i \approx 0$  et  $\sum_i n_i \approx 0$ . D'où le critère d'absence de biais.
- La variance des erreurs d'estimation est plus faible ; i.e.  $\sum_i |e_i| \min$  ou  $\sum_i |e_i|^2 \min$ . d'où la précision de l'estimateur.
- La variance des erreurs réduites est plus proche de 1 ; i.e.  $\left(\frac{1}{n} \sum_i n_i^2\right)^{\frac{1}{2}} \approx 1$ .
- $\rho$  est plus proche de 1 ; i.e.  $\rho \approx 1$ .

La validation croisée ne constitue toutefois qu'une validation partielle du modèle. Elle ne porte que sur les points qui ont servi à construire le variogramme.

## 2.4 La méthode cokrigeage

Le cokrigeage est une extension du krigeage au cas multivarité. Dans le cas simple, l'on souhaite interpoler une variable régionalisée  $Z$  en s'appuyant sur une variable régionalisée  $Y$ .

### 2.4.1 Généralités sur le cokrigeage d'une ou plusieurs variables

Pour simplifier l'écriture, et sans nuire à la généralité, nous considérerons le cas de 2 variables  $Z_1(s)$  et  $Z_2(s)$  informées respectivement sur deux ensembles de points  $s_1$  et  $s_2$  identiques (cas isotopique) ou non (cas hétérotopique). Le cokrigeage permet de tenir compte des informations sur les deux variables en estimant au mieux, linéairement en fonction des données sur  $Z_1$  et  $Z_2$  :  $Z_2(s_0)$  où  $Z_1(s_0)$  d'ailleurs chacune des

deux variables, ou une régularisée de ces variables, ou encore une combinaison linéaire comme  $Z_2(s_0) - Z_1(s_0)$ . Le cokrigage assure la cohérence entre estimations :

$$[Z_2(s_0) - Z_1(s_0)]^{ck} = Z_2(s_0)^{ck} - Z_1(s_0)^{ck}$$

Dans la suite nous commencerons par développer les calculs en considérant l'estimation d'une variable cible notée  $Z(x_0)$  (qui pourra être  $Z_1(s), Z_2(s)$ , ou par exemple  $Z_2(s_0) - Z_1(s_0)$ ...)

D'une manière plus générale. On considère un estimateur constitué par une combinaison linéaire des données :

$$Z^*(s_0) = \sum_{S_1} \lambda_{1\alpha} Z_1(s_\alpha) + \sum_{S_2} \lambda_{2\alpha} Z_2(s_\alpha) + \lambda_0$$

où :  $\sum_{S_1} \lambda_{1\alpha} + \sum_{S_2} \lambda_{2\alpha} = 1$ . Ce qui correspond à l'erreur d'estimation :

$$\xi = Z_0 - Z_0^* = Z_0 - \sum_{S_1} \lambda_{1\alpha} Z_1(s_\alpha) - \sum_{S_2} \lambda_{2\alpha} Z_2(s_\alpha) - \lambda_0$$

On cherchera ensuite les poids tels que l'erreur d'estimation soit une combinaison linéaire autorisée (ce qui permet d'en calculer moyenne et variance), soit d'espérance nulle (estimateur sans biais) et enfin de variance minimale. En terme de covariance, et avec des notations abrégées, la variance qu'il s'agit de minimiser s'écrit :

$$\begin{aligned} Var(\xi) &= C(Z_0, Z_0) - 2 \sum_{S_1} \lambda_{1\alpha} C(Z_0, Z_{1\alpha}) - 2 \sum_{S_2} \lambda_{2\alpha} C(Z_0, Z_{2\alpha}) \\ &\quad + 2 \sum_{S_1} \sum_{S_2} \lambda_{1\alpha} \lambda_{2\beta} C(Z_{1\alpha}, Z_{2\beta}) \\ &\quad + 2 \sum_{S_1} \sum_{S_1} \lambda_{1\alpha} \lambda_{1\beta} C(Z_{1\alpha}, Z_{1\beta}) \\ &\quad + 2 \sum_{S_2} \sum_{S_2} \lambda_{2\alpha} \lambda_{2\beta} C(Z_{2\alpha}, Z_{2\beta}) \end{aligned}$$

Il est clair que si  $Z(x_0)$  figure dans les données, le cokrigage de  $Z(x_0)$  meilleur estimateur linéaire, s'identifie à lui-même : comme le krigage : le cokrigage est un interpolateur exact.

## 2.4.2 Cokrigage Isotopie et hétérotopie

Les mesures disponibles sur deux variables  $Z_1(s)$  et  $Z_2(s)$  dans un domaine donné peuvent être situées, soit aux mêmes points de mesure, soit en des points différents. On distingue les cas suivants :

– **isotopie** : en tous les points d'échantillonnage on dispose d'informations sur la paire de variables.

– **hétérotopie** : les deux variables ont été mesurées sur deux ensembles de points disjoints.

– **hétérotopie partielle** : une partie des points de mesure ne sont pas communs aux deux variables. Un cas particulier d'hétérotopie partielle est celui où l'ensemble des points de mesure d'une variable est inclu dans l'ensemble

Des points de mesure de l'autre variable.

L'estimateur du cokrigeage ordinaire est une combinaison linéaire de pondérateurs  $\lambda_\alpha^i$  avec des variables situées en des points d'un voisinage du domaine

$$Z_{i_0}^*(s_0) = \sum_{i=1}^n \sum_{\alpha=1}^{n_i} \lambda_\alpha^i Z(s_\alpha)$$

où l'index  $i_0$  désigne une variable particulière de l'ensemble de  $N$  variables. On remarque que le nombre d'échantillons  $n_i$  dépend de l'indice  $i$  des variables, afin d'inclure dans la notation le cas de l'hétérotopie, partielle ou totale.

### 2.4.3 Cokrigeage ordinaire

Dans le cadre d'une hypothèse de stationnarité intrinsèque conjointe, on désire estimer une variable particulière d'un ensemble de  $N$  variables, en se basant sur une erreur d'estimation nulle en moyenne. Cette condition est remplie en choisissant des pondérateurs de somme unité pour la variable d'intérêt, et de somme nulle pour les variables auxiliaires.

$$\sum_{\alpha=1}^{n_i} \lambda_\alpha^i \delta_{ii_0} = \begin{cases} 1 & \text{si } i = i_0 \\ 0 & \text{sin on} \end{cases}$$

On développe l'erreur d'estimation

$$\begin{aligned} E[Z_{i_0}^*(s_0) - Z_{i_0}(s_0)] &= E \left[ \sum_{i=1}^n \sum_{\alpha=1}^{n_i} \lambda_\alpha^i Z_i(s_\alpha) - \underbrace{\sum_{\alpha=1}^{n_i} \lambda_\alpha^{i_0} Z_{i_0}(s_0)}_1 - \underbrace{\sum_{\substack{i=0 \\ i \neq i_0}}^N \sum_{\alpha=1}^{n_i} \lambda_\alpha^i Z_i(s_0)}_0 \right] \\ &= \sum_{i=0}^N \sum_{\alpha=1}^{n_i} \underbrace{\lambda_\alpha^i E[Z_i(s_\alpha) - Z_i(s_0)]}_0 \\ &= 0 \end{aligned}$$

Pour la variance de l'erreur d'estimation il vient alors

$$\delta_E^2 = E \left[ \left( \sum_{i=0}^N \sum_{\alpha=1}^{n_i} \lambda_\alpha^i Z_i(s_\alpha) - Z_{i_0}(s_0) \right)^2 \right]$$

En introduisant des pondérateurs  $\lambda_\alpha^i$  définis par

$$\lambda_0^i = -\delta_{ii_0} = \begin{cases} -1 & \text{si } i = i_0 \\ 0 & \text{sin on} \end{cases}$$

ajoutés aux sommations, on condense l'expression de la variance d'estimation

-

Puis, en introduisant des variables aléatoires fictives situées arbitrairement à l'origine, on peut former des incréments

$$\begin{aligned} \delta_E^2 &= E \left[ \left( \sum_{i=0}^N \left( \sum_{\alpha=0}^{n_i} \lambda_\alpha^i Z_i(x_\alpha) - Z_i(0) \sum_{\alpha=0}^{n_i} \lambda_\alpha^i \right) \right)^2 \right] \\ &= E \left[ \left( \sum_{i=0}^N \sum_{\alpha=1}^{n_i} \lambda_\alpha^i \underbrace{\left( Z_i(s_\alpha) - Z_i(0) \right)}_{\text{incrément}} \right)^2 \right] \end{aligned}$$

et une covariance croisée d'incrément  $C_{ij}^I(x_\alpha, x_\beta)$

$$\delta_E^2 = \sum_{i=0}^N \sum_{j=0}^N \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_\alpha^i \lambda_\beta^j C_{ij}^I(x_\alpha, x_\beta)$$

Il est important de voir que l'on est obligé d'introduire, du point de vue physique, l'hypothèse que les covariances croisées d'incrément sont symétriques, pour pouvoir les remplacer par des variogrammes croisés. A cette condition, on a finalement

$$\delta_E^2 = \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_\alpha^i \lambda_\beta^j \gamma_{ii_0}(s_\alpha - s_0) - \gamma_{i_0 i_0}(x_0 - x_0) - \sum_{i=0}^N \sum_{j=0}^N \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_\alpha^i \lambda_\beta^j \gamma_{ij}(s_\alpha - s_\beta)$$

Après minimisation, où les contraintes sur les pondérateurs suscitent  $N$  paramètres de Lagrange  $\mu_i$ , vient le cokrigage ordinaire

$$\begin{cases} \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_\alpha^i \lambda_\beta^j (s_\alpha - s_\beta) + \mu_i = \gamma_{ii_0}(s_\alpha - x_{s_0}) \text{ pour } i = 1, \dots, N; \alpha = 1, \dots, N \\ \sum_{\beta=1}^{n_i} \lambda_\alpha^i = \delta_{ii} \text{ pour } i = 1, \dots, N \end{cases}$$

Et la variance de cokrigage

$$\delta_{Ck}^2 = \sum_{i=0}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i \gamma_{ii_0} (x_{\alpha} - x_0) + \mu_{i_0} - \gamma_{i_0 i_0} (x_0 - x_0)$$

#### 2.4.4 Cokrigage simple

Le cokrigage ordinaire n'a en général pas de sens lorsqu'aucune information n'est disponible sur la variable d'intérêt dans le voisinage où l'on désire le pratiquer. Par contre, le cokrigage simple s'appuie sur une connaissance des moyennes des variables, ce qui lui permet de calibrer l'estimation d'une variable sans posséder aucune donnée sur cette variable dans le voisinage de cokrigage.

L'estimateur du cokrigage simple est constitué de la moyenne de la variable d'intérêt et d'une combinaison linéaire de pondérateurs  $\lambda_{\alpha}^i$  avec les résidus de toutes les variables par rapport à leurs moyennes.

$$Z_{i_0}^*(s_0) = m_{i_0} + \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i (Z_i(s_{\alpha}) - m_i)$$

On lui associe le système de cokrigage simple, écrit sous forme matricielle

$$\begin{pmatrix} W_{11} & \dots & C_{1j} & \dots & C_{1N} \\ \dots & \dots & \dots & \dots & \dots \\ C_{i1} & \dots & W_{ii} & \dots & C_{iN} \\ \dots & \dots & \dots & \dots & \dots \\ C_{N1} & \dots & C_{Ni} & \dots & W_{nn} \end{pmatrix} \begin{pmatrix} 1_1 \\ \dots \\ 1_i \\ \dots \\ 1_N \end{pmatrix} = \begin{pmatrix} c_{1i_0} \\ \dots \\ c_{ii_0} \\ \dots \\ c_{Ni_0} \end{pmatrix}$$

où la matrice du membre gauche est construite avec des blocs symétriques  $W_{ii}$  d'ordre  $n_i \times n_i$  sur la diagonale, et des blocs rectangulaires  $C_{ij}$  d'ordre  $n_i \times n_j$  en dehors de la diagonale avec  $C_{ij} = C_{ji}^T$ .

Dans un contexte d'isotopie, on peut réécrire le vecteur des pondérateurs sous la forme de l'opérateur de sriable avec la variable d'intérêt. En désignant par  $W$  la matrice membre gauche d'ordre  $n \times n$ , le cokrigage simple peut être posé comme

$$W \times L = C_0$$

#### 2.4.5 Cokrigage intrinsèque

En effet, dans le cas intrinsèque, les conditions

$$\sum_{S_1} \lambda_{1\alpha} = 1, \sum_{S_2} \lambda_{2\alpha} = 0$$

sont précisément celles pour lesquelles l'erreur d'estimation

$$\begin{aligned}\varepsilon &= Z_1(s_0) - Z_1^*(s_0) \\ &= Z_1(s_0) - \sum_{S_1} \lambda_{1\alpha} Z_1(s_\alpha) - \sum_{S_2} \lambda_{2\alpha} Z_2(s_\alpha)\end{aligned}$$

est une CLA. Sa moyenne est nulle (ce qui assure le non-biais). Sa variance se développe à l'aide des variogrammes simples et croisés, et conduit à un système de cokrigage identique au précédent, mais écrit en variogrammes (au signe près pour les paramètres de Lagrange).

La prise en compte supplémentaire des données sur  $Z_2$  pour estimer la variable  $Z_1$ , ne peut qu'augmenter la précision : la variance de cokrigage ne peut qu'être inférieure ou égale à la variance de krigeage à partir des seules données sur  $Z_1$  (en effet le krigeage revient à donner un poids nul, en général non optimal, aux données de  $Z_2$ ).

Attention à l'amplitude des poids : les poids de la variable à estimer  $Z_2$  sont sans unité, mais ceux d'une co-variable  $Z_1$  sont en (unité  $Z_1$  / unité  $Z_2$ )

En cokrigage ordinaire, les poids d'une co-variable  $Z_2$  somment à 0. Les poids négatifs, s'ils sont associés à des valeurs fortes, peuvent rendre négative l'estimation d'une variable positive  $Z_1$ .

### 2.4.6 Cokrigage de facteurs

Le modèle linéaire de corégionalisation permet d'estimer localement les facteurs régionalisés  $Y_p^\mu(x)$  par cokrigage. L'estimateur utilisé pour évaluer un facteur  $Y_{p_0}^{\mu_0}(x)$  donné est une combinaison linéaire de pondérateurs  $\lambda_\alpha^i$  avec des variables dans le voisinage d'un point quelconque  $x_0$  du domaine

$$Y_{u_0 p_0}^*(S) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_\alpha^i Z(S_\alpha)$$

En se plaçant dans le cadre d'une hypothèse de stationnarité locale, donnant un sens à des moyennes locales  $m_i(x_0)$ , on construit un estimateur sans biais de facteurs synthétiques (auxquels on a attribué des moyennes nulles)

$$E\left[Y_{u_0 p_0}^*(S) - Y_{p_0}^{\mu_0}(S)\right] = \sum_{i=1}^N m_i(S_0) \sum_{\alpha=1}^{n_i} \lambda_\alpha^i = 0 \text{ pour } \sum_{\alpha=1}^{n_i} \lambda_\alpha^i = 0$$

La variance d'estimation  $\delta_E^2$  vaut

$$E\left[\left(Y_{u_0 p_0}^*(S) - Y_{p_0}^{\mu_0}(S)\right)^2\right] = 1 + \sum_{i=0}^N \sum_{j=0}^N \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_\alpha^i \lambda_\beta^j C_{ij}(s_\alpha - s_\beta) - 2 \sum_{i=0}^N \sum_{\alpha=1}^{n_i} \lambda_\alpha^i a_{u_0 p_0}^{i_0} \rho_{u_0}(s_\alpha - s_0)$$

De sorte que l'on a le système de cokrigeage

$$\left\{ \begin{array}{l} \sum_{i=1}^N \sum_{\beta=1}^{nj} \lambda_{\beta}^i C_{ij}(s_{\alpha} - s_{\beta}) - \mu_i = a_{u_0 p_0}^{i_0} \rho_{u_0}(s_{\alpha} - s_0) \text{ pour } i = 1, \dots, N; \alpha = 1, \dots, N \\ \sum_{\beta=1}^{nj} \lambda_{\beta}^i \text{ pour } i = 1, \dots, N \end{array} \right.$$

Dans le membre droit de ce système apparaissent les coefficients de transformation  $a_{up}^i$ , résultats de la décomposition des matrices de corégionalisation  $B_u$ . Dans le cas d'une analyse en composantes principales de ces matrices, les  $a_{up}^i$  indiquent les covariances entre les variables  $Z_i$  et les facteurs  $Y_p^u$ .

Le cokrigeage de facteurs permet de réaliser des cartes de facteurs en n'utilisant que les données multivariées contenues dans un voisinage local autour de chaque point de la carte.

### 3.1 Introduction

L'application des modèles d'interpolations spatiales présentées dans le chapitre 2, sur une base des données s'appelle "meuse" fourni par le paquet *sp* du logiciel *R*, cet ensemble des données comprenant des métaux lourds mesurés dans le sol supérieur d'une plaine inondable le long de la Meuse (c'est une rivière dans la Hollande), avec une poignée de covariables (les axes). Le processus régissant la distribution des métaux lourds semble que les sédiments pollués sont transportés par la rivière et se déposent principalement près de la rive de la rivière et des zones à basse altitude. Ce document affiche un analyse géostatistique de cet ensemble de données. L'ensemble de données a été introduit par Burrough et McDonnell, 1998

On présente les fonctionnalités du package *gstat* du logiciels *R* utilisé dans conjonction avec le package *sp*.

Le package *gstat* fournit un large éventail de modélisation, de prédiction et de simulation géostatistiques univariables et multivariables fonctions, où le package *sp* fournit des classes et des méthodes à usage général pour définition, importation/exportation et visualisation de données spatiales.

Cette note technique montre comment effectuer le co-krigeage à l'aide du *gstat* package géostatistique de l'environnement *R* pour le calcul statistique et la visualisation. cela est traité dans de nombreux textes . Le principal l'objectif est de montrer diverses techniques *R* pour la manipulation de données, le calcul et graphiques; l'accent est mis sur le package géostatistique **gstat**, le **sp** package de données spatiales et le package de graphiques en treillis.

## 3.2 Packages géostatistiques R

Plusieurs Packages sont utilisés pour l'analyse géostatistique des données, on cite ici les packages suivants :

- Package *gstat* (Pebesma, 2004) est un package R qui fournit des fonctionnalités de base pour l'analyse géostatistique univariable et multivariable, y compris :
  - modélisation du variogramme, modélisation du variogramme résiduel et variogramme croisé modélisation à l'aide de l'ajustement de modèles paramétriques à des variogrammes d'échantillons
  - anisotropie géométrique spécifié pour chaque modèle de variogramme partiel
  - ajustement limité du maximum de vraisemblance des seuils partiels
  - Cartes de variogrammes et de variogrammes croisés dérive (co)krigeage simple, ordinaire, universel et externe(séquentielle) équivalents de (co)simulation gaussienne pour chacun des krigeages Variétés
  - indicateur (co)krigeage et indicateur séquentiel (co)simulation
  - krigeage dans un quartier local ou mondial bloc (co)krigeage ou simulation pour chacune des variétés, pour rectangulaire ou blocs irréguliers
- le package *geoR* et *geoRglm* (par Paulo Ribeiro et Ole Christensen) fournit le cadre géostatistique basé sur un modèle décrit dans Diggle et al. (1998), champs de package (Doug Nychka et autres) fournit une interpolation spline à plaques minces, des fonctions de covariance pour les coordonnées sphériques (données non projetées) et les routines d'échantillonnage spatial optimisation de la conception.
- *sp* :fournit des classes et des méthodes pour les données spatiales dans R.il permet afficher des fond de cartes d'inspecter une table attributaire ect.
- *kriging* :algorithme de krigeage ordinaire simple et hautement optimisé pour tracer géographiquement les données.

## 3.3 Variable cible et covariables

Dans cette section, nous nous intéressons par une variable cible, c'est-à-dire celle à cartographier. Lors du co-krigeage, nous devons également sélectionner une ou plusieurs co-variables pour aider à la cartographie des variables cibles.

Lors de la cartographie avec co-krigeage, il est possible d'utiliser le même ensemble de variables que la cible et les co-variables ; cependant, cela est généralement utilisé dans le cas d'une densité d'échantillonnage plus ou moins égale. Dans ces notes, nous considérons le cas du sous-échantillonnage d'une variable cible, en utilisant des échantillons supplémentaires de co-variables.

### 3.3.1 Variable cible

Nous sélectionnons le plomb (abréviation « Pb ») comme variable cible, c'est-à-dire celle que nous voulons projeter. Ce métal est un grave danger pour la santé humaine. Il peut être inhalé sous forme de poussière provenant du sol perturbé ou absorbé par les plantes et ingéré.

#### 3.3.1.1 Résumer des paramètres descriptives

TABLE 3.1 – Paramètres descriptive de la variable cible Pb

Min	Q <sub>1</sub>	Med	Q <sub>3</sub>	Mean	sd	Max
37	72.5	123.0	207.4	153.4	111.32	654

L'histogramme de cette base de données est donné dans la figure :

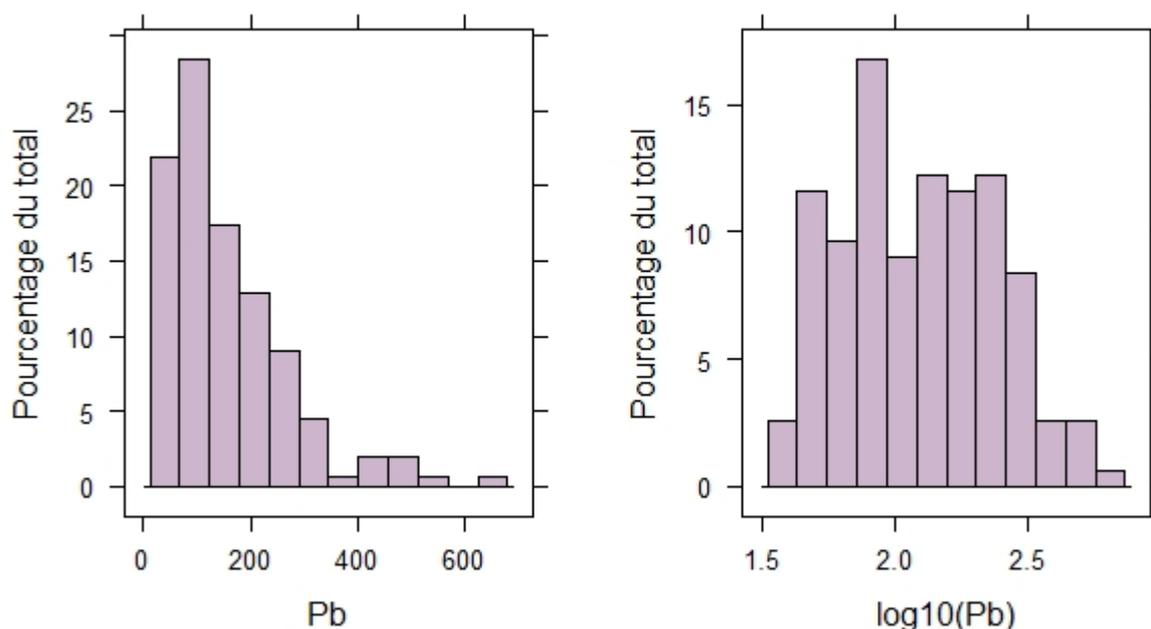


FIGURE 3.1 – Histogramme de la variable cible Pb et  $\log_{10}(Pb)$

### 3.3.2 Sélection des covariables

Une co-variable à utiliser dans le co-krigeage est corrélée à la variable cible à la fois dans l'entité et dans l'espace géographique, et a généralement été plus densément échantillonnée que la variable cible.

Les candidats aux covariables doivent avoir :

1. Une corrélation caractéristique-espace avec la variable cible;
2. Une structure spatiale (c'est-à-dire être modélisée comme une variable régionale);
3. une covariance spatiale avec la variable cible.

Il existe deux manières principales de sélectionner une covariable :

1. théoriquement, à partir de la connaissance du processus spatial à l'origine de la (co-)distribution spatiale observée;
2. empiriquement, en examinant les corrélations caractéristiques-espace (scatter-plots) puis la covariance spatiale (cross-corrélogrammes ou cross variogrammes).

Dans cet étude, nous comparerons deux covariables possibles basées sur la théorie, qui s'avèrent avoir des forces différentes de relation avec la variable cible.

Tout d'abord, nous sélectionnons **la teneur en matière organique** (abréviation « **OM** »). Il est moins cher que le métal lourd à mesurer en laboratoire et peut lier les métaux lourds dans le sol; il peut donc y avoir une corrélation positive dans l'espace des caractéristiques entre cette covariable et la variable cible. Cependant, il est principalement formé en place dans le sol et n'est pas transporté avec les eaux de crue, comme c'est le cas pour le Pb, il n'est donc pas clair s'il s'avérera une covariance spatiale avec le Pb dans la zone d'étude.

Deuxièmement, nous sélectionnons **la teneur en zinc** (abréviation « **Zn** »). De nombreux processus industriels qui entraînent une pollution par le plomb produisent également du zinc, et il est transporté de la même manière, à la fois sous forme de poussière et de sédiments. Il est généralement mesuré avec le plomb sans coût supplémentaire.

### 3.3.2.1 Résumer des paramètres descriptives des deux covariables

TABLE 3.2 – Paramètres descriptive des covariables : OM et Zn

Matière	Min	Q <sub>1</sub>	Med	Q <sub>3</sub>	Mean	sd	Max
OM	1	5.30	6.90	9.00	7.478	3.42	17.0
Zn	113	198.0	326.0	674.5	469.7	367.0738	1839.0

L'histogramme des deux covariables avec logarithme décimal sont présentés dans la figure suivante :

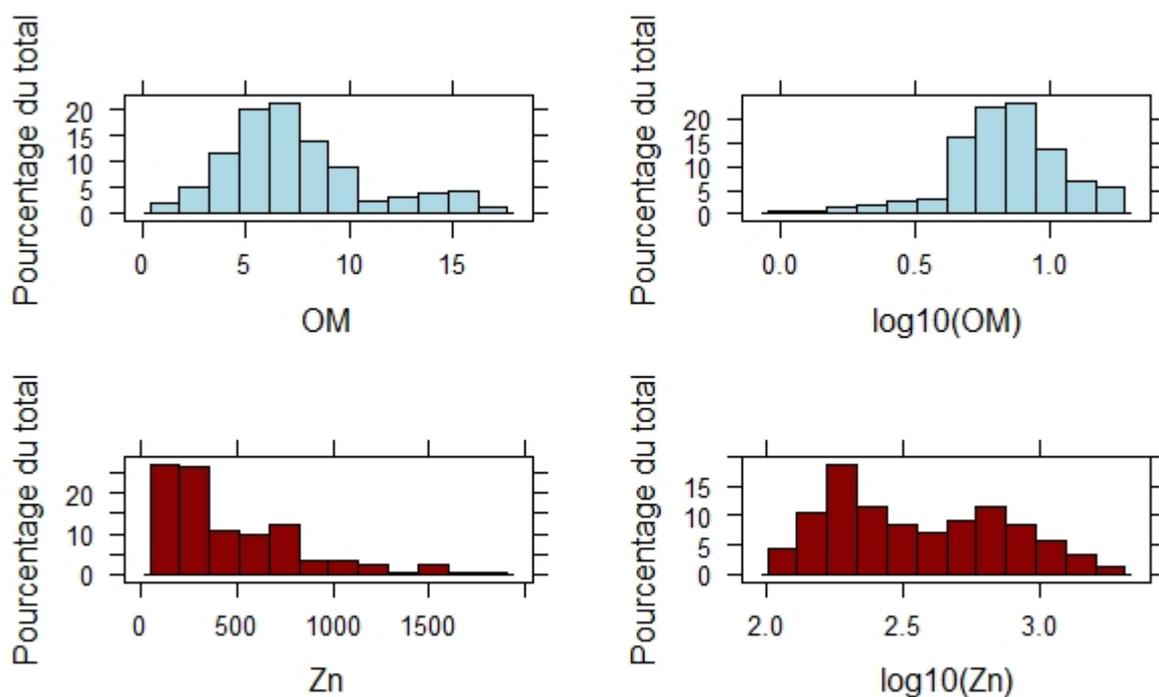


FIGURE 3.2 – Histogramme des deux covariables et leurs  $\log$  decimal

### 3.3.3 Modélisation de la structure spatiale de la variable cible

Commençons maintenant l'analyse spatiale. La première étape consiste à convertir les ensembles de données en jeux de données explicitement spatiaux, à l'aide du package `sp`.

Comme chargé à partir du package `sp`, l'exemple de données meuse est un fichier `data.frame`.

Notez que les coordonnées de chaque point sont répertoriées sous forme de champs. Cependant, `sp` a défini certaines classes pour rendre explicite la nature spatiale des données.

Il fournit également une méthode de coordonnées pour définir les coordonnées spatiales et ainsi créer des données spatiales explicites.

La concentration de la matière de bp dans la rivière Meuse est représentée dans la figure suivante :

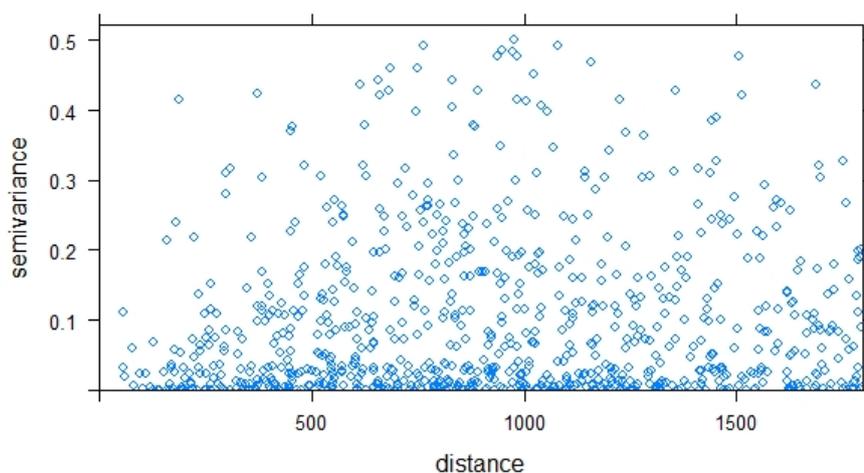


FIGURE 3.3 – La concentration de la matière dans la rivière

### 3.3.4 Le variogramme

Nous examinons d'abord le nuage de variogramme pour voir toutes les paires de points, puis on calcule le variogramme expérimental. Ensuite, nous spécifions un modèle de variogramme et ses paramètres initiaux (déterminés par l'œil regardant le variogramme expérimental)

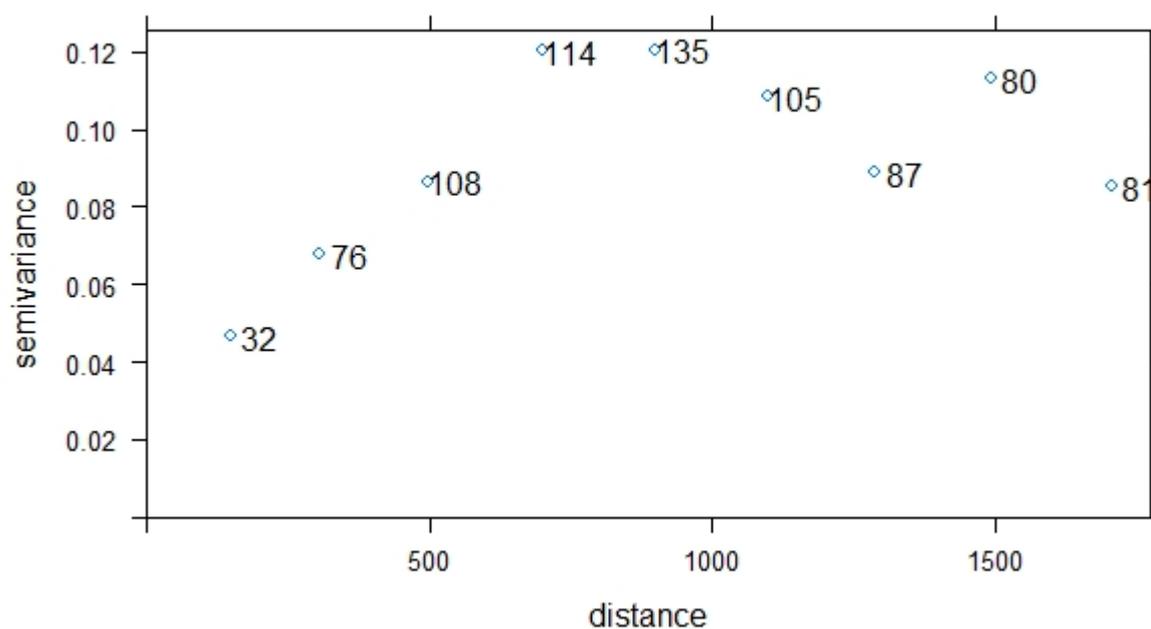


FIGURE 3.4 – Variogramme expérimentale des données de la variable cible

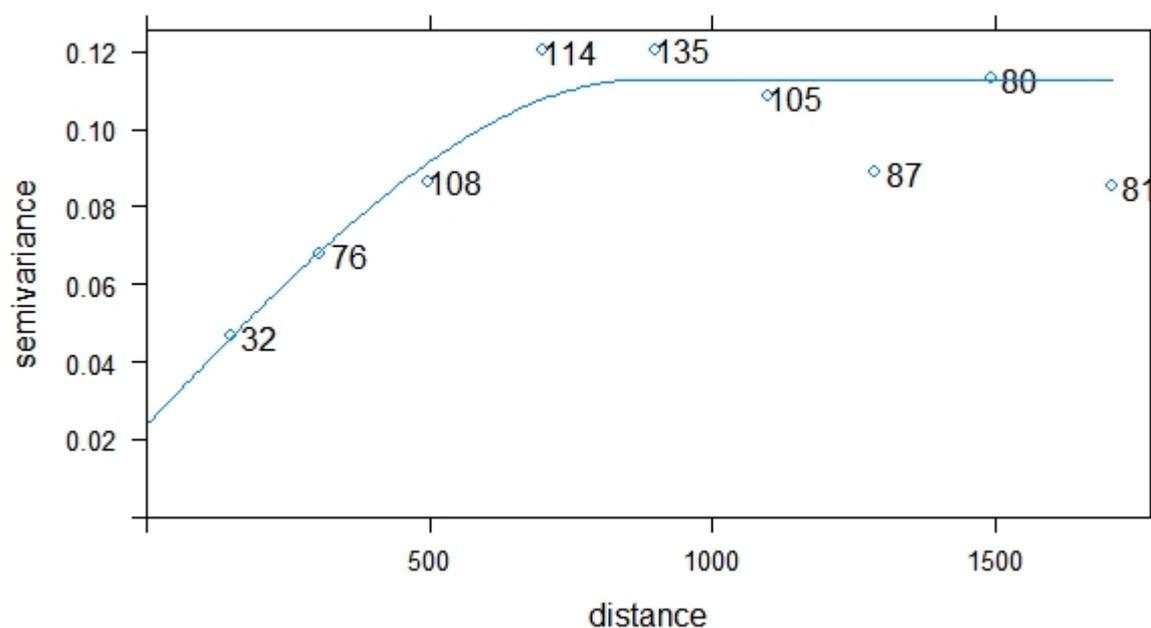


FIGURE 3.5 – Ajustement du variogramme

La méthode du variogramme a des arguments facultatifs pour spécifier la distance maximale, les largeurs de compartiment, les directions, les tolérances directionnelles et si l'estimateur Cressie robuste doit être utilisé. Dans ce cas, nous utilisons l'estimateur par défaut mais spécifions les largeurs de bac et la coupure.

## 3.4 Interpolation par krigeage ordinaire de la variable cible

### 3.4.1 Prédiction de la variable cible

Premièrement, nous prédisons Pb sans les covariables, à partir du petit sous-ensemble. Cela simule la situation où nous avons un petit échantillon de la variable cible coûteuse à mesurer.

Le résumé statistique de la prédiction de la variable cible est donné dans le tableau suivant

TABLE 3.3 – Caractéristiques des prédiction par le krigeage

Min	Q <sub>1</sub>	Med	Q <sub>3</sub>	Mean	Max
1.707	1.897	2.027	2.231	2.060	2.530

Les résultats de prédiction de krigeage de la variable cible sont projetées dans la

figure suivante :

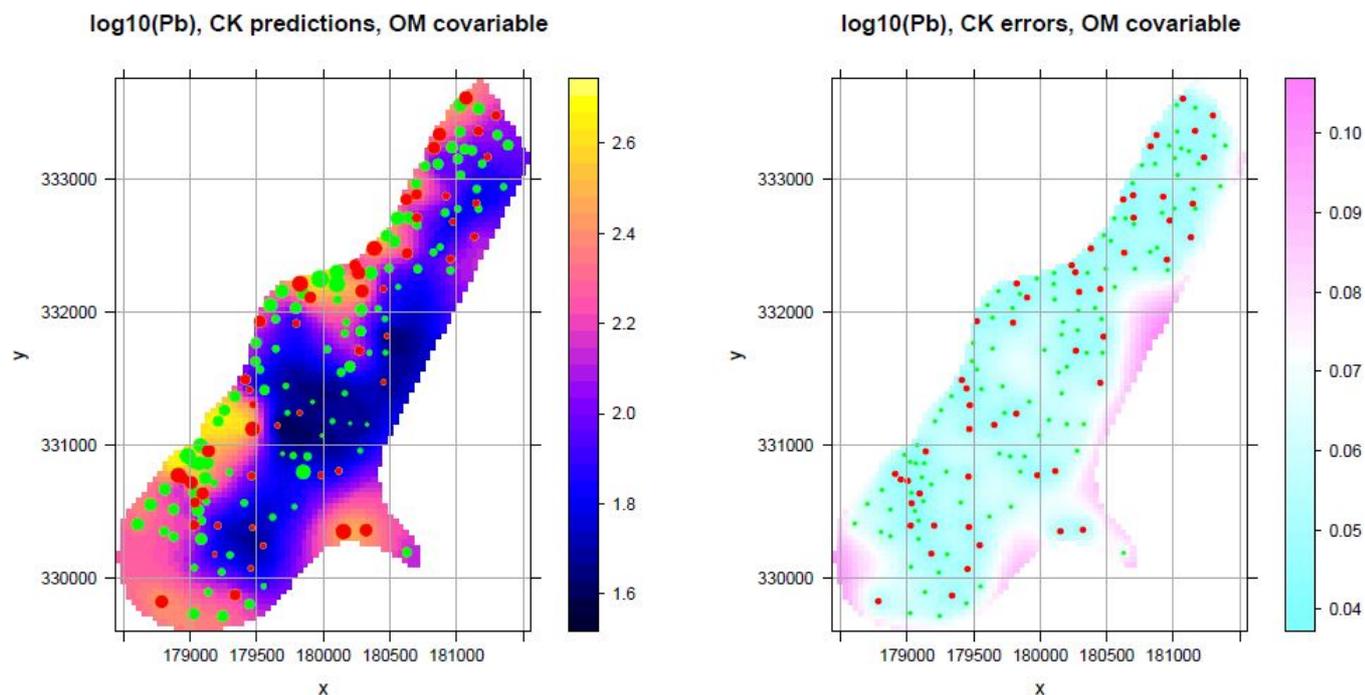


FIGURE 3.6 – Prédications et erreur OK,  $\log_{10}(\text{Pb})$ ; points de sous-échantillon rouges, points supplémentaires verts

### 3.4.2 Evaluation des erreurs

La validation des résultats de prédiction sont résumés dans le tableau suivant

TABLE 3.4 – Evaluation de l'erreur de prédiction par krigeage

Min	Q <sub>1</sub>	Med	Q <sub>3</sub>	Mean	Max
-0.6630	-0.0647	0.0215	0.127	0.0250	0.3720

Le résumé des erreurs statistique sont présentées dans la tableau suivant

TABLE 3.5 – Paramètres d'erreurs de la prédiction

Mean Error	RMSE	Med Error
0.0250	1.4844	2.06818

### 3.4.3 Validation

Les résultats de la validation croisée sont présentées dans le tableau suivant

TABLE 3.6 – Résultats de la validation d’interpolation par le krigeage

Mean Error	RMSE	Med Error
0.00367	0.2367	0.8848

Le graphe des résultats de validation sur une grille régulière

OK validation errors at undersampled points, log<sub>10</sub>(Pb)

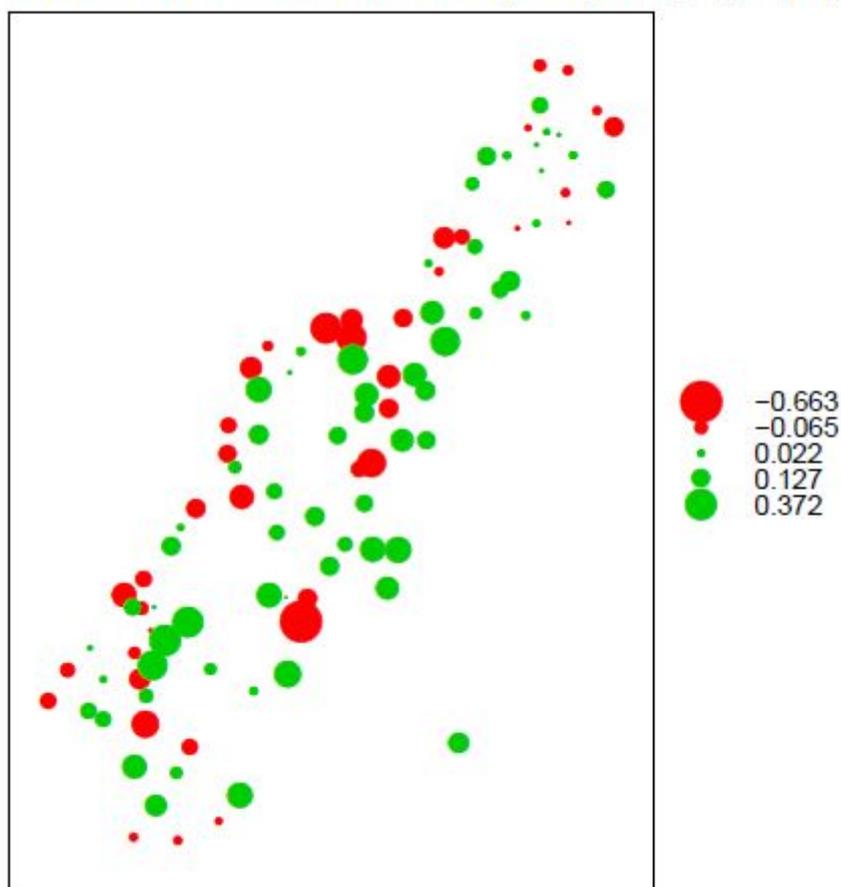


FIGURE 3.7 – Erreur de validation de krigeage ordinaire

### 3.4.4 Évaluation de l’erreur de KO

Pour évaluer la gravité des erreurs, nous devons comparer les valeurs de Pb d’origine avec les valeurs de Pb prédites, et non les valeurs transformées en log (bien que celles-ci aient été utilisées pour l’interpolation).

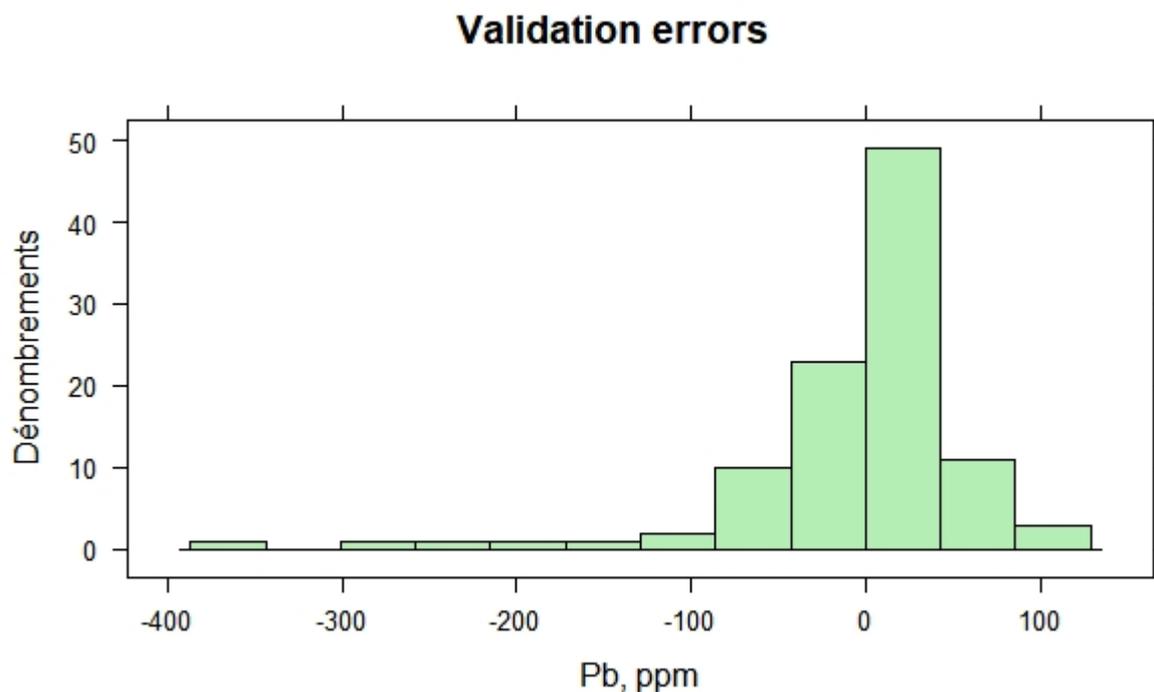


FIGURE 3.8 – Histogramme de l'erreur de validation du KO

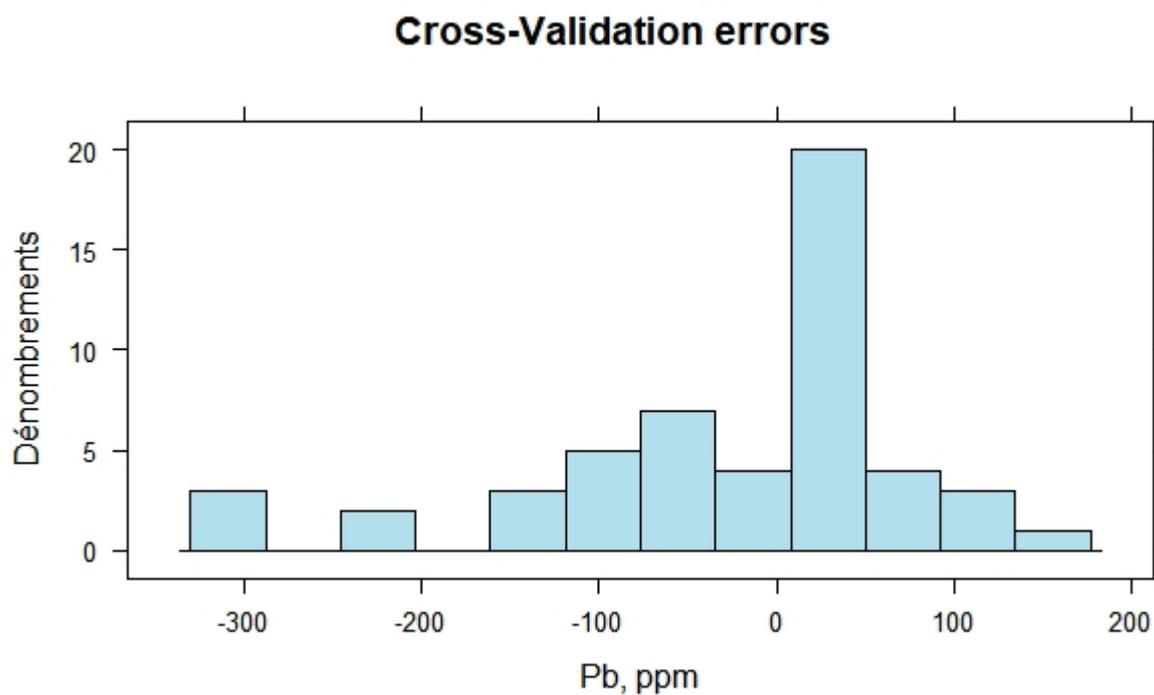


FIGURE 3.9 – Histogramme de l'erreur de la validation croisée du KO

### 3.5 Modélisation des covariables

Le nuage de points entre la covariable OM et la variable cible est présentée dans la figure suivante

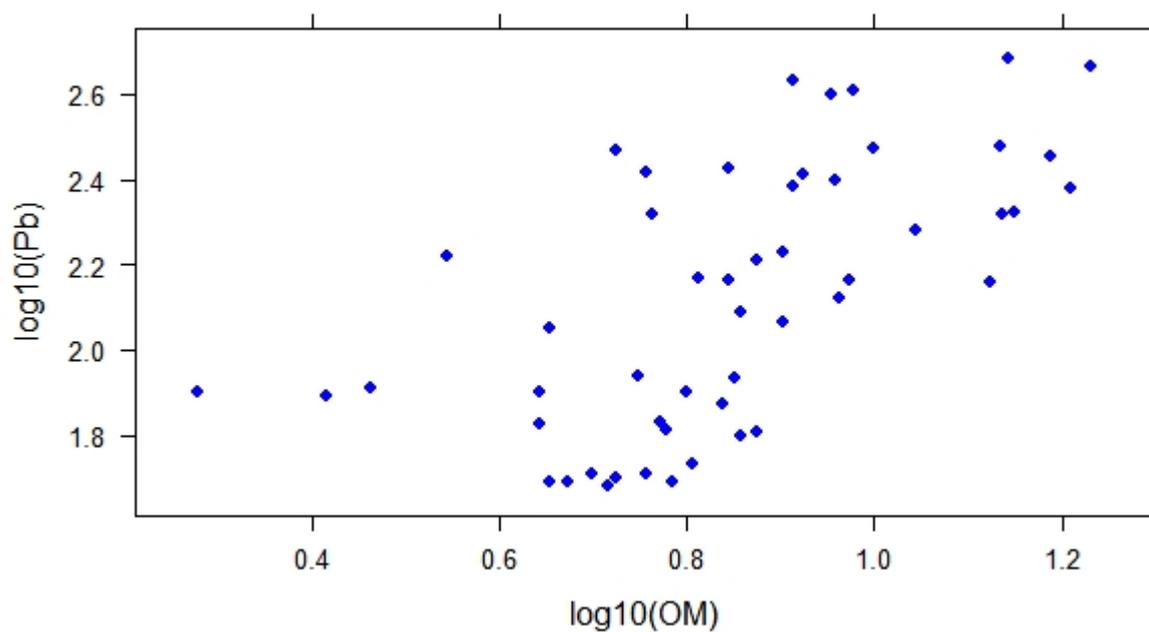


FIGURE 3.10 – Nuage des points, Pb et OM

La corrélation entre les deux variables est  $r = 0.61277$ .

Le variogramme de la covariable est tracé et ajusté, voir la figure

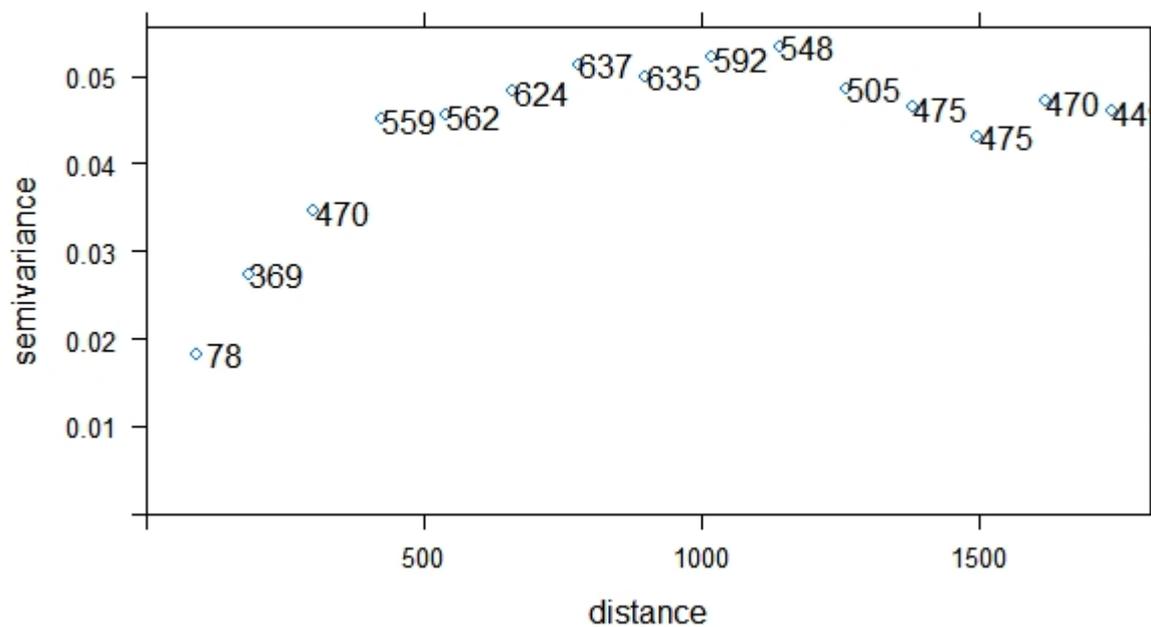


FIGURE 3.11 – Le variogramme expérimentale de la covariable OM

### 3.5.1 Construction des données pour modéliser la co-régionalisation

Le graphe des variogramme de la variable cible et lea covariable OM sont présentés dans la figure suivante

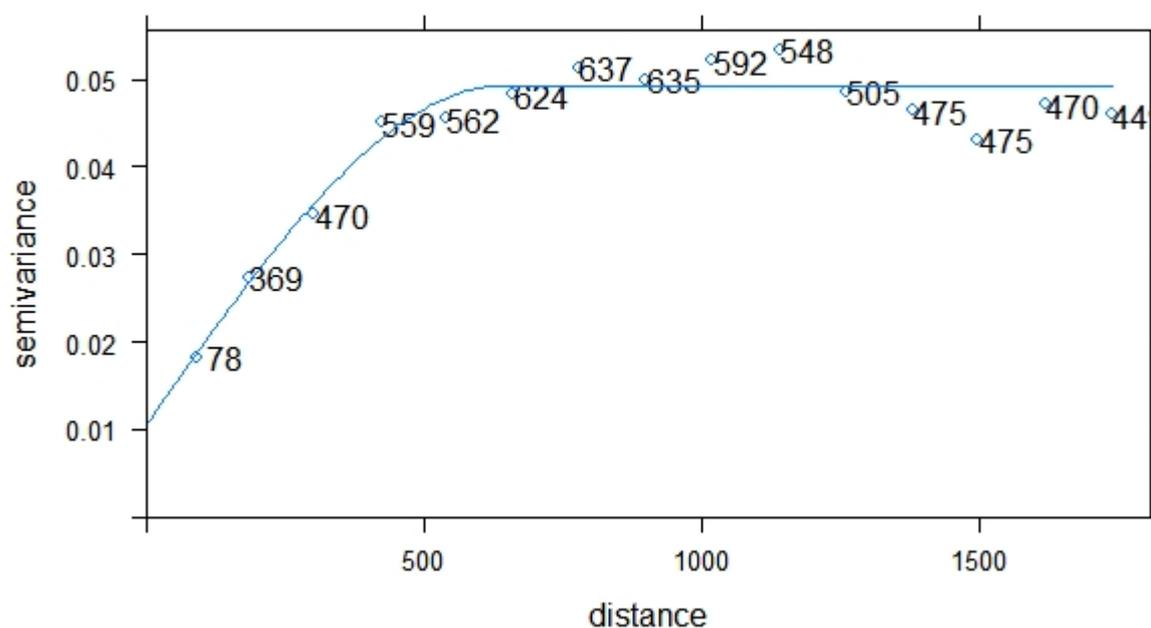


FIGURE 3.12 – Le graphe d’ajustement du variogramme de la covariable OM

### 3.5.2 Comparaison des modèles régionale et co-régionale

## 3.6 Co-krigeage avec une seule co-variable

Maintenant nous pouvons utiliser les échantillons supplémentaires d’OM pour améliorer (nous l’espérons!) les prédictions de la variable cible.

Les résultats de prédictions avec les erreurs de prédictions sont présentées dans le tableau suivant :

Matière	Min	Q <sub>1</sub>	Med	Q <sub>3</sub>	Mean	Max
Préd	1.624	1.835	2.016	2.254	2.043	2.611
Erreurs	0.0416	0.0532	0.0589	0.0679	0.0619	0.105

La projection des erreurs de validation et la validation croisé est données dans la figure

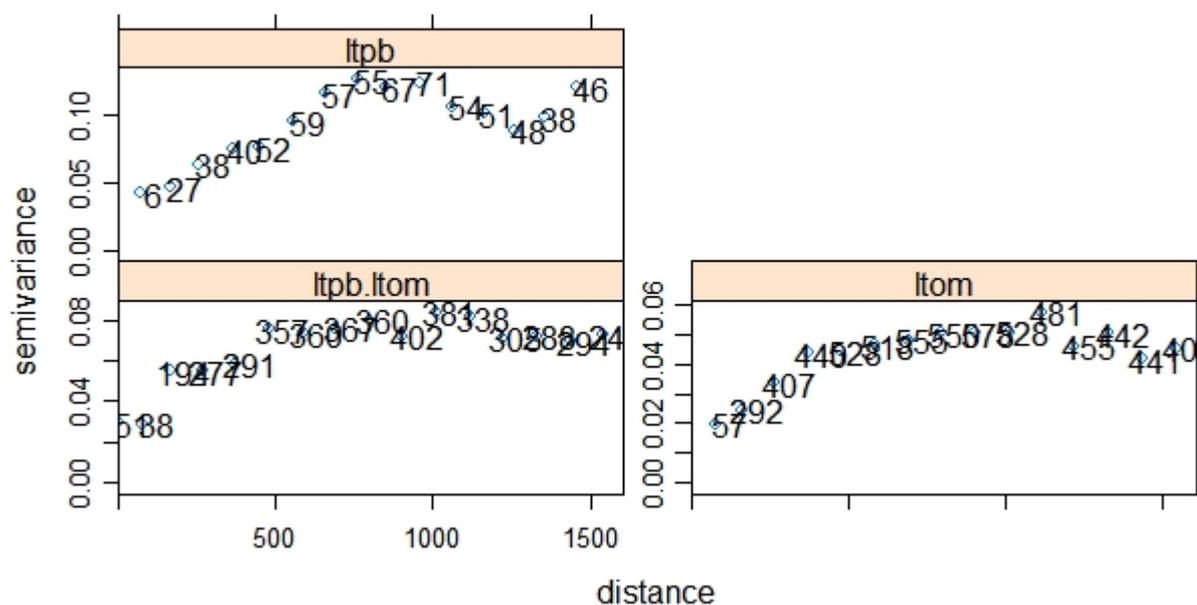


FIGURE 3.13 – Expérimental et ajustement du Variogrammes directs et croisés,  $\log_{10}(Pb)$  et  $\log_{10}(OM)$

Les résultats des validation croisé sont présentées dans le tableau suivant

TABLE 3.8 – Résultats des coefficients statistique des erreurs

ME	RMSE	MedE
0.00486	0.2037	4.9123

### 3.7 Comparaison du co-krigeage et du krigeage ordinaire

Nous avons maintenant utilisé à la fois OK et CK pour prédire la distribution du plomb dans la zone d'étude. Dans cette section, nous comparons les prédictions et leurs erreurs, et étudions les raisons de tout écart.

#### 3.7.1 Prédiction

Nous comparons d'abord les prédictions de la variable cible de OK et CK.

TABLE 3.9 – Résultats de comparaison entre KO et CK

	Min	Q <sub>1</sub>	Med	Q <sub>3</sub>	Moy	Max
Pred KO	1.707	1.897	2.027	2.231	2.060	2.53
Pred CK	1.624	1.835	2.016	2.254	2.043	2.611
Diff	-0.1991	-0.0777	-0.02929	0.03894	-0.01725	0.3133

Les prédictions CK sont en moyenne légèrement inférieures aux prédictions OK ; il existe des différences assez importantes, à la fois positives et négatives.

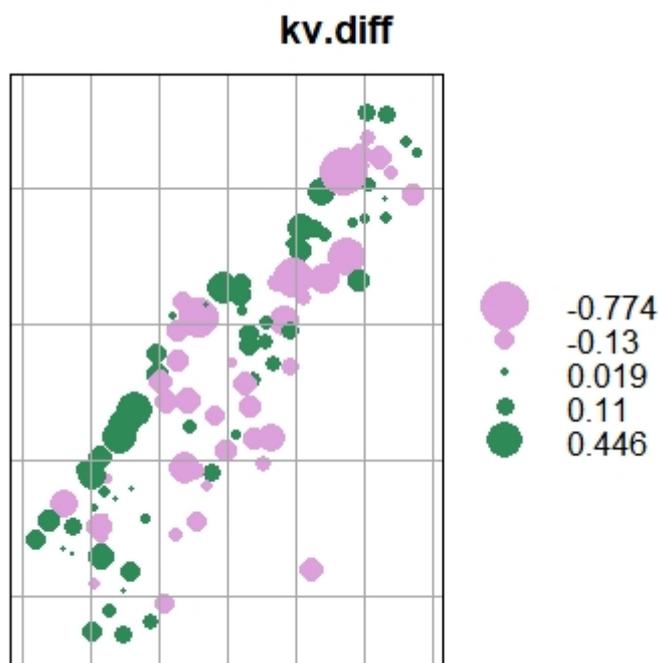
### 3.7.2 Prédiction de l'erreur

Deuxièmement, nous comparons les erreurs de prédiction pour la variable cible de OK et CK.

Un objectif majeur de l'utilisation du co-krigeage est de réduire les variances de prédiction.

Avec autant de nouvelles informations (103 points supplémentaires, plus les 52 points colocalisés) et un assez bon modèle de co-régionalisation, nous nous attendons à moins d'erreurs de prédiction.

FIGURE 3.14 – Projection des erreurs de CK



Les résultats des erreurs de prédiction entre KO et CK sont présentées dans le tableau suivant :

TABLE 3.10 – Comparaison des erreurs des résultats entre KO et CK

	Min	Q <sub>1</sub>	Med	Q <sub>3</sub>	Moy	Max
Pred KO	0.03569	0.05195	0.06078	0.07379	0.06381	0.1119
Pred CK	0.04166	0.05317	0.05891	0.06789	0.06190	0.10503
Diff	-0.02574	-0.00524	-0.00057	0.00246	-0.00191	0.00826

### 3.7.3 Validation

Erreurs de la validation croisé

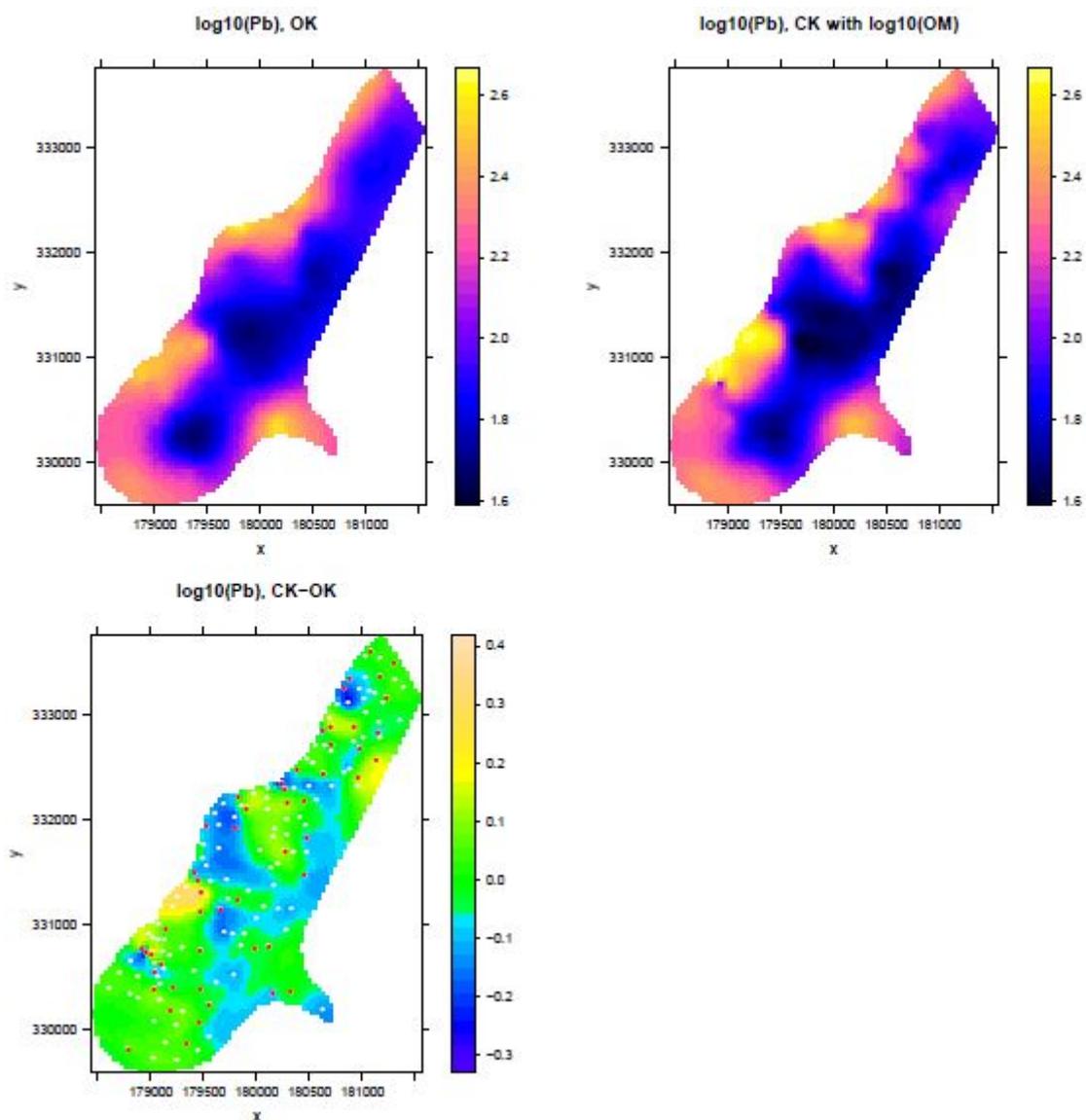


FIGURE 3.15 – Différence entre CK et OK, log10(Pb) en points de validation, co-variable OM

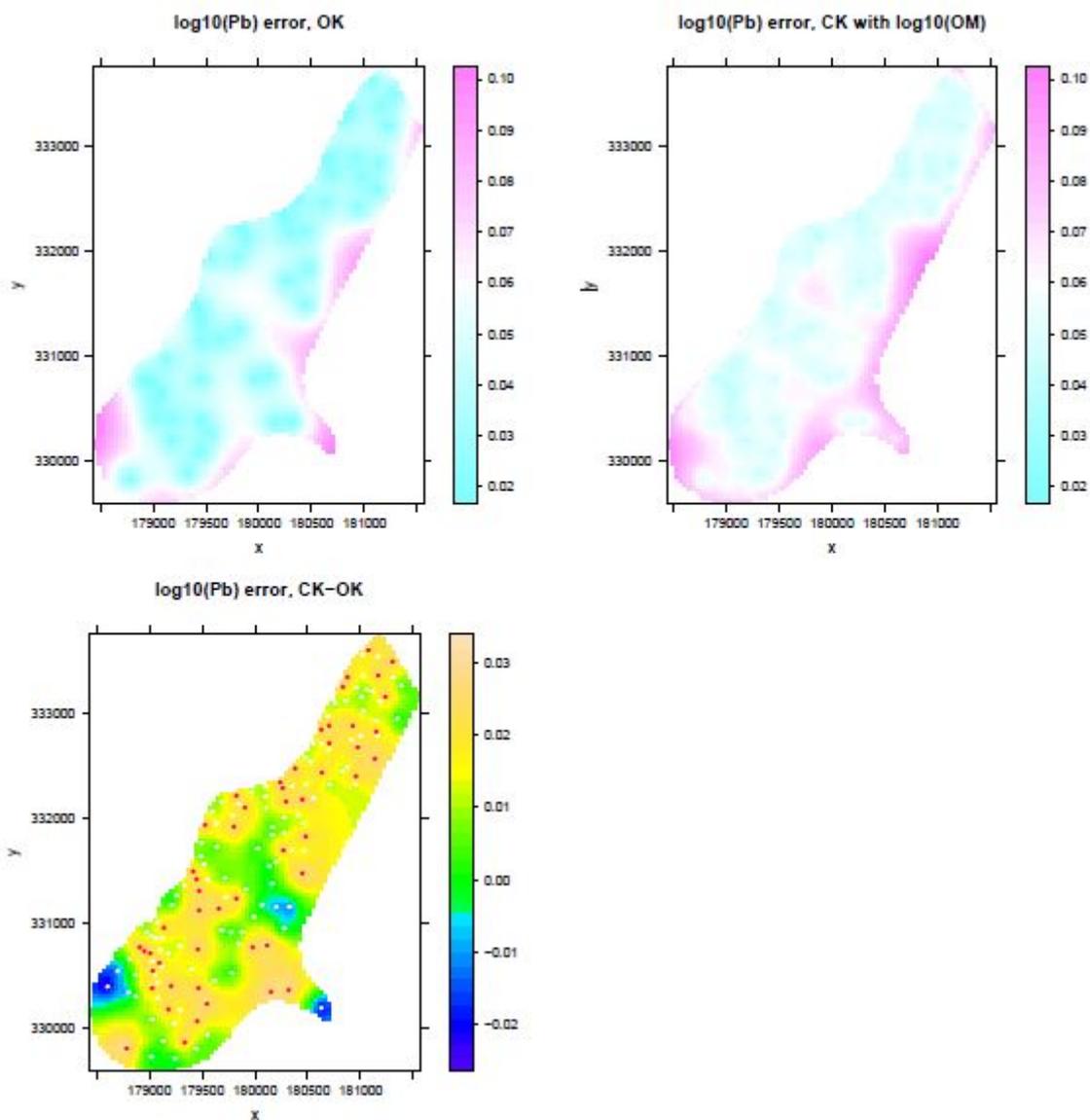


FIGURE 3.16 – Différence entre CK et OK des rrésidus du validation croisée  $\log_{10}(Pb)$

TABLE 3.11 – Comparaison entre CK et KO

	ME	RMSE	MinE	MaxE
KO	0.0249	0.1659	-0.663	0.372
CK avec Zn	<b>0.0043</b>	<b>0.0775</b>	-0.138	0.227

## 3.8 CoKrigage avec une autre covariable

### 3.8.1 Modélisation de la covariable

Le nuage de point de la variable cible(Bp) et la covariable est présentée dans la figure

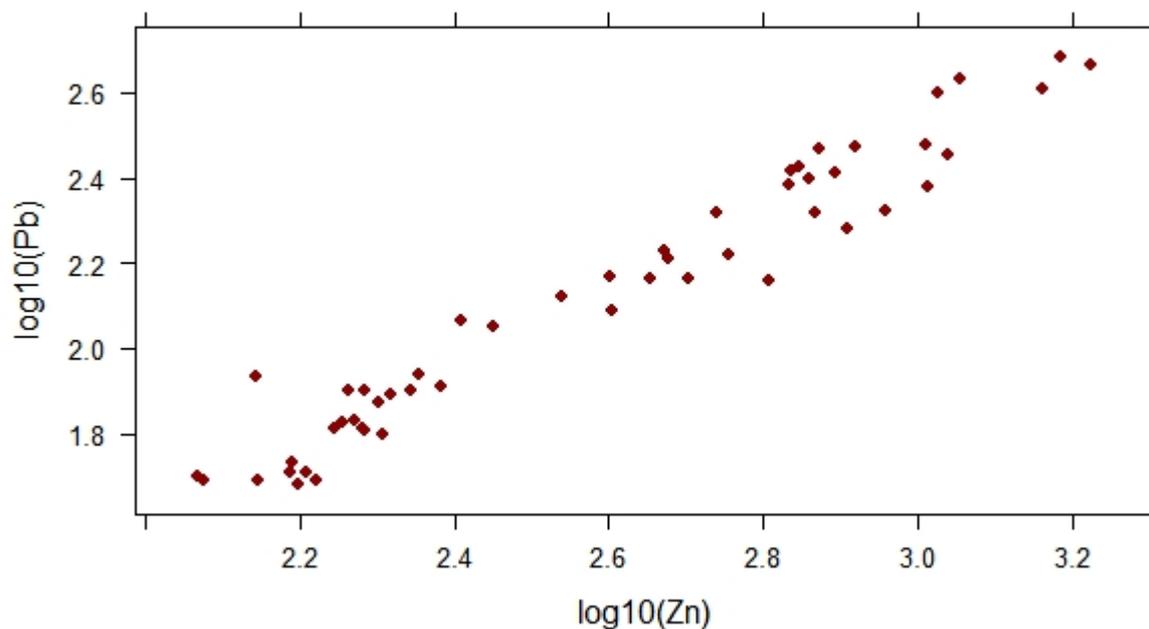


FIGURE 3.17 – Nuage de point de la variable cible Pb avec la covariable Zn

le coefficient de corrélation est très fort qui égale à  $0.9738$ .

Le variogramme est ajusté et présenté dans la figure suivante

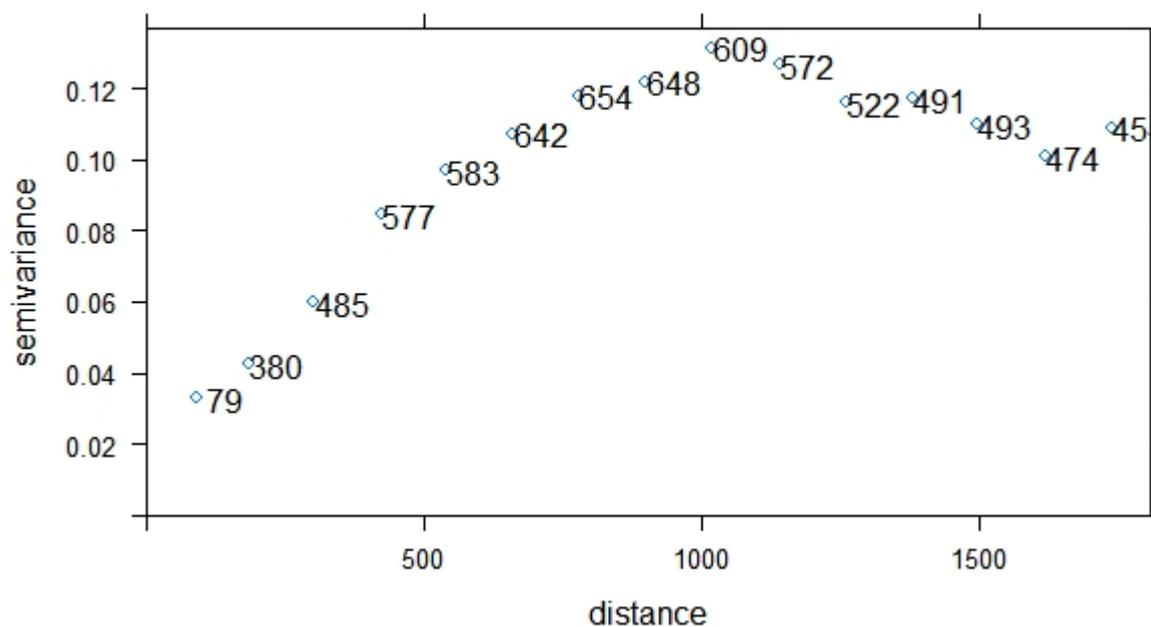


FIGURE 3.18 – Variogramme expérimentale de la covariable Zn

### 3.8.2 Ajustement linéaire de la covariable

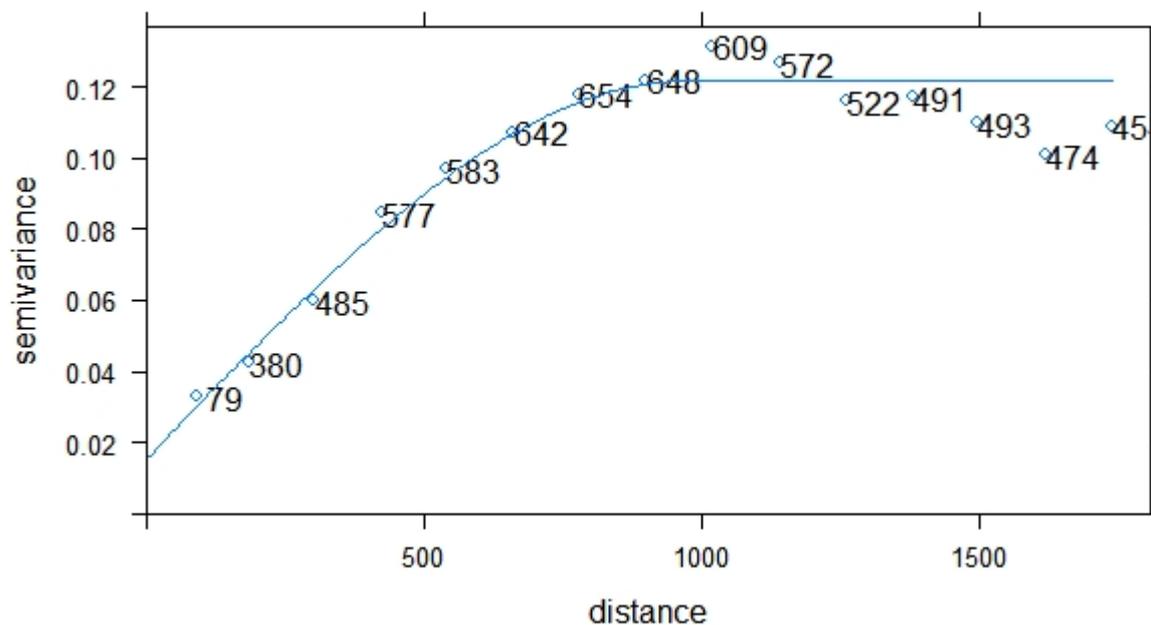


FIGURE 3.19 – Ajustement du variogramme de la covariable Zn

### 3.8.3 CoKrigage avec une covariable

Résultats de prédiction avec le covariable Zn avec l'erreur sont présentées dans le tableau suivant

TABLE 3.12 – Résultats de prédiction du CK avec la covariable Zn

	Min	Q <sub>1</sub>	Med	Q <sub>3</sub>	Moy	Max
Préd	1.663	1.843	1.970	2.201	2.022	2.690
Erreur	0.0319	0.0397	0.0436	0.0511	0.0471	0.0950

### 3.8.4 Validation et validation croisée

Le coKrigage et l'erreur de prédiction sont présentées dans la figure

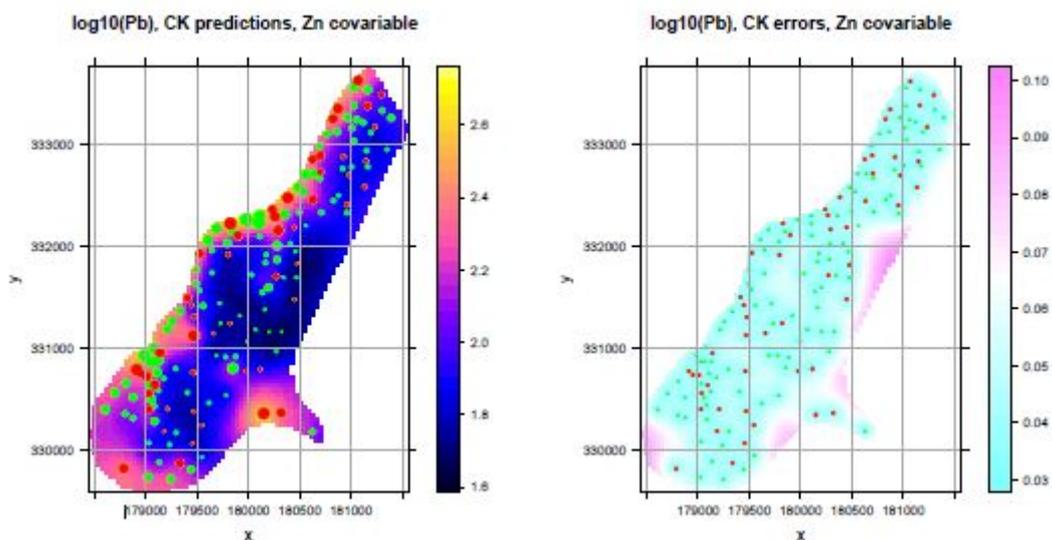


FIGURE 3.20 – CK et prédiction, erreur,  $\log_{10}(Pb)$  et covariable Zn

La validation croisé de Cokrigage de  $\log_{10}(Pb)$  avec la covariable Zn est présentée dans la figure

### CK - OK predictions, validation points, co-variable Zn



FIGURE 3.21 – Différence entre CK et OK prédictions,  $\log_{10}(Pb)$  au points de validation, co-variable Zn

Nous avons utilisé les points d'échantillon supplémentaires pour évaluer les performances de OK et CK avec Zn comme co-variable. Nous comparons ici ces deux

validations; c'est la même chose que nous avons fait pour CK avec OM comme covariable.

TABLE 3.13 – Comparaison entre KO et CK avec la covariable Zn

	ME	RMSE	MinE	MaxE
KO	0.0249	0.1659	-0.663	0.372
CK avec Zn	<b>0.0043</b>	<b>0.0775</b>	-0.138	0.227

La validation CK est bien supérieure dans toutes les mesures; ceci contraste avec le résultat pour CK avec OM comme co-variable. La figure CKZN montre où se situent les différences de prédiction. Il y a beaucoup de changements importants dans les prédictions entre les deux validations, mais pas aussi extrêmes que lorsque OM était utilisé comme co-variable.

### CK – OK cross-validation residuals, co-variable Zn

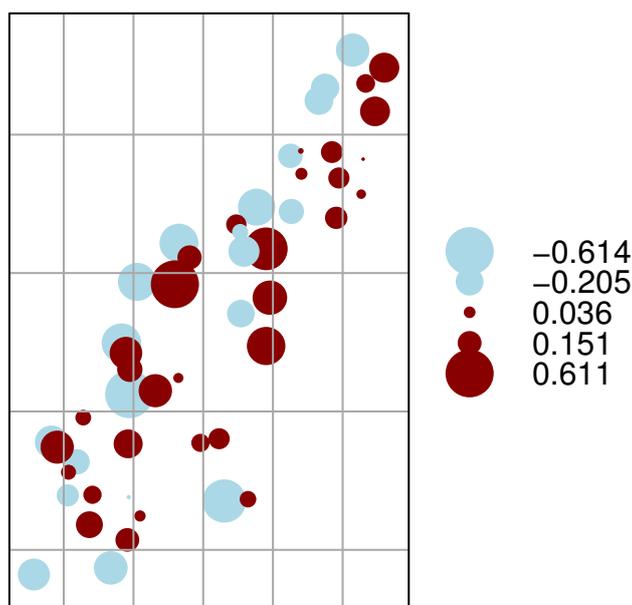


FIGURE 3.22 – Différence entre CK and OK validation croisé résiduels,  $\log_{10}(\text{Pb})$ , co-variable Zn

## 3.9 Comparaison des trois interpolations

### 3.9.1 Prédiction et erreur de prédiction

La figure suivante (à gauche) montre de grandes différences dans les prédictions. Le modèle le plus complexe est pour CK avec la co-variable Zn; cela respecte les points

de l'échantillon complet bien mieux que l'autre covariable, car la corrélation entité-espace et la co-régionalisation spatiale sont beaucoup plus fortes pour Zn que pour OM.

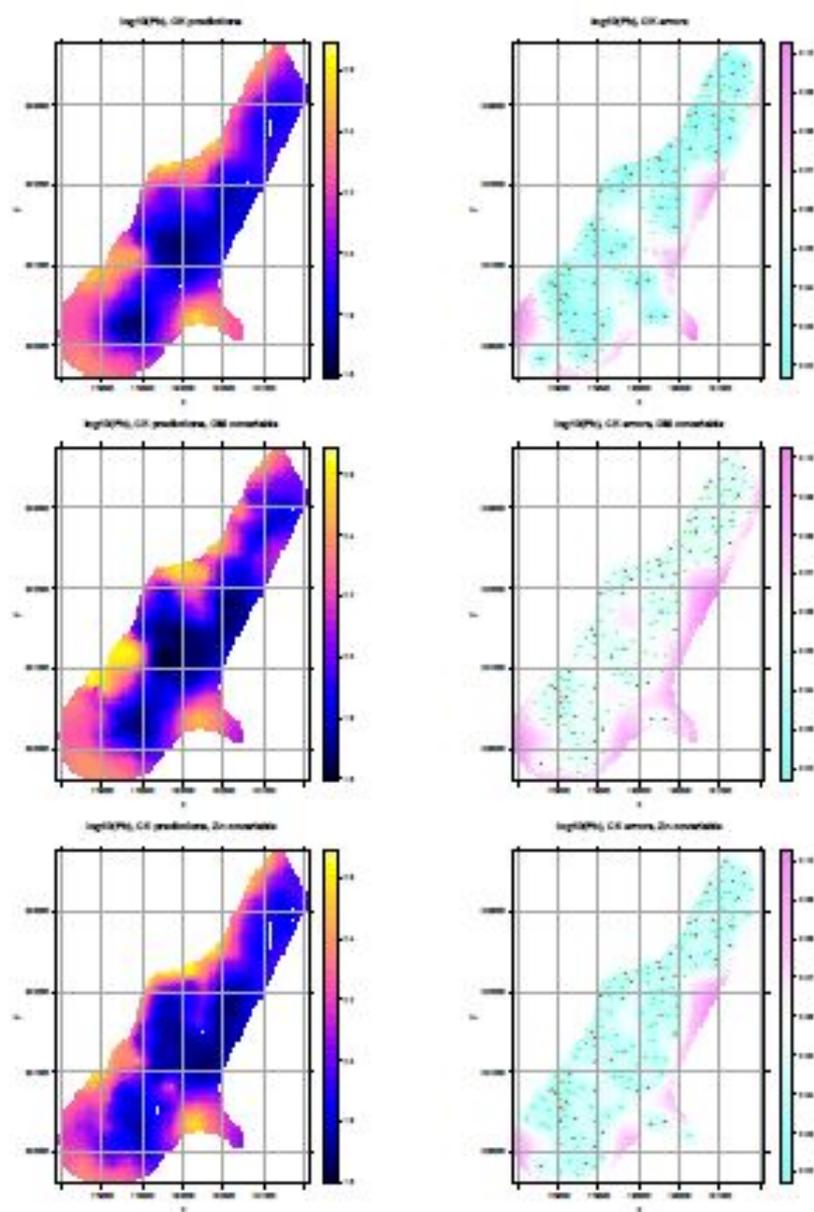


FIGURE 3.23 – Prédications de krigeage et leurs erreurs comparées,  $\log_{10}(\text{Pb})$

### 3.10 Validation

TABLE 3.14 – Comparaison entre les trois interpolation de la variable cible Pb

	ME	RMSE	MinE	MaxE
KO	0.0249	0.1659	-0.663	0.372
CK avec OM	0.0147	0.2226	-0.812	0.555
CK avec Zn	<b>0.0043</b>	<b>0.0775</b>	-0.138	0.227

#### 3.10.1 Validation croisée

La figure suivante (à droite) montre des graphiques à bulles des erreurs de validation croisée; c'est un résumé des Figure 3.23 (à droite), mais avec une seule plage de taille de symbole pour les trois chiffres. La supériorité de CK avec Zn comme covariable est claire.

TABLE 3.15 – Comparaison entre Ko et CK avec les deux covariables

	ME	RMSE	MDSR
KO	0.0037	0.2369	0.088
CK avec OM	0.0059	0.2093	12.62
CK avec Zn	<b>0.0035</b>	<b>0.0807</b>	0.032

CK avec Zn comme co-variable fonctionne bien mieux sur chaque mesure que CK avec OM comme co-variable. CK avec Zn est meilleur que OK sauf qu'il sous-estime l'erreur de prédiction aux points de validation croisée.

## CONCLUSION

L'interpolation est une procédure d'estimation de la valeur d'une variable dans un endroit déterminé à partir de la connaissance des valeurs d'un certain nombre de points pour permettre de les calculer en chaque point de la zone.

Dans ce mémoire, nous illustrons une application courante sur deux méthodes pour la résolution des problèmes d'interpolation spatiale ( **krigage et cokrigage** ) avec l'utilisation du logiciel **R** : l'interpolation des données de la Meuse, notre étude conclut que la méthode de cokrigage est mieux que kigeage lorsque la covariable est bien corrélé avec la variable cible, donc le choix de la variable d'influence est important pour l'interpolation par le cokrigage.

## BIBLIOGRAPHIE

- [Sophie Baillargeon. (2005).] Le krigeage : revue de la théorie et application à l'interpolation spatiale de données de précipitations. Master's thesis, Université de Laval.
- [Matheron, G. (1970).] La théorie des variables régionalisées, et ses applications. Les cahiers du Centre de morphologie mathématique de Fontainebleau, Fascicule 5. Ecole des Mines de Paris, Fontainebleau.
- [Matheron, G. (1965)] Les variable régionalisées et leur estimation. masson, Paris.
- [Matheron, G. (1969).] Le krigeage universel. Les cahiers du Centre de morphologie mathématique de Fontainebleau, Fascicule 1. Ecole de Mines de Paris, Fontainebleau
- [Matheron, G. (1970).] La théorie des variables régionalisées, et ses applications. Les cahiers du Centre de morphologie mathématique de Fontainebleau, Fascicule 5. Ecole des Mines de Paris, Fontainebleau. The
- [Matheron, G. (1973).] The intrinsic random functions and their applications. *Advances in Applied Probability*, 5 :439-468.
- [Matheron, G. (1976).] A simple substitute for conditional expectation : The disjunctive kriging. Dans Guarascio, M., David, M. et Huijbregts, C., éditeurs : *Advanced Geostatistics in the Mining Industry*, pages 221-236. Reidel, Dordrecht.
- [Chauvet. P. (1999).] Aide mémoire de la géostatistique linéaire. Cahiers de Géostatistique, Fascicule 2. Ecole Nationale Su-

- périeur des Mines de Paris, Centre de Géostatistique, Fontainebleau
- [Mitas, L. et Mitsova, H. (1999).] Spatial interpolation. Geographical information systems : principles, techniques, management and applications, 1(2).
- [Hauchecorne, B. (2007).] Les contre-exemples en mathématiques. Ellipses Paris.
- [Krige, D. G. (1951).] A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy, 52(6) :119-139.
- [Arnaud, M. et Emery, X. (2000).] Estimation et interpolation spatiale. Hermes Science Publications, Paris.
- [Myers, D. (1994).] Spatial interpolation : an overview. Geoderma, 62 :17-28.
- [Ripley, B. D. (1981).] Spatial statistics. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.
- [Christakos, G. (1984).] On the problem of permissible covariance and variogram models. Water Resources Research, 20(2) :251-265.
- [Jacques RIVOIRARD]
- [Hans WACKERNAGEL]
- [Ripley, B. D. (1981).] Spatial statistics. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.
- [Matheron, G. (1969).] Le krigeage universel. Les cahiers du Centre de morphologie mathématique de Fontainebleau, Fascicule 1. École de Mines de Paris, Fontainebleau.
- [Cressie, N. A. C. (1993).] Statistics for spatial data. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Inc., New York. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

- [Goovaerts, P. (1997).]                      Geostatistics for Natural Resources Evaluation. Applied Geostatistics Series. Oxford University Press, New York.