

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
**Université Ibn Khaldoun de Tiaret**  
**Faculté des Sciences Appliquées**  
**Département de Génie civil**



***Polycopié***  
**GÉOSTATISTIQUE**  
*(Cours - Exercices)*

Elaboré par:  
Dr. KLOUCHE DJEDID iBrahim

Experts:  
Pr. KRIM Abdallah  
Pr. CHIKH Abdelbaki

*Année universitaire : 2023-2024*

## Table des matières

<i>Introduction</i> .....	2
<i>Préambule</i> .....	2
<i>Historique</i> .....	3
<i>Chapitre 1 : Rappels sur les statistiques</i> .....	4
1.1. <i>Typologie des variables</i> .....	4
1.2. <i>Variables aléatoires</i> .....	4
1.3. <i>Loi de probabilité, Fonction de répartition, Densité de probabilité</i> .....	5
1.4. <i>Moments d'une variable aléatoire X</i> .....	6
1.5. <i>Deux variables quantitatives</i> .....	6
<i>Exercices</i> : .....	8
<i>Chapitre 2 : Bases théoriques de la géostatistique</i> .....	9
2.1. <i>Définition</i> .....	9
2.2. <i>La théorie des variables régionalisées</i> .....	9
2.3. <i>Hypothèse de stationnarité et intrinsèque</i> .....	10
<i>Chapitre 3 : Analyse du variogramme</i> .....	13
3.1. <i>Le variogramme</i> .....	13
3.2. <i>Modèles</i> .....	15
3.3. <i>Propriétés du variogramme</i> .....	15
3.4. <i>Variance d'estimation et de dispersion</i> .....	16
<i>Exercices</i> : .....	19
<i>Chapitre 4 : Théorie du krigeage</i> .....	21
4.1. <i>Le Krigeage</i> .....	21
4.2. <i>Type de Krigeage</i> .....	21
4.2.1. <i>Krigeage stationnaire à moyenne inconnue (krigeage linéaire ordinaire)</i> .....	21
4.2.2. <i>Krigeage stationnaire à moyenne connue (krigeage linéaire simple)</i> .....	22
4.3. <i>Propriétés du krigeage</i> .....	22
4.4. <i>Cokrigeage</i> .....	23
4.4.1. <i>Cokrigeage ordinaire</i> .....	23
4.4.2. <i>Cokrigeage simple</i> .....	24
4.4.3. <i>Cokrigeage sous format matricielle</i> .....	24
<i>Exercices</i> : .....	26
<i>Conclusion</i> .....	28
<i>Références bibliographique</i> : .....	29
 Annexe .....	 31

## INTRODUCTION - Préface

La géostatistique se base sur la continuité spatiale d'un phénomène, elle permet une modélisation des phénomènes naturels surtout pour les sciences de la terre. Il s'agit de tout paramètre reparti dans l'espace dont la connaissance des fonctions de son répartition permettra une meilleure estimation / simulation. Précisément, les étudiants, avec sa dimensionnalité économique est de plus en plus, ont besoin des notions de la géostatistique qui prend en compte la variabilité et les incertitudes des données lors des différentes phases d'une campagne de reconnaissance géotechnique et pour décrire la dispersion des données et leur structuration spatiale.

Les mesures des paramètres géotechniques présentent une variation dans l'espace ce qui marque une incertitude sur leur représentativité dans les calculs géotechniques. Si une valeur est déterministe sur un point, elle sera inconnue sur le volume donné (Exemple : un site adapté), ce qui nécessite un recours aux méthodes statistiques pour son estimation : statistiques classiques (recherche d'ensembles homogènes, analyse de corrélation) et des méthodes regroupées sous le nom de géostatistiques.

Ce support cours avec des études de cas et exercices, est destiné aux étudiants Master en première année Géotechnique a pour objectif d'initier d'abord les étudiants par le contenu de cette matière, en particulier pour les applications de la théorie des variables régionalisées dans le domaine. La géostatistique demeure un outil indispensable depuis les premiers phases d'exploration jusqu'à la reconnaissance des sols

Le polycopié, qui s'associe aussi avec des exercices d'application et des annexes, s'articule essentiellement sur 4 chapitres contenant les axes du nouveau programme ministériel. Après un rappel décrivant les différentes notions de base de la statistique. La deuxième partie concerne l'évaluation de la continuité spatiale de façon expérimentale, donnant une définition d'un variogramme, qui est tracé à partir de la distance des points de mesure et de la variance des valeurs. Celui-ci fait ensuite l'objet d'un ajustement à l'aide d'un modèle mathématique qui permettra de réaliser les calculs d'estimation. Cette étape permet de distinguer l'importance de la variabilité à courte distance et de déterminer si des mesures complémentaires doivent être ou non réalisées. - la troisième partie est l'étude de l'interpolation des données par la méthode de krigeage. Le krigeage se différencie des autres interpolateurs, par la prise en compte, entre les données et la cible, des distances séparant les données entre elles et de la structure spatiale du phénomène (par l'intermédiaire du variogramme)

# Historique

1930-1950 : fondement de la Théorie des fonctions aléatoires par **Kolmogorov** et **Wiener**

1955 : approche empirique (régression) pour corriger les problèmes de biais conditionnel observé dans les mines par **Daniel Krige**.

1960-1970 : **Matheron** dans le domaine des mines, **Matern** (foresterie) et **Gandin** (météorologie) développent un ensemble d'outils pour donner naissance de la géostatistique linéaire stationnaire et aux réponses des questions de Krige. Par la suite Matheron donne le nom **Krigeage** à la méthode d'estimation qu'il développe.

1973 : développement de la géostatistique linéaire non-stationnaire qui est appliquée dans des domaines encore actifs de recherche

1973 : évolution à la géostatistique non-linéaire

1977 : édition du premier livre de géostatistique linéaire en anglais par **M. David**

1980 : développement de la géostatistique linéaire multivariable

1985 : application et simulations

Aujourd'hui la géostatistique est appliquée dans une foule de domaines

Dès le départ il s'agissait de développer les éléments théoriques et méthodologiques permettant de répondre à des besoins de l'industrie. Si initialement les applications concernaient exclusivement la mine, elles se sont rapidement élargies au pétrole et à d'autres domaines comme la météorologie, la cartographie marine, l'halieutique, la pollution de l'air, de l'eau et des sols, etc.

La géostatistique, initiée par Georges Matheron au Bureau minier de l'Algérie puis au BRGM et au CEA, a pris son essor avec la création du Centre de géostatistique et de morphologie mathématique à l'École des Mines de Paris en 1967. Son fort développement a conduit à le scinder en deux entités, le Centre de Géostatistique et le Centre de Morphologie Mathématique.

# Chapitre 1 : Rappels sur les statistiques

---

## 1.1. Typologie des variables

- Variable qualitative : La variable est dite qualitative quand les modalités sont des catégories.
- Variable qualitative nominale : La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées.
- Variable qualitative ordinale : La variable est dite qualitative ordinale quand les modalités peuvent être ordonnées. Le fait de pouvoir ou non ordonner les modalités est parfois discutable. Par exemple : dans les catégories socioprofessionnelles, on admet d'ordonner les modalités : 'ouvriers', 'employés', 'cadres'. Si on ajoute les modalités 'sans profession', 'enseignant', 'artisan', l'ordre devient beaucoup plus discutable.
- Variable quantitative : Une variable est dite quantitative si toutes ses valeurs possibles sont numériques.
- Variable quantitative discrète : Une variable est dite discrète, si l'ensemble des valeurs possibles est dénombrable.
- Variable quantitative continue : Une variable est dite continue, si l'ensemble des valeurs possibles est continu.

## 1.2. Variables aléatoires

On définit une **variable aléatoire** en associant un nombre réel à chaque éventualité d'une expérience aléatoire. Une variable aléatoire  $X$  est une fonction de l'ensemble fondamental  $\Omega$  à valeurs dans  $\mathbb{R}$ ,  $X : \Omega \rightarrow \mathbb{R}$ .

Lorsque la variable  $X$  ne prend que des valeurs discrètes, on parle de variable aléatoire discrète. Un vecteur aléatoire  $X : \Omega \rightarrow \mathbb{R}^d$  est une fonction  $X = (X_1, \dots, X_d)$  à valeurs dans  $\mathbb{R}^d$  telle que les coordonnées  $X_i$  soient des variables aléatoires. Pour tout intervalle  $[a, b] \subset \mathbb{R}$ , l'ensemble  $\{X \in [a, b]\} = \{\omega \in \Omega : X(\omega) \in [a, b]\}$  est un événement.

### Exemple

1. On jette deux dés distincts et on s'intéresse à la somme des points. On note  $X$  cette variable aléatoire, elle est définie par  $X$  :

$\Omega \rightarrow \mathbb{R}$  avec  $\Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$

$(\omega_1, \omega_2) \rightarrow \omega_1 + \omega_2$ .

L'ensemble des valeurs possibles de  $X$  est  $\{2, 3, \dots, 12\}$ .

2. On lance toujours deux dés, mais cette fois on s'intéresse au plus grand chiffre Y obtenu. On a alors Y :

$$\Omega \rightarrow \mathbb{R} \text{ avec } \Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$$

$$(\omega_1, \omega_2) \rightarrow \max(\omega_1, \omega_2).$$

La variable Y est à valeurs dans  $\{1, 2, \dots, 6\}$ .

### 1.3. Loi de probabilité, Fonction de répartition, Densité de probabilité

#### 1- Loi de probabilité d'une variable aléatoire X

Assignation des probabilités sur les différentes valeurs de X (discrète) ou sur des intervalles de valeurs de X (continue) Pour une variable discrète : masses ponctuelles  $P(X = x_i)$  Pour une variable continue : densité de probabilité  $P(a < X < b)$ .

#### 2- Fonction de répartition d'une variable aléatoire X

$$F(x) = P(X < x) \text{ (fonction monotone croissante)}$$

$$\text{d'où: } P(a \leq X < b) = F(b) - F(a)$$

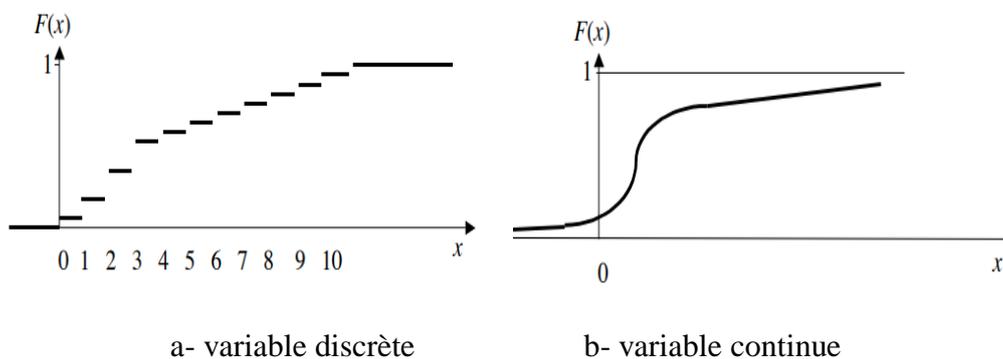


Figure 1 : Fonction de répartition d'une variable.

#### 3- Densité de probabilité (variable continue)

$f(x)$  est la fonction de densité pour une variable X. si pour tout intervalle  $[a, b]$  de  $\mathbb{R}$  on a:

$$P(a < X < b) = \int_a^b f(x) dx = \text{aire sous la courbe } f(x) \text{ au-dessus de } [a, b] ; \text{ d'où}$$

$$\int_{\mathbb{R}} f(x) dx = 1$$

et

$$F(a) = P(X < a) = \int_{-\infty}^{+\infty} f(x) dx$$

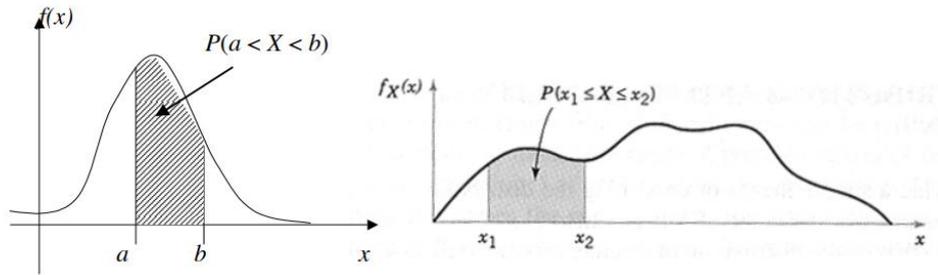


Figure 2 : Densité de probabilité (variable continue).

## 1.4. Moments d'une variable aléatoire X

1- Valeurs typiques :

- Centrales : moyenne (Moments empiriques d'ordre 1).
- De dispersion : variance, écart-type (déviation standard).
- De forme de distribution : coefficient d'asymétrie ('skewness'), d'aplatissement ('kurtosis').

2- Notion d'espérance mathématique :

$E(X)$  = moyenne (= centre de masse):

Variable discrète :

$$\mu = E(X) = \sum_i x_i P(X = x_i) \text{ (Équation 1)}$$

Variable continue de densité  $f(x)$ :

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx \text{ (Équation 2)}$$

$$\text{Variance } \text{Var}(X) = \sigma^2 = E((X) - E(X))^2 = E(X)^2 - \mu^2 \text{ (Équation 3)}$$

$$\text{Écart-type : } \sigma = \sqrt{V(X)} \text{ (Équation 4)}$$

$$\text{Asymétrie : } E \left[ \left( \frac{X-E(X)}{\sigma} \right)^3 \right] \text{ (Équation 5)}$$

$$\text{Aplatissement } E \left[ \left( \frac{X-E(X)}{\sigma} \right)^4 \right] \text{ (Équation 6)}$$

## 1.5. Deux variables quantitatives

Dans ce cas, chaque couple est composé de deux valeurs numériques. Un couple de nombres (entiers ou réels) peut toujours être représenté comme un point dans un plan  $(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)$ .

## Exemple

On mesure le poids Y et la taille X de 20 individus.

Tableau 1 : Exemple d'un couple de valeurs.

yi	60	75	61	76	64	78	67	80	68	85	69	90	70	96	70
xi	155	180	162	175	157	173	170	175	164	179	162	175	169	180	170
yi	96	72	98	73	101										
xi	185	178	189	173	187										

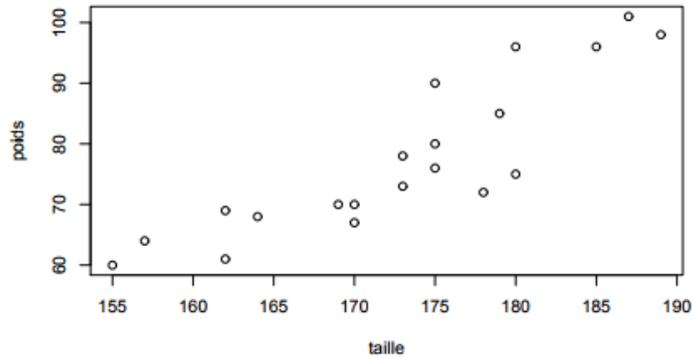


Figure 3 : Le nuage de points.

La fonction de densité est :

$f_{xy}(x, y)$  donne la probabilité que, simultanément  $X = x$  et  $Y = y$  :

$$P(x_1 < X < x_2, y_1 < Y < y_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{xy}(x, y) dx dy \text{ (Équation 7)}$$

La covariance :

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \text{ (Équation 8)}$$

La corrélation mesure la force du lien linéaire entre les variables X et Y. On n'aura que le coefficient de corrélation :

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \text{ (Équation 9)}$$

Avec :

$$-1 \leq \rho_{XY} \leq 1$$

## Exercices :

1- Une variable aléatoire X est établie par la loi de probabilité suivante :

x	-2	-1	0	1	2	3
p (X <sub>i</sub> = x)	0.3	0.05	0.1	0.05	0.2	p

Soit F sa fonction de répartition. Calculer:

a) p. b) F(0,5) c) E(X). d) Var (X).,  $\sigma(X)$ .

**Solution :**

a)  $\sum_i p_x (X_i) = 1$

$$0.3+0.05+0.1+0.05+0.2+p=1$$

$$p=0.3$$

b)  $F(0.5) = p(X \leq 0.5) = 0.3 + 0.05 + 0.1 = 0.45$

c)  $E(X) = 0.3*(-2) + 0.05*(-1) + \dots + 0.3*3 = 0.7$

d)  $Var(X) = 0.3*(-2)^2 + 0.05*(-1)^2 + \dots + 0.3*3^2 - 0.7^2 = 4.3$

$$\sigma(X) = \sqrt{Var(X)} = 2.077$$

2- Une variable aléatoire discrète X présente la fonction de répartition suivante :

F <sub>X</sub> (x)	0	0.4	0.7	0.8	1.0
x	x < 1	1 ≤ x < 2	2 ≤ x < 3	3 ≤ x < 4	4 ≤ x

a) Calculez l'espérance et la variance de X.

b) Quelle est la probabilité que  $1 \leq X < 3$  ?

**Solution :**

a) On calcule p<sub>x</sub>(x) :

x	0	1	2	3	4
p <sub>x</sub> (x)	0	0.4	0.3	0.1	0.2

$$E(X) = 0*0 + 0.4*1 + 0.3*2 + 0.1*3 + 0.2*4 = 2.1$$

$$E(X^2) = 0*0 + 0.4*1 + 0.3*2^2 + 0.1*3^2 + 0.2*4^2 = 5.7$$

$$Var(X) = E(X^2) - E(X)^2 = 5.7 - 2.1^2 = 3.6$$

b)  $P(1 \leq X < 3) = P(X=1) + P(X=2) = 0.4 + 0.3 = 0.7$

Voir Annexe sur Rappel de quelques notions statistiques à la fin du polycopié

## ***Chapitre 2 : Bases théoriques de la géostatistique***

---

### **2.1. Définition**

La géostatistique est une application de la théorie des fonctions aléatoires à des données localisées dans un espace géographique. Elle étudie les phénomènes naturels répartis dans l'espace (Phénomènes régionalisés) et/ou dans le temps (Minéralisation, pollution, propriété physique).

Les méthodes géostatistiques, telles le krigeage, ont été initialement proposées en exploration minière et pétrolière et elles ont retrouvé leur place en statistiques il y a plus d'une décennie. La géostatistique est classiquement subdivisée en géostatistique linéaire et multivariée, géostatistique non-linéaire, simulations géostatistiques.

"La géostatistique est l'application du formalisme des fonctions aléatoires à la reconnaissance et à l'estimation des phénomènes naturels. À ce propos Matheron écrit : « un phénomène est régionalisé s'il se déploie dans l'espace et y montre une certaine structure ».

Les principaux domaines d'applications de la Géostatistique sont :

- Mine
- Pétrole
- Hydrogéologie et Géotechnique
- Stockage (CO<sub>2</sub>, gaz, déchets)
- Halieutique
- Environnement
- Océanographie
- Santé
- Divers (Matériaux, ...)

### **2.2. La théorie des variables régionalisées**

Le problème classique en géostatistique est la prédiction d'une grandeur physique d'intérêt (variable régionalisée) sur un domaine d'étude à partir d'un ensemble fini d'observations éventuellement espacées irrégulièrement. Les variables régionalisées ont une structure d'auto-corrélation qui dépend du module et de la direction du vecteur séparant deux points de mesure. Mathématiquement, une variable régionalisée est une fonction du point  $x$ . Cette fonction est généralement irrégulière et montre deux aspects complémentaires :

- un aspect aléatoire qui explique les irrégularités locales ;
- un aspect structuré qui reflète les tendances du phénomène à grande échelle.

Le problème épistémologique lié au fait qu'habituellement

dans la pratique seule une réalisation partiellement échantillonnée est disponible a été résolu par Matheron dans son livre Estimer et Choisir (1989) en distinguant entre des quantités objectives qui sont essentiellement des intégrales de la réalisation unique et des paramètres conventionnels qui sont associés au modèle de fonction aléatoire. Les premiers peuvent être estimés alors que les seconds doivent être choisis.

Pour résoudre ce problème, l'approche usuelle est de considérer la variable régionalisée comme une réalisation d'une fonction aléatoire  $Z(x)$ , c'est-à-dire une famille de variables aléatoires  $z$  dépendant de la localisation  $x$  définie sur un domaine fixe et continu  $D$  :

$$Z = (Z(x), x \in D)$$

En chaque point  $x_i$ , on ne dispose d'une seule réalisation  $Z(x_i)$  et l'ensemble fini des points de mesure  $x_i$  est donc considéré comme une réalisation particulière de la fonction aléatoire  $Z(x_i)$ .

Si au point  $x_i$  de l'espace, la variable régionalisée  $z(x_i)$  est considéré comme valeur unique (valeur vraie) alors la géostatistique étudiera la corrélation spatiale de la variable régionalisée et la structure de cette variable dans l'espace. C'est la géostatistique transitive.

Le choix constitutif, de la géostatistique minière consiste à interpréter chaque valeur de la variable régionalisée  $z(x_i)$ , comme une réalisation particulière d'une variable aléatoire  $z(x_i)$  implanté au point  $x_i$ , donc plusieurs réalisations sont possibles. C'est la géostatistique intrinsèque.

Le modèle utilisé pour représenter les ensembles de valeurs va consister à trouver ce qu'il peut y avoir de commun entre ces ensembles. Le problème de l'étude de la fonction  $Z(x_i)$ , vu sous l'angle probabiliste, se ramènera à ajuster une loi de probabilité sur les données de manière qu'elle rende compte de l'échantillonnage. Par exemple, on avait effectué le même nombre d'essais en des points localisés différemment, on aurait trouvé des valeurs différentes.

### **2.3. Hypothèse de stationnarité et intrinsèque**

Une fonction aléatoire est stationnaire, au sens strict si la loi spatiale (la variance des incréments) est invariante par translation du vecteur  $h$ , de la même manière, une fonction aléatoire stationnaire se répète elle-même dans l'espace. Dans ce cas les deux variables aléatoires vectorielles à  $k$  composants :

$(Z(x_1), \dots, Z(x_k))$  et  $(Z(x_1+h), \dots, Z(x_k+h))$  présentent la même loi de distribution à  $k$  variables, quel que soit le vecteur translation  $h$ . Cette hypothèse permet de résoudre le problème posé par

l'inférence statistique, car à partir d'une réalisation on peut obtenir plusieurs. Sous l'hypothèse que les incréments sont stationnaires, on a :

$$\text{Var} (Z(x + h) - Z(x)) = E(|Z(x + h) - Z(x)|^2) \forall x \text{ (Équation 10)}$$

Il existe différents degrés de stationnarité, selon le ou les moments qui existent. On parle de stationnarité d'ordre 1 quand l'espérance  $E(Z(x))$  de la variable  $z$  existe est restée constante sur tout le domaine étudié.

$$\mu = E(Z(x)) \forall x \text{ (Équation 11)}$$

On parle de stationnarité d'ordre 2 si la covariance  $\text{Cov}((Z(x + h), Z(x)))$  est fonction seulement du vecteur  $h$  qui sépare les points  $((Z(x + h), Z(x)))$ . On aura :

$$\text{Cov} (Z(x), Z(x + h)) = E[(Z(x + h) - \mu)(Z(x) - \mu)] \forall x \text{ (Équation 12)}$$

Dans son sens le plus strict, la stationnarité nécessaire que tous les moments soient invariants par translation, mais comme cela ne peut être vérifié avec un nombre limité de données expérimentales, on se contente généralement de demander que les deux premiers moments (la moyenne et la covariance) soient invariants par translation. On parle alors de stationnarité faible ou d'ordre 2.

Des fonctions aléatoires dont les incréments sont stationnaires d'ordre deux et de moyenne nulle sont réputés satisfaire l'hypothèse intrinsèque, ce qui par abus de langage a été qualifié de *stationnarité intrinsèque* dans la littérature statistique.

La fonction aléatoire  $Z(x_i)$  est très souvent modélisée sous une hypothèse de stationnarité de second ordre ou de stationnarité intrinsèque. C'est l'hypothèse la plus courante, laquelle est définie pour des fonctions aléatoires stationnaires d'ordre deux de moyenne  $\mu$  (Wackernagel, 2004).  $Z(x_i)$  étant la fonction aléatoire représentant la variable étudiée. Elle est dite intrinsèque si :

- L'espérance mathématique existe et ne dépend que du point  $x : E(Z(x)) = \mu, \forall x$
- Pour tout vecteur  $h$ , l'accroissement  $(Z(x + h) - Z(x))$  aura une variance finie qui ne dépend pas de  $x$ , décrit sur l'équation 10.

De la sorte, la moyenne est constante à travers le domaine d'intérêt et la structure de dépendance spatiale (covariance ou variogramme) entre deux emplacements ne dépend que du vecteur de distance les séparant.

L'existence et la stationnarité de la covariance impliquent l'existence et la stationnarité de la variance. En effet :

$$\text{Var} (Z(x)) = E(|Z(x) - \mu|^2) = C(0) \forall x \text{ (Équation 13)}$$

Mais certains phénomènes physiques présentent une capacité de dispersion illimitée, c'est-à-dire qu'ils ne présentent ni covariance ni variance à priori finie. Pour les traiter, il convient de considérer leur accroissement ce qui conduit à l'hypothèse intrinsèque qui ne suppose que l'existence du variogramme.

## Chapitre 3 : Analyse du variogramme

---

### 3.1. Le variogramme

Le variogramme est défini pour toute fonction aléatoire intrinsèque et dépendant uniquement de l'interdistance  $h$ , alors que la fonction de covariance ne l'est que pour le cas d'une fonction aléatoire stationnaire d'ordre 2. De plus, l'estimation du variogramme n'est pas biaisée par la moyenne, au contraire de la covariance.

L'étude de la structure par le variogramme consiste à suivre l'évolution de « variation quadratique moyenne » de l'accroissement de la fonction  $Z(x)$  en fonction de  $h$  d'amplitudes, croissante, où  $h$  est le vecteur reliant deux points dans  $D$ . On obtient ainsi le variogramme dans une direction donnée.

On suppose habituellement que l'espérance des accroissement est stationnaire et nulle. Dans le cas de l'hypothèse intrinsèque, la fonction semi-variogramme  $\gamma(h)$  est définie par la relation :

$$\gamma(x+h, x) = \gamma(h) = \frac{1}{2} E(|Z(x+h) - Z(x)|^2) \forall x \text{ (Équation 14)}$$

Et

$$\text{Var}(Z(x+h) - Z(x)) = E(|Z(x+h) - Z(x)|^2) = 2\gamma(h) \forall x \text{ (Équation 15)}$$

Pour des fonctions aléatoires stationnaires d'ordre deux le variogramme est borné et peut être calculé à partir de la fonction de covariance (la variance à priori est finie  $C(0)$ ) :

$$\gamma(h) = \frac{1}{2} E(|Z(x) - \mu|^2) = C(0) - C(h) \forall x \text{ (Équation 16)}$$

Où :

$C(h)$  : est la covariance stationnaire

$C(0)$  : est la variance stationnaire  $C(0) = \text{Var}(Z(x))$

Noter que  $h$  est un vecteur. Les fonctions  $C(h)$  et  $\gamma(h)$  dépendent et de sa longueur (distance entre  $x$  et  $x+h$ ) et de sa direction. Quand elles ne dépendent que de sa longueur, ces fonctions sont dites isotropes.

Conditions limites :

$$\begin{array}{ll} C(h) \rightarrow 0 & |h| \rightarrow \infty \\ \gamma(h) \rightarrow C(0) & |h| \rightarrow \infty \end{array}$$

On peut également calculer une fonction de corrélation entre deux V. A. séparées par la distance  $h$ , qu'on appelle Corrélogramme  $\rho(h)$  :

$$\rho(h) = \frac{\text{Cov}(Z(x), Z(x+h))}{\sqrt{\text{Var}(z(x)) * \text{Var}(z(x+h))}} = \frac{C(h)}{C(0)} = \frac{C(0) - \gamma(h)}{C(0)} \quad (\text{Équation 17})$$

$$\rho(h) = 1 - \frac{\gamma(h)}{C(0)} \quad (\text{Équation 18})$$

Si pour tout couple :  $Z(x + h)$ ,  $Z(x)$ , la covariance existe et ne dépend que de la distance  $h$ . Le variogramme n'est pas borné et est pour cela préférable à la fonction de covariance :

$$\text{Cov}(Z(x), Z(x + h)) = C(h) = E[(Z(x + h) - \mu)(Z(x) - \mu)] \quad \forall x \quad (\text{Équation 19})$$

Le variogramme théorique sert d'une part à l'analyse structurale du phénomène étudié (effet de pépite, portée, existence de palier,...) et d'autre part à aborder certains problèmes de variabilité spatiale et d'estimation. En principe, pour estimer le variogramme théorique  $\gamma(h)$  à partir des données disponibles, on utilise la formule suivante :

$$\gamma(h) = \frac{1}{2} \sum_{i=1}^N [Z(x_i + h) - Z(x_i)]^2 \quad (\text{Équation 20})$$

Dans laquelle  $N$  représente le nombre de couples de valeurs de  $Z(x_i)$  mesurées en des points distincts de  $h$ .

Le variogramme expérimental étant obtenu, on détermine le variogramme théorique qui s'ajuste le mieux aux points du variogramme expérimental. On calcule le variogramme expérimental à l'aide de :

$$\gamma_e(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (\text{Équation 21})$$

où  $N$  nombre de paires dont les points sont espacés de  $h$  et  $Z(x_i + h) - Z(x_i)$  et l'écart entre leur valeur.

À titre d'exemple, la portée du variogramme dans la direction verticale est en général différente de celle obtenue dans la direction horizontale.

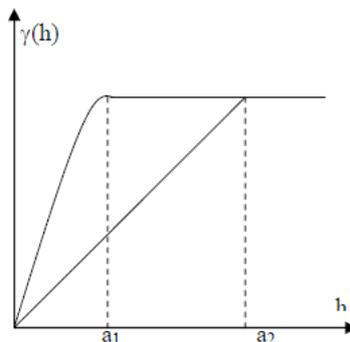


Figure 4 : Ajustement du variogramme- Modèle théorique.

### 3.2. Modèles

Différents modèles théoriques ont été élaborés pour tenir compte des traits caractéristiques du comportement du variogramme. Les composantes sont définies par un palier  $C$  et éventuellement une portée  $a$  et des paramètres de formes. On distingue les modèles sans paliers et modèles avec palier. Les composantes  $\gamma_i$  les plus fréquemment utilisées sont :

Effet de pépite :  $\gamma(h) = 0$  si  $h = 0$

$$\gamma(h) = C_0 \text{ si } h > 0$$

Sphérique :  $\gamma(h) = C [1.5 \frac{h}{a} - 0.5 (\frac{h}{a})^3]$  si  $0 < h < a$

$$\gamma(h) = C \text{ si } h \geq a$$

Gaussien :  $\gamma(h) = C [1 - \exp(-3(\frac{h^2}{a}))]$

Exponentiel :  $\gamma(h) = C [1 - \exp(-3(\frac{h}{a}))]$

Puissance :  $\gamma(h) = C h^b$   $0 < b < 2$  (linéaire :  $b = 1$ )

### 3.3. Propriétés du variogramme

Le variogramme est une fonction paire, à valeurs positives. Il est souvent une fonction croissante bornée. La figure ci-dessous représente une courbe de la variation typique du variogramme en fonction de la distance  $h$ . Ainsi, pour les modèles de variogramme montrant un seuil, on a :

- Portée  $a$  : On nomme *portée* la distance à partir de laquelle le variogramme atteint, respectivement, son palier ; la *portée pratique* (parfois *facteur d'échelle*) est la distance à partir de laquelle le variogramme reste dans un intervalle de 5 % autour de son palier. La *norme* est le rapport de la portée sur la portée pratique. La portée représente la distance où deux observations ne se ressemblent plus du tout en moyenne, elles ne sont plus liées (covariance nulle) linéairement. À cette distance, la valeur du variogramme correspond à la variance de la variable aléatoire.

- Palier : la limite du variogramme à l'infini.  $\sigma^2 = C_0 + C$ : Variance de la variable aléatoire. (Var(Z(x)))

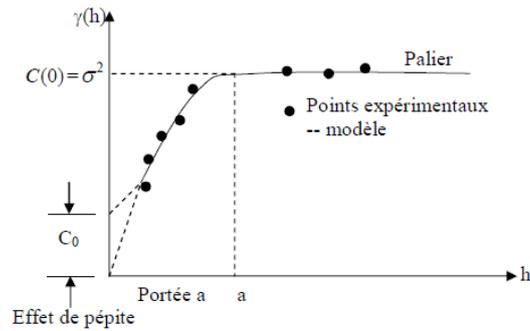


Figure 5 : Exemple de variogramme.

- Effet de pépité :  $C_0$ : Variation à très courte échelle, erreurs de localisation, erreurs d'analyse et précision analytique.

Le variogramme croît avec  $h$ . ceci provient du fait que plus les points sont éloignés, plus les valeurs des paramètres en ces points ont des chances d'être différentes. En absence de dérive et lorsque la capacité de dispersion du milieu est finie, le variogramme se stabilise autour d'une valeur limite  $\gamma(\infty)$  pour des distances  $h$  supérieures à une certaine limite à appeler portée.

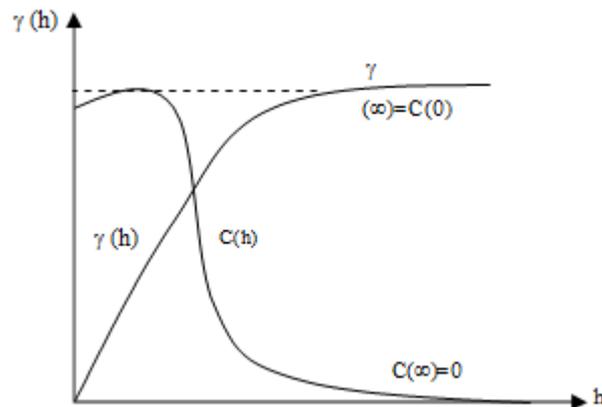


Figure 6 : Propriétés du variogramme.

Lorsque  $h = 0$

$$\gamma(0) = \frac{1}{2} \text{Var} (Z(x+h) - Z(x)) = 0 \text{ (Équation 22) et non pas } C_0$$

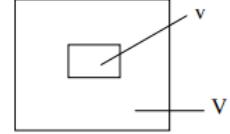
### 3.4. Variance d'estimation et de dispersion

En fait la vraie valeur de la variable  $Z$  n'est connue qu'en certains points où l'on dispose de mesures ponctuelles par des sondages. Pour connaître la vraie valeur de  $Z$  en tout autre point, on doit procéder à une estimation à partir des données disponibles. Il convient alors de connaître l'erreur commise lorsqu'on utilise la valeur estimée  $Z^*$  en un point au lieu de la vraie valeur  $Z$  inconnue. Pour caractériser cette erreur, on fait appel aux notions de variance d'estimation et de variance de dispersion.

Variance d'estimation :

Dans cette section, on cherche à établir les résultats permettant de fournir une mesure de la précision des estimés effectués par une méthode d'estimation quelconque (linéaire).

Soit une variable aléatoire  $Z_v$  que l'on veut estimer, d'une façon ou d'une autre, en formant une combinaison linéaire des valeurs observées en différents endroits. Considérons un sous domaine  $v$  de l'espace  $V$ . Par exemple,  $V$  est un dépôt d'argile et  $v$  l'ensemble fini des points de mesure.



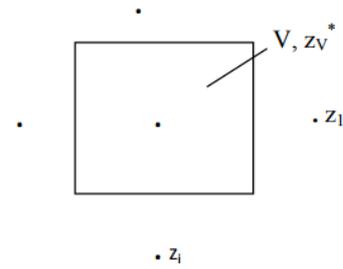
$$v = \{x_i / i = 1, 2, \dots, N\}$$

où  $N$  est le nombre de points de mesure.

$Z_i$  : valeur observée au point  $x_i$  (variable aléatoire)

$Z_v^*$  : estimateur de  $Z_v$

Par exemple si l'on dispose de  $N$  informations  $v_i$  de teneurs moyennes  $z(x_i)$ ,  $\lambda_i$  étant le pondérateur associé à l'information  $v_i$ .



On définit l'erreur d'estimation, par suite, la moyenne des valeurs mesurées est :

$$Z_v^* = \frac{1}{N} \sum_{i=1}^N Z(x_i) = \sum_{i=1}^N \lambda_i Z(x_i) \text{ (Équation 23)}$$

Lorsqu'on estime la vraie valeur de la moyenne  $Z$  par  $Z^*$  on commit une erreur d'estimation :

$$\bar{V} = Z_v - Z_v^* \text{ (Équation 24)}$$

$$Var(Z_v - Z_v^*) = Var \bar{V} = Var(Z_v) + Var(Z_v^*) - 2 Cov(Z_v, Z_v^*) \text{ (Équation 25)}$$

Substituant  $Z_v^*$  par son expression, en fonction des  $Z_i$ , on obtient:

$$\sigma_e^2 = Var(Z_v) + \sum_i \sum_j \lambda_i \lambda_j Cov(Z_i, Z_j) - 2 \sum_i \lambda_i Cov(Z_i, Z_v) \text{ (Équation 26)}$$

Elle peut être réécrit en fonction du variogramme:

$$\sigma_e^2 = (\sigma^2 - \bar{\gamma}(v, v)) + \sum_i \sum_j \lambda_i \lambda_j (\sigma^2 - \gamma(x_i - x_j)) - 2 \sum_i \lambda_i (\sigma^2 - \bar{\gamma}(x_i, v)) \text{ (Équation 27)}$$

Puis finalement, puisqu'on a habituellement:

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

$$\sigma_e^2(V, v) = 2 \sum_i \lambda_i \bar{\gamma}(V, v) - \bar{\gamma}(V, V) - \sum_i \sum_j \lambda_i \lambda_j \gamma(v_i, v_j) \text{ (Équation 28)}$$

Le krigeage consiste à déterminer les pondérateurs  $\lambda_i$ .

Variance de dispersion

La variance d'estimation dépend de la forme et de la taille de  $v$  et  $V$ , ainsi que de la position relative de  $v$  par rapport à  $V$ .

De même, la variance de la valeur moyenne  $Z_v^*$  des  $z(x_i)$  lorsque  $v$  occupe toutes les positions possibles dans un domaine peu vaste  $V$  a une valeur moyenne appelée variance de dispersion de  $v$  dans  $V$  et est désignée par :  $\sigma^2 (v/V)$ . On démontre que :

$$\sigma_D^2 \left( \frac{v}{V} \right) = \gamma(V, V) - \gamma(v, v) \text{ (Équation 29)}$$

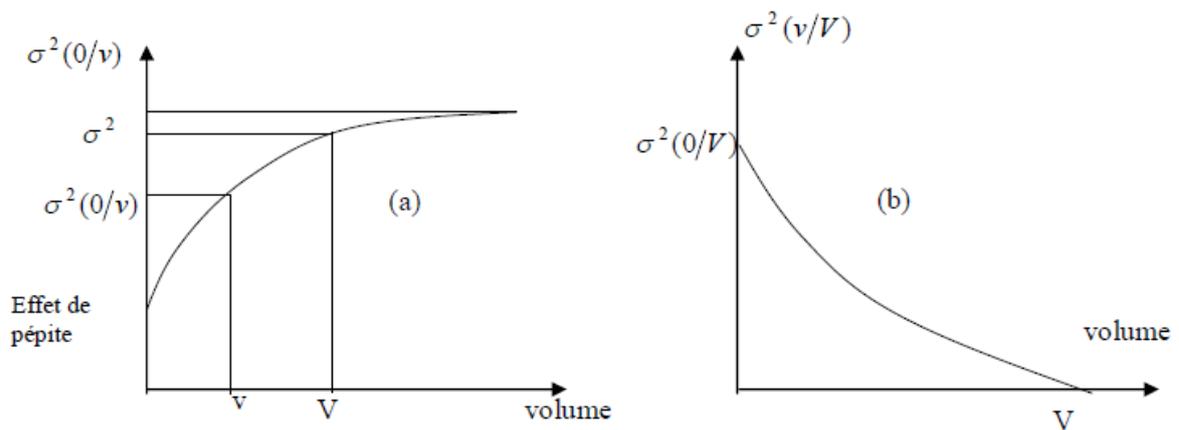


Figure 7 : Variance d'estimation (a) et variance de dispersion (b).

## Exercices :

1- Le tableau suivant donne une série de levés topographique de 4 points distants de 10m. Calculer le variogramme expérimental de cette série.

Point	1	2	3	4	5
Levé (m)	13	15	16	14	18

Pour  $h = 10\text{m}$  :

- Pt1#Pt2 :  $g(x) = g(\text{Pt1}) = 13\text{m}$

$$g(x+h) = g(\text{Pt2}) = 15\text{m}$$

Pt2#Pt3 :  $g(x) = g(\text{Pt2}) = 15\text{m}$

$$g(x+h) = g(\text{Pt3}) = 16\text{m}$$

$$\gamma(10) = \frac{1}{2 * 2} \sum [(13 - 15)^2 + (15 - 16)^2] = 1.25$$

- Pt4#Pt2 :  $g(x) = g(\text{Pt4}) = 14\text{m}$

$$g(x+h) = g(\text{Pt2}) = 15\text{m}$$

Pt2#Pt5 :  $g(x) = g(\text{Pt2}) = 15\text{m}$

$$g(x+h) = g(\text{Pt5}) = 18\text{m}$$

$$\gamma(10) = \frac{1}{2 * 2} \sum [(14 - 15)^2 + (15 - 18)^2] = 2.75$$

Pour  $h = 20\text{m}$  :

- Pt1#Pt3 :  $g(x) = g(\text{Pt1}) = 13\text{m}$

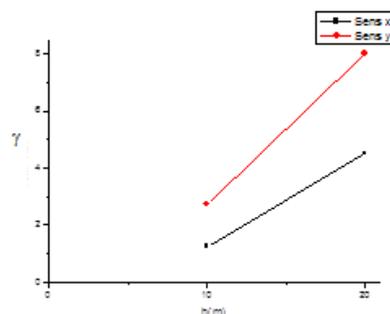
$$g(x+h) = g(\text{Pt3}) = 16\text{m}$$

$$\gamma(20) = \frac{1}{2 * 1} \sum [(13 - 16)^2] = 4.5$$

- Pt4#Pt5 :  $g(x) = g(\text{Pt4}) = 14\text{m}$

$$g(x+h) = g(\text{Pt5}) = 18\text{m}$$

$$\gamma(20) = \frac{1}{2 * 1} \sum [(14 - 18)^2] = 8$$



3- On a un variogramme sphérique avec :

$C_0=1$  et  $C=10$ ,  $a=20$ .

Soit les points  $x_1=(0,0)$  et  $x_2=(10,0)$

a) Quelle est la variance de  $Z(x_1)$ ? de  $Z(x_2)$ ?

b) Quelle est la covariance entre  $Z(x_1)$  et  $Z(x_2)$ ?

c) On forme  $Z_3=0.8*Z(x_1)+0.2*Z(x_2)$ .

Quelle est la variance de  $Z_3$ ? Quelle est la covariance de  $Z_3$  avec  $Z(x_1)$ ?

d) On forme  $Z_4=0.4*Z(x_1)+0.6*Z(x_2)$ . Quelle est la covariance entre  $Z_3$  et  $Z_4$ ?

**Solution :**

1- 1 a)  $\text{Var}(Z(x_1))=\text{Var}(Z(x_2))=C_0+C=11$

b)  $\gamma(h)=1+10*(1.5(10/20)-0.5*(10/20)^3)=7.875$   $\text{Cov}(Z(x_1),Z(x_2))=11-7.875=3.125$

c)  $\text{Var}(Z_3)=0.8^2\text{Var}(Z(x_1))+0.2^2\text{Var}(Z(x_2))+2*0.8*0.2*\text{Cov}(Z(x_1),Z(x_2))$   
 $=0.64*11+0.04*11+2*0.8*0.2*3.125=8.48$

$\text{Cov}(Z_3,Z(x_1))=0.8*\text{Var}(Z(x_1))+0.2*\text{Cov}(Z(x_1),Z(x_2))=0.8*11+0.2*3.125=9.425$

d)  $\text{Cov}(Z_3,Z_4)=\text{Cov}(0.8*Z(x_1)+0.2*Z(x_2),0.4*Z(x_1)+0.6*Z(x_2))=$   
 $0.8*0.4*\text{Var}(Z(x_1))+0.8*0.6*\text{Cov}(Z(x_1),Z(x_2))+0.2*0.4*C$

## Chapitre 4 : Théorie du krigeage

---

### 4.1. Le Krigeage

Le principal objectif de la géostatistique est la prédiction spatiale, encore appelée krigeage, consistant à prédire une variable régionalisée d'intérêt sur un domaine d'étude, à partir des données observées à certains emplacements. Le krigeage repose fondamentalement sur la modélisation et l'estimation de la structure de dépendance spatiale. La description de cette dernière se fait couramment à l'aide d'outils statistiques tels que le variogramme ou la covariance, calculés sur l'ensemble du domaine d'intérêt et sous une hypothèse de stationnarité.

Seule la méthode d'interpolation par krigeage repose sur une méthode statistique satisfaisante et permet d'obtenir la vraie variance d'estimation.

Puisqu'on peut calculer la variance d'estimation pour tout estimateur linéaire, pourquoi ne pas choisir celui qui assure la variance d'estimation minimale? C'est précisément ce qu'effectue le krigeage. Dans le cas stationnaire, on en reconnaît 2 types principaux, selon que la moyenne du processus est connue ou non, soit le krigeage simple et le krigeage ordinaire. Ce dernier est, de loin, le plus fréquemment utilisé.

### 4.2. Type de Krigeage

#### 4.2.1. Krigeage stationnaire à moyenne inconnue (krigeage linéaire ordinaire)

Le krigeage ordinaire est une opération qui est répétée en chaque noeud  $x_0$  d'une grille régulière recouvrant le domaine étudié. Pour un ensemble de  $n$  points de données  $x$  d'un voisinage centré autour d'un point  $x_0$  de la maille d'estimation, on peut construire, en minimisant la variance d'estimation. Ce système permet donc de retrouver les  $N$  pondérateurs  $\lambda_i$ . Ces derniers, dit pondérateurs de krigeage, donnent la variance d'estimation la plus petite possible et elle est appelée variance de Krigeage. Où les  $\lambda_i$  sont des pondérateurs à affecter aux points de données et où  $\mu$  est un paramètre de Lagrange qui intervient pour des raisons algébriques. Le membre gauche du système contient les covariances entre les points de données, tandis que le membre droit contient les covariances entre chaque point de donnée et le point d'estimation  $x_0$ . Ce système, une fois résolu, permet de transférer au point  $x_0$  de l'information en provenance des points de données voisins, par le calcul d'une moyenne pondérée

En effectuant la multiplication matricielle, on peut réécrire le système sous la forme :

$$\sum_{i=1}^N \lambda_i C(x_\alpha - x_i) + \mu = C(x_\alpha - x_0), \quad \alpha = 1 \text{ à } N \text{ (Équation 30)}$$

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

Le système de krigeage ordinaire suivant :

$$\begin{vmatrix} C(x_1 - x_1) & C(x_1 - x_2) & \dots & C(x_1 - x_n) & 1 \\ C(x_2 - x_1) & \dots & \dots & C(x_2 - x_n) & 1 \\ \dots & \dots & \dots & \dots & \dots \\ C(x_n - x_1) & C(x_n - x_2) & \dots & C(x_n - x_n) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix} \begin{vmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ \mu \end{vmatrix} = \begin{vmatrix} C(x_1 - x_0) \\ C(x_2 - x_0) \\ \dots \\ C(x_n - x_0) \\ 1 \end{vmatrix}$$

$$Z_v^*(x_0) = \sum_{i=1}^N \lambda_i Z(x_i) \text{ (Équation 31)}$$

La variance d'estimation en krigeage ordinaire est :

$$\sigma_{ord}^2 = C(0) - \sum_{i=1}^N \lambda_i C(x_0, x_i) - \mu \text{ (Équation 32)}$$

#### 4.2.2. Krigeage stationnaire à moyenne connue (krigeage linéaire simple)

Parfois on connaît la moyenne " $\mu$ " du champ à estimer ou du moins on en possède un estimé fiable. On peut alors former un estimateur sans biais :

$$Z_v^* = \sum_{i=1}^N \lambda_i Z(x_i) + (1 - \sum_{i=1}^N \lambda_i) \mu \text{ (Équation 33)}$$

La variance d'estimation en krigeage simple est :

$$\sigma_{simp}^2 = C(0) - \sum_{i=1}^N \lambda_i C(x_0, x_i) \text{ (Équation 34)}$$

### 4.3. Propriétés du krigeage

Les principales propriétés et caractéristiques associées au krigeage sont: Linéaire, sans biais, à variance minimale, par construction. Interpolateur exact. : Si l'on estime un point connu, on retrouve la valeur connue (le point de point connue sera égale 1 et le poids des autres point sera égal à 0). Présente un effet d'écran : les points les plus près reçoivent les poids les plus importants. Cet effet d'écran varie selon la configuration et selon le modèle de variogramme utilisé pour le krigeage. Plus l'effet de pépite est important, moins il y a d'effet d'écran. Tient compte de la taille du champ à estimer et de la position des points entre eux. Par l'utilisation du variogramme, tient compte de la continuité du phénomène étudié (effet de pépite, anisotropie, etc.). Effectue généralement un lissage, i.e. les estimations sont moins variables que les teneurs réelles (point ou bloc) que l'on cherche à estimer. Presque sans biais conditionnel. Ceci signifie que lorsqu'on applique une teneur de coupure à des valeurs estimées, on récupérera

approximativement la teneur prévue. C'est une propriété très importante pour les mines. Cette propriété implique que l'estimateur utilisé soit plus lisse que la valeur qu'il cherche à estimer, ce qui est le cas pour le krigeage. Si l'on observe en un point une valeur coïncidant avec la valeur krigée pour ce point, alors les valeurs krigées en d'autres points ne sont pas modifiées par l'inclusion de ce nouveau point dans les krigeage. Par contre les variances de krigeage, elles, sont diminuées. De même, si l'on krige un certain nombre de points et que l'on utilise les valeurs krigées comme si c'étaient de nouvelles données, alors les krigeages subséquents ne s'en trouvent pas modifiés (sauf pour la variance de krigeage).

#### 4.4. Cokrigeage

Souvent l'on a plusieurs variables mesurées, soit aux mêmes points échantillons, soit en des points différents. Sans perte de généralité, l'on va considérer le cas où une des variables est identifiée comme prioritaire (variable principale  $Z$ ), et les autres sont des variables secondaires. Pour simplifier l'écriture, on va considérer que l'on a une seule variable secondaire ( $Y$ ). Toutefois l'extension à plusieurs variables est immédiate et ne pose aucun problème théorique particulier.

##### 4.4.1. Cokrigeage ordinaire

On veut former une estimation linéaire de la variable principale  $Z(x)$  (avec  $n_z$  observations) à partir d'observations de la variable principale et de la variable secondaire  $Y(x)$  (avec  $n_y$  observations).

Les moyennes de  $Z(x)$  et  $Y(x)$  sont inconnues :

$$Z_0^* = \sum_{i=1}^{n_z} \lambda_i Z(x_i) + \sum_{i=1}^{n_y} \alpha_i Y(x_i) \text{ (Équation 35)}$$

L'estimateur doit être sans biais, ceci est assuré en imposant :

$$\sum_{i=1}^{n_z} \lambda_i = 1, \sum_{i=1}^{n_y} \alpha_i = 0$$

La variance d'estimation s'écrit :

$$\begin{aligned} \text{Var}(Z_0 - Z_0^*) &= \text{Var}(Z_0) + \sum_{i=1}^{n_z} \sum_{j=1}^{n_z} \lambda_i \lambda_j \text{Cov}(Z_i, Z_j) + \sum_{i=1}^{n_y} \sum_{j=1}^{n_y} \alpha_i \alpha_j \text{Cov}(Y_i, Y_j) + \\ &+ 2 \sum_{i=1}^{n_z} \sum_{j=1}^{n_y} \lambda_i \alpha_j \text{Cov}(Z_i, Y_j) - 2 \sum_{i=1}^{n_z} \lambda_i \text{Cov}(Z_0, Z_i) - 2 \sum_{i=1}^{n_y} \alpha_i \text{Cov}(Z_0, Y_i) \text{ (Équation 36)} \end{aligned}$$

On forme le Lagrangien et l'on dérive par rapport aux poids inconnus et aux 2 multiplicateurs de Lagrange introduits pour tenir compte des contraintes de non-biais. On obtient le système de cokrigage ordinaire :

$$\sum_{j=1}^{nz} \lambda_j Cov(Z_i, Z_j) + \sum_{j=1}^{ny} \alpha_j Cov(Z_i, Y_j) + \mu_z = Cov(Z_0, Z_i) \quad \forall i = 1 \dots \dots \dots nz \quad (\text{Équation 37})$$

$$\sum_{j=1}^{nz} \lambda_j Cov(Y_i, Z_j) + \sum_{j=1}^{ny} \alpha_j Cov(Y_i, Y_j) + \mu_y = Cov(Z_0, Y_i) \quad \forall i = 1 \dots \dots \dots nz \quad (\text{Équation 38})$$

$$\sum_{i=1}^{nz} \lambda_i = 1, \quad \sum_{i=1}^{ny} \alpha_i = 0$$

$$\sigma_{ck}^2 = Var(Z_0) - \sum_{i=1}^{nz} \lambda_j Cov(Z_0, Z_i) - \sum_{i=1}^{ny} \alpha_i Cov(Z_0, Y_i) - \mu_z \quad (\text{Équation 39})$$

#### 4.4.2. Cokrigage simple

Les moyennes  $\mu_z$  et  $\mu_y$  de  $Z(x)$  et  $Y(x)$  sont connues. On estime alors en  $x_0$  un résidu auquel on rajoute la moyenne  $m_z$ . Les conditions de non-biais ne sont plus requises. La matrice de cokrigage est alors de taille  $(nz+ny) \times (nz+ny)$ . Il résulte :

$$Z_0^* = m_z + \sum_{i=1}^{nz} \lambda_i (Z_i - \mu_z) + \sum_{i=1}^{ny} \alpha_i (Y_i - \mu_y) \quad (\text{Équation 40})$$

$$\sum_{j=1}^{nz} \lambda_j Cov(Z_i, Z_j) + \sum_{j=1}^{ny} \alpha_j Cov(Z_i, Y_j) = Cov(Z_0, Z_i) \quad \forall i = 1 \dots \dots \dots nz \quad (\text{Équation 41})$$

$$\sum_{j=1}^{nz} \lambda_j Cov(Y_i, Z_j) + \sum_{j=1}^{ny} \alpha_j Cov(Y_i, Y_j) = Cov(Z_0, Y_i) \quad \forall i = 1 \dots \dots \dots nz \quad (\text{Équation 42})$$

$$\sigma_{ck}^2 = Var(Z_0) - \sum_{i=1}^{nz} \lambda_j Cov(Z_0, Z_i) - \sum_{i=1}^{ny} \alpha_i Cov(Z_0, Y_i) \quad (\text{Équation 43})$$

#### 4.4.3. Cokrigage sous format matricielle

Le cokrigage ordinaire :

$$k\lambda = k$$

$$\sigma_{ck}^2 = Var(Z_0) - \lambda'k \quad (\text{Équation 44})$$

$$\begin{array}{cccccc} k_{zz} & k_{zy} & 1 & 0 & \lambda & k_{zz} \\ k_{yz} & k_{yy} & 0 & 1 & \alpha & k_{yz} \\ 1' & 0' & 0 & 0 & \mu_z & = & 1 \\ 0' & 1' & 0 & 0 & \mu_y & & 0 \end{array}$$

Le cokrigeage simple :

$$\begin{array}{ccc} k_{zz} & k_{zy} & \lambda \\ k_{yz} & k_{yy} & \alpha \end{array} = \begin{array}{c} k_{zz} \\ k_{yz} \end{array}$$

$$k_{zz} = nz * nz$$

$$k_{zy} = nz \text{ par } ny$$

$$k_{yz} = ny * nz$$

$$k_{yy} = ny \text{ par } ny$$

$$k_{zz} = nz * 1$$

$$k_{yz} = ny * 1$$

Les matrices de cokrigeage sont toujours symétriques :

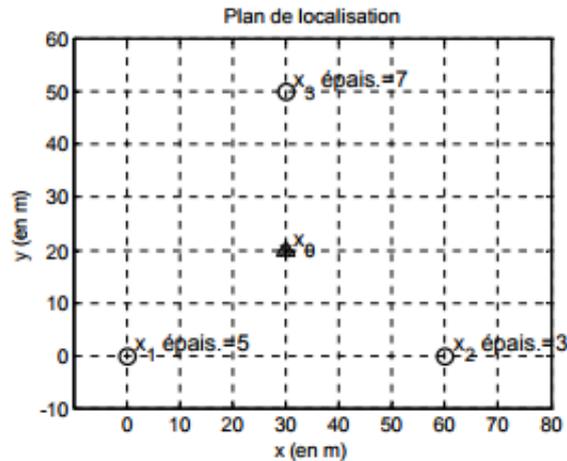
$$k_{zz} = k'_{yz} \text{ (Équation 45)}$$

Toutefois, les fonctions de covariances croisées, elles, ne sont pas nécessairement symétriques, en effet :

$$C_{zy}(h) = Cov(Z(x), Y(x+h)) = Cov(Y(x+h), Z(x)) = C_{yz}(-h) \text{ (Équation 46)}$$

### Exercices :

L'épaisseur (en m) d'un dépôt de sable présente un variogramme sphérique isotrope avec  $a=100\text{m}$ ,  $C_0=1\text{m}^2$  et  $C=10\text{m}^2$ . La figure suivante montre les épaisseurs observées en 3 points échantillons, on désire fournir une estimation par krigeage ordinaire de l'épaisseur de sable au point  $x_0$ .



Donnez le système de krigeage ordinaire sous forme matricielle.

### Solution :

Calcul de h :

	$X_0$	$X_1$	$X_2$	$X_3$
$X_0$	0	36	36	30
$X_1$	36	0	60	58
$X_2$	36	60	0	58
$X_3$	30	58	58	0

Calcul de  $\gamma(h)$  :  $\gamma(h) = C_0 + C[0.5 \frac{h}{a} - 0.5 (\frac{h}{a})^3]$

	$X_0$	$X_1$	$X_2$	$X_3$
$X_0$	0	6.16672	6.16672	5.365
$X_1$	6.16672	0	8.92	8.72444
$X_2$	6.16672	8.92	0	8.72444
$X_3$	5.365	8.72444	8.72444	0

Calcul de  $C(h)$  :  $C(h) = Var(z(x)) - \gamma(h)$

	$X_0$	$X_1$	$X_2$	$X_3$
$X_0$	11	4.83328	4.83328	5.635
$X_1$	4.83328	4.83328	2.08	2.27556
$X_2$	4.83328	2.08	11	2.27556
$X_3$	5.635	2.27556	2.27556	11

Calcul matrice de krigeage :

	X <sub>0</sub>	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>
X <sub>0</sub>	11	4.83328	4.83328	5.635
X <sub>1</sub>	4.83328	4.83328	2.08	2.27556
X <sub>2</sub>	4.83328	2.08	11	2.27556
X <sub>3</sub>	5.635	2.27556	2.27556	11

$$\begin{array}{|cccc|c|c|c|}
 \hline
 4.83328 & 2.08 & 2.27556 & 1 & \lambda_1 & & 4.83328 \\
 2.08 & 11 & 2.27556 & 1 & \lambda_2 & = & 4.83328 \\
 2.27556 & 2.27556 & 11 & 1 & \lambda_3 & & 5.635 \\
 1 & 1 & 1 & 0 & \mu & & 1 \\
 \hline
 \end{array}$$

## ***Conclusion***

---

Ce polycopié regroupe les développements des outils et des notions de la statistique met à la disposition de l'ingénierie, en particulier pour l'ingénieur géotechnicien. Leurs utilisations permettent d'évaluer l'influence des incertitudes affectant les propriétés des sols sur la validité et la fiabilité des résultats de ce type d'analyse numérique, qui devraient être appliquée à toute reconnaissance géotechnique.

À la fin, je serais ouvert à toutes propositions pour améliorer la qualité de ce cours.

### Les variables régionalisées

Dans chacun de ces domaines, on s'intéresse à une "variable régionalisée" qui évolue à l'intérieur d'un certain champ de l'espace.

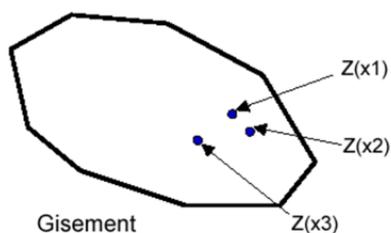
Cette variable peut être définie dans les 3 dimensions de l'espace : porosité d'une roche, perméabilité, teneur en métal... ou en 2 dimensions : densité d'arbres, densité de poissons (rapportée à la surface de la mer), profondeur d'un horizon pédologique... ou à 1 dimension (teneurs par transect).

Du phénomène étudié sur son terrain d'investigation, le géostatisticien (ou le praticien) ne connaît que les mesures de la variable en certains points ou le long de lignes.

Toutefois, il dispose également d'une information qualitative importante (milieu physique, conditions d'échantillonnage, expérience...).

Ces informations seront très utiles lors de l'application de la géostatistique et l'interprétation des résultats de celle-ci

### Point de vue de la géostat



Gisement => infinité de points ou très grand nombre de « quasi-points »  
x : emplacement géographique  
chaque point -> teneur -> Z(x)

Il cherche à prévoir une quantité qui dépend, bien évidemment, de la variable en dehors des points de données.

Or les variables régionalisées étudiées en géostatistique présentent très souvent un comportement irrégulier, aléatoire, rendant la prévision incertaine.

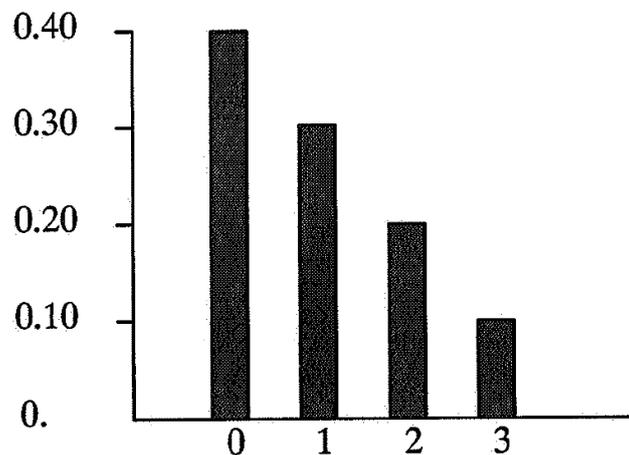
C'est en étudiant ce comportement, en dégagant la structure, puis en faisant usage de celle-ci, que le géostatisticien parvient à une meilleure connaissance de son phénomène.

**Imaginons l'expérience suivante :**

Nous creusons 10 trous dans le sol, et comptons le nombre de pierres dans chacun d'eux.

- dans quatre des dix trous nous n'en trouvons aucune,
- trois autres trous contiennent chacun une pierre,
- nous en comptons deux dans chacun des deux autres trous,
- le dernier trou donne trois pierres.

L'histogramme de la figure montre graphiquement les fréquences des valeurs 0, 1, 2 et 3 prises par la variable "nombre de pierres dans un trou".



**La moyenne** de cette variable vaut :  $m = (0+0+0+0+1+1+1+2+2+3)/10=1$  pierre

Ainsi, en moyenne, un trou contient 1 pierre. Mais certains trous en comptent davantage, d'autres aucune. Cette variabilité se mesure par la **variance** des valeurs :

$$\delta^2 = ((-1)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (0)^2 + (0)^2 + (0)^2 + (1)^2 + (1)^2 + (2)^2) / 10$$

C'est la moyenne des carrés des écarts à la moyenne.

Il s'agit d'une quantité nécessairement positive (à la rigueur nulle si toutes les valeurs sont égales), qui après calcul se trouve ici égale aussi à 1.

Comme la variable s'exprime en nombre de pierres, la variance, homogène au carré de la variable, s'exprime en pierres carrées, ici 1 (pierre)<sup>2</sup>.

Pour avoir une quantité homogène à la variable, on utilise **l'écart-type**  $\delta$ , racine carrée de la variance, égal ici à 1 pierre.

Variance ou écart-type indiquent de combien en moyenne les valeurs s'écartent de leur moyenne (sur l'histogramme, étalement des valeurs autour de leur moyenne).

### **Une Autre Vision**

Considérons maintenant le nombre de pierres par trou comme une **variable aléatoire** (VA en abrégé) notée X.

Les fréquences précédentes deviennent des probabilités X vaut 0 avec la probabilité 0.4 ("4 chances sur 10"), 1 avec la probabilité 0.3, 2 avec 0.2 et enfin 3 avec 0.1.

En moyenne, la VA X vaut :  $0 \times 0.4 + 1 \times 0.3 + 2 \times 0.2 + 3 \times 0.1 = 1$

Cette **moyenne** s'appelle aussi **l'espérance mathématique**, ou la valeur probable, de X, notée E(X).

On peut calculer de la même façon **l'espérance du carré de la variable** :

$$E(X^2) = 0^2 \times 0.4 + 1^2 \times 0.3 + 2^2 \times 0.2 + 3^2 \times 0.1 = 2$$

Et La variance de X:

$$\delta_x^2 = \text{var}X = E(X - E(X))^2 = E(X^2) - (E(X))^2 = (-1)^2 \times 0.4 + 0^2 \times 0.3 + 1^2 \times 0.2 + 2^2 \times 0.1 = 1$$

Cette variance est égale à E(X<sup>2</sup>) si E(X) est nulle.

### **Propriétés**

- ✓ De plus, rajouter une constante b à X donne une nouvelle espérance E(X)+b mais ne change pas sa variance  $\text{var}(X+b) = \text{var}X$ .
- ✓ Noter qu'une variance ne peut jamais être négative.

### **Entre deux variables aléatoires X et Y, on utilise souvent la covariance :**

$$\text{cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Cette covariance ne peut dépasser, en valeur absolue, le produit des écarts-types  $\delta x \delta y$ ,  
et on appelle **corrélation** entre X et Y, le coefficient sans dimension, compris entre -1 et 1, égal  
à  $\text{cov}(X,Y)/\delta x \delta y$

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \text{ (Équation 9)}$$

$$-1 \leq \rho_{XY} \leq 1$$

Considérons maintenant une combinaison linéaire de X et Y :  $aX+bY$

Son espérance vaut :  $aE(X) + bE(Y)$

L'espérance de son carré s'écrit :  $E(aX+bY)^2 = a^2 E(X^2) + b^2 E(Y^2) + 2abE(XY)$

et la variance:  $\text{var}(aX+bY)=a^2 \text{var}X + b^2 \text{var}Y + 2ab \text{cov}(X,Y)$

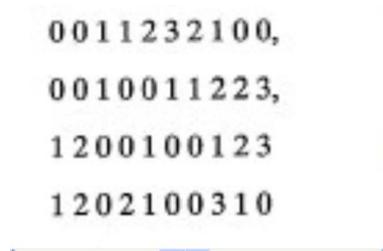
Ainsi les « **moments d'ordre 1** »,  $E(X)$  et  $E(Y)$ ,

et les « **moments d'ordre 2** »,  $E(X^2)$ ,  $E(Y^2)$ ,  $E(XY)$  (et donc variances et covariance) de X et Y, suffisent à calculer espérance et variance de leurs combinaisons linéaires.

Ce seront de tels outils que l'on utilisera dans l'application de la géostatistique

### Revenons un instant à nos trous et à leurs pierres,

et supposons que ces trous soient creusés le long d'une ligne de telle sorte que le nombre de pierres trouvées le long de la ligne soit, dans cet ordre :



0	0	1	1	2	3	2	1	0	0
0	0	1	0	0	1	1	2	2	3
1	2	0	0	1	0	0	1	2	3
1	2	0	2	1	0	0	3	1	0

Malgré les mêmes moyennes et mêmes variances, ces lignes ne possèdent pas la même structure spatiale.

C'est précisément de l'identification de telles structures (analyse structurale) et de leur utilisation, qu'il va être question dans la première partie de la géostatistique.

En géostatistique, le premier niveau de modélisation mathématique considère que les valeurs  $\{z_i (i = 1 \dots n)\}$  observées aux temps  $\{t_i (i = 1 \dots n)\}$  sur un ensemble de supports  $\{x_i (i = 1 \dots n)\}$  dans un domaine D sont structurées selon une fonction déterministe  $z(x,t)$ .

Le concept de fonction est utilisé uniquement pour indiquer qu'en un point x et à un instant t, il ne peut y avoir qu'une seule valeur  $z(x,t)$ .

Nous considérons implicitement que les valeurs sont mesurées ou observées en même instant t de sorte que nous ne traitons pas la dimension temporelle : nous écrivons simplement :  $z(x)$  pour faire référence à la valeur en x et z pour désigner la fonction elle-même.

### Modèle probabiliste

Afin de traduire la variation spatiale de la variable régionalisée sous la forme d'un modèle probabiliste ou stochastique, il faut considérer que :

- chaque valeur  $z(x)$  est une réalisation de la variable aléatoire  $Z(x)$  pour tout point  $x \in D$ ,
- que les variables aléatoires  $Z(x) \in \text{domaine } D$  ne sont pas indépendantes les unes des autres mais liées entre elles par une structure de corrélation.

**L'objectif central de la géostatistique est de**

Proposer des modèles de FA (fonction aléatoire) qui puissent rendre compte de façon suffisamment fine de la variation spatiale exhibée par la VR, tout en autorisant des opérations mathématiques les moins restrictives possibles.

**Ces deux objectifs étant généralement contradictoires.**

Le choix d'un type de fonction aléatoire étant fait, la géostatistique propose d'opérer dans l'espace probabilisé, puis d'exprimer les résultats de ces opérations sur le même plan que la VR.

**L'espérance d'une VA** en un point  $x$  est généralement considérée comme une fonction de  $x$  :

$$E(Z(x)) = m(x)$$

avec  $m(x)$  la fonction de moyenne, ou plus simplement la moyenne de la fonction aléatoire  $Z(x)$ .

Lorsqu'elle est définie, **la variance de  $Z(x)$**  est également exprimée comme une fonction de  $x$

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

et si la variance est définie, **la covariance entre deux VA**  $Z(x_1)$  et  $Z(x_2)$  l'est également :

$$\text{Cov}(Z(x_1), Z(x_2)) = E((Z(x_1) - m(x_1))(Z(x_2) - m(x_2)))$$

On peut toujours considérer une réalisation  $y(x)$  d'une F.A.  $Y(x)$  comme une variable régionalisée.

Inversement, on peut interpréter une V.R. donnée  $y(x)$  comme une réalisation d'une certaine F.A.  $Y(x)$  : cette interprétation permet d'appliquer aux V.R. les résultats de la théorie probabiliste des F.A.

**Remarques :**

1/ On ne peut jamais dire qu'une V.R. donnée  $y(x)$  est une F.A. Cela n'aurait pas plus de sens que de dire : le nombre 98 est une V.A.

L'énoncé correct de l'hypothèse probabiliste de base que nous désirons introduire est :  $y(x)$  est une réalisation d'une F.A.  $Y(x)$ .

2/ Pour que cette hypothèse probabiliste ait une signification réelle, il faut que l'on puisse reconstituer au moins en partie la loi de la F.A.  $Y(x)$  dont la V.R.  $y(x)$  est supposée être une réalisation.

Cela suppose que l'inférence statistique soit possible.

Or, l'inférence statistique n'est en général pas possible si l'on ne dispose que d'une seule réalisation  $y(x)$  de  $Y(x)$  (de même, on ne peut pas reconstituer la loi d'une V.A.  $Y$  à partir du résultat numérique  $y = 98$  d'une épreuve unique).

Pour que l'inférence statistique soit possible, il est nécessaire d'introduire des hypothèses supplémentaires sur la F.A.  $Y(x)$ , de manière à réduire le nombre des paramètres dont dépend sa loi.

Tel est le but de l'hypothèse stationnaire que nous allons définir :

Une F.A. stationnaire se répète en quelque sorte elle-même dans l'espace, et cette répétition rend à nouveau possible l'inférence statistique à partir d'une réalisation unique.

Ne disposant que d'une seule VR, pour réaliser l'inférence des paramètres de la FA, il faut trouver l'information nécessaire parmi les valeurs  $(z(x_i), i=1 \dots n)$ .

Faire l'hypothèse de la stationnarité revient à compenser l'absence de plusieurs réalisations de la FA par une forme de redondance de l'information au sein d'une seule réalisation (la VR).

Il convient toutefois de distinguer plusieurs formes de stationnarité d'une FA

## **Stationnarité d'ordre 2**

Une fonction aléatoire est dite stationnaire à l'ordre 2 (FAST-2) si la covariance existe et ne dépend que du vecteur inter-support  $h$ , ce qui implique que l'espérance et la variance existent et ne dépendent pas de  $x$  soit :

$$E(Z(x)) = m$$

$$\text{Var}(Z(x)) = E((Z(x) - m)^2) = C(0)$$

$$\text{Cov}(Z(x), Z(x+h)) = E(Z(x), Z(x+h)) - m^2 = C(h)$$

## **Hypothèse intrinsèque**

Dans le cas des VR qui présentent une variation spatiale qui n'apparaît pas bornée, au moins au sein du domaine d'étude  $D$ . Il n'est pas réaliste d'employer une FAST-2 et il convient d'affaiblir encore davantage l'hypothèse de stationnarité.

Une fonction aléatoire est intrinsèque à l'ordre 0 (FAI-0) si ses accroissements d'ordre 1 sont stationnaires d'ordre 2. Autrement dit, si les espérances et les variances des incréments

$Z(x+h) - Z(x)$  existent et ne dépendent pas de  $x$ , soit :

$$E(Z(x+h) - Z(x)) = m$$

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

Avec  $\gamma(h)$  une fonction nommée demi-variogramme, ou selon l'usage le plus répandu « variogramme ».

Afin de simplifier les développements mathématiques, on considère généralement que la dérive constante est nulle, soit  $m = 0$ .

Ce qui donne la définition du variogramme la plus fréquemment rencontrée dans la littérature :

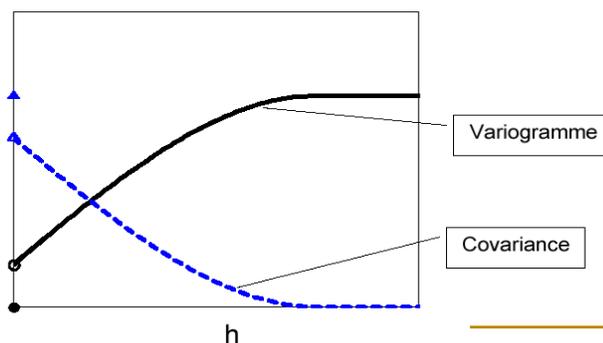
$$\gamma(h) = \frac{1}{2} E \left[ \{Z(x+h) - Z(x)\}^2 \right]$$

Lorsque les variogrammes montrent un palier alors on peut facilement établir le lien entre la valeur du variogramme pour la distance  $h$  et la covariance pour deux observations séparées de  $h$ .

$$\begin{aligned} \gamma(h) &= \frac{1}{2} \text{Var} ( Z(x) - Z(x+h) ) \\ &= \frac{1}{2} [ \text{Var} ( Z(x) ) + \text{Var} ( Z(x+h) ) - 2 \text{Cov} ( Z(x) , Z(x+h) ) ] \\ &= \sigma^2 - \text{Cov} ( Z(x) , Z(x+h) ) = \sigma^2 - C(h) \end{aligned}$$

#### Lien entre variogramme et covariance

$$\gamma(h) = \sigma^2 - \text{Cov}(h)$$



Certaines propriétés du milieu naturel peuvent présenter un gradient se traduisant par une tendance. C'est le cas lorsque ni la variance ni la covariance ne peuvent être définies.

L'espérance mathématique n'est donc plus constante et elle est fonction de la position des points d'appui de  $x$  ; on aura donc :

$$E(Z(x)) = m(x) \quad \text{et} \quad E(Z(x) - Z(x+h)) \neq 0$$

Dans ce cas, l'estimation du variogramme par l'équation habituelle n'est plus valable et devient :

$$\gamma_{\text{réel}}(h) = \gamma_{\text{estimé}}(h) - 1/2[m(x) - m(x+h)]^2$$

Le résultat essentiel à retenir est que l'affaiblissement progressif de l'hypothèse de stationnarité s'accompagne d'une abstraction croissante du modèle topo-probabiliste tout en autorisant la modélisation de classes de phénomènes plus vastes.

## **Notion de variable régionalisée**

La géostatistique est l'application de la théorie des variables régionalisées à un phénomène qui se déploie dans l'espace et y manifeste une certaine structure, qu'il est régionalisé.

Si  $f(x)$  désigne la valeur au point  $z$  d'une caractéristique  $f$  de ce phénomène, nous dirons que  $f(x)$  est une variable régionalisée, en abrégé V.R.

C'est là un terme neutre, purement descriptif, antérieur, en particulier, à toute interprétation probabiliste.

Du point de vue mathématique, une V.R. est donc simplement une fonction  $f(z)$  du point  $z$ , mais c'est, en général, une fonction fort irrégulière.

Elle se présente sous deux aspects contradictoires (ou complémentaires) :

- ✓ un aspect aléatoire (haute irrégularité, et variations imprévisibles d'un point à l'autre).
- ✓ un aspect structuré (elle doit cependant refléter à sa manière les caractéristiques structurales du phénomène régionalisé).

La théorie des V.R. se propose donc deux objectifs principaux :

- sur le plan théorique, imprimer ces caractéristiques structurales sous une forme mathématique adéquate ;

- sur le plan pratique, résoudre le problème de l'estimation d'une V.R. à partir d'un échantillonnage fragmentaire.

### **Idée fondamentale :**

La nature n'est pas entièrement "imprévisible". Deux observations situées l'une près de l'autre devraient, en moyenne, se ressembler davantage que deux observations éloignées.

Exemple :

Soit trois localisations  $x_0$ ,  $x_1$  et  $x_2$ , Dans un domaine D. On mesure la teneur (valeur) en chacun de ces points.

$x_1$                        $x_0$                        $x_2$

La teneur au point  $x_1$  devrait ressembler plus (en moyenne) à celle observée en  $x_0$  qu'à celle en  $x_2$ .

On a peut-être intérêt à utiliser l'information contenue en  $x_1$  et  $x_2$  pour fournir un meilleur estimé de  $x_0$  que si l'on n'utilisait que  $x_1$ .

**Notion de "continuité"** : Implicitement toutes les méthodes d'estimation reposent sur ce concept plus ou moins défini.

En géostatistique, on cherche à quantifier cette continuité préalablement à tout calcul effectué dans le domaine D.

Soit deux points  $x$  et  $x+h$  séparés d'une distance  $h$  :  $x$    $x+h$

La teneur en  $x$  est une variable aléatoire  $Z(x)$ .

La teneur en  $x + h$  aussi,  $Z(x+h)$ .

La différence entre les valeurs prises par ces deux V.A. est  $Z(x) - Z(x+h)$ .

C'est également une V.A. dont on peut calculer la variance. Cette variance devrait être plus petite lorsque les points sont rapprochés (les valeurs se ressemblent plus en moyenne) et plus grande lorsque les points sont éloignés.

On appelle variogramme la demi-variance de cette différence :

$$\gamma(x, x+h) = 0.5 * \text{Var}(Z(x) - Z(x+h))$$

**Comment obtenir cette formule :**

$$(x_1 - x_m)^2 + (x_2 - x_m)^2$$

$$(x_1^2 + x_m^2 - 2x_1 x_m) + (x_2^2 + x_m^2 - 2x_2 x_m)$$

$$x_m = (x_1 + x_2) / 2$$

$$(x_1^2 + ((x_1 + x_2) / 2)^2 - 2x_1 (x_1 + x_2) / 2) + (x_2^2 + ((x_1 + x_2) / 2)^2 - 2x_2 (x_1 + x_2) / 2)$$

ce qui donne

$$\frac{1}{2} (x_1 + x_2)^2 \text{ comme } x_2 = x_1 + h \text{ et } x \text{ est exprimé par } z(x)$$

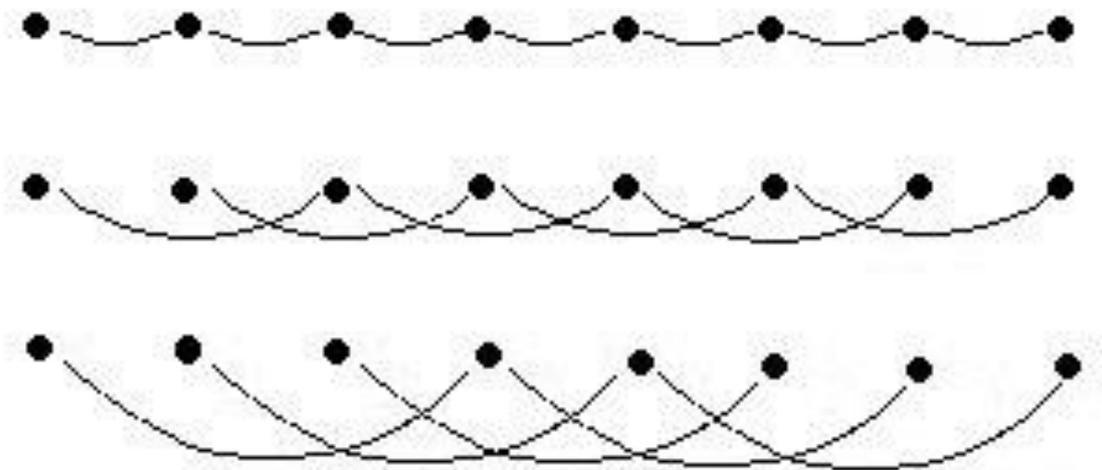
$$\frac{1}{2} (z(x) - z(x+h))^2 \text{ pour } n \text{ couples on aura}$$

$$\frac{1}{2} n (z(x_i) - z(x_i+h))^2 = \gamma(h)$$

**Calcul du Variogramme :**

On cherche à construire un graphique représentant en abscisse les distances h séparant les points et en ordonnée les semi-variances.

La construction du variogramme est illustrée ci-dessous par des schémas établis à partir de 8 points d'observation répartis à distance égale de 1 mètre le long d'un transect



8 individus

Distance 1 mètre : 7 couples

Distance 2 mètres : 6 couples

Distance 3 mètres : 5 couples

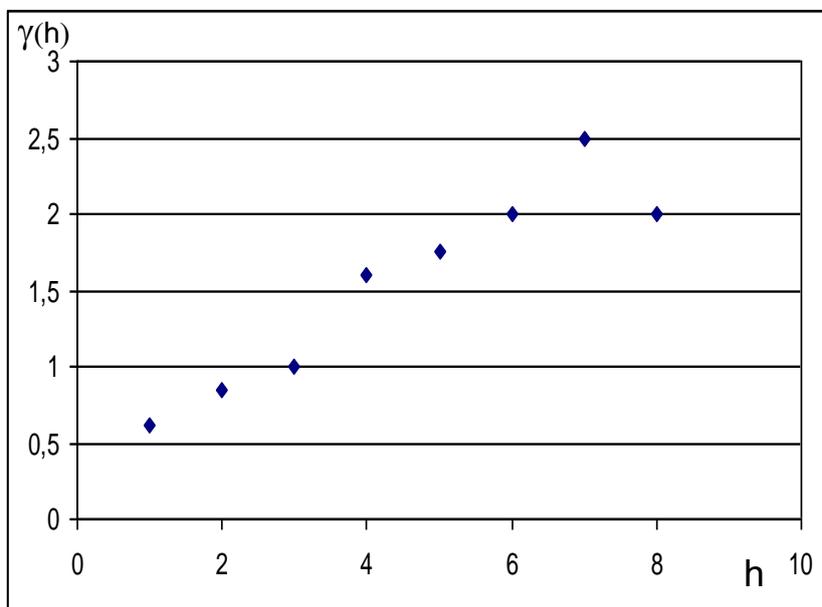
Ce schéma montre que le nombre de points participant au calcul du variogramme diminue au fur et à mesure que la distance augmente. Les valeurs de semi-variance risquent donc d'être moins précises pour les grandes valeurs de  $h$ .

### Exercice :

Soit un exemple fictif correspondant à des observations disposées le long d'un transect à des intervalles réguliers de 1 mètre.

Calculer  $\gamma(h)$  (variogramme)

1 2 1 1 2 1 3 4 3



h	$\gamma(h)$
1	0,62
2	0,85
3	1
4	1,6
5	1,75
6	2
7	2,5
8	2

## Structure d'un variogramme :

Exemples : Calculer moyennes, variances, variogrammes

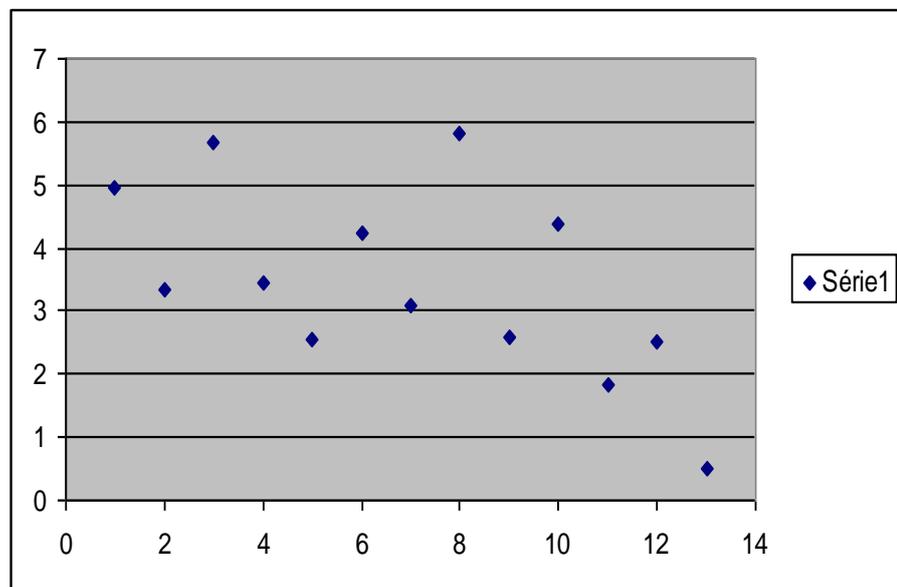
5 7 3 4 1 5 7 2 3 1 6 2 6 4

5 2 3 1 1 2 4 3 4 5 6 7 6 7

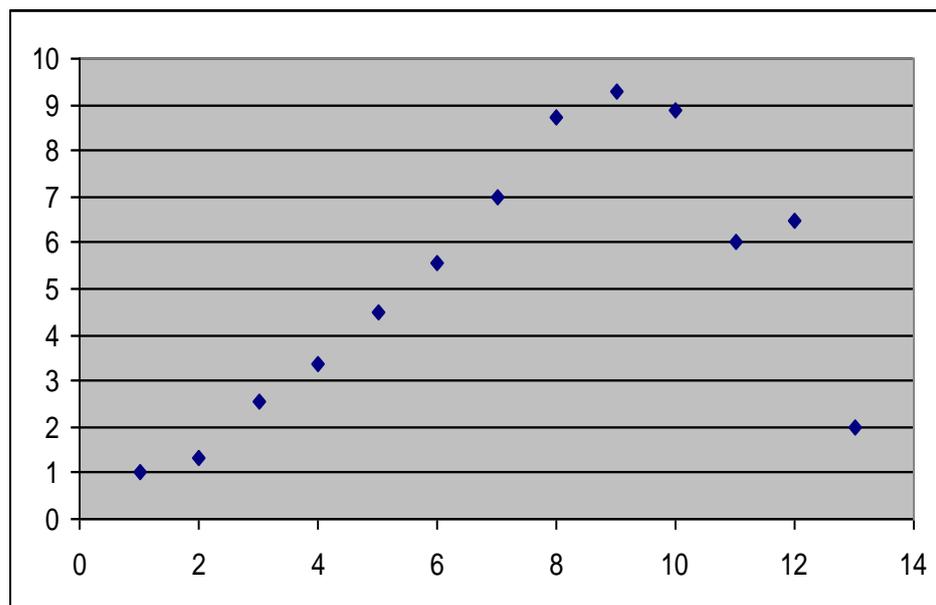
Moyenne = 4,0

Variance = 4,31

1	4,96154
2	3,33333
3	5,68182
4	3,45
5	2,55556
6	4,25
7	3,07143
8	5,83333
9	2,6
10	4,375
11	1,83333
12	2,5
13	0,5



1	1
2	1,33
3	2,54
4	3,35
5	4,5
6	5,56
7	7
8	8,75
9	9,3
10	8,87
11	6
12	6,5
13	2



## Variogrammes directionnels :

Pour un champ donné, rien n'assure que la continuité soit identique dans toutes les directions.

Par exemple, il se pourrait que des teneurs montrent une meilleure continuité parallèlement à la stratigraphie que perpendiculairement à celle-ci. De même, pour la pluviométrie, on pourrait observer une meilleure continuité parallèlement à la pente que verticalement en raison de l'influence de l'altitude.

Si le nombre d'observations le permet (typiquement au moins 50, préférablement 100), on peut chercher à vérifier ce point en calculant le variogramme expérimental dans différentes directions.

On peut aussi calculer le variogramme selon certaines directions spécifiques :

$$\gamma_e(h, \theta) = \frac{1}{2 N(h, \theta)} \sum_{i=1}^{N(h, \theta)} [Z(x_i) - Z(x_i + h)]^2$$

Où :  $N(h, \theta)$  = nombre de paires séparées de  $h$  dans la direction  $\theta$ .

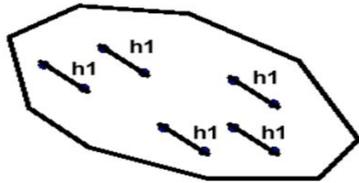
En pratique on s'accorde une tolérance sur  $h$  et sur  $\theta$  afin d'avoir suffisamment de paires pour chaque  $h$  et chaque  $\theta$ .

Pour chacune des classes ainsi formées, on calcule la distance moyenne séparant les extrémités des paires (abscisse) et on évalue le variogramme expérimental pour chaque classe.

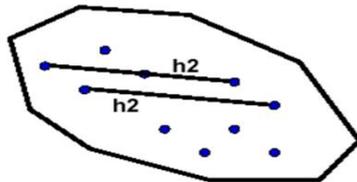
On obtient donc une série de points expérimentaux auxquels on cherche à ajuster un modèle théorique.

Une fois le modèle adopté, toute la suite des calculs se fait avec les **valeurs obtenues du modèle** et non avec les valeurs expérimentales.

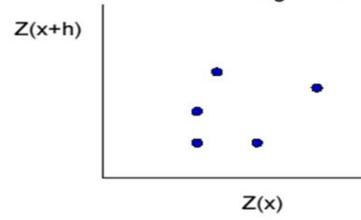
Faire varier « h »



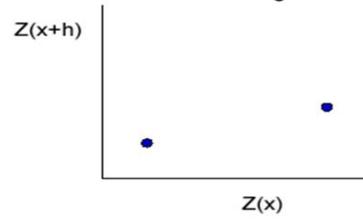
Gisement



« h1 scattergram »

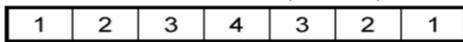


« h2 scattergram »

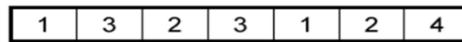
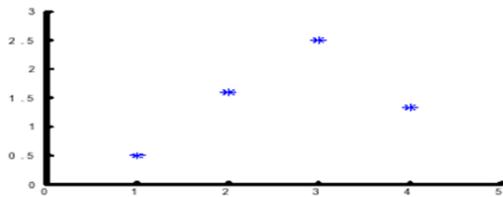


## Exemple

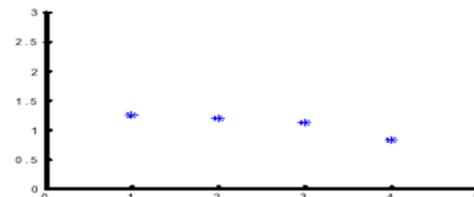
h=1



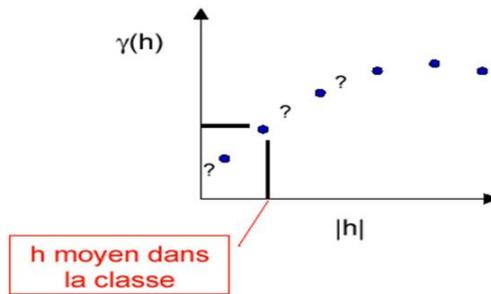
h	N(h)	$\gamma(h)$
1	6	0.5
2	5	1.6
3	4	2.5
4	3	1.33



h	N(h)	$\gamma(h)$
1	6	1.25
2	5	1.25
3	4	1.125
4	3	0.83



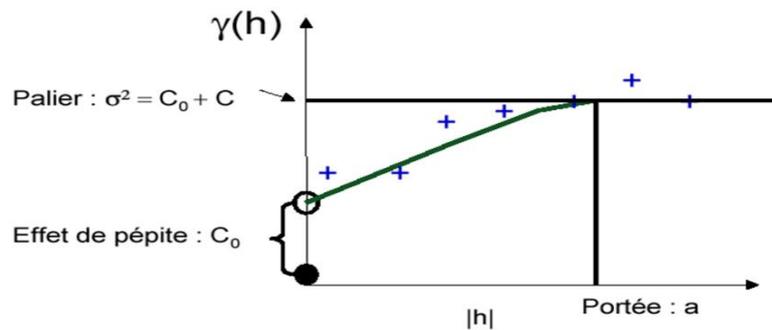
## Variogramme expérimental



Dans les calculs géostat, on doit connaître Cov ou  $\gamma$  pour **tout** « h »

↓  
**Modèle**

Généralement,



Cette fonction, habituellement croissante en fonction de  $h$ , synthétise beaucoup d'informations concernant le comportement conjoint des variables aléatoires et concernant "la continuité" du phénomène.

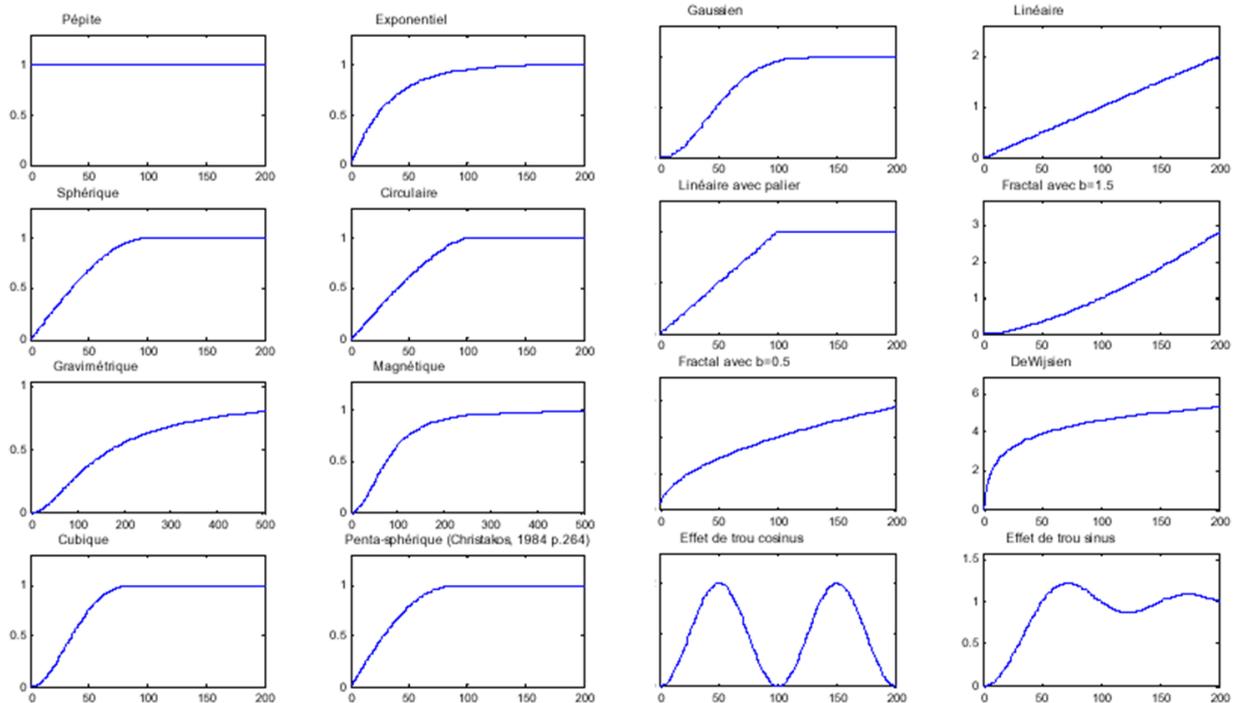
**Ainsi, pour les modèles de variogramme montrant un seuil, on a :**

Portée a : Distance (au-delà de laquelle) deux observations ne se ressemblent plus du tout en moyenne, elles ne sont plus liées (covariance nulle) linéairement. À cette distance, la valeur du variogramme correspond à la variance de la variable aléatoire.

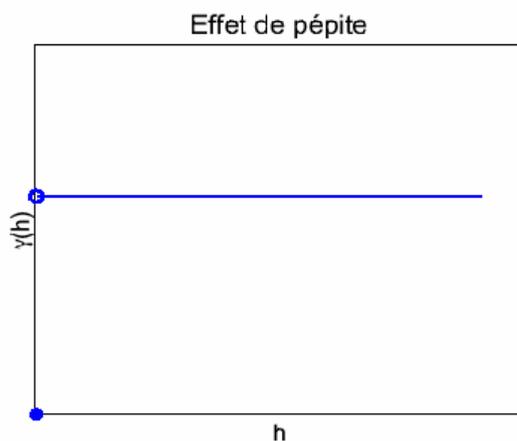
Palier ;  $\hat{\sigma}^2 = C_0 + C$  : Variance de la v.a. ( $\text{Var}(Z(x))$ ); Écart les plus grands, en moyenne, entre deux v.a.

Effet de pépite C0 : Variation à très courte échelle, erreurs de localisation, erreurs d'analyse et précision analytique.

## Exemples de modèle



### Les principaux modèles :



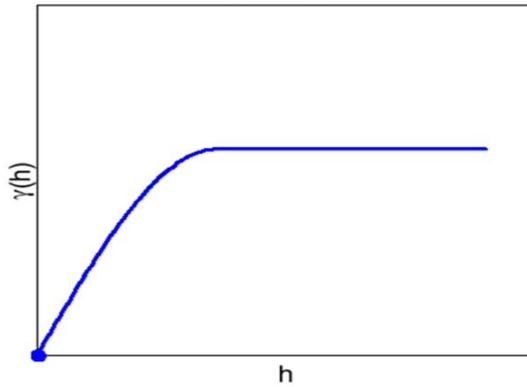
$$\gamma(h) = 0 \quad \text{si } h = 0$$

$$C_0 \quad \text{si } h > 0$$

- Erreurs de mesure
- Erreurs de localisation
- Microstructure non-identifiable dû au manque de données

Presque toujours présent mais rarement seul  
Effet de pépité pur => estimation impossible

### Sphérique



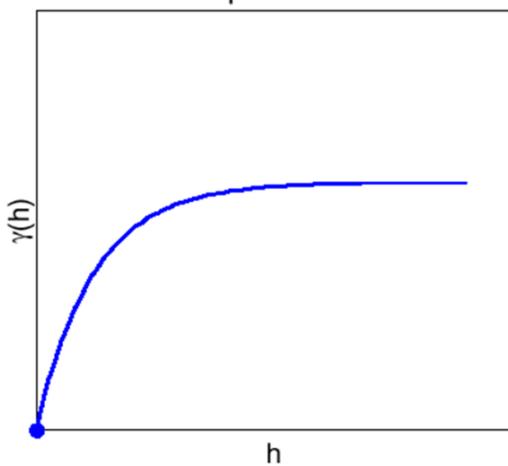
Modèle le + fréquent

e.g. teneur, épaisseur, ...

Combiné avec effet pépité

$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C \left( 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 \right) & \text{si } 0 < h < a \\ C & \text{si } h \geq a \end{cases}$$

### Exponentiel

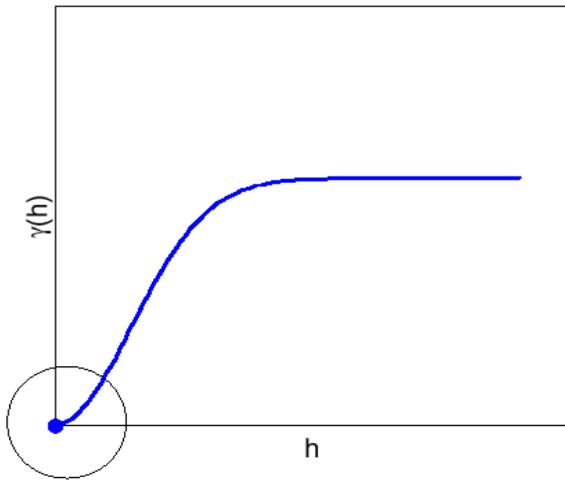


Assez commun

Semblable au sphérique

$$\gamma(h) = C \left( 1 - \exp\left(\frac{-|h|}{a'}\right) \right) \text{ ou } C \left( 1 - \exp\left(\frac{-3|h|}{a}\right) \right) \left. \begin{array}{l} a: \text{ portée effective } \gamma(h)=0.95 \cdot C \\ a'=a/3 \end{array} \right\}$$

### Gaussien

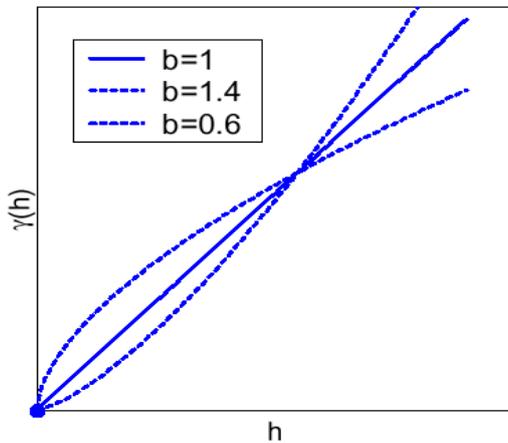


- Peu fréquent en mine
- Variables très continues : e.g. topographie, gravimétrie, magnétisme, épaisseur, ...
- Problèmes numériques si absence d'effet de pépite

a: portée effective  $\gamma(h)=0.95 \cdot C$   
 $a' = a/3^{0.5}$

$$\gamma(h) = C \left( 1 - \exp \left( - \left( \frac{|h|}{a'} \right)^2 \right) \right) \text{ ou } C \left( 1 - \exp \left( - 3 \left( \frac{|h|}{a} \right)^2 \right) \right)$$

### $h^b$



- Modèles sans palier
- Moyenne, variance et covariance non définies

$$\gamma(h) = C \left( \frac{|h|}{a} \right)^b \quad |h| > 0, \quad 0 \leq b < 2$$

## Références bibliographique :

- Armstrong, M., & Carignan, J. (1997). *Géostatistique linéaire: application au domaine minier*: Presses des MINES.
- Benamghar, A. (2007). *Polycopié du cours de géostatistique*. Ecole Nationale Polytechnique.
- Chauvet, P. (2008). *Aide-mémoire de géostatistique linéaire*: Presses des MINES.
- Chiles, J., & Delfiner, P. (1999). Kriging. *Geostatistics: Modeling Spatial Uncertainty*, 150-230.
- Chilès, J., & Delfiner, P. (2012). Wiley series in probability and statistics. *Geostatistics: Modeling Spatial Uncertainty*, 705-714.
- Cressie, N. A. (1993). Statistics for spatial data: Wiley series in probability and mathematical statistics. *Find this article online*.
- Koneshloo, M. (2007). *Caractérisation, estimation et valorisation de gisements d'argiles kaoliniques du bassin des Charentes*. École Nationale Supérieure des Mines de Paris.
- Marcotte, D. (2003). *Ph.D.(Poly) Géostatistique*. École Polytechnique de Montréal. <http://www.groupes.polymtl.ca/geo/marcotte/>.
- Matheron, G. (1965). *Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature*: Masson et CIE.
- Matheron, G. (1971). The theory of regionalized variables and its applications, volume 5 of Les cahiers du Centre de Morphologie Mathématique de Fontainebleau. *École Nationale Supérieure des Mines de Paris*.
- Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in applied probability*, 5(3), 439-468.
- Matheron, G. (1989). Estimating and choosing. An essay on probability in practice. Translated from the French and with a preface by AM Hasofer: Springer-Verlag.
- Tillé, Y. (2010). Résumé du Cours de Statistique Descriptive.
- Wackernagel, H. (2004). *Géostatistique et assimilation séquentielle de données*. Université Pierre et Marie Curie-Paris VI.

## **Sites internet**

- <https://perso.univ-rennes1.fr/helene.guerin/enseignement/MSB/varBio.pdf>
- [http://www.irisa.fr/prive/kadi/Cours\\_LR2V/Divers\\_Cours/cours-probabilit%E9.pdf](http://www.irisa.fr/prive/kadi/Cours_LR2V/Divers_Cours/cours-probabilit%E9.pdf)
- [http://www.irisa.fr/prive/kadi/Cours\\_LR2V/Divers\\_Cours/cours-probabilit%E9.pdf](http://www.irisa.fr/prive/kadi/Cours_LR2V/Divers_Cours/cours-probabilit%E9.pdf)
- <http://www.geosciences.mines-paristech.fr/fr/equipes/geostatistique/historique-1>
- <https://elearn2013.univouargla.dz/main/document/showinframes.php?cidReq=GEOSTATISTIQUE&&curdirpath=/&file=/geostatistique.html>
- <http://www.geosciences.mines-paristech.fr/fr/equipes/geostatistique/themes-de-recherche-1>
- <https://fr.wikipedia.org/wiki/Variogramme>
- <http://www.groupe.polymtl.ca/geo/marcotte/glq3401geo/chapitre2.pdf>
- <http://www.groupe.polymtl.ca/geo/marcotte/mth2302c/chapitre9.pdf>