

CHAPITRE 2 : ANALYSE DU VARIOGRAMME

2.1. Généralités

La première loi de Tobler :

" All things are similar, but nearby things are more similar than distant things."

C'est à dire :

" Toutes choses sont similaires, mais les choses proches sont plus semblables que les choses lointaines."

La mesure géostatistique qui décrit le taux de variation de la variable régionalisée est connue comme la semi-variance ou semi-variogramme.

La semi-variance est utilisée pour l'analyse descriptive où la structure spatiale des données est étudiée en utilisant le semi-variogramme et pour des applications prédictives où le semi-variogramme est équipé d'un modèle théorique, paramétré et utilisé pour prédire la variable régionalisée à d'autres points non mesurés (krigeage).

2.2. Variables régionalisées

Une variable qui prend des valeurs en fonction de sa **localisation** spatiale est connue comme une variable régionalisée.

Pour une variable z mesurée à l'emplacement i , on peut partitionner la variabilité totale de z en trois composantes :

$$z(i) = s(i) + f(i) + \varepsilon$$

Où:

$f(i)$ effet d'échelle ; $s(i)$ dépendance spatiale locale ; ε variance d'erreur

Les variables régionalisées sont des variables qui sont comprises entre des variables aléatoires et les variables déterministes.

À l'inverse des variables aléatoires, les variables régionalisées présentent une continuité spatiale.

Le variogramme est utilisé pour décrire des variables régionalisées.

2.3. Variogramme

D'après Marcotte¹, 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. La différence entre les valeurs prises par deux variables aléatoires est $Z(s) - Z(s+h)$. C'est également une variable aléatoire 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.

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

L'outil mesure la « variabilité spatiale », c'est-à-dire la dissemblance entre les valeurs en fonction de leurs séparations. Il décrit la continuité spatiale de la variable régionalisée.

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.

¹ Denis **Marcotte**, professeur de **géostatistique** au département de génie minéral de l'École Polytechnique de Montréal

Le variogramme : à quoi sert-il ?

Le variogramme est une fonction mathématique utilisée en géostatistique pour **analyser la dépendance spatiale** entre deux observations. Il permet de déterminer si la distribution de la mesure, teneur, etc. possède une **structuration spatiale** sur la parcelle étudiée.

2.3.1 Estimation du semi-variogramme

Variogramme théorique :

Le **variogramme théorique** d'une fonction aléatoire a pour expression :

Le variogramme théorique est défini comme :

$$\gamma(h) = \frac{1}{2} \text{Var}(Z(s) - Z(s+h)) = \frac{1}{2} E\{[Z(s) - Z(s+h)]^2\} = C(0) - C(h)$$

.avec $C(0)=\text{Var}(Z(s))$ et $C(h)=\text{Cov}(Z(s),Z(s+h))$

Remarque :

- Un variogramme peut se calculer non seulement pour une distance donnée mais aussi pour direction θ donnée : $\gamma_\theta(h)$
- La covariance mesure la ressemblance entre les valeurs en fonction de leur éloignement alors que le variogramme mesure la dissemblance entre les valeurs en fonction de leur éloignement.
- Dans l'hypothèse de stationnarité d'ordre 2, covariance et variogramme existent et sont liés par la relation $\gamma(h) = C(0) - C(h)$. Dans l'hypothèse intrinsèque, seul le variogramme existe. C'est pourquoi il est généralement préféré à la covariance pour décrire et interpréter la structure spatiale du phénomène étudié.
- Le variogramme réel d'une fonction aléatoire est généralement inconnu, mais il peut être évalué à partir des données d'échantillonnages. On obtient ainsi le variogramme expérimental proposé par Matheron (1962).

Le variogramme expérimental :

Le variogramme **expérimental** ou **empirique** est un estimateur du variogramme théorique à partir des données.

Le variogramme réel d'une fonction aléatoire est inconnu mais il peut être évalué à partir des données d'échantillonnage. On obtient ainsi le **variogramme expérimental** :

$$\hat{\gamma}(s, s') = \hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{N(h)} [Z(s_i) - Z(s_i + h)]^2$$

Où $N(h)$ nombre de couples de points de mesure distants de h
 $z(s_i)$ valeur au point de mesure s_i

Ce variogramme expérimental n'est pas défini partout, notamment aux distances h pour lesquelles il n'existe aucune paire de points de mesure. Aussi lui est-il ajusté un **modèle de variogramme**, qui est la somme d'une ou plusieurs fonctions mathématiques.

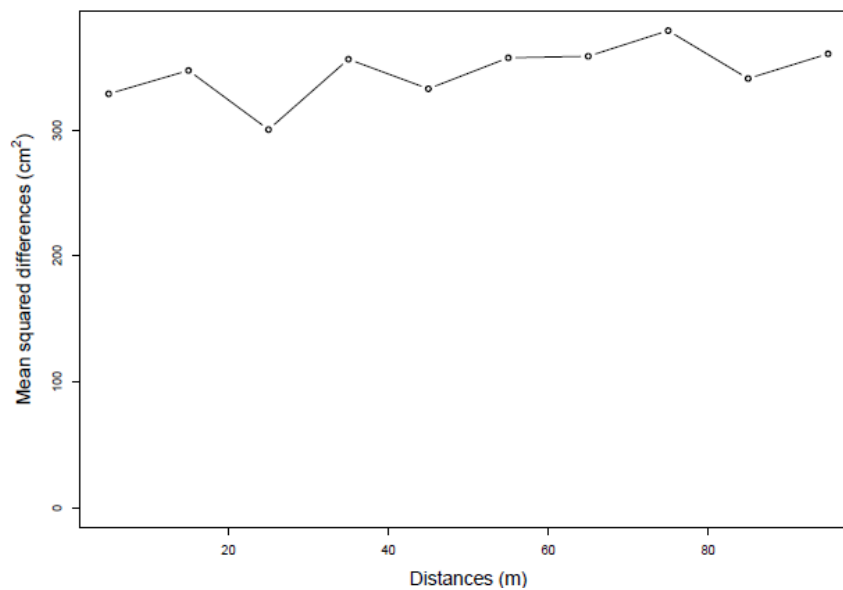
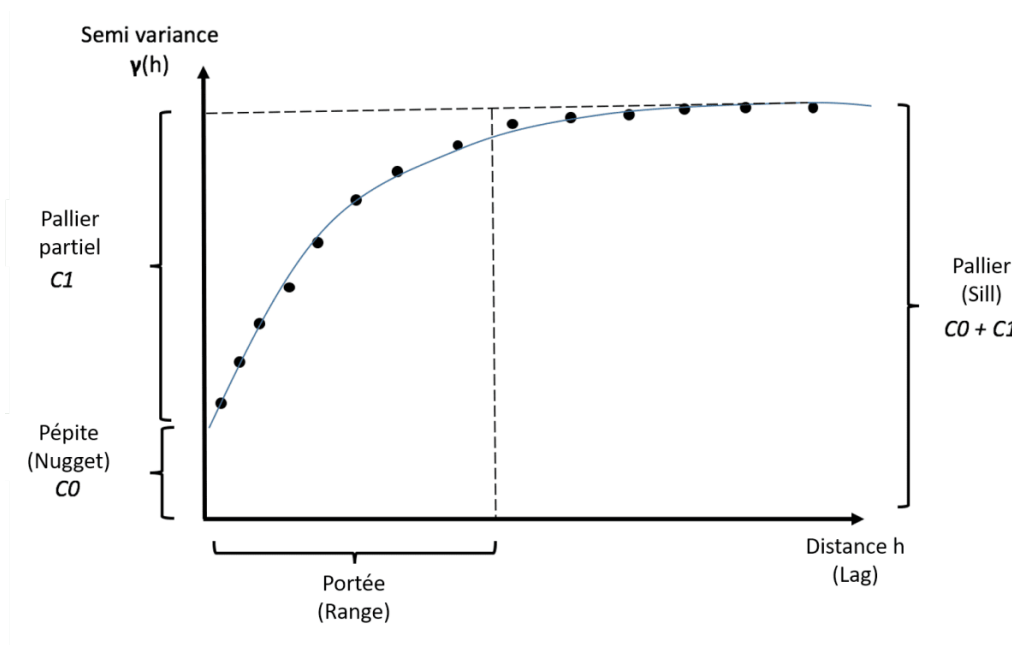


Figure 2.1 : Variogramme empirique des profondeurs de sol. L'écart quadratique moyen ne semble pas dépendre de la distance qui sépare les points. La variable n'est pas structurée spatialement.

2.3.3 Propriétés du semi-variogramme

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

La figure suivante présente un modèle **théorique** de variogramme (sphérique) où la semi-variance $\gamma(h)$ est fonction de l'intervalle d'échantillonnage (h).



Remarque : plus la fonction croît, moins les observations se ressemblent.

Le variogramme est une fonction de h , croissante et souvent caractérisé par les paramètres suivants :

- la portée : a
- le palier : $C_1 + C_0$
- l'effet pépite : C_0

1-La **Portée "a"** : 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.

C'est aussi la distance en deçà de laquelle la dépendance spatiale est apparente (l'étendue spatiale de la structure dans les données)

2-Le **seuil** ou **Palier** $\sigma^2 = C_0 + C_1$: est la valeur à laquelle les niveaux du semi-variogramme s'annulent (valeur asymptotique).

C'est aussi : la variance de la v.a. ($\text{Var}(Z(s))$) ou encore Écarts les plus grands, en moyenne entre deux v.a.

3-Effet de **pépite** C_0 : est la semi-variance à une distance (0, 0).

C'est aussi : la variation à très courte échelle, erreurs de localisation, erreurs d'analyse et précision analytique.

Exemple : Une carotte fendue en deux et dont chaque partie est analysée séparément ne fournira pas exactement les mêmes valeurs pour les deux moitiés. Un même paquet de poudre, séparé en deux parties pour analyse ne donnera pas exactement la même teneur.

N.B. :

- Lorsque $h = 0$ on a $\gamma(0) = 0$ et non c_0

Par contre, $\lim_{\varepsilon \rightarrow 0+} \gamma(\varepsilon) = c_0$, C'est-à-dire on a une discontinuité à l'origine du variogramme.

- Parfois les variogrammes ne montrent pas de palier (dans ce cas, la covariance et la variance n'existent pas).
- 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 .

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

Donc

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

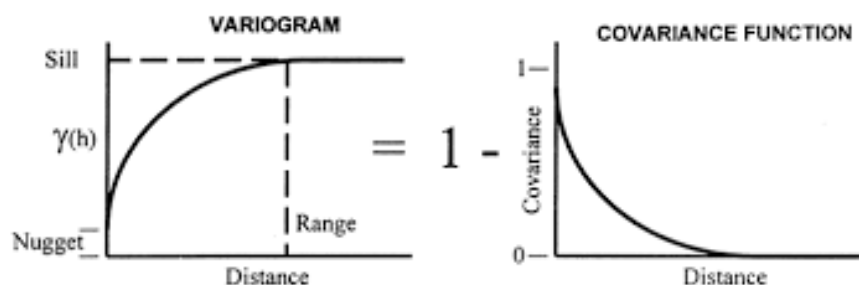
$C(h)$ est appelé le covariogramme de Z . Cette relation est importante et elle est continuellement utilisée en géostatistique.

On voit que lorsque la portée est atteinte, il n'y a plus de covariance entre les v.a., i.e. $C(h) = 0$ si $h \geq a$.

Lorsqu'il y a un palier, les deux fonctions sont équivalentes en ce sens qu'elles fournissent la même information sur le processus.

Le variogramme possède toutefois deux avantages sur le covariogramme.

- Le variogramme est défini même s'il n'y a pas de palier.
- Dans l'expression du variogramme, la constante "m" n'apparaît pas et l'on n'a donc pas besoin de l'estimer comme c'est le cas lorsqu'on veut calculer directement le covariogramme.



Graettinger (1998) Figure 2.4

Chaque phénomène géologique possède un variogramme qui lui est propre. Ainsi,

- Un gisement d'or présentera un variogramme erratique avec un fort effet de pépite et une faible portée.
- Un gisement de cuivre porphyrique montrera un variogramme linéaire à l'origine avec faible effet de pépite et grande portée.
- Un gisement sédimentaire de fer présentera une portée plus grande parallèlement à la stratification que perpendiculairement à celle-ci (anisotropie géométrique).
- La topographie pourra présenter un variogramme très continu avec comportement parabolique à l'origine et absence d'effet de pépite.

2.4 Hypothèses de l'Autocorrélation ?

2.4.1 Stationnarité : L'autocorrélation suppose la stationnarité, ce qui signifie que la structure spatiale de la variable est homogène sur l'ensemble des données.

Concrètement La stationnarité veut dire que deux paires de points espacés d'un même vecteur h ont des caractéristiques (moyenne et covariance) semblables. Ou encore, la variable régionalisée ne dépend pas de sa position dans l'espace, elle garde les mêmes caractéristiques où que l'on se place.

Cela nous permet de nous détacher de la localisation et de nous restreindre uniquement à la distance qui sépare les points d'observations, ce qui nécessite que :

- 1 - la moyenne est constante dans la région
 - 2 - la variance est constante et finie, et
 - 3 - la covariance ne dépend que de la distance entre les échantillons,
- Dans de nombreux cas, ceci n'est pas vrai à cause des grandes tendances dans les données
 - Dans ces cas, les données sont souvent redressées avant l'analyse.
 - Une façon de redressement de données consiste à ajuster une régression à la tendance, et utiliser uniquement les résidus pour l'analyse de l'autocorrélation

2.4.2 Isotropie : L'autocorrélation suppose également l'**isotropie**, ce qui signifie que la structure spatiale de la variable est uniforme dans toutes les directions.

Ceci n'est pas souvent le cas, et la variable montre une **anisotropie**, ce qui signifie qu'il y a une tendance de dépendance à la direction dans les données.

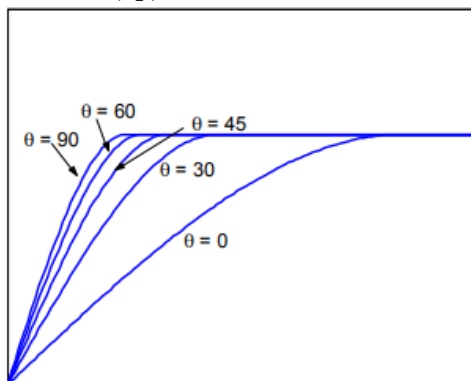
Une **anisotropie** caractérise un phénomène qui se déploie préférentiellement dans certaines directions de l'espace.

- Si un paramètre présente différentes variations avec la direction, ceci implique qu'il y a une **anisotropie géométrique**.

Par exemple, dans un dépôt de dunes, le plus large éventail coïncide avec la direction du vent comparé à celui qui lui est perpendiculaire.

Caractéristiques de l'Anisotropie géométrique

- On observe dans **diverses directions** des paliers et des composantes pépitiques **identiques** mais des portées **différentes**.
- Les portées maximales (ag) et minimales (ap) s'observent selon **deux directions orthogonales**.

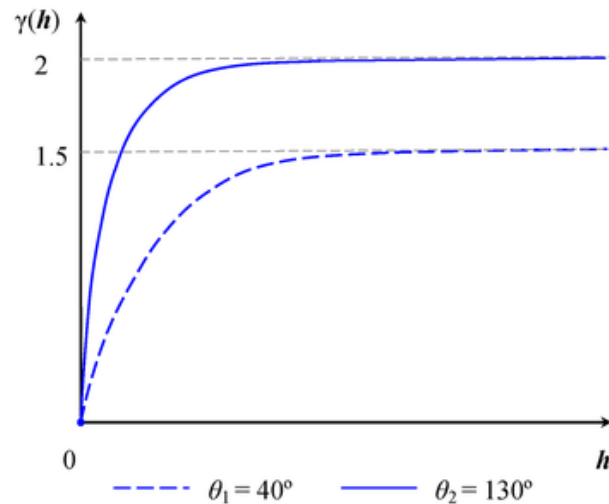


Note: Pour l'anisotropie géométrique, on peut toujours, par simple rotation et dilatation se ramener à un modèle isotrope, c'est ce qui est fait dans la méthode utilisant le calcul de h_g .

- **anisotropie zonale** : toute anisotropie qui n'est pas géométrique.

Donc on est dans le cas d'une somme de composantes isotropes et avec anisotropies géométriques.

On observe des paliers différents selon les directions. Il n'existe pas de modèle d'ajustement pratique pour traiter ce type d'anisotropie.



2.5 L'ajustement d'une fonction de structure théorique

2.5.1 Introduction

On a vu que les fonctions $\hat{C}$ et $\hat{\gamma}$ calculées empiriquement permettaient d'estimer les fonctions théoriques sous-jacentes C et γ . Mais ces fonctions empiriques ne sont pas totalement satisfaisantes pour plusieurs raisons :

Ces courbes empiriques sont souvent assez erratiques, il est donc difficile de résumer en quelques mots comment elles se comportent

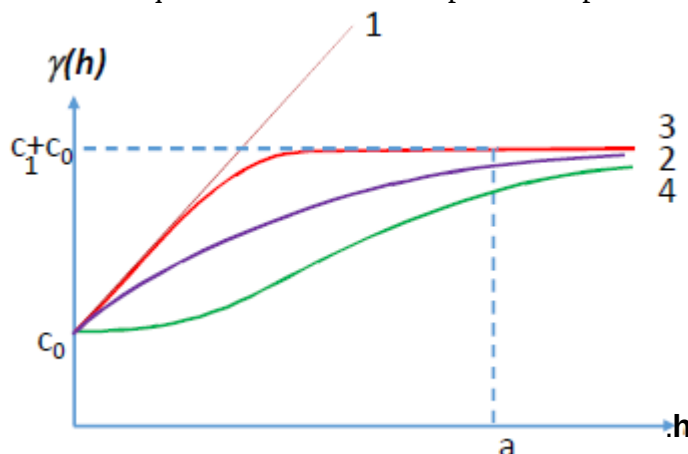
On a vu qu'une fonction de covariance théorique satisfaisait la propriété de défini positivité, condition qui n'est pas assurée par la covariance empirique.

On résout ces deux problèmes en substituant à la fonction de structure empirique une fonction appartenant à une famille paramétrique simple de fonctions défini-positives.

2.5.2 Modélisation du semi-variogramme

Une fois le variogramme **expérimental** est calculé, il faut alors déterminer un **modèle mathématique** qui lui correspond et qui doit être opérationnel et simple à l'emploi : C'est **l'ajustement** du variogramme expérimental. Les schémas théoriques d'usage courant sont classés en schémas à **palier**, schémas **sans palier** et schémas à **effet de trous**.

Illustration des quatre modèles utilisés pour s'adapter aux semi-variogrammes :

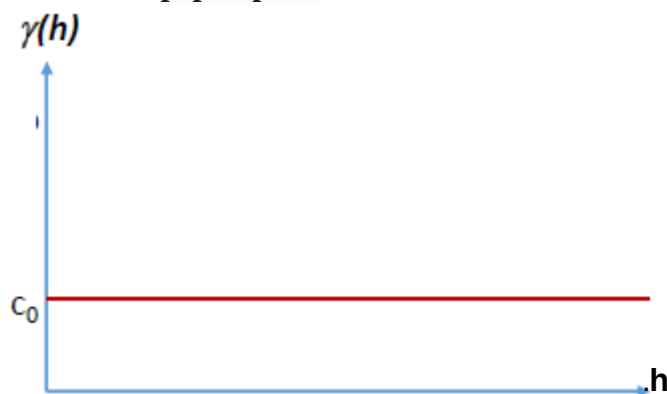


Où	Where
c_0 = pépité	nugget
b = pente de la régression	regression slope
a = portée	range
$c_0 + c_1$ = variance totale	sill

Comportement	Nom du modèle	Formule de la composante $\gamma(h)$	Remarque	N°
Composantes classiques à palier C, et portée a	Sphérique	$\gamma(h) = \begin{cases} c_0 + c \left(\frac{3h}{2a} - \frac{h^3}{2a^3} \right), & 0 \leq h \leq a \\ c_0 + c, & h > a \end{cases}$	Le plus fréquent	3
	Gaussien	$\gamma(h) = c_0 + c \left[1 - e^{-\left(\frac{h}{a}\right)^2} \right]$	Peu fréquent en mine	4
	Exponentiel	$\gamma(h) = c_0 + c \left[1 - e^{-\frac{ h }{a}} \right]$	Assez commun	2
	Linéaire	$\gamma(h) = c_0 + bh$	Existe mais rarement seul	1

N.B. : Si $c_0=0$ cela sous-entend que le processus est régulier, et que la mesure n'est pas entachée d'une variabilité instrumentale ou de l'effet de la micro distribution. Par contre une ordonnée à l'origine indique ($c_0=0$) un effet de pépité.

Illustration du modèle de semi-variance **pépité pur** :



. à pépité pur: la fonction fluctue autour d'une valeur constante ce qui signifie qu'il y a une indépendance totale des observations.

En général, seules des variables ayant cette propriété sont susceptibles d'être traitées par la statistique classique.

Lorsqu'il n'y a aucune dépendance entre des observations, faire une cartographie à base d'interpolation n'a pas plus de sens que de tracer une régression non-significative ($r \approx 0$).

2.5.3 Méthodes d'ajustement

Pour estimer les paramètres de la covariance théorique on commence par choisir une famille paramétrique simple, (en général des combinaisons linéaires à coefficients positifs du type, $C(h) = \sigma^2 I_{pép}(h) + C'(h)$, puis on estime les deux ou trois paramètres réels mis en jeu par l'une des trois manières suivantes :

Ajustement à la main

C'est la méthode la plus simple, on choisit les paramètres du modèle de manière à obtenir le meilleur ajustement visuel, en portant une attention particulière (par ordre décroissant d'importance)

- au comportement à l'origine
- au comportement aux petites distances (linéaire, parabolique)
- à la distance à laquelle le comportement se stabilise.

Moindres carrés

On minimise par rapport au paramètre inconnu θ la somme des erreurs quadratiques

$$\sum_{hy} (\gamma^*(h) - \gamma(h))^2$$

Comme cette méthode donne un poids égal à toutes les classes de distances, elle donne de mauvais résultats. On lui préfère la minimisation du critère suivant :

$$\sum_{hy} (\gamma^*(h) - \gamma(h))^2 \omega(h)$$

Où $\omega(h)$ est une fonction plus ou moins arbitraire qui donne beaucoup de poids à l'origine et peu aux grandes distances.

Vraisemblance

Si on se donne une loi de probabilité paramétrique pour le vecteur aléatoire $(z_1, \dots, z_n)^t$, on peut écrire la vraisemblance de l'échantillon $(z_1, \dots, z_n / \theta)$ et la maximiser par rapport à θ .

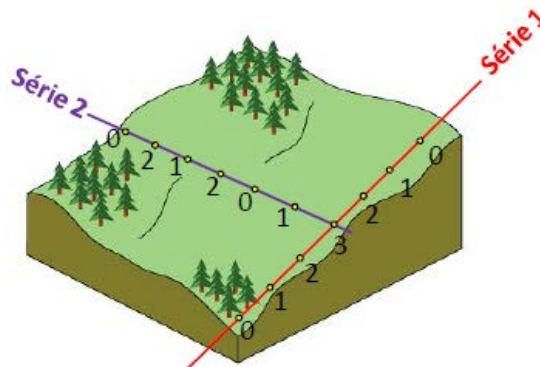
Conceptuellement c'est la méthode la plus satisfaisante, néanmoins elle a trois inconvénients :

- La méthode repose fortement sur l'hypothèse distributionnelle (en général la multi normalité),
- elle produit parfois des solutions bizarres,
- de plus on n'a pas de solution explicite, il faut maximiser numériquement, enfin la solution est parfois biaisée.

On est parfois surpris du manque de formalisation qui prévaut dans les méthodes d'estimations de paramètres Il y a deux raisons à cela :

- tout d'abord les méthodes habituelles (moindres carrés, maximum de vraisemblance) ne marchent pas toujours très bien.
- ensuite les développements des méthodes géostatistiques ont été fortement guidés par des problèmes concrets au cours desquels les géostatisticiens disposaient d'information extérieure (ou a priori) sur le phénomène qu'ils étudiaient, d'où la nécessité de conserver une certaine flexibilité aux techniques d'inférence pour pouvoir inclure cette information a priori plus ou moins subjective.

2.6. Exercices :



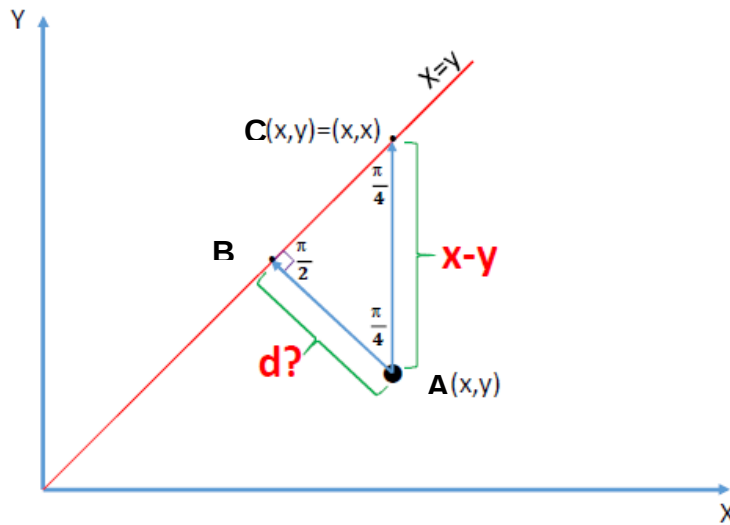
1. Comparer les deux séries en calculant la moyenne et la variance.
2. Représenter et ajuster le semi-variogramme pour les deux séries. Que peut-on conclure

2.7. Référence bibliographique :

1. Introduction à la géostatistique, notes de cours de Mr Gilles Guillot, Institut National Agronomique de Paris-Grignon janvier 2004.
2. Practical Geostatistics, Isobel Clark, Geostokos limited, Alloa Business center, Scotland 2001.
3. Géostatistique, cours d'Aloui T., Faculté des sciences mathématiques physiques et naturelles de Tunis.
4. <http://books.geojamal.com>
5. www.geojamal.com
6. Site internet : Wikipédia.

Annexe

Le semi-variogramme peut être considéré comme tant le moment d'inertie par rapport à la droite $y=x$.
 Pour un point A défini par ses coordonnées cartésiennes (x,y)



On projette le point M sur la droite ($y=x$) ;

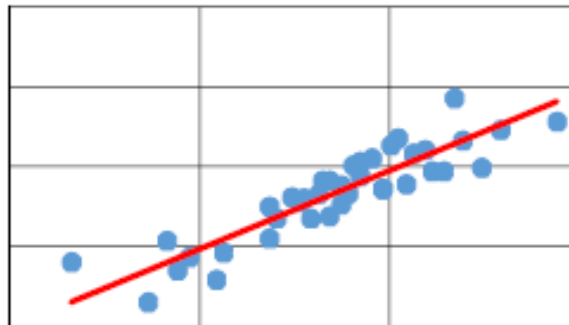
On trace la verticale qui passe par A et coupe la droite ($y=x$) en C

Du triangle rectangle isocèle ABC, on écrit que : $\cos \frac{\pi}{4} = \frac{1}{\sqrt{2}} = \frac{d}{x-y}$

$$d = \frac{1}{\sqrt{2}}(x - y)$$

$$\text{D'où } d = \frac{1}{\sqrt{2}}(x - y) \rightarrow d^2 = \frac{1}{2}(x - y)^2, \quad d \geq 0$$

Pour un nuage de n points connues par les coordonnées cartésiennes (x, y)

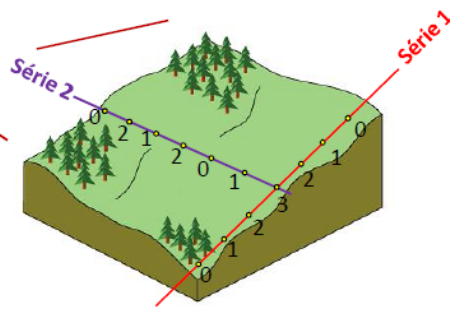
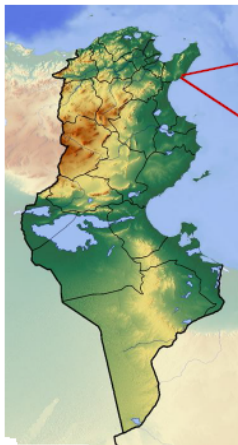


Calculons la moyenne des carrés des distances par rapport à la droite ($y=x$)

$$\bar{d}^2(h) = \gamma(h) = \frac{1}{2n} \sum_{i=1}^n (z_i - z_{i+h})^2$$

n: Nombre de paires dont les points sont espacés de h

La prospection d'un minéral a donné les résultats suivants :



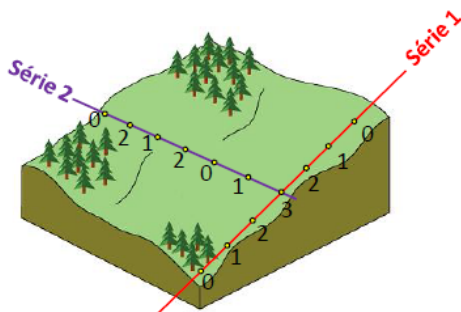
h (km)	0	1	2	3	4	5	6
Série 1 (g/tonne)	0	1	2	3	2	1	0
Série 2 (g/tonne)	3	1	0	2	1	2	0

	Série 1	Série 2
Moyenne	1,285714	1,285714
Variance	1,238095	1,238095

$$\text{Moyenne} = \bar{z} = \frac{1}{n} \cdot \sum_{i=0}^n z_i \quad \text{Variance} = V = \frac{1}{n} \cdot \sum_{i=0}^n (z_i - \bar{z})^2$$

Semi-variogramme

Ces deux séries ont même moyenne et même variance, toutefois clairement elles n'ont pas le même degré de continuité spatiale, la 1ère série étant nettement plus continue que la seconde. Voyons leur semi-variogramme:



Série 1

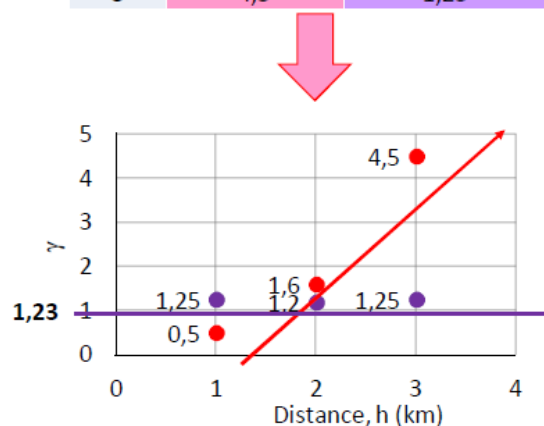
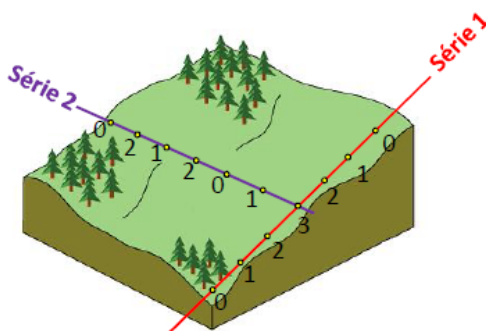
h=1 km	z _i	z _{i+h}	a=½(z _{i+h} -z _i)²
	0	1	0,5
	1	2	0,5
	2	3	0,5
	3	2	0,5
	2	1	0,5
	1	0	0,5
	γ=α̂		0,5
h=2 km	z _i	z _{i+h}	a=½(z _{i+h} -z _i)²
	0	2	2
	2	2	0
	2	0	2
	1	3	2
	3	1	2
	γ=α̂		1,6
h=3 km	z _i	z _{i+h}	a=½(z _{i+h} -z _i)²
	0	3	4,5
	3	0	4,5
	γ=α̂		4,5

Série 2

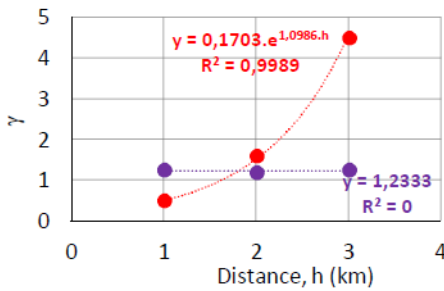
h=1 km	z _i	z _{i+h}	a=½(z _{i+h} -z _i)²
	3	1	2
	1	0	0,5
	0	2	2
	2	1	0,5
	1	2	0,5
	2	0	2
	γ=α̂		1,25
h=2 km	z _i	z _{i+h}	a=½(z _{i+h} -z _i)²
	3	0	4,5
	0	1	0,5
	1	0	0,5
	1	2	0,5
	2	2	0
	γ=α̂		1,2
h=3 km	z _i	z _{i+h}	a=½(z _{i+h} -z _i)²
	3	2	0,5
	2	0	2
	γ=α̂		1,25

Le semi-variogramme de la 1ère série montre **une croissance soutenue** alors que la seconde série montre un semi-variogramme à peu près constant à un niveau près de la variance expérimentale (V=1.23).

h (km)	Série 1	Série 2
1	0,5	1,25
2	1,6	1,2
3	4,5	1,25



Theoretical semivariogram



Semi-variogramme théorique

Série 1 $\rightarrow \gamma = 0,1703 \cdot e^{1,0986 \cdot h}$

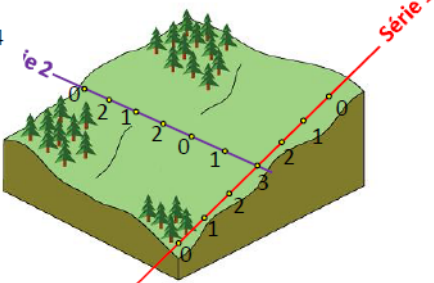
$$\gamma(h) = c_o + c \cdot \left(1 - e^{\frac{-|h|}{a}}\right)$$

$C_o=0$, $C=0,1703$ et $a=-0,91025$

Série 2 $\rightarrow \gamma = V=1,23$

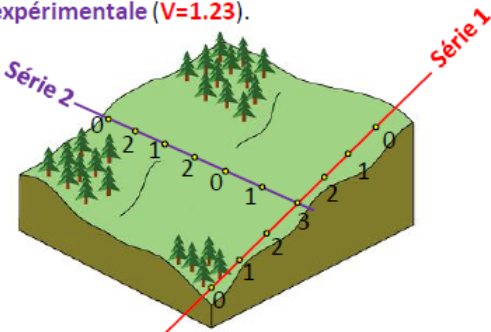
$$\gamma(h) = c_o + b \cdot h$$

$C_o=V=1,2333$ et $b=0$

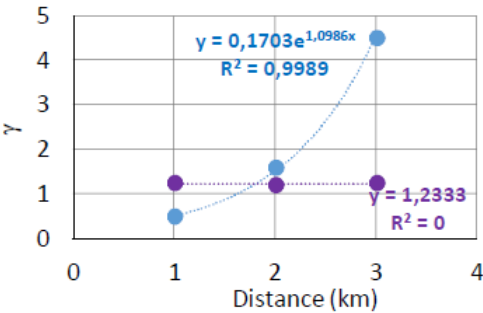


55

Le semi-variogramme de la 1ère série montre **une croissance soutenue** alors que la seconde série montre un semi-variogramme à peu près constant à un niveau près de la variance expérimentale ($V=1.23$).



h (km)	Série 1	Série 2
1	0,5	1,25
2	1,6	1,2
3	4,5	1,25



Variogrammes théoriques

Série 1 $\rightarrow \gamma = 0,1703e^{1,0986 \cdot h}$

Série 2 $\rightarrow \gamma = V=1,23$