II. Théorie Des Variables Régionalisées

II.1 introduction

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

Les méthodes géostatistiques, ont été initialement appliquées en exploration minière et pétrolière et elles ont retrouvé leur place en statistiques il y plus de 50ans. La géostatistique est l'application du formalisme des fonctions aléatoires à la reconnaissance et à l'estimation des phénomènes naturels (exploration minière et pétrolière, hydrologie et Hydrogéologie, météorologie, Environnement et pollution...).

La géostatistique, connue sous le titre de "Théorie des Variables Régionalisées". Cette méthode a été initiée en 1965 par G. Matheron, professeur à l'Ecole des Mines de Paris, et a connu un développement très important depuis lors.

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.

II.2 Analyse de la structure spatiale de la variable

II.2.1 Notion de variogramme

Considérons une propriété du sol notée Z connue en n points de l'espace géographique, chacun de ces points étant repérés par le vecteur $\vec{\mathbf{x}}$ de ses coordonnées géographiques (longitude et latitude). De la sorte, la notation $\mathbf{Z}(\mathbf{x_i})$ représente la valeur observée de la propriété Z au i ème point d'échantillonnage de coordonnées xi.

Deux observations situées l'une près de l'autre devraient, en moyenne, se ressembler davantage (être plus corrélées) que deux observations éloignées. Cette idée peut être exprimée simplement par la comparaison de ces deux valeurs : prenons deux points pour lesquels on connaît des valeurs Z(x1) et Z(x2) de la propriété Z. On cherche à comparer ces deux valeurs. Une façon simple est d'utiliser la variance entre les observations de ces deux sites, notée s^2. Elle est par définition égale à :

$$s^2 = [Z(x1) - \bar{Z}]^2 + [Z(x2) - \bar{Z}]^2 \qquad (1)$$

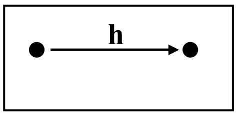
Où : $\bar{\mathbf{Z}}$ est la moyenne entre ces deux observations.

Cette variance s^2, qui traduit l'importance des écarts à la moyenne, est d'autant plus grande que les observations sont différentes. On peut d'ailleurs développer l'équation (1) pour obtenir une autre expression de cette valeur s^2 :

$$s^2 = \frac{1}{2}[Z(x1) - Z(x2)]^2 \qquad (2)$$

Cette équation peut être écrite pour tout couple de sites.

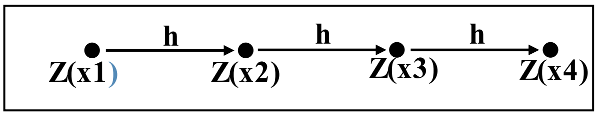
Pour cela, considérons deux sites Z(xi) et Z(xi+h) où xi représente les coordonnées géographiques d'un des sites et h est un vecteur caractérisant la distance entre les sites.



L'équation (2) s'écrit alors :

$$s^2 = \frac{1}{2}[Z(xi) - Z(xi + h)]^2 \qquad (3)$$

Calculons à présent la distance géographique séparant chacun des points d'observation et considérons les m couples de points séparés par une même distance géographique h.



On peut comme précédemment, calculer la variance des observations pour les sites pris deux à deux. La moyenne $\hat{s}^2$ de ces m variances s'écrit en employant (3) :

$$\hat{s}^2 = \frac{1}{2m}\sum_{i=1}^{m}\frac{1}{2}[Z(xi) - Z(xi + h)]^2 \qquad (4)$$

Pour une distance h séparant deux points d'observation, $\hat{s}^2$ rend compte de la ressemblance des observations faites en ces deux points: il sera d'autant plus grand que ces observations sont différentes. $\hat{s}^2$ dans ce cas est qualifiée de "semi-variance".

Cette semi-variance est appelée aussi variogramme, généralement donné par l'expression suivante :

$$\gamma(h) = \frac{1}{2\,m}\sum_{i=1}^{m}[Z(xi + h) - Z(xi)]^2) \qquad (5)$$

De ce fait, on conçoit que deux observations soient en général d'autant plus semblables qu'elles sont proches géographiquement l'une de l'autre.

Le calcul de $\boldsymbol{\gamma}(\mathbf{h})$ pour différentes distances h, va permettre de quantifier cette idée, il permet de suivre l'évolution des écarts entre des observations en fonction de la distance qui les sépare.

II.2.2 Hypothèses de base et définition

L'intérêt de cette notion simple et les conditions de sa généralisation ont été définis par la théorie des variables régionalisées (Matheron, 1965). Cette théorie montre que la généralisation de l'équation (5) suppose deux conditions, regroupées sous le terme d'hypothèsee intrinsèque :

- L'espérance mathématique ne dépend pas de x, $E[Z(x)]=m$
  ou
  L'espérance des écarts est zéro, $E[Z(x) - Z(x+h)] = 0$

- La covariance entre $Z(x)$ et $Z(x+h)$ ne dépend que de h $Cov(Z(x), Z(x+h)) = C(h)$; stationnarité du second ordre, $C(h)$ appelé fonction de covariance ou covariogramme.
  ou
  Le variogramme $\gamma(h)$ ne dépend pas de la localisation x, seulement de h (soit en module, soit en module et en direction) ; C à d la différence $[Z(x) - Z(x+h)]$ a une variance finie, qui ne dépend que de la distance h séparant les points.

$$\mathbf{Var}\,[\mathbf{Z}(\mathbf{xi}+\mathbf{h})-\mathbf{Z}(\mathbf{xi})] = \mathbf{2}\,\boldsymbol{\gamma}(\mathbf{h})$$
$$= \mathbf{E}[\mathbf{Z}(\mathbf{xi}+\mathbf{h})-\mathbf{Z}(\mathbf{xi})]^{\mathbf{2}}$$

Cette dernière hypothèse est légèrement moins restrictive que la stationnarité du second ordre est appelée hypothèse intrinsèque.

Conclusion :

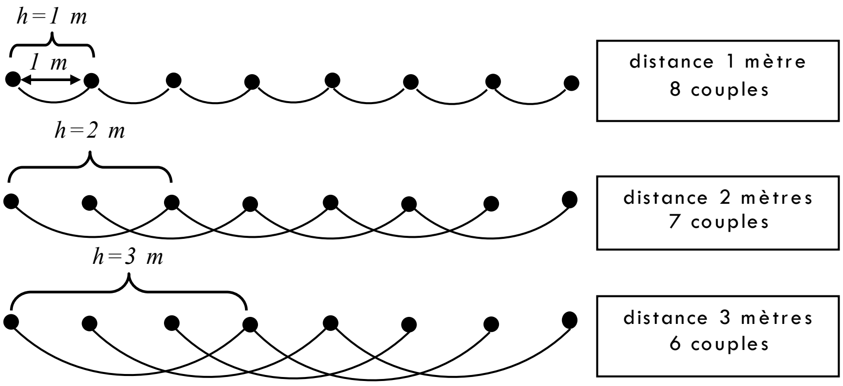
Quand ces deux conditions sont vérifiées, la valeur $\hat{s}^2$ définie dans l'équation (4) constitue un estimateur non biaisé de la fonction $\gamma(h)$ définie en (5). Cette fonction $\gamma(h)$ est nommée variogramme.

En étudiant l'évolution du variogramme $\gamma(h)$ en fonction de la distance h séparant des couples d'observation, on va analyser la façon dont se détériore l'information acquise en un point au fur et à mesure que l'on s'éloigne de ce point.

II.2.3 Calcul du variogramme (Cas d'échantillonnage à une Dimension « 1D »)

On cherche à construire un graphique représentant en abscisse les distances h séparant les points et en ordonnée les valeurs du variogramme (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.



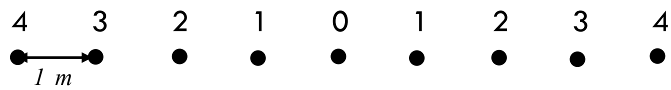
Ce schéma montre que le nombre de points participant au calcul du variogramme diminue au fur et à mesure que la distance augmente. Les valeurs de semi-variance risquent donc d'être moins précises pour les grandes valeurs de h.

Exemple :

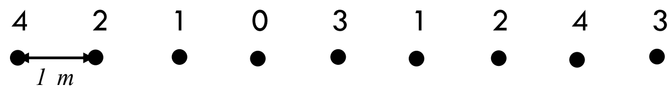
Soit deux exemples fictifs correspondant à des observations disposées le long d'un transect à des intervalles réguliers de 1 mètre.

Calculez pour chacun des exemples: - la moyenne; - la variance; - ☐(1), ☐(2) et ☐(3)

1er Cas :



1er Cas :



Solution :

1. **Calcule de la moyenne et de la variance pour les deux cas :**

	Formule	Cas 01	Cas 2
Moyenne	$\bar{Z} = \frac{\sum Z(xi)}{N}$	2.22	2.22
Variance (s2)	$\hat{s}^2 = \frac{\sum (Z(xi) - \bar{Z})^2}{N-1}$	1.94	1.94

2. Calcule du variogramme $\gamma(h)$ pour le premier cas :

La valeur du variogramme $\gamma(h)$ est donnée par l'expression suivante :

$$\gamma(h) = \frac{1}{2\,m}\sum_{i=1}^{m}[Z(xi + h) - Z(xi)]^2$$

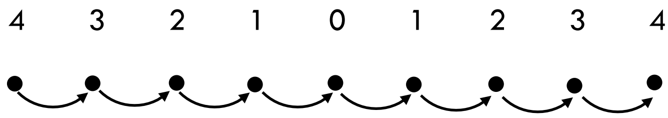
Ou :

- m est le nombre possible de couples qu'on peut former par les valeurs d'observations disponibles, utilisés pour le calcule de $\gamma(h)$;
- Z(xi) et Z(xi+h) sont les valeurs des points d'observations formant les différents couples.

1er Cas :

Alors, on doit d'abord fixer la distance h pour laquelle on veut calculer $\gamma(h)$;

Pour h =1 m les couples de points utilisés sont comme le montre la figure :



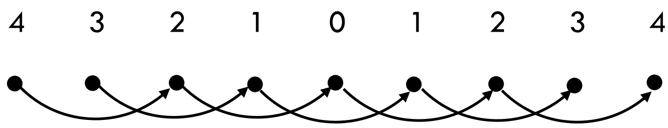
On compte huit (08) couple c à d m =8 et le variagramme est :

$$\gamma(h) = \frac{1}{2\ (8)}[(4-3)^2 + (3-2)^2 + (2-1)^2 + (1-0)^2 + (0-1)^2 + (1-2)^2 + (2-3)^2 + (3-4)^2]$$

$$\gamma(h) = \frac{1}{16}[(1)^2 + (1)^2 + (1)^2 + (1)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (-1)^2]$$

$$\gamma(h) = \frac{8}{16} = 0.5$$

Pour h =2 m les couples de points utilisés sont comme le montre la figure :



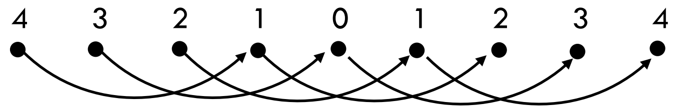
On compte huit (07) couple c à d m =7 et le variagramme est :

$$\gamma(h) = \frac{1}{2\ (7)}[(4-2)^2 + (3-1)^2 + (2-0)^2 + (1-1)^2 + (0-2)^2 + (1-3)^2 + (2-4)^2]$$

$$\gamma(h) = \frac{1}{14}[(2)^2 + (2)^2 + (2)^2 + (0)^2 + (-2)^2 + (-2)^2 + (-2)^2]$$

$$\gamma(h) = \frac{24}{14} = 1.71$$

Pour h =3 m les couples de points utilisés sont comme le montre la figure :



On compte huit (06) couple c à d m =6 et le variagramme est :

$$\gamma(\mathrm{h}) = \frac{1}{2\ (6)}[(4-1)^2 + (3-0)^2 + (2-1)^2 + (1-2)^2 + (0-3)^2 + (1-4)^2]$$

$$\gamma(\mathrm{h}) = \frac{1}{12}[(3)^2 + (3)^2 + (1)^2 + (-1)^2 + (-3)^2 + (-3)^2]$$

$$\gamma(\mathrm{h}) = \frac{38}{12} = 3.16$$

Les résultats du premier cas se résument dans le tableau ci-contre :

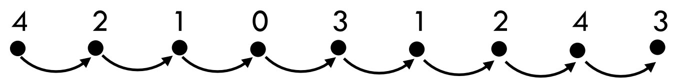
h	m	γ(h) (Cas 01)
1	8	0.5
2	7	1.71
3	6	3.16

3. **Calcule du variogramme $\gamma(h)$ pour le deuxième cas :**

2ème Cas :

On reprend la même procédure mais avec les valeurs de l'échantillonnage du deuxième cas. Alors, on fixe la distance h pour laquelle on veut calculer γ(h) ;

Pour h =1 m les couples de points utilisés sont comme le montre la figure :



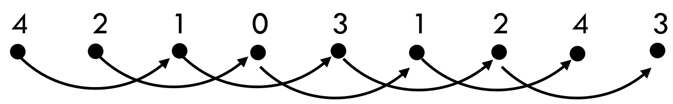
On compte huit (08) couple c à d m =8 et le variagramme est :

$$\gamma(\mathrm{h}) = \frac{1}{2\ (8)}[(4-2)^2 + (2-1)^2 + (1-0)^2 + (0-3)^2 + (3-1)^2 + (1-2)^2 + (2-4)^2 + (4-3)^2]$$

$$\gamma(\mathrm{h}) = \frac{1}{16}[(2)^2 + (1)^2 + (1)^2 + (-3)^2 + (2)^2 + (-1)^2 + (-2)^2 + (1)^2]$$

$$\gamma(\mathrm{h}) = \frac{25}{16} = 1.56$$

Pour h =2 m les couples de points utilisés sont comme le montre la figure :



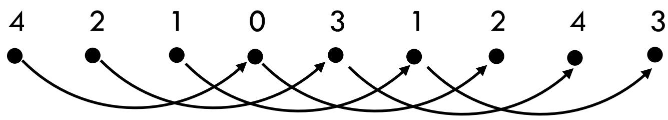
On compte huit (07) couple c à d m =7 et le variagramme est :

$$\gamma(\mathrm{h}) = \frac{1}{2\ (7)}[(4-1)^2 + (2-0)^2 + (1-3)^2 + (0-1)^2 + (3-2)^2 + (1-4)^2 + (2-3)^2]$$

$$\gamma(\mathrm{h}) = \frac{1}{14}[(3)^2 + (2)^2 + (-2)^2 + (-1)^2 + (1)^2 + (-3)^2 + (-1)^2]$$

$$\gamma(\mathrm{h}) = \frac{29}{14} = 2.07$$

Pour h =3 m les couples de points utilisés sont comme le montre la figure :



On compte huit (06) couple c à d m =6 et le variagramme est :

$$\gamma(\mathrm{h}) = \frac{1}{2\ (6)}[(4-0)^2 + (2-3)^2 + (1-1)^2 + (0-2)^2 + (3-4)^2 + (1-3)^2]$$

$$\gamma(\mathrm{h}) = \frac{1}{12}[(4)^2 + (-1)^2 + (0)^2 + (-2)^2 + (-1)^2 + (-2)^2]$$

$$\boldsymbol{\gamma(\mathrm{h}) = \frac{26}{12} = 2.16}$$

Les résultats du premier cas se résument dans le tableau ci-contre :

h	m	γ(h) (Cas 02)
1	8	1.56
2	7	2.07
3	6	2.16

Conclusion :

Les résultats obtenus permettent de conclure que :

- Dans les deux cas les échantillons présentent les mêmes observations et par conséquent la même moyenne et même variance. Mais, on voit clairement qu'ils n'ont pas le même degré de continuité spatiale, la 1ère série étant nettement plus continue que la seconde.
- Par contre l'analyse des résultats du variogramme montrent que la 1ère série présente une croissance continue alors que la seconde série montre un variogramme à peu près constant à un niveau près de la variance expérimentale

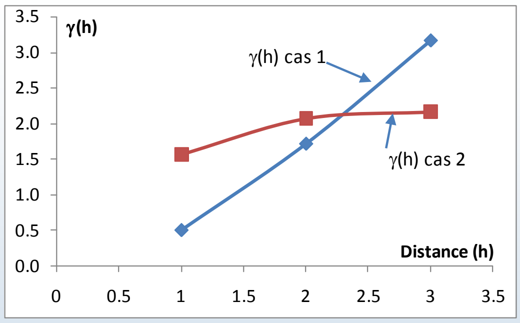
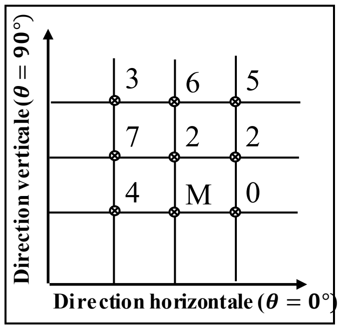


Fig. 1 : Représentation graphique du variogramme

II.2.4 Calcul du variogramme (Cas d'échantillonnage à deux Dimension «2D»)

Soit une matrice de données 3 x 3 ayant les valeurs suivantes (la distance horizontale et verticale entre 2 éléments consécutifs est de 1 m et M indique une donnée manquante).



Dans ce cas le variogramme peut être calculé selon certaines directions spécifiques et on note :

$$\gamma(h,\theta) = \frac{1}{2\,m(h,\theta)} \sum_{i=1}^{m} [Z(xi+h) - Z(xi)]^2$$

Ou :

- m est le nombre possible de couples qu'on peut former à partir des valeurs d'observations disponibles, suivant la direction θ ;
- θ est la direction (angle) suivant laquelle le variogramme est calculé (θ est choisi pour déterminer la meilleure continuité du phénomène dans l'espace, son choix se fait on se basant sur les conventions trigonométrique ou géographique).
- h est la distance séparant deux points Z(xi) et Z(xi+h) dans la direction θ.

En pratique on accorde une tolérance sur h et sur θ pour avoir suffisamment de paires pour chaque distance h et chaque angle θ.

Concéderons les données de cet exemple, et on va calculer le variogramme pour différente directions :

- **Calcule du variogramme suivant la direction horizontale $\theta = 0°$:**

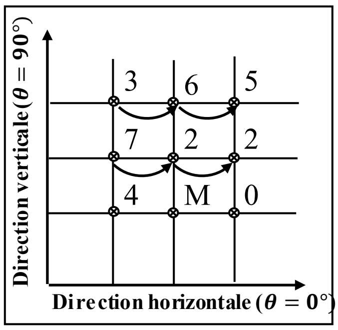
Pour h=1m :

Les pairs d'observations possibles dans ce cas sont comme le montre la figure ;

$$\gamma(1{,}0°) = \frac{1}{2\,(4)}[(7-2)^2 + (2-2)^2 + (3-6)^2 + (6-5)^2]$$

$$\gamma(1, 0°) = \frac{1}{(8)}[(5)^2 + (0)^2 + (-3)^2 + (1)^2]$$

$$\gamma(1, 0°) = \frac{35}{8} = 4.375$$



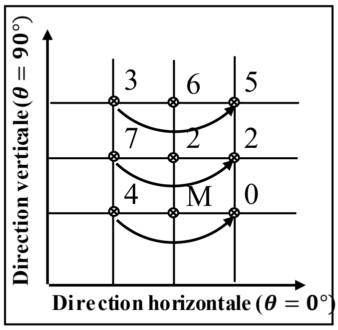
Pour h=2m :

Les pairs d'observations possibles dans ce cas sont comme le montre la figure ;

$$1{,}0°) = \frac{1}{2\,(3)}[(4-0)^2 + (7-2)^2 + (3-5)^2]$$

$$\gamma(1, 0°) = \frac{1}{(6)}[(4)^2 + (5)^2 + (-2)^2]$$

$$\gamma(1, 0°) = \frac{45}{6} = 7.5$$



Les résultats de calcule du variogramme sont :

h	N(h)	γ(h)
1	4	4.375
2	3	7.5

- **Calcule du variogramme suivant la direction verticale $\theta = 90°$:**

Pour h=1m :

Les pairs d'observations possibles dans ce cas sont comme le montre la figure ;

$$\gamma(1{,}0°) = \frac{1}{2\,(5)}[(3-7)^2 + (7-4)^2 + (6-2)^2 + (5-2)^2 + (2-0)^2]$$

$$\gamma(1, 0°) = \frac{1}{(10)}[(4)^2 + (-3)^2 + (4)^2 + (3)^2 + (2)^2]$$

$$\gamma(1, 0°) = \frac{54}{10} = 5.4$$

Pour h=2m :

Les pairs d'observations possibles dans ce cas sont comme le montre la figure ;

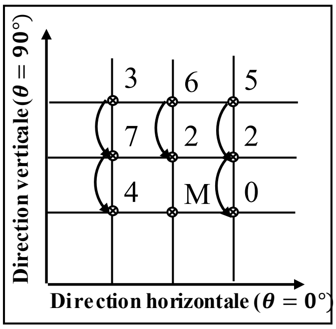
$$1{,}0°) = \frac{1}{2\,(2)}[(3-4)^2 + (5-0)^2]$$

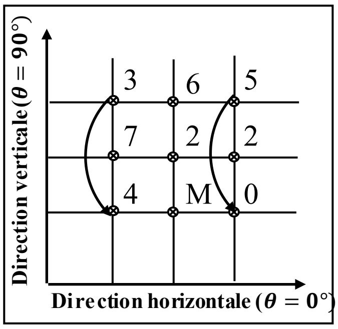
$$\gamma(1, 0°) = \frac{1}{(4)}[(1)^2 + (5)^2]$$

$$\gamma(1, 0°) = \frac{26}{4} = 6.5$$

Les résultats de calcule du variogramme sont :

h	N(h)	γ(h)
1	5	5.4
2	2	6.5





- **Calcule du variogramme suivant la direction verticale $\theta = 45°$:**

Pour h=1.41m :

Les pairs d'observations possibles dans ce cas sont comme le montre la figure ;

$$\gamma(1{,}0°) = \frac{1}{2\,(3)}[(4-2)^2 + (2-5)^2 + (7-6)^2]$$

$$\gamma(1, 0°) = \frac{1}{(6)}[(2)^2 + (-3)^2 + (1)^2]$$

$$\gamma(1, 0°) = \frac{14}{6} = 2.33$$

Pour h=2.82m :

Les pairs d'observations possibles dans ce cas sont comme le montre la figure ;

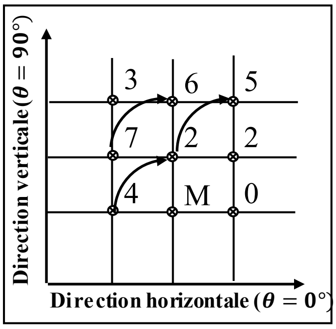
$$1{,}0°) = \frac{1}{2\,(1)}[(4-5)^2]$$

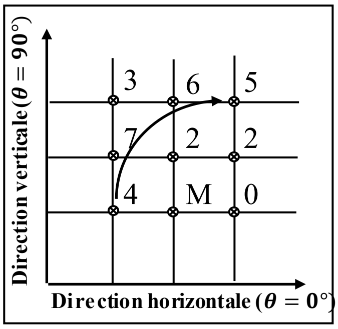
$$\gamma(1, 0°) = \frac{1}{(2)}[(1)^2]$$

$$\gamma(1, 0°) = \frac{1}{2} = 0.5$$

Les résultats de calcule du variogramme sont :

h	N(h)	γ(h)
1.41	3	2.33
2.82	1	0.5





II.2.5 Caractéristiques d'un variogramme

Un certain nombre de termes sont utilisés pour décrire un variogramme :

- **Portée (range) 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.

- **Palier (sill) σ^2 = Co + C :** Variance de la variable aléatoire (Var(Z(x)), il représente les écarts les plus grands, en moyenne entre deux variables aléatoires.

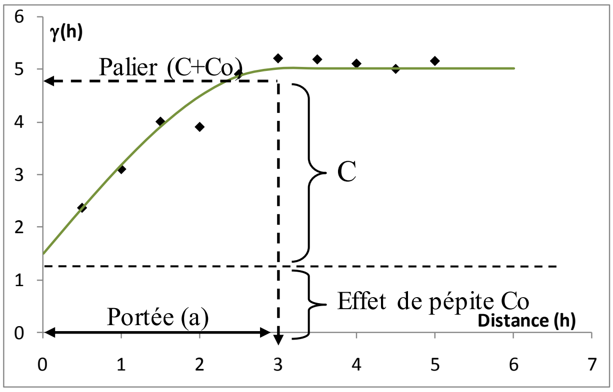


Fig.2 : Exemple de variogramme type

- **Effet de pépite (nugget effect) Co :** c'est la variation à très courte échelle, erreurs de localisation, erreurs d'analyse et précision analytique. (Un échantillon séparé en deux parties pour analyse ne donnera pas exactement la même teneur.

Lorsque h = 0 on a : $\gamma(0) = \frac{1}{2}\mathrm{Var}[Z(x) - Z(x)] = 0$ et non Co.

Par contre, $\lim_{\varepsilon \to 0^+} \gamma(\varepsilon) = C_0$ Cela signifie qu'on a une discontinuité à l'origine du variogramme.

Dans la pratique il existe deux catégories de variogrammes :

- Les variogrammes qui ne montrent pas de palier (dans ce cas, la covariance et la variance n'existent pas).
- Les variogrammes qui montrent un palier, dans ce cas on peut facilement établir le lien entre la valeur du variogramme pour la distance h et la covariance pour deux observations séparées de h.

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

$$\boldsymbol{\gamma(h) = \frac{1}{2}\left[\mathrm{Var}\big(Z(x)\big) + \mathrm{Var}\, Z(x+h) - 2\mathrm{Cov}(Z(x), Z(x+h))\right]}$$

$$\gamma(h) = [\sigma^2 - \mathrm{Cov}(Z(x), Z(x+h))] = \sigma^2 - C(h)$$

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

Ou, C(h) est appelé la fonction de covariance de Z.

- Lorsque la portée est atteinte (h = a), il n'y a plus de covariance entre les variables aléatoires, C(h) = 0 si h >= a.
- Lorsqu'il y a un palier, les deux fonctions sont équivalentes et 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.

Conclusion :

L'analyse du variogramme porte sur différents éléments. Ces éléments permettent de décrire la structure spatiale d'une variable :

- le variogramme atteint-il un palier ou croît-il sans cesse dans le domaine d'étude ?
- quelle est la portée éventuelle du variogramme?
- existe-t'il un effet de pépite et comment peut-il s'expliquer (erreurs de mesure...) ?