

2. LE VARIOGRAMME

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.

Ex.

Soit trois localisations x_0 , x_1 et x_2 , que l'on promène dans le gisement. On mesure la teneur en chacun de ces points.

$x_1 \quad x_0 \quad x_2$

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" de la minéralisation.

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 sur le gisement.

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

$x \longleftarrow \text{-----} \longrightarrow x+h$

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

Si l'on considère n localisations différentes x_1, x_2, \dots, x_n , la meilleure description que l'on puisse faire des n variables aléatoires $Z(x_1), Z(x_2), \dots, Z(x_n)$ est d'établir la fonction de distribution conjointe (multivariable). Clairement, ceci n'est pas possible puisqu'on ne peut disposer généralement que d'une seule observation à chacun de ces n points. On pourrait formuler une hypothèse très forte du genre: le vecteur des v.a. suit une loi multinormale de moyennes et variances-covariances spécifiées. Ceci serait beaucoup trop restrictif.

La géostatistique a des visées plus modestes. On veut estimer des paramètres statistiques à partir des données et non imposer un modèle a priori qui aurait toutes les chances de s'avérer inadéquat. Les paramètres que l'on cherchera à estimer ne sont pas la fonction de distribution conjointe, ni même la fonction de distribution bivariable (i.e. les v.a. considérées deux à deux) mais simplement les deux premiers moments (moyenne, variance, covariance) des v.a. prises deux à deux. Même réduit à cela, on ne dispose toujours que d'une seule paire d'observations situées précisément aux points x et $x+h$. On ne peut donc estimer les paramètres statistiques sans formuler certaines hypothèses. Ces hypothèses ont uniquement pour but de permettre

l'estimation des paramètres statistiques de notre modèle à partir des données. On les appelle hypothèses de stationnarité du second ordre; elles visent essentiellement à "détacher" les deux premiers moments de localisations précises en permettant des translations des emplacements x et $x+h$. La covariance (et le variogramme) deviennent donc des fonctions dépendant uniquement de la distance séparant les points d'observation et non plus de leur localisation exacte.

2.1 Hypothèses de base et définition:

Bref, on suppose que:

- i. L'espérance mathématique ne dépend pas de x ,
i.e. $E[Z(x)] = m$
ou
L'espérance des écarts est zéro
i.e. $E[Z(x) - Z(x+h)] = 0$
- ii. La covariance entre $Z(x)$ et $Z(x+h)$ ne dépend que de h
i.e. $Cov(Z(x), Z(x+h)) = C(h)$; stationnarité du second ordre, $C(h)$ est 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).
i.e. $1/2 \text{Var}(Z(x) - Z(x+h)) = \gamma(h)$; hypothèse intrinsèque (cette dernière hypothèse est légèrement moins restrictive que la stationnarité du second ordre)

Évidemment, ces hypothèses supposent une certaine régularité, une certaine homogénéité du gisement étudié. Si on peut reconnaître des zones très différentes géologiquement, on a habituellement intérêt à les traiter séparément.

La fonction la plus utilisée en géostatistique pour décrire la continuité de la minéralisation est le variogramme, et ce surtout parce qu'elle est plus simple à estimer que la covariance (qui demande l'estimation préalable de l'espérance mathématique), mais également parce qu'elle permet d'accommoder les situations où $\text{Var}(Z(x))$ n'est pas définie.

Le variogramme théorique est défini comme:

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

où x est le vecteur de coordonnées (1, 2 ou 3 coordonnées selon le cas)
 h est le vecteur distance .

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

- i. 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.
- ii. Palier $\sigma^2 = C_0 + C$: Variance de la v.a. ($\text{Var}(Z(x))$)
Écarts les plus grands, en moyenne entre deux v.a.
- iii. Effet de pépite : C_0 : Variation à très courte échelle, erreurs de localisation, erreurs d'analyse et précision analytique.

Ex. 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.

Notes : i. Lorsque $h = 0$ on a

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

par contre,

$$\lim_{\varepsilon \rightarrow 0^+} \gamma(\varepsilon) = C_0$$

i.e. on a une discontinuité à l'origine du variogramme.

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

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

donc,

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

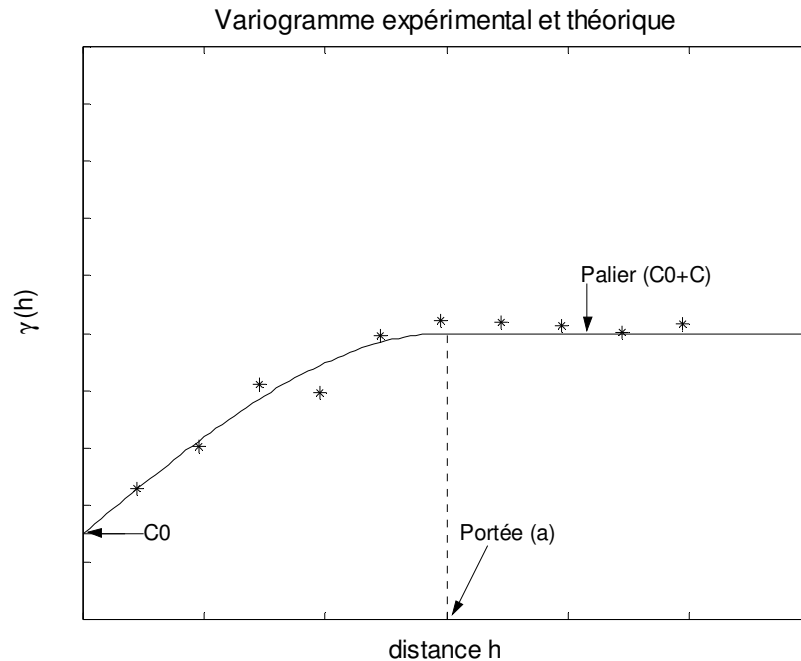
$C(h)$ est appelé le covariogramme de Z . Cette relation est importante et elle est continuellement utilisée en géostatistique.

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.

- i. Le variogramme est défini même s'il n'y a pas de palier.
- ii. 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.



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

- Un gisement d'or présentera un variogramme erratique avec un fort effet de pépité et une faible portée.
- Un gisement de cuivre porphyrique montrera un variogramme linéaire à l'origine avec faible effet de pépité 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 pépité.

Variogramme => outil descriptif puissant utilisable dans une multitude de domaines.

2.3 Estimation du variogramme

On estime le variogramme à l'aide de

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

où $N(h)$ nombre de paires dont les points sont espacées de h .

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 contamination par des hydrocarbures, on pourrait observer une meilleure continuité horizontalement que verticalement en raison de la gravité. 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.

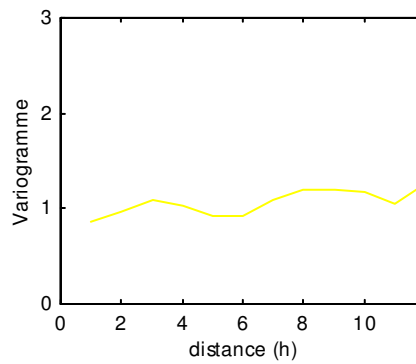
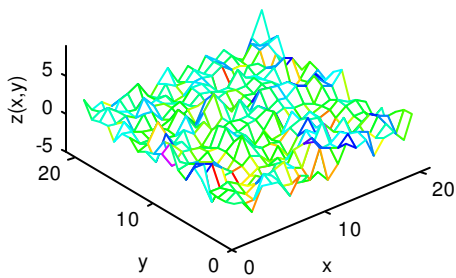
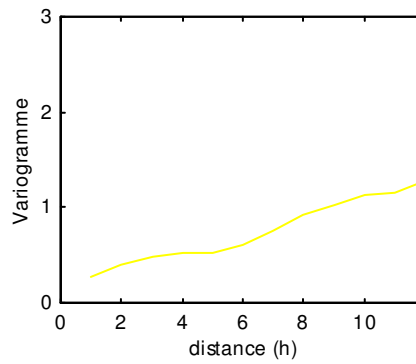
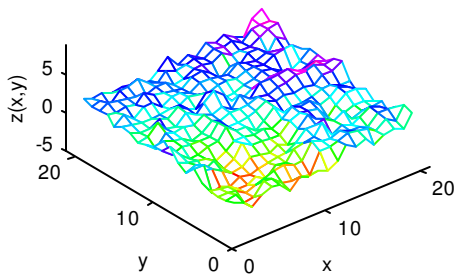
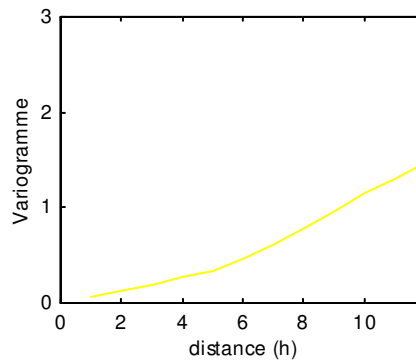
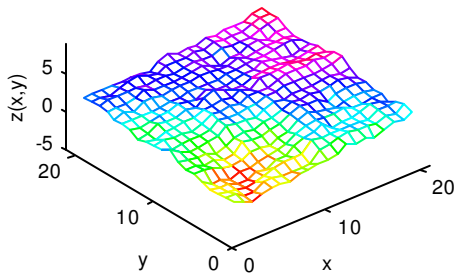
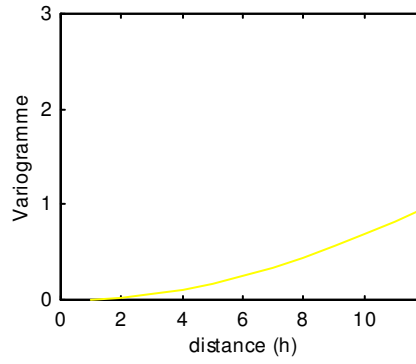
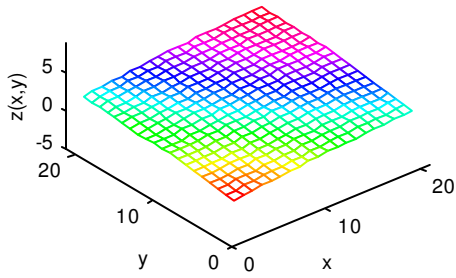
On peut aussi calculer le variogramme selon certaines directions spécifiques:

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

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

En pratique on s'accorde une tolérance sur h et sur θ afin d'avoir suffisamment de paires pour chaque h et chaque θ . 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 (i.e. expression analytique) permettant de déduire la covariance entre deux points quelconque en fonction de leur espacement géographique (et, éventuellement, de la direction qu'ils définissent). Une fois le modèle adopté, toute la suite des calculs se fait avec les **valeurs obtenues du modèle** et non avec les valeurs expérimentales.

La figure suivante illustre quelques exemples de surface et le variogramme expérimental correspondant. Les simulations ont été réalisées avec GSLIB-SGSIM, en imposant les valeurs 0, 2, 2 et 4 aux 4 coins. De haut en bas, on a simulé un gaussien de portée 25, un sphérique de portée 25, un sphérique avec 20% d'effet de pépite et portée 25, un sphérique avec 80% d'effet de pépite et portée 25. Comme on le voit, le variogramme expérimental décrit bien le degré d'irrégularité des surfaces.



Note: Sur les 4 figures précédentes, les points (0,0), (20,0), (0,20) et (20,20) ont des valeurs identiques de respectivement 0,2,2 et 4.

2.3.1 Exemple numérique

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 NaN indique une donnée manquante).

3	6	5
7	2	2
4	NaN	0

Le calcul du variogramme selon la direction horizontale donne:

h	$g(h)$	$N(h)$
1	4.375	4
2	7.5	3

Note: $g(1)=0.5/4*[(3-6)^2+(6-5)^2+(7-2)^2+(2-2)^2]$

Dans la direction verticale, on calcule:

h	$g(h)$	$N(h)$
1	5.4	5
2	6.5	2

Dans la direction 45, on calcule:

h	$g(h)$	$N(h)$
1.41	2.33	3
2.82	0.5	1

2e Exemple numérique (1D)

Soit les séquences 1D suivantes :

0 1 2 3 2 1 0

3 1 0 2 1 2 0

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 variogramme:

h	$g(h)$ - 1ère série	$g(h)$ - 2e série
1	0.5	1.25
2	1.6	1.2
3	2.5	1.13

Le variogramme de la 1ère série montre une croissance soutenue alors que la seconde série montre un variogramme à peu près constant à un niveau près de la variance expérimentale (1.06).

2.4 Modélisation

Les modèles sont des expressions analytiques que l'on tente d'ajuster le mieux possible aux points des variogrammes expérimentaux.

Condition d'admissibilité des modèles:

Toute fonction ne peut être utilisée comme modèle. Soit une somme quelconque de variables aléatoires (plus généralement, une combinaison linéaire de telles v.a.), la variance de cette combinaison est nécessairement positive (une variance est, par définition, toujours positive). Or cette variance peut s'exprimer en fonction du covariogramme (modèles avec palier) ou du variogramme (modèles avec palier ou sans palier pourvu que la somme des poids de la combinaison linéaire donne 0). Il faut donc que le covariogramme ou le variogramme assure des variances positives quelle que soit la combinaison des v.a. considérée.

Bref, soit une combinaison linéaire $\sum_i \lambda_i Z_i$. Dans le cas stationnaire (variogramme avec palier),

$$\text{Var}\left(\sum_i \lambda_i Z_i\right) = \sum_i \sum_j \lambda_i \lambda_j \text{Cov}(Z_i, Z_j) = \sum_i \sum_j \lambda_i \lambda_j C(h_{i,j}) \geq 0$$

Dans le cas intrinsèque (variogramme sans palier)

$$\text{Sous la condition } \sum_i \lambda_i = 0, \text{ on a } \text{Var}\left(\sum_i \lambda_i Z_i\right) = -\sum_i \sum_j \lambda_i \lambda_j \gamma(h_{i,j}) \geq 0$$

La vérification de l'admissibilité d'un modèle donné est relativement complexe et dépasse le cadre de ce cours. Dans la pratique on se limite à des modèles éprouvés et à des modèles construits à partir de modèles éprouvés en utilisant des propriétés comme :

- une combinaison linéaire (avec coefficients positifs) de variogrammes admissibles donne un modèle admissible;
- un produit de modèles de covariance admissibles donne un modèle de covariance admissible;
- un modèle admissible en \mathbb{R}^p est admissible en \mathbb{R}^{p-1} (l'inverse n'est pas nécessairement vrai).

Types de modèles courants

En géologie, les modèles les plus courants sont :

- Effet de pépite.
- Puissance (cas particulier : linéaire).
- Sphérique.
- Gaussien.
- Exponentiel.

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

	C_0	si $h > 0$	
Sphérique :	$\gamma(h) =$	$C [1.5 h/a - 0.5 (h/a)^3]$	si $0 < h < a$
		C	si $h \geq a$
Gaussien:	$\gamma(h) =$	$C [1 - \exp(-3(h/a)^2)]$	
Exponentiel	$\gamma(h) =$	$C [1 - \exp(-3h/a)]$	
Puissance	$\gamma(h) =$	$C h^b$	$0 < b < 2$ (linéaire : $b=1$)

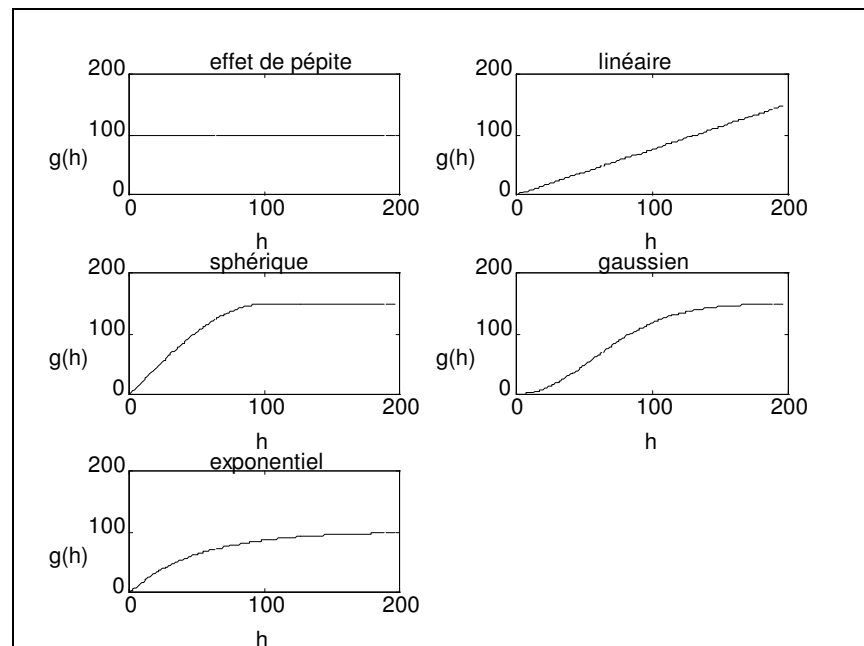
On peut combiner plusieurs modèles en les additionnant. Ainsi, l'effet de pépite est presque toujours présent en association avec un ou plusieurs des autres modèles décrits plus haut. Il est important de noter que ces cinq modèles ne sont pas les seuls que l'on peut utiliser, en réalité, il en existe un très grand nombre.

Il y a un lien étroit entre la nature de la variable étudiée et le type de modèle que l'on est susceptible de rencontrer. Ainsi, le modèle gaussien exprime une très grande continuité typique d'une variable comme la topographie, l'épaisseur d'une formation, le champ gravimétrique, la charge hydraulique. Ce modèle n'est pratiquement jamais rencontré pour des variables comme les teneurs de gisement, les propriétés mécaniques des roches, les analyses géochimiques en général. Pour ces variables, les modèles sphérique et exponentiel sont beaucoup plus courants. Pour des variables discrètes, le modèle gaussien est même à proscrire complètement.

Notes :

- i. Lorsque $h = 0$, par définition $\gamma(0)=0$.
Lorsque $h = 0+$, alors $\gamma(h)=C_0$.
L'effet de pépite se présente donc comme une discontinuité à l'origine du variogramme. L'effet de pépite peut représenter des erreurs d'analyse (voir théorie d'échantillonnage de P. Gy), de réelles micro-structures ou/et des structures d'une certaine taille non-détectées par un échantillonnage insuffisant.
- ii. Parfois les variogrammes ne montrent pas de palier (cas du modèle linéaire). D'autre fois ils ne montrent qu'un palier atteint asymptotiquement (cas des modèles exponentiel et gaussien). Dans ce dernier cas, l'on définit la portée effective comme la distance où est atteint 95% du palier. Ainsi, pour les modèles exponentiel et gaussien, la portée effective est "a". Lorsque le variogramme ne montre pas de palier et que sa croissance s'effectue à un taux supérieur à h^2 alors il y a lieu de suspecter une dérive de la moyenne (i.e. l'hypothèse stationnaire ou intrinsèque ne tient pas).
- iii. 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 . Lorsqu'il y a un palier, covariogramme et variogramme sont équivalents en ce sens que les deux fonctions fournissent la même information spatiale sur le phénomène étudié.
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. On n'a donc pas besoin de l'estimer comme c'est le cas lorsqu'on veut calculer directement le covariogramme.

Note: Lorsqu'on suppose que la moyenne et le covariogramme ne dépendent pas des emplacements précis considérés (x), ceci implique une certaine homogénéité du champ étudié. En termes statistique, on formule une hypothèse dite de **stationnarité du 2e ordre**. Si on suppose plutôt que le variogramme ne dépend pas de l'emplacement x , alors ce cas inclut les cas sans paliers, et est donc légèrement plus général que le précédent. Cette hypothèse est nommée **hypothèse intrinsèque**. Elle est en fait une hypothèse de stationnarité de 2e ordre postulé pour les incréments de Z plutôt que pour Z lui-même. On peut généraliser ce modèle en supposant la stationnarité d'ordre 2 pour des incréments d'ordre supérieur de Z (dans ce cas, de nouvelles fonctions de structure spatiale, les covariances généralisées, peuvent être utilisées). À l'opposé, des modèles plus restrictifs (stationnarité stricte au niveau univariable ou même multivariable) sont utilisés en géostatistique non-linéaire.



2.4.1 Anisotropies

La continuité spatiale n'est pas nécessairement la même dans toutes les directions.

- ex.
- gisement présentant une forme lenticulaire; on peut avoir une meilleure continuité selon l'allongement principal des lentilles;
 - gisement stratiforme; meilleure continuité parallèlement aux strates que perpendiculairement.
 - placer; meilleure continuité le long des paléochenaux que perpendiculairement.
 - etc.

Bien que dans la nature il existe une très grande variété d'anisotropies, en géostatistique, on ne peut modéliser aisément que les anisotropies géométriques.

Anisotropie géométrique

Caractéristiques :

- On observe dans diverses directions des **paliers et des composantes pépitiqes identiques mais des portées différentes.**
- Les portées maximales (a_g) et minimales (a_p) s'observent selon deux directions orthogonales.
- On peut rendre les portées identiques (et égales à a_g suivant toutes les directions en multipliant la composante de la portée parallèle à a_p par le facteur (a_g/a_p). Bref, les portées décrivent une ellipse dont l'axe majeur est orienté parallèlement à a_g .

i.e.

$$\frac{(a_g \cos \theta)^2}{a_g^2} + \frac{(a_p \sin \theta)^2}{a_p^2} = 1$$

Connaissant a_g et a_p , on peut trouver a_θ , où θ désigne l'angle mesuré par rapport à la direction où est rencontré a_g .

$$a_\theta = \frac{a_g a_p}{\{a_p^2 \cos^2 \theta + a_g^2 \sin^2 \theta\}^{1/2}}$$

On peut ainsi évaluer $\gamma(h, \theta)$ soit en utilisant a_θ , soit en corrigeant la distance h pour tenir compte de l'anisotropie:

$$\gamma(h_\theta, \theta) = \gamma(h_g)$$

modèle isotrope avec portée a_g

avec,

$$h_g = \sqrt{(h_\theta \cos \theta)^2 + \left(\frac{a_g}{a_p} h_\theta \sin \theta\right)^2}$$

Typiquement on retrouve l'anisotropie géométrique là où le corps étudié montre des allongements préférentiels. (lentilles, paléochenaux, strates...).

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 .

Exemple:

Un gisement 2D est modélisé par un modèle avec anisotropie géométrique. Le modèle est sphérique avec $C=17\%^2$ et effet de pépité $C_0=13\%^2$ et les portées sont de 100m dans la direction (convention trigonométrique) de plus grande continuité (30°) et 60m dans la direction de plus petite continuité (120°). Quelle est la valeur du variogramme entre deux observations situées aux coordonnées $(x_1, y_1)=(10, 30)$ et $(x_2, y_2)=(40, 20)$

1ère méthode:

On calcule la distance séparant les deux points et la direction qu'ils définissent:

$$h = ((20-30)^2 + (40-10)^2)^{1/2} = 31.62m$$

$$\theta = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) = \arctan\left(\frac{-10}{30}\right) = -18.4^\circ$$

Cette direction forme un angle de 48.4° avec la direction de plus grande continuité.

On calcule la portée dans cette direction en utilisant la formule plus haut :

$$a_\theta = \frac{100 * 60}{\left\{60^2 \cos^2(48.4) + 100^2 \sin^2(48.4)\right\}^{1/2}} = 70.81m$$

On calcule la valeur du variogramme en utilisant l'équation du modèle sphérique pour la distance plus haut et avec la portée 70.81m:

$$\gamma_\theta(31.62m) = 13\%^2 + 17\%^2 * \left(1.5 * \frac{31.62m}{70.81m} - 0.5 * \left(\frac{31.62m}{70.81m}\right)^3\right) = 23.63\%^2$$

2^e méthode

On calcule la distance équivalente dans la direction de meilleure continuité avec la formule précédente, où θ représente l'angle entre la direction de meilleure continuité et la direction définie par les deux points (48.4°).

$$h_g = \sqrt{(31.62m \cos(48.4))^2 + \left(\frac{100m}{60m} 31.62m \sin 48.4\right)^2} = 44.65m$$

On calcule la valeur du variogramme en utilisant l'équation du sphérique pour la distance 44.65m et avec la portée $a_g = 100m$.

$$\gamma_g(44.65m) = 13\%^2 + 17\%^2 * \left(1.5 * \frac{44.65m}{100m} - 0.5 * \left(\frac{44.65m}{100m}\right)^3\right) = 23.63\%^2$$

Remarques importantes concernant la détection d'anisotropies géométriques:

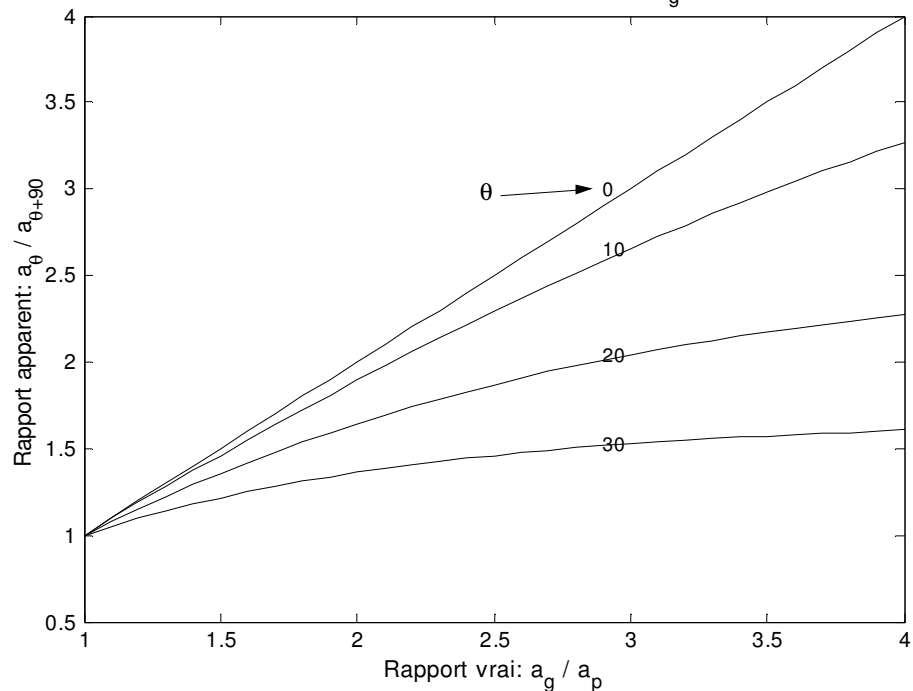
a) Le facteur d'anisotropie géométrique obtenu avec les variogrammes expérimentaux sous estime en général le véritable facteur d'anisotropie en raison de l'utilisation d'une fenêtre angulaire et du fait que les variogrammes expérimentaux ne sont pas nécessairement orientés exactement selon les directions principales de l'ellipse d'anisotropie.

b) L'estimation correcte et à la limite, la détection, d'anisotropie géométrique n'est possible, en pratique, qu'à quatre conditions (fortement liées) devant être remplies simultanément:

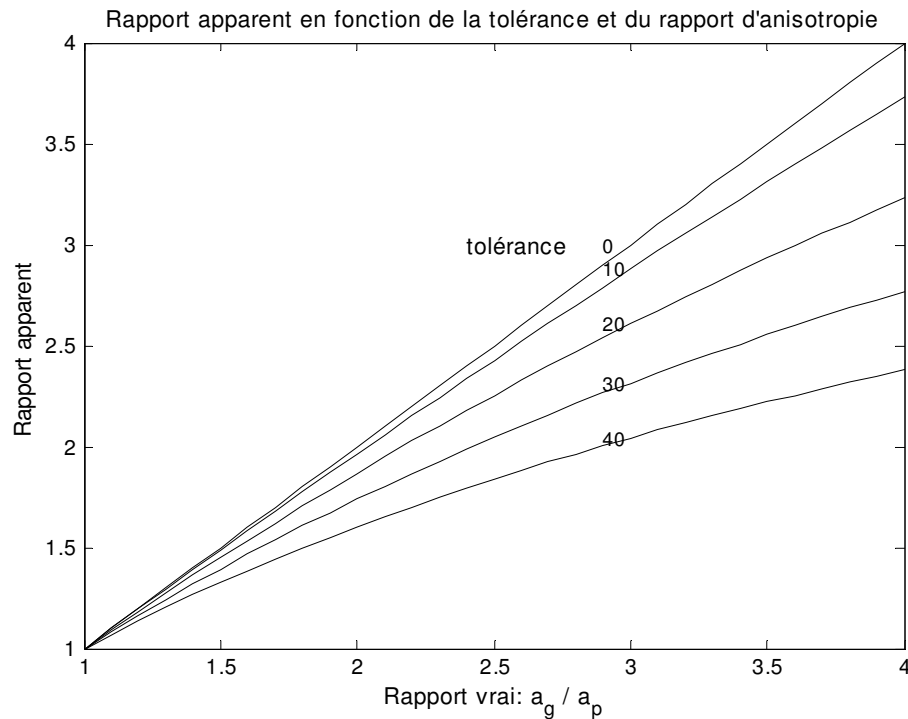
- Le nombre de données est suffisant (au moins 50)
- Le facteur d'anisotropie est important (au moins 1.5)
- Une des directions utilisées dans le calcul du variogramme est près de la direction de plus grande portée.
- La fenêtre angulaire utilisée est suffisamment étroite.

La figure suivante montre le rapport d'anisotropie apparent observé en considérant les directions θ et $\theta+90$ (θ : angle avec la direction de plus grande portée) en fonction du rapport d'anisotropie (plus grande portée/plus petite portée). On constate qu'un rapport d'anisotropie peut facilement être sous-estimé si on n'identifie pas correctement la direction de plus grande portée.

Rapport apparent en fonction de la direction (θ : angle avec a_g) et du rapport d'anisotropie



La figure suivante montre le rapport d'anisotropie apparent que l'on devrait observer en fonction de l'angle de tolérance (fenêtre angulaire) adopté dans le calcul du variogramme. Le tableau suivant illustre les résultats pour un rapport d'anisotropie réel de 5.



Rapport d'anisotropie apparent en fonction de la tolérance pour un rapport vrai de 5

Tolérance angulaire	Rapport apparent
45°	2.4
22.5°	3.6
10°	4.5
5°	4.9

Anisotropie zonale

Parfois, une simple correction géométrique ne suffit pas à rendre les modèles isotropes. C'est le cas par exemple si l'on observe des paliers différents ou si les portées ne décrivent pas une ellipse. On peut alors tenter d'ajuster les variogrammes expérimentaux directionnels à l'aide d'une somme (ou éventuellement d'un produit de covariances) de modèles isotropes ou avec anisotropie géométrique. Parfois la physique du phénomène peut aider à déterminer le modèle. Ainsi, en hydrogéologie, la charge hydraulique est une quantité anisotrope par sa nature même; en effet, dans le sens de l'écoulement on observe les variations maximales alors que perpendiculairement à l'écoulement la charge est constante.

La modélisation d'anisotropies zonales est généralement assez délicate et nécessite une certaine expérience. Le modèle le plus simple d'anisotropie zonale consiste à combiner une ou plusieurs composantes isotropes à une composante avec anisotropie géométrique dont a_g est infinie.

$$\gamma_{zonal}(h, \theta) = \gamma_{isotrope}(h) + \gamma_p(h \sin \theta)$$

où l'indice p réfère au modèle anisotrope suivant la direction de portée minimale

Cas 3D:

En 3D, l'ellipse d'anisotropie devient un ellipsoïde. Pour entièrement spécifier le modèle, il faut fournir les trois portées principales (axes de l'ellipsoïde) et les 3 angles de rotation qui permettent de faire coïncider le système de référence avec les axes de l'ellipsoïde. Souvent la géologie dictera les directions où calculer le variogramme pour tenter de détecter une éventuelle anisotropie (ex. perpendiculairement à la stratigraphie et dans le plan de la stratigraphie).

Lors de l'utilisation d'un programme de calcul de variogramme ou de krigeage, il est très important de bien comprendre les conventions utilisées pour le système de référence et les rotations afin de spécifier correctement les modèles. Habituellement, le système de référence utilisé est le système « main droite » (pouce pointe vers "z", la main droite repliée va de "x" vers "y").

La modélisation en 3D est parfois très difficile en raison d'une disposition défavorable des observations. Si l'on prend l'exemple d'une grille régulière de forages verticaux, on dispose de beaucoup de paires pour toute distance selon la verticale. Par contre, dans le plan horizontal, aucune paire ne peut être formée pour des distances autres qu'un multiple du pas de grille. Si le pas est large, ce qui est souvent le cas, on aura très peu de points sur le variogramme expérimental et la détermination des portées dans ces directions sera difficile. La situation se complique davantage lorsque la géologie ne suit pas les directions du système de référence.

Finalement, il faut noter qu'en 3D, la spécification des paramètres de recherche de paires pour le calcul du variogramme nécessite une bonne dose de réflexion afin de s'assurer que la zone spécifiée correspond bien à celle désirée. Par exemple, une tolérance angulaire de 10° sur la direction et sur le pendage du vecteur souhaité ne représentent pas du tout la même enveloppe si le vecteur considéré est horizontal ou vertical. Ce ne sont pas tous les programmes qui permettent de spécifier un cône de tolérance autour de l'orientation du vecteur distance souhaitée.

2.5 Remarques concernant le calcul de variogrammes et l'ajustement de modèles

- On accorde plus de poids aux points du variogramme expérimental calculés avec beaucoup de paires.
- On essaie d'avoir $N(h) \geq 30$ pour chaque point expérimental du variogramme. Si ce n'est pas possible pour certaines classes, on accorde moins d'importance à ces points. Si le nombre de paires est très faible (≤ 10), on ne considère plus du tout le point.
- On accorde plus de poids aux premiers points du variogramme (h petit) car ce sont ces valeurs qui ont le plus d'impact dans les calculs géostatistiques.
- Lorsque h dépasse environ $d_{\max}/2$, on ne tient pas compte des valeurs du variogramme. d_{\max} est la taille du phénomène étudié dans la direction considérée.
- On cherche à obtenir des modèles les plus simples possible qui rendent bien compte des valeurs expérimentales.

Stratégie de modélisation (cas 2D)

- Calculer les variogrammes directionnels selon différentes directions (ex. 0° , 45° , 90° , 135°) ainsi que le variogramme omnidirectionnel (i.e. sans tenir compte de la direction).

La géologie peut apporter une information précieuse dans le choix des directions et la présence ou non d'anisotropies.

- Vérifier les critères ci-dessus : $N(h) \geq 30$, $h < d_{\max}/2$
- Si nécessaire, augmenter la tolérance angulaire ou le pas de calcul de façon à augmenter $N(h)$.
- Déterminer s'il y a anisotropie (différences de palier ou de portées qui ne peuvent raisonnablement être imputées à des fluctuations aléatoires du variogramme). Une bonne méthode consiste d'abord à ajuster le variogramme omnidirectionnel et de vérifier si ce modèle est acceptable pour les différents variogrammes directionnels. L'effet de pépite et le palier en particulier devraient être estimés à l'aide du variogramme omnidirectionnel et gardés constants lors de l'ajustement des variogrammes directionnels. Si les paliers changent d'une direction à l'autre, on peut soit essayer de modéliser une anisotropie zonale, soit adopter un palier compromis, surtout si l'ajustement est adéquat à courte distance.
- Procéder à l'ajustement d'un modèle anisotrope ou isotrope selon le cas (habituellement par essai et erreur, bien que l'on puisse aussi obtenir ces ajustements de façon automatique par régression (pondérée, et souvent, non-linéaire).
- Chercher à respecter la règle de la parcimonie: adopter les modèles les plus simples possibles qui permettent un ajustement adéquat. Comparer des modèles concurrents à l'aide de la *technique de validation croisée*.

Note:

- Plus les classes sont larges, plus il y a de paires dans chaque classe, et plus le variogramme expérimental est lisse (et donc facile à modéliser) mais moins on a de définition pour connaître le comportement du variogramme, surtout à faible distance. On cherche habituellement à avoir au moins trois ou quatre classes, et si possible davantage, avant d'atteindre le palier.
- Pour les variogrammes directionnels, plus l'angle de tolérance est grand, plus on a de paires pour chaque point du variogramme mais moins le variogramme expérimental permettra de déceler les anisotropies. On ne devrait pas excéder 22.5 degrés de part et d'autre de la direction considérée. On peut descendre jusqu'à 0+ degrés si les données sont abondantes et sur une grille parfaitement régulière. Une valeur typique pourrait être de 10 degrés de part et d'autre de la direction considérée. On spécifie le calcul du variogramme omnidirectionnel en utilisant un angle de tolérance de 90 degrés de part et d'autre d'une direction arbitraire, le choix de la direction n'ayant dès lors aucune importance.
- Souvent en 3D les seules directions pour lesquelles on peut véritablement calculer un variogramme fiable sont les directions prises *le long des forages*. Une des raisons pour cela est que les trous de forage peuvent dévier considérablement et que l'arpentage des trous de forage, même s'il a été réalisé, est souvent fort imprécis. Les positions véritables des observations n'étant pas connues l'effet de pépite est accru, la forme du variogramme expérimental est altérée, et les variogrammes deviennent plus erratiques. Cet effet est particulièrement important dans le cas de gisement à faible continuité spatiale.
- Plusieurs problèmes 3D peuvent être simplifiés à des problèmes 2D. C'est le cas notamment de gisements se présentant sous la forme de veines minces. Habituellement on travaille alors avec l'accumulation de métal (produit épaisseur x teneur) et l'épaisseur de la veine qui sont 2 variables additives (alors que la teneur ne l'est pas dans ce cas). On obtient la teneur estimée par division de l'accumulation estimée par l'épaisseur estimée. Certains praticiens préfèrent toutefois travailler avec la teneur directement même si celle-ci n'est pas additive. (Par additif, on entend que la valeur de la variable sur une zone est donnée par la moyenne des valeurs des points de la zone. Si l'épaisseur d'une veine varie alors la teneur pour une portion de la veine n'est plus égale à la teneur moyenne des points correspondants mais plutôt à une moyenne pondérée par les épaisseurs).

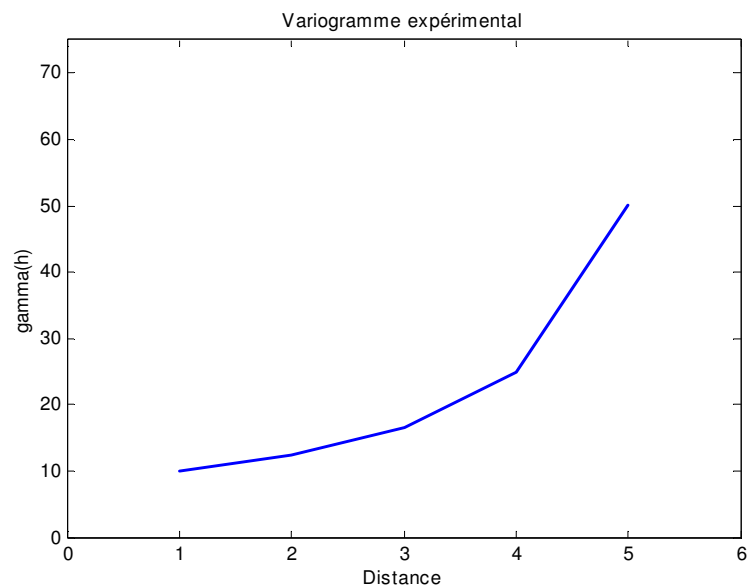
2.6 Problèmes courants avec les variogrammes et solutions possibles

2.6.1 Données extrêmes

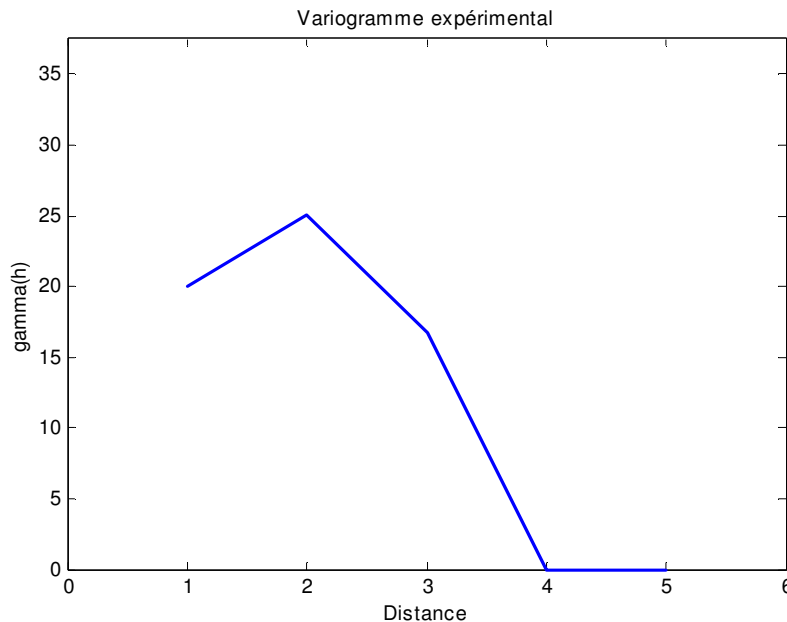
Le variogramme étant une moyenne de différences au carré, il est clair que la contribution d'une donnée extrême peut être déterminante. Si la localisation d'une donnée extrême est telle qu'elle apparaît plus souvent dans certaines classes de distance que d'autres, alors le variogramme sera très bruité. Si elle est située en périphérie du domaine, elle introduira une tendance croissante sur le variogramme. Si elle est située au centre, elle introduira plutôt une tendance décroissante.

Exemple:

10	0	0	0	0	0
----	---	---	---	---	---



0	0	10	0	0	0
---	---	----	---	---	---



Comme on le voit, la position de la valeur extrême dans le champ a une influence prépondérante sur la forme du variogramme.

Solutions possibles:

- Si la donnée extrême est une erreur, on l'enlève tout simplement.
- Enlever la donnée extrême pour le calcul et la modélisation du variogramme afin de mieux cerner la structure spatiale sous-jacente. Toutefois, il faut remettre cette donnée au moment de l'estimation. Généralement, il faut aussi modifier le modèle de façon à ce que son palier reflète mieux la variance des données lorsque les données extrêmes s'y trouvent (par exemple, ajout d'un effet de pépite ou multiplication de C et C_0 par une constante appropriée).
- Transformer les données de façon à diminuer l'influence des données extrêmes (ex. couper les valeurs extrêmes à un seuil maximal, prendre le logarithme, la racine carrée, etc.). On peut par la suite identifier les grandes caractéristiques du variogramme (modèle, isotropie-anisotropie, importance approximative de $C_0/(C_0+C)$) avec les données transformées puis on cherche à retrouver ces caractéristiques sur les variogrammes expérimentaux. Normalement on ne peut estimer les valeurs transformées car la transformation inverse pose un problème de biais difficile à contourner.
- Utiliser un estimateur robuste aux données extrêmes (ex. au lieu de prendre la moyenne des écarts-carrés, on pourrait en prendre la médiane). Toutefois cet estimateur sous-estime la variabilité spatiale et il doit être modifié pour tenir compte de ce fait.

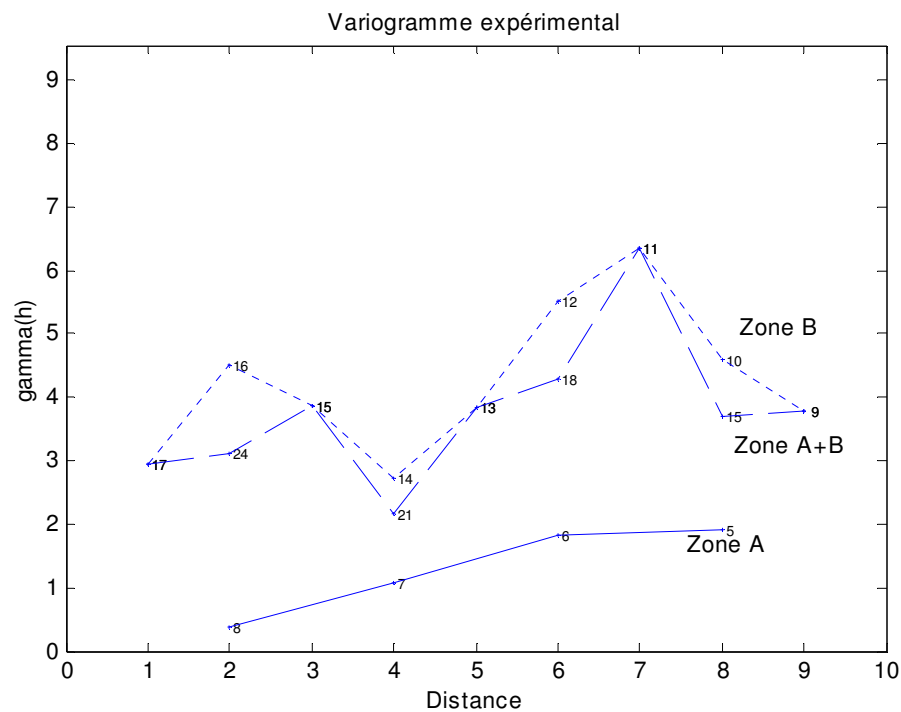
2.6.2 Pas d'échantillonnage variable selon les zones d'un gisement

Zone A: (pas de 2m)

4	4	5	6	6	7	6	5	4
---	---	---	---	---	---	---	---	---

Zone B: (pas de 1m)

8	6	8	10	12	8	10	12	14	10	8	6	12	8	10	10	8	10
---	---	---	----	----	---	----	----	----	----	---	---	----	---	----	----	---	----



Le variogramme A est plus bas car la zone A est moins variable. Le variogramme B est le plus élevé car la zone B est la plus variable et le variogramme A+B est un mélange des 2 zones. Toutefois, comme les pas d'ordre impair (1,3,5...) n'apparaissent pas dans la zone A, le variogramme A+B est identique au variogramme B pour ces pas.

Solutions possibles:

- Séparer en 2 zones d'étude distinctes si possible, sinon
- Uniformiser l'échantillonnage, par exemple en prenant 1 point sur 2 dans la zone B.

2.6.3 Ré-échantillonnage des zones riches

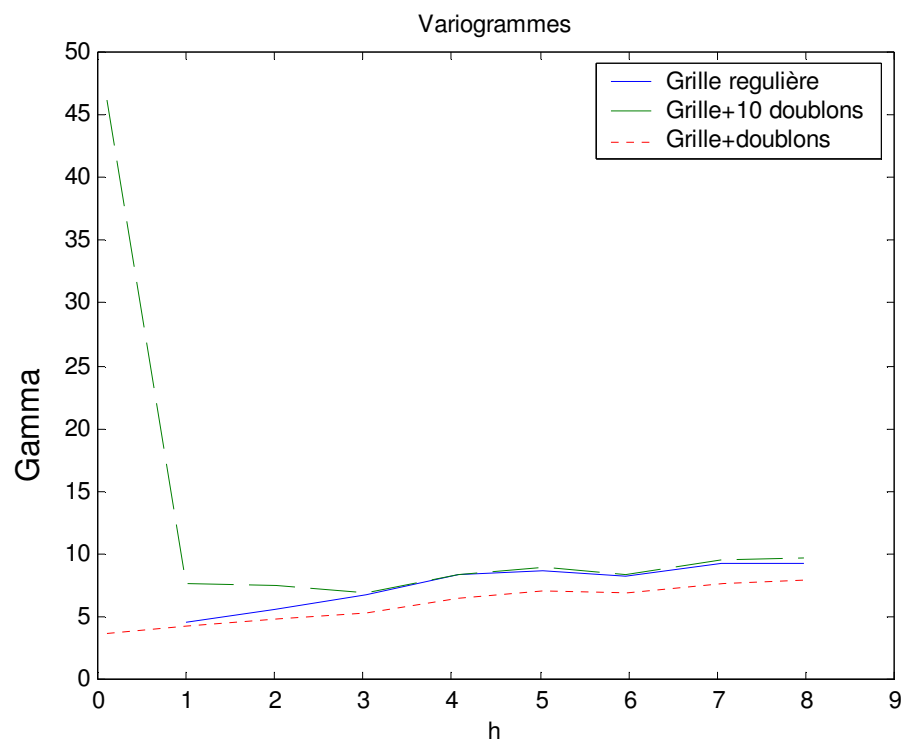
Un problème semblable à celui rencontré précédemment survient lorsqu'on échantillonne préférentiellement à proximité des valeurs fortes que l'on rencontre (pour confirmer disent certains). Comme les distributions des teneurs des gisements typiques sont fortement asymétriques avec peu de valeurs fortes, les chances sont très grandes qu'une valeur forte ne sera pas "confirmée". On aura donc ainsi plusieurs valeurs fortes accompagnées de valeurs nettement plus faibles à proximité. Les seules paires de données à petite distance peuvent provenir précisément de ces ré-échantillonnages. Conséquemment, ceci aura pour effet de faire paraître la continuité spatiale beaucoup moins forte qu'elle ne l'est réellement.

Exemple: On simule 225 valeurs sur une grille régulière de pas 1 (15*15). On décide d'échantillonner les 10 valeurs les plus fortes en se plaçant à 0.1 (en x) du point. Voici les 3 variogrammes obtenus en utilisant:

1- les 225 points

2- les 225 points + les 10 "doublons"

3- les 225 points et les 225 "doublons" (i.e. chaque point est ré-échantillonné à 0.1 de façon systématique).



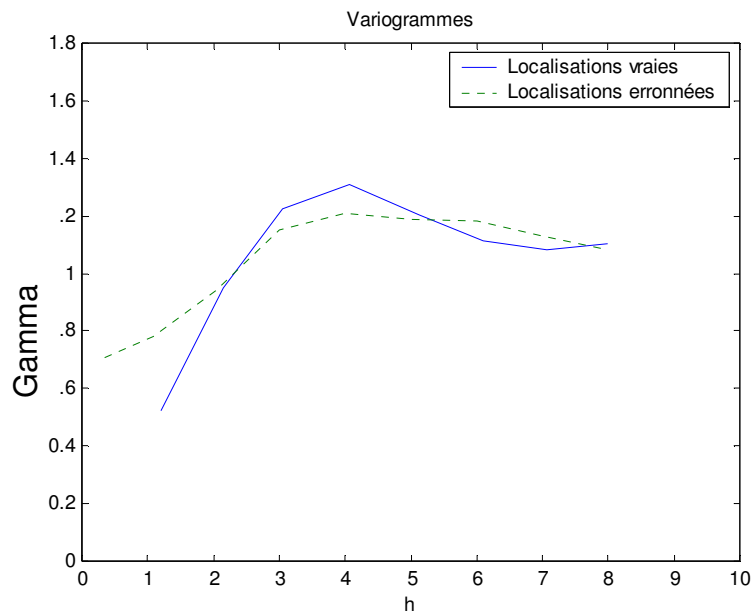
Solutions possibles:

- Éviter les stratégies d'échantillonnage biaisées vers les valeurs fortes
- Décimer l'échantillon pour assurer une couverture uniforme partout

2.6.4 Erreurs de localisation

Les erreurs de localisation viennent fausser les distances et donc fausser le variogramme expérimental. Ainsi, certaines paires à petites distances seront considérées comme à des distances plus grandes qu'elles ne le sont réellement et vice-versa. L'effet net est d'augmenter l'effet de pépite apparent sur le variogramme.

Exemple: On simule 225 données sur une grille régulière de 15 x 15 et de pas 1, puis on suppose que chaque point a été mal localisé et que la position rapportée se situe à ± 1 en x et y. Voici les variogrammes obtenus en utilisant les vraies localisations et les localisations rapportées:



Solutions possibles:

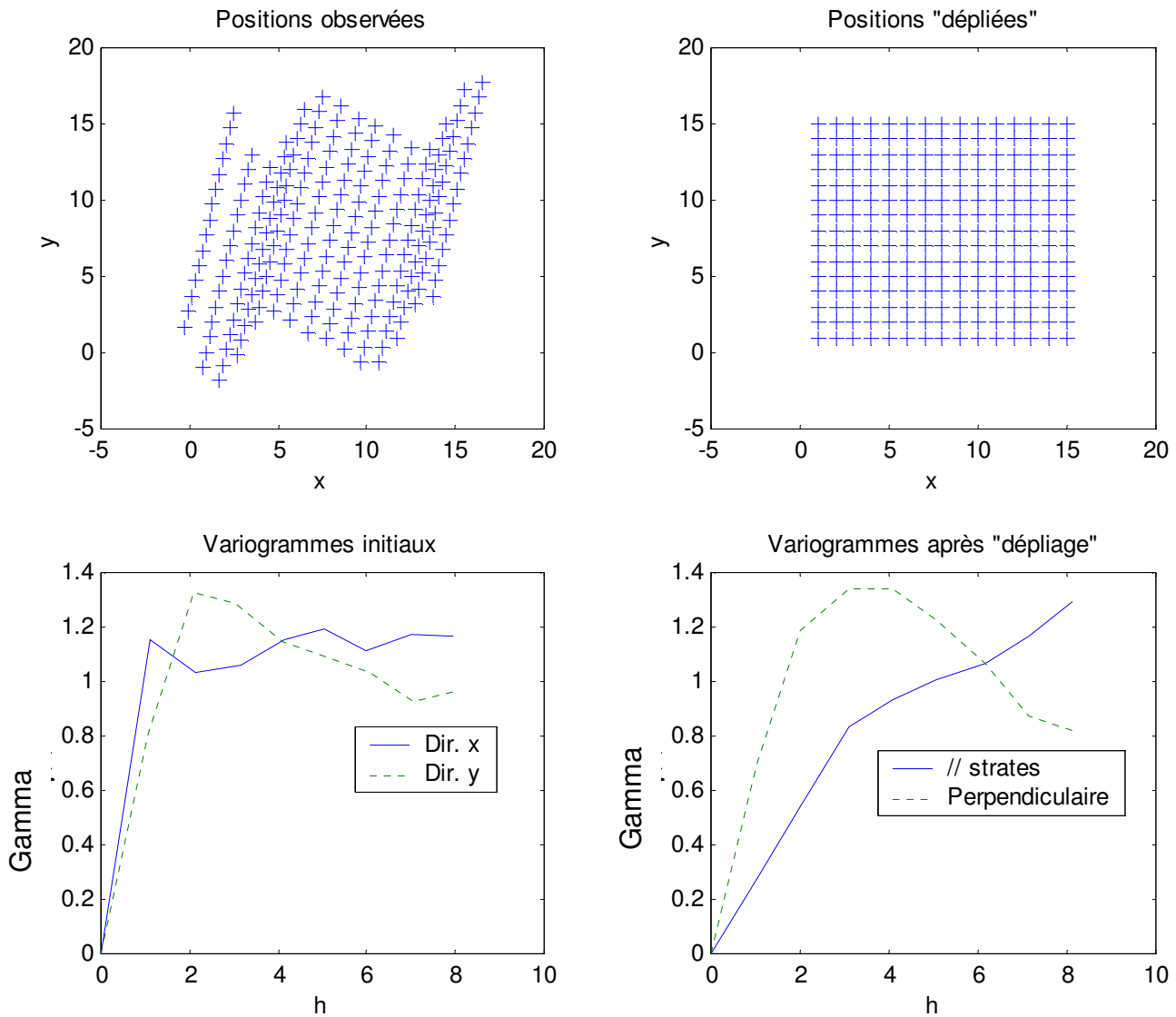
- Localiser avec soin les données
- Si on a une idée sur la distribution des erreurs de localisation, on peut essayer d'en tenir compte lors de l'estimation (il subsiste quand même une perte d'information substantielle)

Note: Ce problème est particulièrement aigu avec des données de forages (3D) qui dévient souvent de façon importante et pour lesquels les méthodes de mesure de déviations ne sont pas toujours précises.

2.6.5 Systèmes de coordonnées géologiques

Une bonne connaissance de la géologie peut parfois permettre de définir un système de coordonnées plus naturel pour le phénomène étudié. Pour l'illustrer, imaginons un gisement qui se forme dans des roches sédimentaires (ex. skarn). Les teneurs montreront une continuité spatiale dictée par le phénomène de minéralisation.

Supposons qu'un événement tectonique tardif vienne plisser les roches. Les distances entre les points, après plissement, seront modifiées altérant ainsi la structure spatiale. Dans certains cas, lorsque la tectonique n'est pas trop compliquée et que la géologie est bien connue, il est possible de "déplisser" le gisement pour retrouver les positions originales des points et ainsi mieux décrire la structure spatiale de la minéralisation.



Exemples géologiques:

- Structures plissées
- Minéralisation ou hydrocarbures liés à la présence de chenaux
- Contamination se déplaçant dans l'eau souterraine suivant les directions d'écoulement

Remarques:

- La transformation des coordonnées demande une très bonne connaissance géologique du phénomène.
- Certains logiciels (coûteux) incluent des fonctions pour effectuer cette transformation