

**UNIVERSITE TUNIS EL MANAR
FACULTE DES SCIENCES DE TUNIS
DEPARTEMENT DE GEOLOGIE**

Cours de

**Géostatistique et risques
environnementaux**

Elaboré par

Feyda SRARFI

LSGA3-GAST

Chapitre 1

NOTIONS DE BASE ET NOMENCLATURE

1. Historique

La "géostatistique" a été inventée par Georges Matheron en 1962.

Le préfixe "géo" fait référence au domaine des sciences de la terre, qui est historiquement celui où la géostatistique s'est développée. Les champs d'application actuels concernent des domaines très variés (géologie minière, hydrogéologie, climatologie, sciences sociales, ... etc.).

La terminaison "statistique" se réfère à l'utilisation de méthodes probabilistes.

L'originalité de la géostatistique par rapport à la statistique classique est la prise en compte de la dépendance entre les observations lorsque celles-ci sont situées dans l'espace. Toutefois, les méthodes géostatistiques ne sont pas exclusivement probabilistes : il existe une branche, connue sous le nom de "géostatistique transitive", qui ne fait pas appel au concept des processus aléatoires.

La géostatistique est connue depuis près de 40 ans dans les mines d'or du Witwatersrand où Daniel Krige proposa une correction statistique à la manière traditionnelle d'estimer la teneur d'un bloc de minerai à partir d'un nombre limité d'échantillons pris autour du bloc à exploiter. La théorie était formulée 10 ans plus tard par Georges Matheron (1971), qui introduisit un outil pour analyser la continuité spatiale des teneurs appelé " le variogramme " et une méthode d'estimation basée sur le variogramme appelée " le krigeage ".

L'avenir de la géostatistique semble brillant. Avec toutes les expériences et les développements des trente dernières années, la géostatistique est devenue une alternative possible aux méthodes géométriques traditionnelles de l'estimation des gisements. D'autres disciplines qui utilisent des données distribuées spatialement (pétrole, environnement, hydrologie, océanographie, foresterie) ont commencé à l'adopter.

2. Variables aléatoires

On définit une variable aléatoire en associant un nombre réel à chaque éventualité d'une expérience aléatoire.

3. Moments d'une variable aléatoire X

1- Valeurs typiques :

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

2- Notion d'espérance mathématique :

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

4. Variable régionalisée

Dans le domaine de la géostatistique, une variable régionalisée (VR) est toute fonction mathématique déterministe destinée à modéliser un phénomène présentant une structure plus ou moins prononcée dans l'espace et/ou le temps : phénomène physique ou abstrait (financiers, par exemple).

Historiquement les premières utilisations du vocabulaire et du concept de « variable régionalisée » concernaient presque exclusivement la répartition des teneurs minéralisées dans un gisement minier ; mais cet outil a par la suite trouvé des applications dans des domaines aussi variés que la météorologie, la bathymétrie et la topographie (MNT), l'environnement, l'agriculture de précision, l'épidémiologie, le génie civil, toute cartographie quantitative en général, etc.

Sous un vocabulaire différent, une variable régionalisée est strictement équivalente à la notion physique de champ, et plus précisément de champ déterministe.

5. Stationnarité

En statistique classique, l'inférence des paramètres est rendue possible par la répétition indépendante des données. En statistiques spatiales, on observe très souvent une réalisation unique des données, par exemple un épisode de pollution à l'ozone, une région agricole particulière, une épidémie végétale, etc... Pour pouvoir réaliser l'inférence statistique pour un événement unique, il faut donc en quelques sortes remplacer l'hypothèse sur les répétitions indépendantes par une hypothèse sur le champ aléatoire qui considère d'une part que certaines de ses caractéristiques sont identiques d'un point à l'autre de l'espace. On pose donc des hypothèses de stationnarité et d'ergodicité.

Un champ aléatoire $Z(x; \omega)$ sur R_d est dite ergodique pour la moyenne (ou ergodique au premier ordre) si la moyenne spatiale sur un domaine D de $Z(x; \omega)$ converge vers son espérance mathématique lorsque $D \rightarrow R_d$.

6. Notation

Z : la variable régionalisée

D : le champ (domaine sur lequel la variable régionalisée est définie) $s \rightarrow D$: une position dans le champ

$Z(s)$: une valeur prise par la variable régionalisée au point s

$Z^*(s)$: une estimation de $Z(s)$

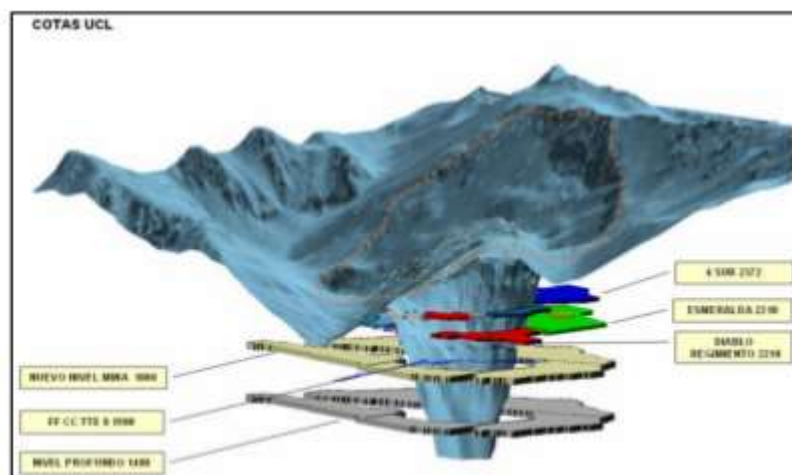
h : la distance qui sépare deux points

7. Domaines d'application de la géostatistique

Initialement appliquée au domaine minier, la géostatistique est maintenant une discipline utilisée dans l'analyse et la gestion des ressources minérales ou renouvelables, l'environnement, la climatologie, la météorologie, l'écologie mais également la finance, les technologies de pointes ou encore la santé.

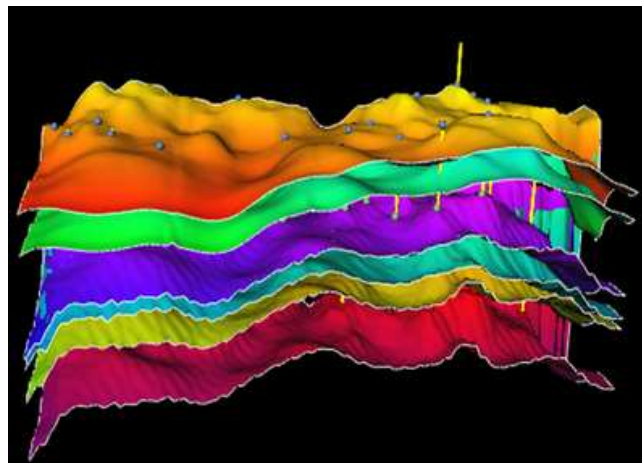
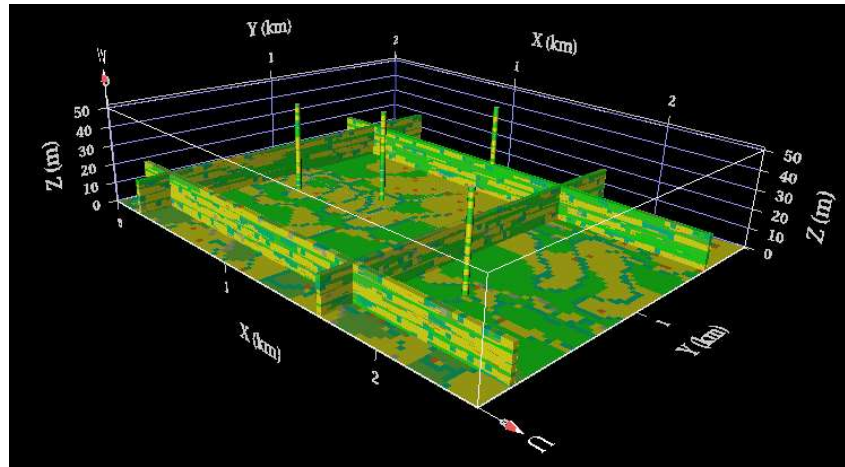
Mine : estimation des ressources *in situ*, prévision des réserves récupérables au-dessus de coupures, classification des ressources, incertitudes sur les limites de domaines, prise en compte des valeurs extrêmes de teneurs, impact de l'échantillonnage spatial sur la caractérisation des gisements, incertitudes sur les réserves ...

<<Modélisation du Dyke de la mine d'El Teniente (Chili)



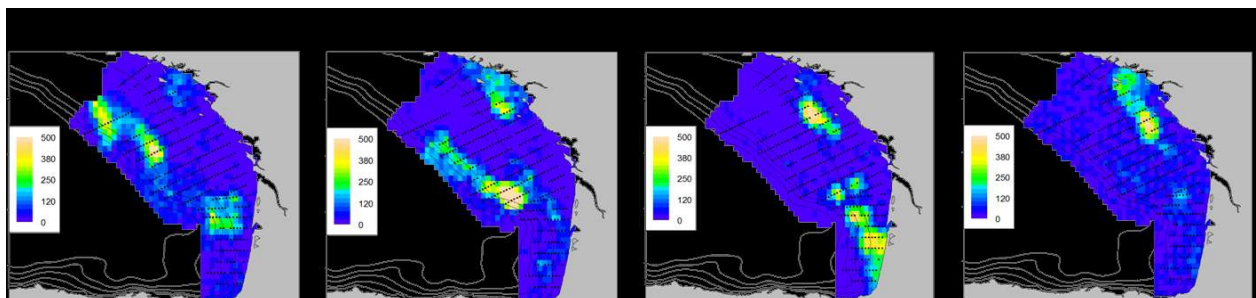
Pétrole : modélisation géométrique, caractérisation de réservoirs, ingénierie réservoir, sismique

<< Exemple de réservoir simulé avec Flumy –Empilement de couches géologiques >>

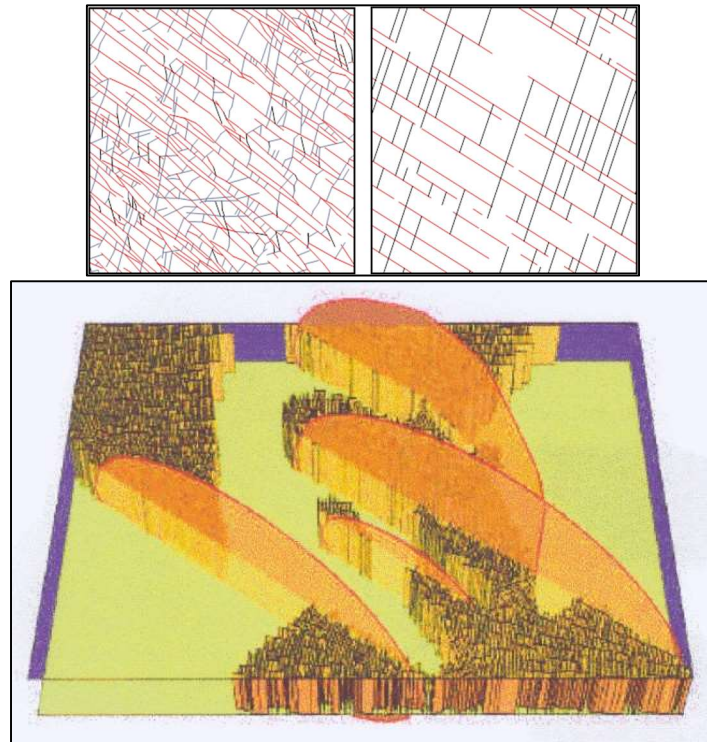


Halieutique : densité de population, migration des espèces

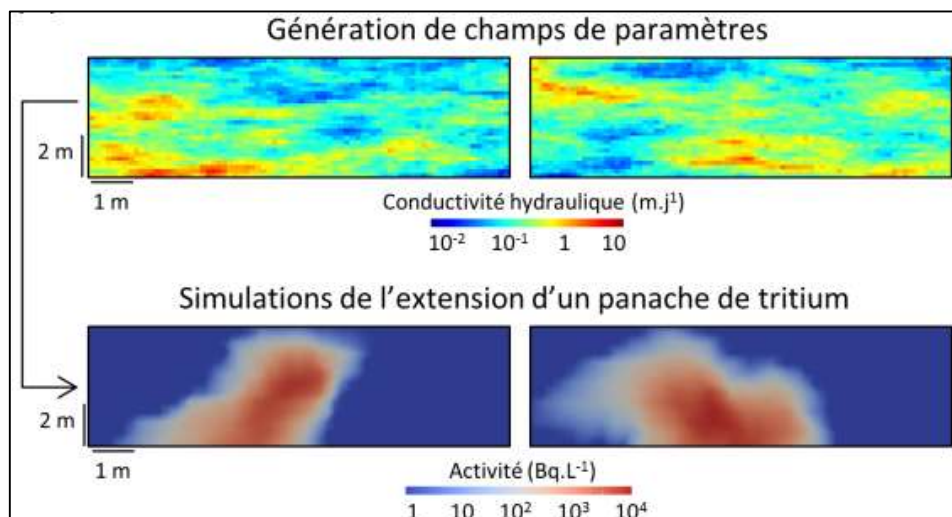
« Population de sardines dans le Golfe de Gascogne »



Milieu fracturés : blocométrie, détermination des seuils de percolation, calcul de perméabilité équivalente

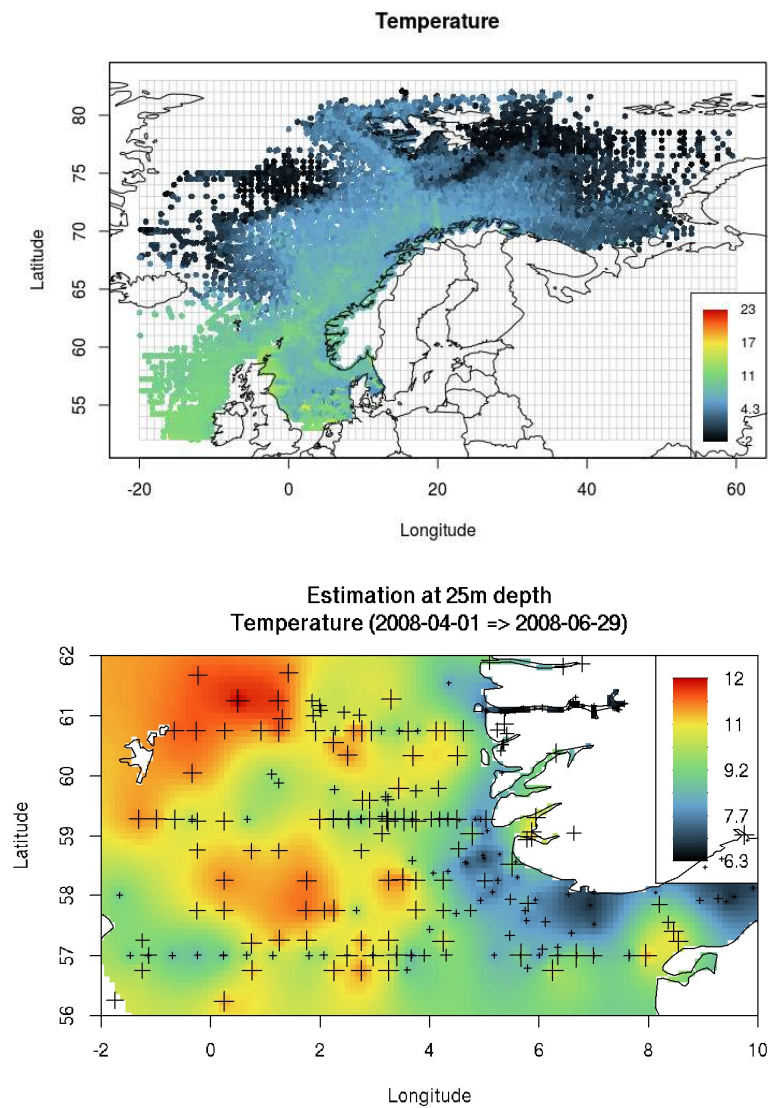
<< *Modèle de fractures hiérarchique de Tayma-Réservoir carbonaté fracturé* >>

Environnement : qualité de l'air, cartes de pollution (fond ou proximité), rivières et nappes (concentrations en polluants le long des cours d'eau), sols et sédiments de retenues, climatologie, concentration en CO₂

<< *Simulation de l'extension d'un panache de tritium* >>

Océanographie : cartographie des fonds marins, cartes de température ou de salinité, épaisseur de glaces

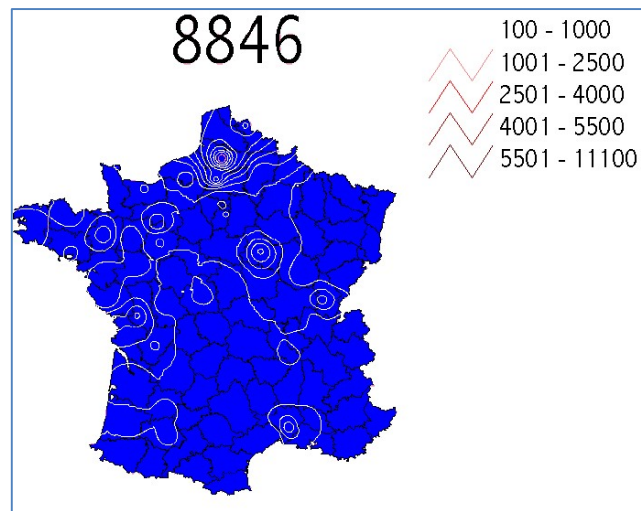
« Cartographie de la température à 25m de profondeur au large de la Norvège »>>



Santé : épidémiologie, hygiène industrielle, exposition aux ondes radioélectriques

<< Carte d'exposition aux ondes radioélectriques dans un quartier de Grenoble- Epidémie de grippe 1988/1989 en France >>





Et aussi :

- Hydrogéologie et Géotechnique,
- Agronomie,
- Modélisation démographique,
- Modélisation et évaluation des performances d'un réseau de distribution,
- Estimation des risques.

Chapitre II

MODELISATION DES CORPS GEOLOGIQUES : LE VARIOGRAMME

1. 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 , 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 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 . C'est la 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 .

$$x < \dots \dots \dots > x+h$$

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 :

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

La géostatistique veut estimer des paramètres statistiques à partir des données. Les paramètres que l'on cherchera à estimer sont la fonction de distribution des deux premiers moments (moyenne, variance, covariance) des v.a. prises deux à deux. On ne dispose 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. Hypothèses de base et définition :

On suppose que :

- i. L'espérance mathématique ne dépend pas de x , $E[Z(x)] = m$ ou L'espérance des écarts est zéro: $E[Z(x) - Z(x+h)] = 0$
- ii. 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, $C(h)$ est appelé fonction de covariance ou le variogramme $\gamma(h)$ ne dépend pas de la localisation x , seulement de h : $1/2 \text{Var}(Z(x) - Z(x+h)) = \gamma(h)$.

Évidemment, ces hypothèses supposent une certaine régularité, une certaine homogénéité de la structure géologique étudiée. 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 :

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

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 = 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.

Note :

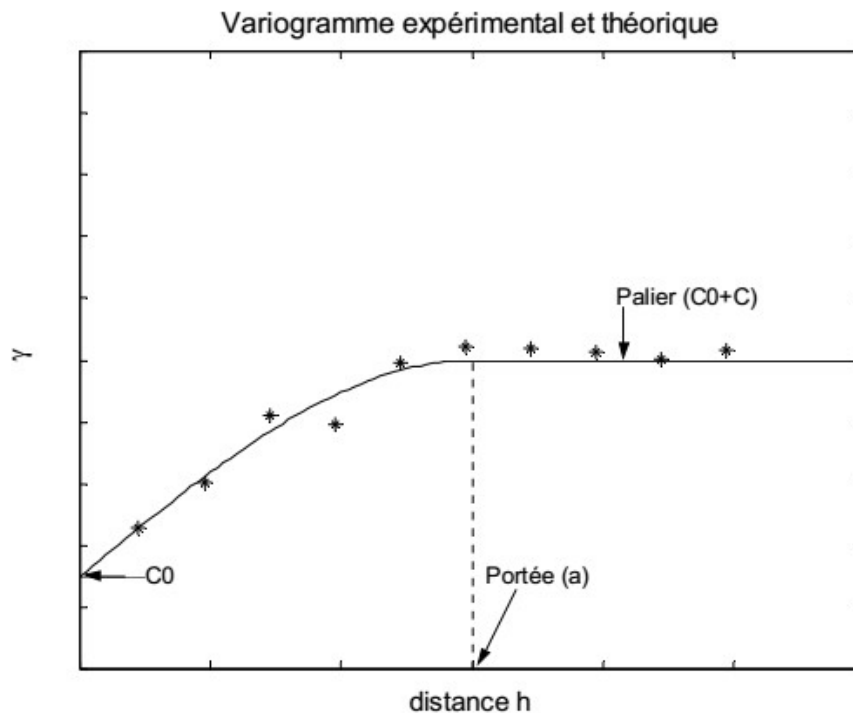
- i. Lorsque $h = 0$ on a $\text{Var}(Z(x) - Z(x)) = 0$ et non C_0
- ii. Parfois les variogrammes ne montrent pas de palier (dans ce cas, la covariance et la variance n'existent pas).
- iii. Lorsque les variogrammes montrent un palier alors on peut facilement établir le lien entre la valeur du variogramme pour la distance h et la covariance pour deux observations séparées de h .

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

$C(h)$ 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, $C(h) = 0$ 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 un grand avantage sur le covariogramme. Le variogramme est défini même s'il n'y a pas de palier.



Généralement, la stabilisation du variogramme est marquée au-delà d'une certaine valeur de h notée est appelée **portée** du variogramme. Ce dernier reste inchangé au-delà de $h = a$ et ne dépend plus de l'interdistance entre les valeurs : les mesures ne sont plus corrélées. La portée "a" donne alors un sens à la notion de **zone d'influence** d'un échantillon ou d'un sondage. Cette portée n'est pas forcément la même dans toutes les directions de l'espace. Elle permet donc de détecter une éventuelle **anisotropie** ou **isotropie** dans la structure de la minéralisation. Dans la pratique (expérimental) quand $h=0$, $\gamma(h)=C_0$ appelée **effet de pépité**.

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 .

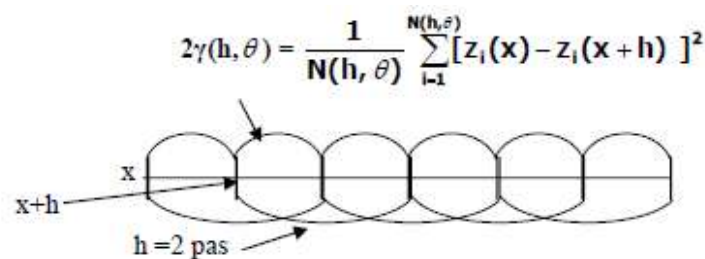
Pour un champ donné, rien n'assure que la continuité soit identique dans toutes les directions. Par exemple, il se pourrait que des teneurs montrent une meilleure continuité parallèlement à la stratigraphie que perpendiculairement à celle-ci. De même, pour la contamination par des hydrocarbures, on pourrait observer une meilleure continuité horizontalement que

verticalement en raison de la gravité. Si le nombre d'observations le permet (au moins 50), 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.



Méthode de calcul du variogramme expérimental

4. Modélisation variographique

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

Types de modèles courants En géologie, les modèles les plus courants sont :

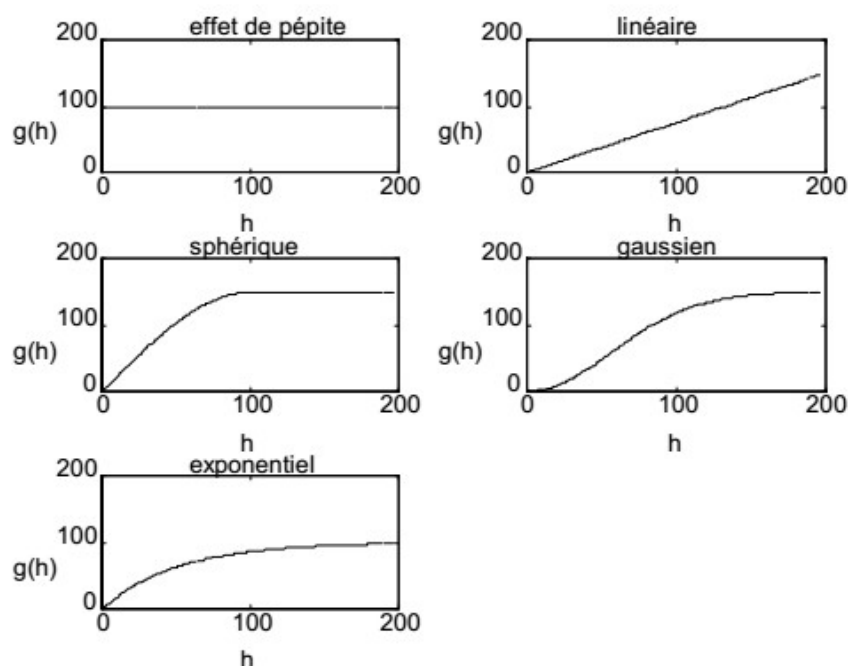
- Effet de pépite. $\gamma(h) = 0$ si $h = 0$; C_0 si $h > 0$.
- Puissance (cas particulier : linéaire). $\gamma(h) = C h^b$ $0 < b < 2$ (linéaire : $b=1$)

- 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. $\gamma(h) = C [1 - \exp(-3h/a)]$

On peut combiner plusieurs modèles en les additionnant. Ainsi, l'effet de pépité 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.

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

- Un gisement d'or présentera un variogramme irrégulier avec un fort effet de pépité et une faible portée.

- Un aquifère alluvionnaire multicouches 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.

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

Comparer les trois variogrammes obtenus, quelle information peut-on déduire ?

Note :

- Lorsque $h = 0$, par définition $\gamma(0)=0$.
- Lorsque pour $h = 0$, $\gamma(h)=C_0$, l'effet de pépité se présente donc comme une discontinuité à l'origine du variogramme. L'effet de pépité peut représenter des erreurs, de réelles microstructures ou/et des structures d'une certaine taille non-détectées par un échantillonnage insuffisant.
- 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.
- 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.

4. Anisotropies

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

Exemples :

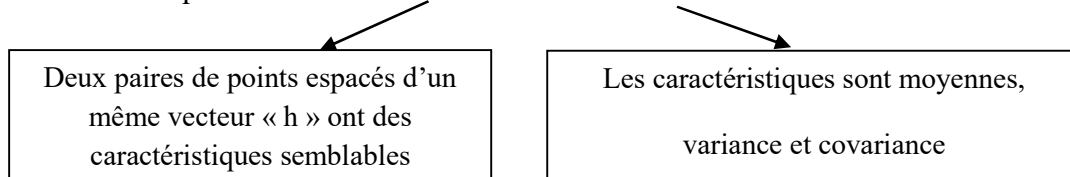
- Un corps géologique présentant une forme lenticulaire : on peut avoir une meilleure continuité selon l'allongement principal des lentilles ;
- Un corps stratiforme : meilleure continuité parallèlement aux strates que perpendiculairement.
- Méandres et chenaux : meilleure continuité le long des paléochenaux que perpendiculairement.

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

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.

NB : Géostatistique linéaire et Stationnarité du second ordre



Précautions pour l'estimation du variogramme

*Sensibilité aux choix des classes de distances

Le choix des classes (la tolérance) et leur nombre influence l'allure du variogramme. Une règle simple : le nombre de paires > 30 et $h < D/2$ ($D = \max(d_{ij})$)

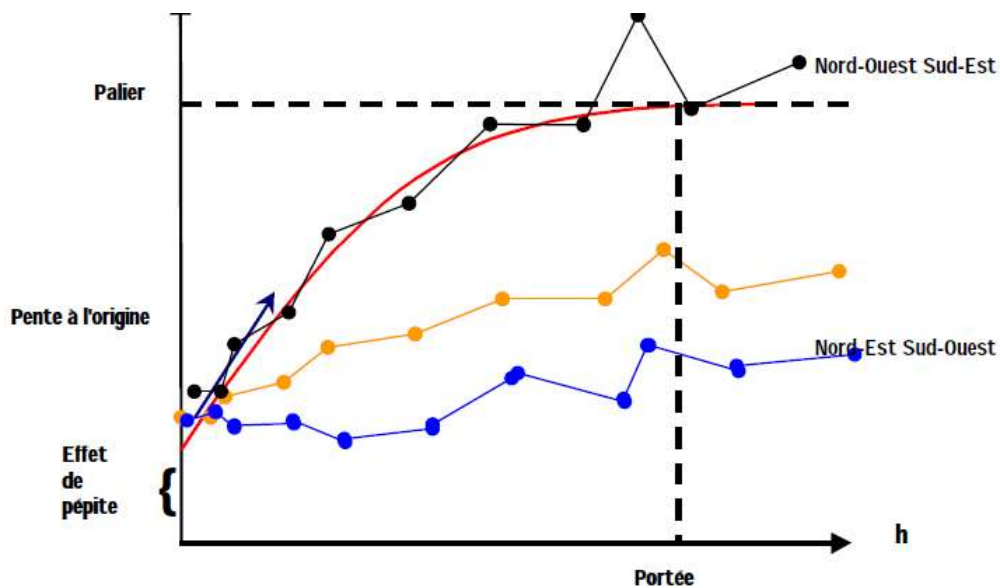
*Sensibilité au domaine d'étude

La stationnarité du phénomène doit conduire à découper la zone d'étude si elle n'est pas observée sur le domaine global, des valeurs extrêmes (outliers) aussi.

*Sensibilité à l'anisotropie

La présence d'anisotropie dans les données est très importante pour ajuster un modèle

Modèle de variogramme multidirectionnel

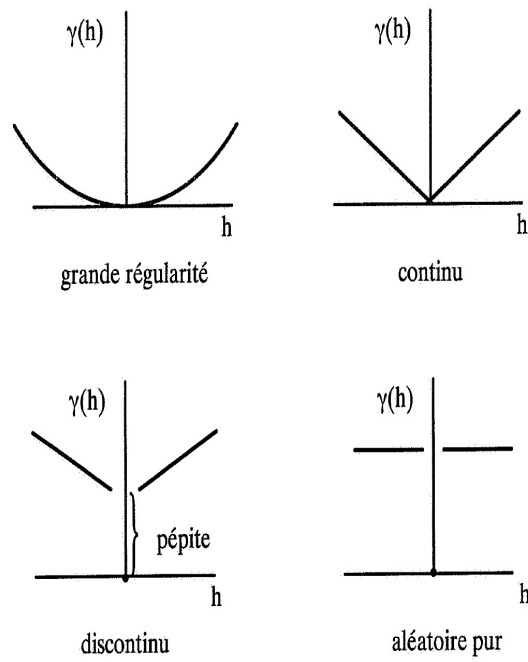


Analyse du variogramme Caractéristiques :

Comportement à l'origine ($h=0$), Comportement à longue distance (palier et portée), Comportement directionnel.

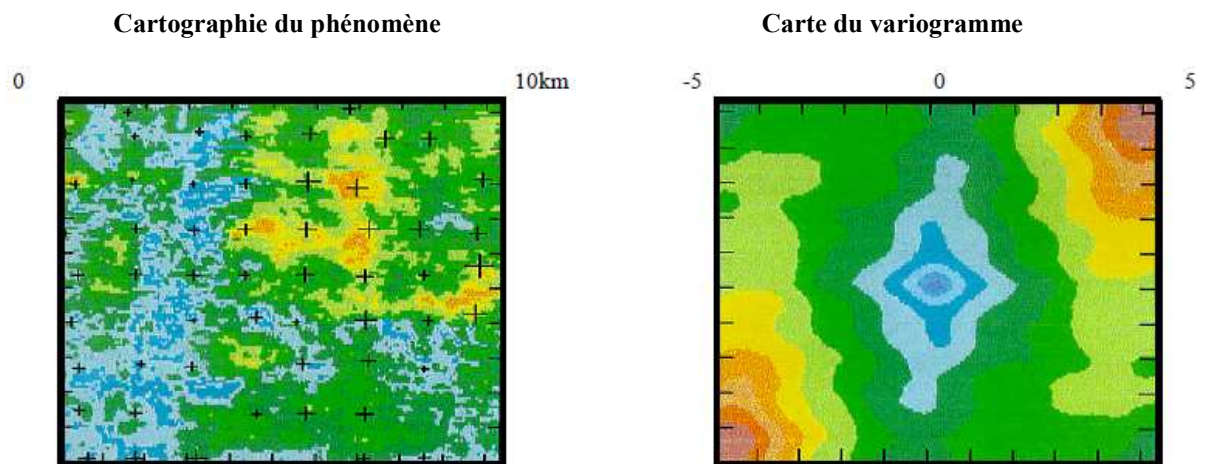
A l'origine : Mesure de la continuité du phénomène. Le degré d'irrégularité : tangente verticale = grande irrégularité, tangente horizontale = grande régularité · discontinuité à l'origine purement aléatoire · erreurs de mesures · structures au rang inférieur à l'échelle de l'échantillon.

A longue distance : borné : les données ont un maximum d'hétérogénéité (variance = palier), la portée = distance à laquelle le palier est atteint (théoriquement deux points séparés par $d > portée$ sont non corrélés).



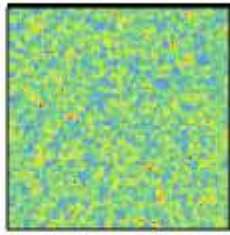
Comportement du variogramme à l'origine

Illustration d'une anisotropie

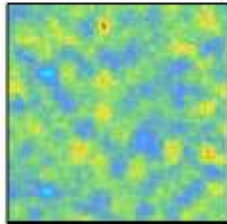


Influence de la portée

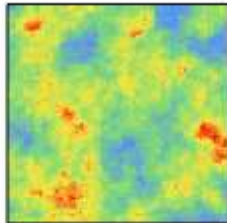
3 unités



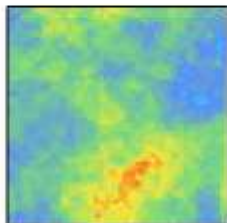
10 unités



25 unités

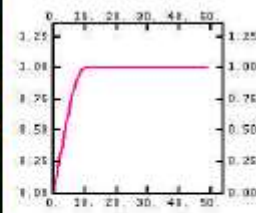
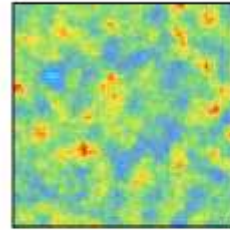


50 unités

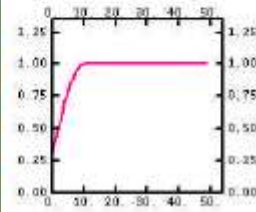
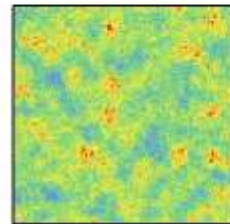


Influence de l'effet de pépite

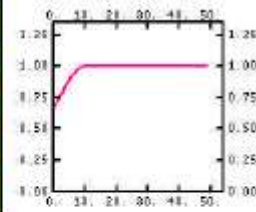
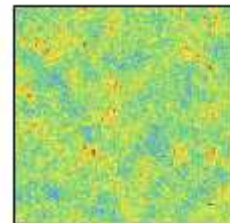
pas de pépite



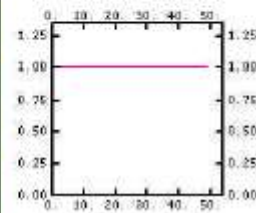
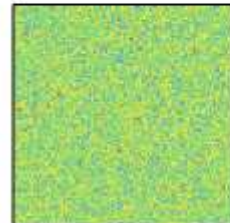
33% pépite



66% pépite

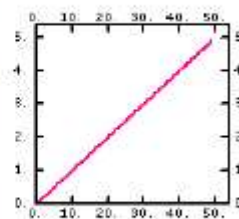
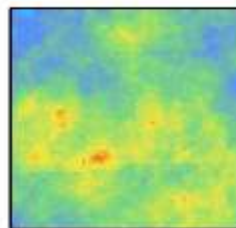


100% pépite



Exemple de variogrammes stationnaire

Linéaire



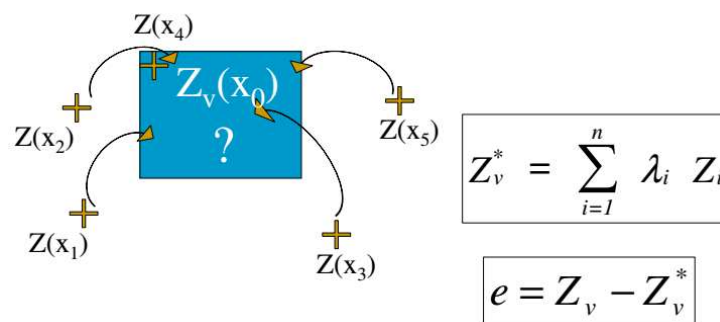
Chapitre III

INTERPOLATION OPTIMALE ET CARTOGRAPHIE : LE KRIGEAGE

1. Problématique

Le géologue a besoin de connaître la valeur de la variable à des endroits non échantillonnés. Il procède à une estimation à partir des données disponibles. C'est la procédure d'interpolation.

L'estimation d'une valeur d'une variable en un point, en une surface ou en un volume, nécessite une combinaison linéaire des informations disponibles avoisinantes.



L'interpolation repose sur un principe important en géographie appelé, l'autocorrélation spatiale. Selon ce principe, des objets rapprochés dans l'espace tendent à posséder des caractéristiques similaires.

$Z(x_i)$ sont connus, il reste à déterminer λ_i .

λ_i représente le pourcentage d'influence de chaque x sur x_0 , appelé pondérateur.

Pour la plupart des méthodes d'estimation $\sum \lambda_i = 1$

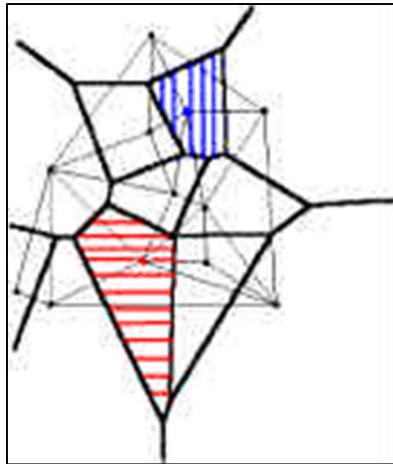
2. Méthodes d'interpolation

2.1 L'approche déterministe :

Elle regroupe les méthodes d'interpolation dont la fonction de structure (fonction de pondération) est choisie a priori. Elles ne fournissent pas d'informations sur la variance d'estimation.

VORONOÏ : Basée sur des critères de voisinages simples elle construit une parcellisation du domaine d'étude.

Chaque cellule contient un et un seul point de l'échantillon, l'ensemble des points de l'espace appartenant à la cellule a pour plus proche voisin le point d'échantillonnage associé à la cellule. La valeur du point échantillonné est associée à tous les points de la parcelle ou cellule. Cette approche est semblable à la triangulation. Les limites sont évidentes, car il y a de brusques sauts de discontinuité.



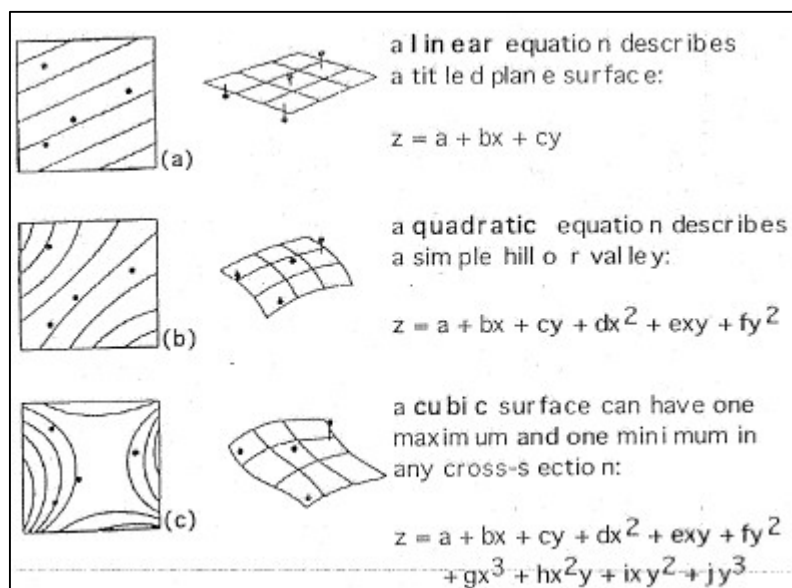
Surface trend

Cette méthode utilise la combinaison linéaire de fonctions de bases. Il s'agit d'appréhender le phénomène par l'utilisation d'un polynôme dont le degré est laissé à l'estimation du thématicien. D'une manière générale on écrit : $Z = a + b_1x + b_2y + b_3xy + b_4x^2 + b_5y^2 + \dots$

Les paramètres du modèle sont estimés par la méthode des moindres carrés. Celle-ci vise à minimiser l'erreur entre la valeur prédite et la valeur observée.

$$\sum_i (f(X_i) - Z_i)^2$$

Cette méthode s'intéresse peu aux changements locaux, aux irrégularités, mais seulement à la tendance. Le choix du degré fixe la complexité de l'interpolation dans la pratique on utilise des degrés de 3 à 5. Cette méthode d'interpolation a l'inconvénient de s'accroître ou décroître rapidement dans les zones où l'échantillonnage est faible et particulièrement sur les bords de la zone.

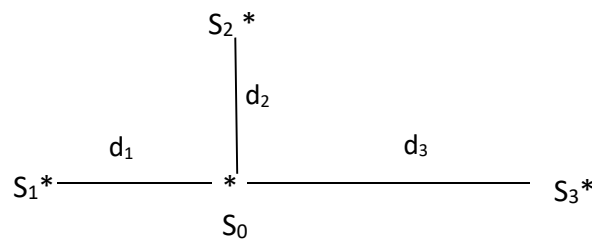


Méthode inverse des distances

C'est une méthode d'interpolation simple, basée sur la mesure des distances d_i entre le point de mesure et le point où en va estimer la variable ; avec $\lambda_i = \frac{\frac{1}{d_i}}{\sum \frac{1}{d_i}}$ et $Z^*(x) = \sum \lambda_i Z(x_i)$; $Z^*(x)$ est la valeur estimée.

Exemple numérique

Soient trois sondages S1, S2 et S3 disposés dans un lac selon le schéma suivant :



Sondage	Distance (m)	1/di	λ_i	Z(xi)
S1	1			22%
S2	2			20%
S3	3			16%

Calculer $Z^*(S_0)$

La méthode des inverses des distances ou les autres méthodes d'interpolation simples ne prennent pas en compte la nature de la variable (concentration, topographie, profondeur,...). Elles tiennent compte seulement des distances entre les points et non pas l'emplacement exacte de ces points (coordonnées). Ce qui incite à des erreurs importantes.

2.2 Krigage et spatialisation des données

Puisqu'on peut calculer la variance d'estimation pour tout estimateur linéaire, on choisit celui qui assure la variance d'estimation minimale. C'est le krigage. Dans le cas stationnaire, on en reconnaît 2 types principaux, soit le krigage simple et le krigage ordinaire. Ce dernier est, de loin, le plus fréquemment utilisé.

Supposons que l'on veuille estimer un bloc v centré au point x_0 . Notons Z_v la vraie valeur (inconnue) de ce bloc et Z_v^* l'estimateur que l'on obtient.

L'estimateur est linéaire :

$$Z_v^* = \sum_{i=1}^n \lambda_i Z_i$$

Pour que l'estimateur soit sans biais, il faut que :

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

Comme la variance d'estimation s'écrit directement en termes de variogramme, on peut aussi écrire le système de krigeage en fonction du variogramme.

Ceci tient au fait que $C(h) = \sigma^2 - \gamma(h)$ et que $\sum \lambda_i = 1$.

$$\sum_{j=1}^n \lambda_j \gamma(x_i, x_j) - \mu = \bar{\gamma}(v, x_i) \quad \forall i = 1 \dots n$$

et

$$\sigma_k^2 = \sum_{i=1}^n \lambda_i \bar{\gamma}(v, x_i) - \bar{\gamma}(v, v) - \mu$$

En effet, dans ce cas,

$$E[Z_v^*] = \sum_i \lambda_i E[Z_i] = \sum_i \lambda_i m = m$$

On a un problème de minimisation d'une fonction quadratique sous contrainte d'égalité que l'on résout par la méthode de Lagrange.

Où μ est le multiplicateur de Lagrange. Le minimum est atteint lorsque toutes les dérivées partielles par rapport à chacun des λ_i et par rapport à μ s'annulent. Ceci conduit au système de krigeage ordinaire. Il est intéressant de visualiser le système de krigeage ordinaire et la variance de krigeage ordinaire sous forme matricielle :

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

Par rapport à d'autres méthodes d'interpolation le krigeage se distingue par le calcul d'erreur d'estimation associée.

- Le système de krigeage ordinaire est un système à (N+1) équations à (N+1) inconnues qui sont les N pondérateurs (λ_i) et μ qui est le paramètre de Lagrange.
- Le krigeage ordinaire est un estimateur linéaire sans biais. C'est un interpolateur exact.

3. Validation croisée

Une pratique intéressante pour valider le modèle de variogramme et le voisinage utilisé pour le krigeage consiste à effectuer une validation croisée. Le principe est d'éliminer à tour de rôle chaque observation et de l'estimer à l'aide de ses voisins. En chaque point, on obtient donc une valeur vraie et une valeur estimée que l'on peut comparer pour déterminer si le modèle fournit des estimations se comportant comme prévu, si le voisinage utilisé est adéquat, etc.

Plus précisément, soit Z_i^* l'estimation obtenue par krigeage au point "i" (en enlevant la valeur observée Z_i) ainsi que la variance de krigeage σ_{ki}^2 .

On peut définir un résidu $e_i = Z_i - Z_i^*$ et un résidu normalisé

$$n_i = \frac{e_i}{\sigma_{ki}}$$

4. Variances de blocs, de dispersion, d'estimation

4.1 Variances de blocs :

Décrit l'amplitude théorique des variations des teneurs de bloc pour un domaine infini. C'est l'analogue de la variance ponctuelle pour des blocs. Cette notion n'est définie que pour les modèles de variogramme avec palier.

4.2 Variance de dispersion

La variance de bloc permet de calculer la variance théorique de la teneur de bloc dans un domaine d'extension infinie. Bien sûr les gisements réels ne sont jamais infinis et il est souhaitable de pouvoir prévoir l'amplitude des variations des teneurs de bloc pour un domaine fini correspondant au gisement ou à une partie du gisement.

Considérons un grand bloc V_j découpé en petits blocs v_i .

On a :

$$Z(V_j) = \frac{1}{n} \sum_{i=1}^n Z(v_i)$$

	v_i		

V_j

On peut vouloir déterminer l'importance de la variation de v_i dans V_j , en moyenne pour l'ensemble des blocs V . C'est ce que l'on appelle la variance de dispersion de v dans V que l'on note $D^2(v|V)$.

La variance de dispersion n'est autre qu'une différence de variabilité de teneurs mesurées sur 2 volumes différents. On notera en particulier que, pour les modèles croissants de variogramme, on aura :

$v \rightarrow 0$	$D^2(v V) \rightarrow \bar{\gamma}(V, V)$
$v \rightarrow V$	$D^2(v V) \rightarrow 0$
$V \rightarrow \infty$	$D^2(v V) \rightarrow \sigma_v^2$

4.3 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 v.a. Z_v que l'on veut estimer, en formant une combinaison linéaire des valeurs observées en différents endroits,

$$Z_v^* = \sum_{i=1}^n \lambda_i Z_i$$

où:

Z_i : valeur observée au point x_i (v.a.)

Z_v^* : estimateur de Z_v

On définit l'erreur d'estimation : $e = Z_v - Z_v^*$

4.4 Calcul des quantités $\bar{\gamma}(v, v)$ ou $\bar{\gamma}(x_i, v)$

Les termes $\gamma(v,v)$ ou $\gamma(x_i,v)$ sont requis pour obtenir les variances de blocs, les variances de dispersions et les variances d'estimation. Ces valeurs peuvent être calculées de trois façons différentes :

- i. Intégration de $\gamma(h)$ pour obtenir une expression analytique : praticable surtout en 1-D.
- ii. Abaques pour certains modèles (sphérique ; exponentiel).
- iii. Approximation numérique par ordinateur.

Exemple numérique de krigeage

Soit les points suivants :

x_1

x_2

x_0

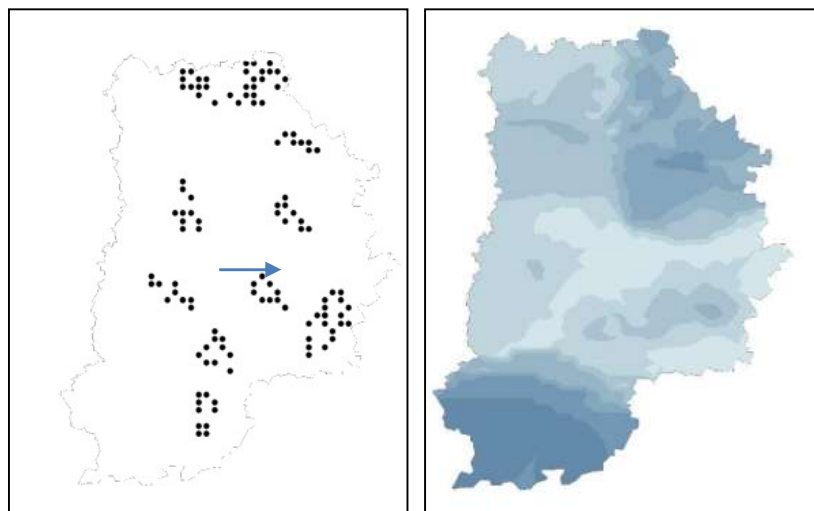
x_3

$$x_1=(0,1) \quad Z_1=9$$

$$x_2=(0,0) \quad Z_2=3$$

$$x_3=(3,0) \quad Z_3=4$$

On veut estimer le point x_0 situé à (1,0). Supposons que l'on a un modèle sphérique, avec effet de pépité 1, palier 11 et portée 3.



Données ponctuelles

Données surfaces