

Chapitre 1

Notions de géostatistique

Ce premier chapitre introduit pour certains et redéfinit pour d'autres les notions de géostatistique nécessaires à la bonne compréhension de certaines étapes de la méthode stochastique de délimitation des zones de protection proposée dans ce travail, et notamment les concepts de fonction aléatoire, de krigeage, de cokrigeage et de simulation stochastique.

1.1 Géostatistique linéaire

La variabilité spatiale du milieu souterrain naturel induit que celui-ci ne peut être objectivement décrit de façon quantitative par des modèles déterministes. Toutefois, de par les structures géologiques qui le composent, il n'est en général pas non plus totalement "chaotique" et aléatoire. Il présente une structure spatiale que les méthodes géostatistiques permettent de révéler par l'analyse des données disponibles.

A la base de toute analyse géostatistique, il faut nécessairement un jeu de données réparties dans l'espace. Notons : $(z^{\text{exp}}(\underline{x}_i))_{i=1,p}$ les p valeurs expérimentales de la variable d'étude z mesurée en différents points de coordonnées généralisées $(\underline{x}_i)_{i=1,p} = (x_i, y_i)_{i=1,p}$.

1.1.1 Concept de Fonction Aléatoire

1.1.1.1 Variable Régionalisée

Les quelques valeurs mesurées aux points $(\underline{x}_i)_{i=1,p}$ sont celles d'une fonction $z(\underline{x})$, supposée continue, définie sur un espace borné Ω , et prenant ses valeurs dans \mathfrak{R} . Cette fonction est appelée **variable régionalisée** [MATHERON, 1970].

1.1.1.2 Variable Aléatoire

Une autre approche est de considérer le jeu des p valeurs expérimentales $(z^{\text{exp}}(\underline{x}_i))_{i=1,p}$ comme le résultat d'un mécanisme aléatoire : la valeur échantillonnée $z^{\text{exp}}(\underline{x}_i)$ représente une des valeurs possibles de la **variable aléatoire** $Z(\underline{x}_i)$. En chaque point $(\underline{x}_i)_{i=1,p}$, la variable aléatoire $(Z(\underline{x}_i))_{i=1,p}$ peut *a priori* prendre différentes valeurs $(z(\underline{x}_i))_{i=1,p}$.

1.1.1.3 Fonction Aléatoire

La variable régionalisée $z(\underline{x})$ est supposée être UNE réalisation particulière d'une **fonction aléatoire** $Z(\underline{x})$, continue, et obéissant à une **loi spatiale**. CHAUVET [1992] insiste sur le fait que cette démarche constitue "un choix méthodologique, librement consenti". Choix qui, une fois réalisé, nous permet d'aller puiser dans la trousse à outils géostatistiques pour tenter de faire apparaître cette loi spatiale. L'intérêt est de passer de mesures discrètes – données ponctuelles réparties dans l'espace – à un phénomène continu. La figure 1.1 schématise ce concept de fonction aléatoire qui combine le mécanisme de régionalisation et l'approche probabiliste.

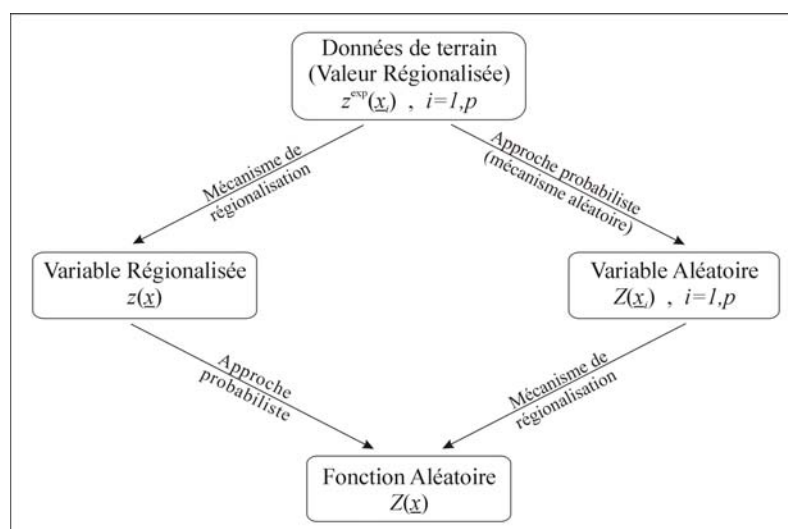


Figure 1.1 : Concept de fonction aléatoire [inspiré de WACKERNAGEL, 1995]

1.1.2 Loi spatiale

Par définition, la loi spatiale F associée à la fonction aléatoire $Z(\underline{x})$ permet de définir celle-ci en terme de probabilité. C'est la loi de probabilité conjointe des $(Z(\underline{x}_1), Z(\underline{x}_2), \dots, Z(\underline{x}_n))$, quel que soit l'ordre n et quels que soient les points $(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$.

$$F(s_1, s_2, \dots, s_n) = P\{ (Z(\underline{x}_1) \leq s_1) \text{ et } (Z(\underline{x}_2) \leq s_2) \text{ et } \dots \text{ et } (Z(\underline{x}_n) \leq s_n) \} \quad (1.1)$$

avec $P\{ A \}$, la probabilité d'occurrence de l'événement A .

Par définition, la loi spatiale complète inclut la loi monovariante ($n = 1$) et la loi bivariable ($n = 2$). La géostatistique linéaire se contente de l'ordre 2 de la loi spatiale.

1.1.2.1 Loi spatiale monovariante (loi d'ordre 1)

La loi spatiale monovariante permet de définir la distribution des probabilités de $Z(\underline{x})$ en tout point. Cette fonction, appelée classiquement fonction de répartition (ou de distribution), est définie comme la probabilité que $Z(\underline{x})$ soit inférieur à une valeur numérique s donnée.

$$F_{Z(\underline{x})}(s) = P\{ Z(\underline{x}) \leq s \} \quad (1.2)$$

Pour une fonction aléatoire continue, la densité de probabilité est définie selon :

$$f_{Z(\underline{x})} = \frac{dF_{Z(\underline{x})}(s)}{ds} \quad (1.3)$$

Cette fonction permet de définir la probabilité que $Z(\underline{x})$ soit inclus dans un petit intervalle δs :

$$f_{Z(\underline{x})}(s) \cdot \delta s = P\{ s \leq Z(\underline{x}) < s + \delta s \} \quad (1.4)$$

La fonction de distribution des probabilités se déduit de la fonction de densité par intégration :

$$F_{Z(\underline{x})}(s) = \int_{-\infty}^s f_{Z(\underline{x})}(s) \cdot ds \quad (1.5)$$

$$F_{Z(\underline{x})}(\infty) = \lim_{s \rightarrow +\infty} F_{Z(\underline{x})}(s) = 1 \quad (1.6)$$

La densité de probabilité intervient dans le calcul de moments :

Moment d'ordre m : Le moment d'ordre m , $m \in \mathbb{N}$, de la densité de probabilité $f_{Z(\underline{x})}(s)$ au point \underline{x} , $M_{Z(\underline{x})}^m$, est défini comme suit :

$$M_{Z(\underline{x})}^m = \int_{-\infty}^{+\infty} s^m \cdot f_{Z(\underline{x})}(s) \cdot ds \quad (1.7)$$

Moment centré d'ordre m :

$$Mc_{Z(\underline{x})}^m = \int_{-\infty}^{+\infty} s^m \cdot f_{Z(\underline{x})}(s) \cdot ds - (M_{Z(\underline{x})}^1)^m \quad (1.8)$$

Les moments d'ordre 1 et 2 permettent de définir l'**espérance mathématique** et la **variance** :

L'espérance mathématique de Z au point \underline{x} , $E[Z(\underline{x})]$, est définie comme le moment d'ordre 1 de la densité de probabilité $f_{Z(\underline{x})}(s)$.

$$E[Z(\underline{x})] = M_{Z(\underline{x})}^1 = \int_{-\infty}^{+\infty} s \cdot f_{Z(\underline{x})}(s) \cdot ds \quad (1.9)$$

La variance de Z au point \underline{x} , $\sigma_Z^2(\underline{x})$, est définie comme le moment centré d'ordre 2 de la densité de probabilité $f_{Z(\underline{x})}(s)$.

$$\begin{aligned} \sigma_Z^2(\underline{x}) &= \int_{-\infty}^{+\infty} s^2 \cdot f_{Z(\underline{x})}(s) \cdot ds - (E[Z(\underline{x})])^2 \\ &= E[(Z(\underline{x}) - E[Z(\underline{x})])^2] \end{aligned} \quad (1.10)$$

Sans hypothèse particulière, ces grandeurs statistiques dépendent du point pour lequel elles sont calculées.

1.1.2.2 Loi spatiale bivariable (loi d'ordre 2)

La loi spatiale bivariable de $Z(\underline{x})$, $F_{Z(\underline{x}_1), Z(\underline{x}_2)}(s_1, s_2)$, est définie comme la probabilité conjointe sur deux points $(\underline{x}_1, \underline{x}_2)$.

$$F_{Z(\underline{x}_1), Z(\underline{x}_2)}(s_1, s_2) = P\{ (Z(\underline{x}_1) \leq s_1) \text{ et } (Z(\underline{x}_2) \leq s_2) \} \quad (1.11)$$

La densité de probabilité conjointe $f_{Z(\underline{x}_1), Z(\underline{x}_2)}(s_1, s_2)$ est définie comme suit :

$$f_{Z(\underline{x}_1), Z(\underline{x}_2)}(s_1, s_2) = \frac{\partial^2 F_{Z(\underline{x}_1), Z(\underline{x}_2)}(s_1, s_2)}{\partial s_1 \partial s_2} \quad (1.12)$$

La **covariance** et le **variogramme**, deux des outils de base de la géostatistique, sont entièrement déterminés par la loi spatiale bivariable (la réciproque étant fausse). Ils permettent d'estimer le degré de corrélation entre deux points $(\underline{x}_1, \underline{x}_2)$, par mesure quadratique de l'écart statistique entre les valeurs $Z(\underline{x}_1)$ et $Z(\underline{x}_2)$.

Covariance non centrée :

$$\begin{aligned} C_Z^{nc}(\underline{x}_1, \underline{x}_2) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} s_1 \cdot s_2 \cdot f_{Z(\underline{x}_1), Z(\underline{x}_2)}(s_1, s_2) \cdot ds_1 ds_2 \\ &= E[Z(\underline{x}_1) \cdot Z(\underline{x}_2)] \end{aligned} \quad (1.13)$$

Covariance centrée (définition la plus utilisée) :

$$C_Z(\underline{x}_1, \underline{x}_2) = C_Z^{nc}(\underline{x}_1, \underline{x}_2) - E[Z(\underline{x}_1)] \cdot E[Z(\underline{x}_2)] \quad (1.14)$$

Variogramme :

$$\gamma_Z(\underline{x}_1, \underline{x}_2) = \frac{1}{2} \cdot E \left[(Z(\underline{x}_1) - Z(\underline{x}_2))^2 \right] \quad (1.15)$$

1.1.3 Hypothèses de base

Le but avoué de toute l'étude géostatistique est de construire un modèle mathématique de la fonction aléatoire $Z(\underline{x})$, en partant du jeu de données expérimentales $(z^{\text{exp}}(\underline{x}_i))_{i=1,p}$, unique réalisation de $Z(\underline{x})$ à disposition. Pour pouvoir déduire des lois de probabilité au départ d'une réalisation unique, deux hypothèses sont nécessaires.

1.1.3.1 Hypothèse d'ergodicité

Un processus est dit ergodique si son information statistique peut être obtenue à partir d'une réalisation quelconque définie sur un domaine spatial Ω , quand celui-ci tend vers l'infini. Les moyennes spatiales convergent vers leur espérance mathématique quand Ω augmente indéfiniment. Cette hypothèse très forte implique que les moyennes spatiales réalisées sur la variable régionalisée $z(\underline{x})$ sont équivalentes aux espérances mathématiques calculées sur la fonction aléatoire $Z(\underline{x})$.

$$m_z = E[Z(\underline{x})] = \lim_{\Omega \rightarrow \infty} \frac{1}{\Omega} \iint_{\Omega} z(\underline{x}) \cdot d\underline{x} \quad (1.16)$$

$$C_z^{nc}(\underline{h}) = E[Z(\underline{x}) \cdot Z(\underline{x} + \underline{h})] = \lim_{\Omega \rightarrow \infty} \frac{1}{\Omega} \iint_{\Omega} z(\underline{x}) \cdot z(\underline{x} + \underline{h}) \cdot d\underline{x} \quad (1.17)$$

Intuitivement, la convergence des moyennes spatiales vers les moyennes statistiques est atteinte lorsque le domaine Ω est suffisamment grand pour que toutes les structures spatiales puissent s'y trouver.

1.1.3.2 Hypothèse de stationnarité

Définition [MATHERON, 1970] : La fonction aléatoire $Z(\underline{x})$ est dite stationnaire si sa loi spatiale est invariante par translation. Si $(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$ sont n points d'appui arbitraires, et si \underline{h} est un vecteur quelconque, les n variables aléatoires $(Z(\underline{x}_1), Z(\underline{x}_2), \dots, Z(\underline{x}_n))$ ont la même loi spatiale (à n variables) que les variables aléatoires $(Z(\underline{x}_1 + \underline{h}), Z(\underline{x}_2 + \underline{h}), \dots, Z(\underline{x}_n + \underline{h}))$.

Espérance mathématique et variance : Si elles existent, l'espérance mathématique et la variance d'une fonction aléatoire stationnaire sont des constantes indépendantes du point d'appui.

$$F_{Z(\underline{x})}(s) = F(s) \quad (1.18)$$

$$E[Z(\underline{x})] = m_z \quad (1.19)$$

$$\sigma_z^2(\underline{x}) = E[(Z(\underline{x}) - E[Z(\underline{x})])^2] = \sigma_z^2 \quad (1.20)$$

Covariance : Considérons deux points d'appui, \underline{x}_0 et $\underline{x}_0 + \underline{h}$. Si les deux fonctions aléatoires $Z(\underline{x}_0)$ et $Z(\underline{x}_0 + \underline{h})$ sont stationnaires et admettent une espérance et une variance, elles admettent aussi une covariance qui ne dépend pas du point d'appui.

$$C_Z^{nc}(\underline{x}_0, \underline{x}_0 + \underline{h}) = C_Z^{nc}(\underline{h}), \forall \underline{x}_0 \quad (1.21)$$

Pour $\underline{h} = \underline{0}$, $C_Z^{nc}(\underline{0}) = E[(Z(\underline{x}))^2] = \sigma_Z^2$, variance de $Z(\underline{x})$. Pour qu'une fonction aléatoire stationnaire admette une covariance, il faut et il suffit qu'elle admette une variance finie.

Hypothèse de stationnarité du deuxième ordre : Une fonction aléatoire $Z(\underline{x})$ sera dite stationnaire d'ordre 2 si en tout point, $Z(\underline{x}_0)$ admet une espérance m_Z indépendante du point d'appui, et si pour tout vecteur \underline{h} , la covariance :

$$C_Z(\underline{h}) = E[Z(\underline{x}_0 + \underline{h}) \cdot Z(\underline{x}_0)] - m_Z^2 \quad (1.22)$$

existe et ne dépend pas du point d'appui. Cette hypothèse n'entraîne pas la stationnarité au sens strict, mais elle est suffisante pour la théorie des variables régionalisées et les développements qui nous seront utiles par la suite.

La structure spatiale d'un processus stationnaire est donc conservée sur l'entièreté du domaine. Il en résulte d'une part, que des observations dans différentes portions de l'espace pourront être considérées comme autant de réalisations différentes de la fonction aléatoire $Z(\underline{x})$, et d'autre part, que les calculs de probabilité ne dépendent pas du point d'appui.

1.1.3.3 Hypothèse intrinsèque

Cette hypothèse est moins restrictive que l'hypothèse de stationnarité du deuxième ordre. Elle consiste à dire que même si la variance de la fonction aléatoire $Z(\underline{x})$ n'est pas finie, la variance des premiers accroissements de $Z(\underline{x})$ est finie et ces premiers accroissements sont eux-mêmes stationnaires au deuxième ordre. On satisfait alors les deux relations suivantes :

$$E[Z(\underline{x}_0 + \underline{h}) - Z(\underline{x}_0)] = m(\underline{h}) \quad (1.23)$$

$$Var[Z(\underline{x}_0 + \underline{h}) - Z(\underline{x}_0)] = 2\gamma(\underline{h}) \quad (1.24)$$

Les équations (1.23) et (1.24) sont moins restrictives que l'équation (1.22) et suffisent à définir le variogramme :

$$\gamma_Z(\underline{h}) = \frac{1}{2} \cdot E \left[(Z(\underline{x}_0 + \underline{h}) - Z(\underline{x}_0))^2 \right] \quad (1.25)$$

Dans le cadre de l'hypothèse de stationnarité du deuxième ordre, le variogramme existe, ne dépend pas du point d'appui et est défini par la relation suivante :

$$\gamma_Z(\underline{h}) = C_Z(0) - C_Z(\underline{h}) \quad (1.26)$$

1.1.4 Définition des longueurs de corrélation

Plaçons-nous dans le cadre de l'hypothèse de stationnarité du deuxième ordre. La covariance $C_Z(\underline{h})$ de la fonction aléatoire $Z(\underline{x})$, définie par l'équation (1.22), ne dépend pas du point d'appui, mais seulement de la distance $\underline{h} = (h_x, h_y, h_z)$ entre les points. Dans la plupart des cas rencontrés, la covariance tend vers zéro quand la distance entre les deux points tend vers l'infini, ce qui traduit l'absence de corrélation statistique entre deux points très éloignés. Pour une décroissance suffisamment rapide, les intégrales suivantes, appelées échelles intégrales ou portées de la fonction aléatoire $Z(\underline{x})$ [DAGAN, 1989], sont définies :

$$I_{Z,x} = \frac{1}{\sigma_Z^2} \cdot \int_0^{+\infty} C_Z(h_x, 0, 0) \cdot dh_x \quad (1.27)$$

$$I_{Z,y} = \frac{1}{\sigma_Z^2} \cdot \int_0^{+\infty} C_Z(0, h_y, 0) \cdot dh_y \quad (1.28)$$

$$I_{Z,z} = \frac{1}{\sigma_Z^2} \cdot \int_0^{+\infty} C_Z(0, 0, h_z) \cdot dh_z \quad (1.29)$$

$$I_{Z,xy} = \left[\frac{4}{\pi} \int_0^{+\infty} \int_0^{+\infty} \frac{1}{\sigma_Z^2} \cdot C_Z(h_x, h_y, 0) \cdot dh_x \cdot dh_y \right]^{1/2} \quad (1.30)$$

$$I_Z = \left[\frac{6}{\pi} \int_0^{+\infty} \int_0^{+\infty} \int_0^{+\infty} \frac{1}{\sigma_Z^2} \cdot C_Z(h_x, h_y, h_z) \cdot dh_x \cdot dh_y \cdot dh_z \right]^{1/3} \quad (1.31)$$

Intuitivement, $I_{Z,x}$ est une mesure de la distance entre deux points le long de l'axe x , au-delà de laquelle les valeurs de Z en ces points ne sont plus corrélées. $I_{Z,xy}$ représente le rayon du cercle à l'intérieur duquel les points sont corrélés dans le plan xy . I_Z est une mesure du rayon de la sphère à l'intérieur de laquelle les points sont corrélés. $I_{Z,x}$, $I_{Z,y}$ et $I_{Z,z}$ sont les échelles intégrales les plus souvent utilisées. Elles définissent les longueurs de corrélation le long des axes. Dans le cas isotrope, elles sont égales.

Dans l'exemple de covariance isotrope définie par une fonction exponentielle décroissante $C_z(h) = \sigma_z^2 \cdot \exp\left(-\frac{h}{\lambda}\right)$, illustré à la figure 1.2, λ correspond à la longueur de corrélation :

$$I_{Z,x} = I_{Z,y} = I_{Z,z} = \lambda.$$

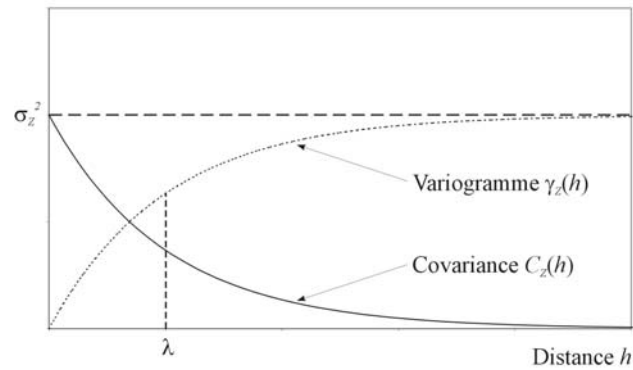


Figure 1.2 : Covariance et variogramme exponentiels

1.1.5 Estimation pratique de la loi spatiale

Replaçons-nous dans le contexte réel : nous ne disposons que des valeurs expérimentales de la variable z mesurées en quelques points : $(z^{\text{exp}}(\underline{x}_i))_{i=1,p}$. La loi spatiale devra donc être inférée par l'étude géostatistique de ces données expérimentales, en admettant les hypothèses d'ergodicité et de stationnarité du deuxième ordre ou intrinsèque. Il n'y a pas d'autre moyen d'appréhender la loi spatiale. Par ailleurs, à ce stade, il n'est pas certain que la loi estimée soit "la bonne" ni même qu'il existe effectivement une "bonne loi". Pour garder notre enthousiasme à poursuivre l'étude géostatistique, nous reprendrons la définition optimiste de MATHERON [1970] : *la "bonne loi" est celle qui nous donne de bons résultats !*

Dans la pratique, le nombre de données étant limité, il n'est en général pas possible de déterminer la loi spatiale au-delà de son moment d'ordre 2, c'est-à-dire la loi statistique de comportement des couples – ou paires – $(Z(\underline{x}), Z(\underline{x} + \underline{h}))$. Faute de mieux, on se contente donc de la reproduction de la loi spatiale bivariable.

1.2 Estimation par krigeage

Le krigeage [MATHERON, 1970 ; DE MARSILY, 1986] est une procédure d'interpolation qui fournit, à partir des données disponibles, une estimation optimale, linéaire et non-biaisée de la propriété étudiée, dont l'erreur d'estimation est minimisée. Par définition, l'estimation par krigeage est unique, contrairement à l'infinité des réalisations possibles pouvant être obtenues par simulations stochastiques (voir paragraphe 1.3). Cependant, elle n'est pas l'image de la réalité : elle est beaucoup plus "lisse", c'est-à-dire qu'elle ne reproduit pas la variabilité réelle de la propriété.

Les avantages du krigeage par rapport à d'autres méthodes d'interpolations plus classiques sont les suivants :

- le krigeage intègre, dans son processus d'estimation, la structure spatiale des données sous forme d'un modèle de variogramme ;
- c'est un interpolateur exact puisque la surface obtenue passe par les points expérimentaux ;
- il fournit une information sur l'erreur d'estimation, permettant donc d'évaluer la précision de l'estimation en chaque point du domaine étudié (carte des incertitudes).

1.2.1 Principe du krigeage

Le principe du krigeage est d'estimer la valeur de la variable aléatoire $z(\underline{x})$, en tout point \underline{x}_0 d'une grille de calcul, par interpolation linéaire à partir des valeurs $(z^{\text{exp}}(\underline{x}_i))_{i=1,p}$, également notées $(z_i)_{i=1,p}$, connues aux points expérimentaux $(\underline{x}_i)_{i=1,p}$.

Notons z_0^* l'estimateur linéaire de $z(\underline{x}_0)$, défini comme suit :

$$z_0^* = z^*(\underline{x}_0) = \sum_{i=1}^p \lambda_0^i z_i \quad (1.32)$$

Il s'agit de trouver, en tout point \underline{x}_0 , la valeur des coefficients inconnus $(\lambda_0^i)_{i=1,p}$, appelés coefficients de pondération du krigeage. Deux conditions portant sur l'erreur d'estimation $\varepsilon(\underline{x}_0) = z_0^* - z_0$, sont imposées à cet estimateur z_0^* .

1. La condition de non-biais

$$E[z_0^* - z_0] = 0 \quad (1.33)$$

2. La condition de minimum de la variance de l'erreur d'estimation

$$Var [z_0^* - z_0] = E \left[(z_0^* - z_0)^2 \right] \text{ minimum} \quad (1.34)$$

Le développement de (1.33) conduit à la condition suivante, appelée condition d'universalité :

$$\sum_{i=1}^p \lambda_0^i = 1 \quad (1.35)$$

Le développement des expressions (1.34) et (1.35) conduit, pour tout point \underline{x}_0 , au système d'équations linéaires suivant :

$$\begin{cases} \sum_{j=1}^p \lambda_0^j \gamma_{ij} + \mu_0 = \gamma_{i0} & , \quad \forall i = 1, p \\ \sum_{i=1}^p \lambda_0^i = 1 \end{cases} \quad (1.36)$$

où les coefficients $(\lambda_0^i)_{i=1,p}$ et μ_0 (multiplicateur de Lagrange) sont les inconnues ;

γ_{ij} est la valeur du variogramme calculé pour la distance entre les deux points expérimentaux \underline{x}_i et \underline{x}_j : $\gamma_{ij} = \gamma_Z(\underline{x}_i, \underline{x}_j)$;

γ_{i0} est la valeur du variogramme calculé pour la distance entre le point estimé \underline{x}_0 et le point expérimental \underline{x}_i : $\gamma_{i0} = \gamma_Z(\underline{x}_i, \underline{x}_0)$.

Le système linéaire de krigeage (1.36), pour un modèle de variogramme nul à l'origine (c'est-à-dire sans effet de pépite^(*)), s'écrit sous la forme matricielle suivante :

$$\begin{pmatrix} 0 & \gamma_{12} & \gamma_{13} & \dots & \gamma_{1n} & 1 \\ \gamma_{21} & 0 & \gamma_{23} & \dots & \gamma_{2n} & 1 \\ \vdots & \vdots & & & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \dots & 0 & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_0^1 \\ \lambda_0^2 \\ \vdots \\ \lambda_0^n \\ \mu_0 \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix} \quad (1.37)$$

(*) Il arrive en pratique que les variogrammes montrent une ordonnée non nulle à l'origine, que l'on appelle effet de pépite. Cet effet de pépite peut provenir de fortes variations locales des valeurs de la propriété ou être attribué soit à des erreurs de mesure, soit au fait que les données ne sont pas récoltées à un intervalle de mesure suffisamment petit pour pouvoir observer la structure sous-jacente continue du phénomène.

Pour résoudre ce système, il suffit d'inverser la matrice une seule fois pour tout point \underline{x}_0 puisqu'elle dépend uniquement de la distance entre les points expérimentaux $(z^{\text{exp}}(\underline{x}_i))_{i=1,p}$ et du modèle de variogramme choisi.

La variance de l'erreur d'estimation au point \underline{x}_0 s'écrit :

$$\text{Var}[z_0^* - z_0] = \sigma_0^2 = \sum_{i=1}^p \lambda_0^i \gamma_{i0} + \mu_0 \quad (1.38)$$

Les équations (1.36) et (1.38) sont les équations habituelles du krigeage, que l'on utilise dans le cadre de l'hypothèse intrinsèque, ou même de stationnarité du deuxième ordre.

L'estimation par krigeage en un point est d'autant meilleure que ce point est proche des points de mesure. La zone d'influence des données est conditionnée par le modèle de variogramme : elle dépend de la longueur de corrélation. Tracer la carte des écart-types de krigeage $\sqrt{\sigma^2}$ – expression (1.38) – permet de visualiser la zone d'influence des données, celle où cet écart-type est le plus faible.

Remarque : le krigeage par "blocs" permet d'estimer des valeurs moyennisées sur des blocs au lieu de valeurs ponctuelles [DE MARSILY, 1986].

1.2.2 Cokrigeage

Le cokrigeage est une technique d'estimation simultanée de deux ou plusieurs variables régionalisées, corrélées entre elles. C'est une extension du krigeage lorsque plusieurs variables régionalisées sont disponibles. Ces variables peuvent être ou ne pas être connues aux mêmes points de mesure, appelées alors respectivement variables isotopiques ou hétérotopiques. Le cokrigeage est un moyen d'améliorer l'estimation de la variable étudiée par l'apport de variables auxiliaires mieux échantillonnées.

Seules deux variables distinctes seront considérées dans la suite des développements (condition de notre cas d'étude), mais plus de deux variables peuvent être cokrigées. Dans ce cas, les équations sont reprises par WAKERNAGEL [1995].

Soit deux jeux de variables, $(z_1^{\text{exp}}(\underline{x}_i))_{i=1,p}$ et $(z_2^{\text{exp}}(\underline{x}_i))_{i=1,m}$, de tailles différentes ($p \neq m$) et dont les mesures ne sont pas nécessairement situées aux mêmes points (cas hétérotopique).

L'estimateur de la variable d'intérêt $z_1(\underline{x}_0)$ s'écrit par cokrigage,

$$z_1^*(\underline{x}_0) = \sum_{i=1}^p \lambda_1^i z_1(\underline{x}_i) + \sum_{l=1}^m \lambda_2^l z_2(\underline{x}_l) \quad (1.39)$$

Pour une meilleure compréhension, les notations ont été légèrement modifiées par rapport à celles utilisées dans les équations du krigeage (point 1.2.1) : λ_j^i correspond au poids de krigeage pour la variable z_j mesurée au point \underline{x}_i . L'indice 0 utilisé précédemment (λ_0^i), indiquant que l'estimation est réalisée au point \underline{x}_0 , est omis. Il est cependant évident que les poids varieront en fonction du point \underline{x}_0 pour lequel l'estimation est réalisée.

Comme pour le krigeage, la condition de non-biais doit être vérifiée :

$$E[z_1(\underline{x}_0)] = m_1 = E[z_1^*(\underline{x}_0)] = \sum_{i=1}^p \lambda_1^i m_1 + \sum_{l=1}^m \lambda_2^l m_2 \quad (1.40)$$

ce qui implique les conditions suivantes :

$$\begin{cases} \sum_{i=1}^p \lambda_1^i = 1 \\ \sum_{l=1}^m \lambda_2^l = 0 \end{cases} \quad (1.41)$$

Le cokrigage s'obtient en minimisant la variance de l'erreur d'estimation $E[(z^*(\underline{x}_0) - z(\underline{x}_0))^2]$, sous réserve que celle-ci est bien une combinaison linéaire autorisée, respectant la condition (1.41). Par la méthode des multiplicateurs de Lagrange, le système d'équations suivant est obtenu [DE MARSILY, 1986] :

$$\begin{cases} \sum_{i=1}^p \lambda_1^i \gamma_{ij}^1 + \sum_{l=1}^m \lambda_2^l \gamma_{lj}^{12} + \mu_1 = \gamma_{0j}^1, & \forall j = 1, p \\ \sum_{i=1}^p \lambda_1^i \gamma_{ij}^{12} + \sum_{l=1}^m \lambda_2^l \gamma_{lj}^2 + \mu_2 = \gamma_{0j}^{12}, & \forall j = 1, m \\ \sum_{i=1}^p \lambda_1^i = 1 \\ \sum_{l=1}^m \lambda_2^l = 0 \end{cases} \quad (1.42)$$

avec γ^1 et γ^2 , respectivement les variogrammes simples des variables z_1 et z_2 ;
 γ^{12} , le variogramme croisé de z_1 et z_2 , défini comme suit :

$$\gamma^{12}(\underline{h}) = \frac{1}{2} E \{ [z_1(\underline{x}_0 + \underline{h}) - z_1(\underline{x}_0)] [z_2(\underline{x}_0 + \underline{h}) - z_2(\underline{x}_0)] \} \quad (1.43)$$

Dans l'hypothèse de stationnarité du deuxième ordre, c'est la covariance croisée que l'on utilise et qui est définie par :

$$C^{12}(\underline{h}) = E \{ [z_1(\underline{x}_0) - m_1] [z_2(\underline{x}_0 + \underline{h}) - m_2] \} \quad (1.44)$$

On peut démontrer que [JOURNEL & HUIJBREGTS, 1978],

$$C^{12}(\underline{h}) = C^{21}(-\underline{h}) \neq C^{21}(\underline{h}) \quad (1.45)$$

tandis que

$$\gamma^{12}(\underline{h}) = \gamma^{21}(\underline{h}) = \gamma^{21}(-\underline{h}) = \gamma^{12}(-\underline{h}) \quad (1.46)$$

1.3 Simulations stochastiques

L'hypothèse primordiale est de considérer la variable régionalisée $z(\underline{x})$ comme UNE réalisation particulière d'une fonction aléatoire $Z(\underline{x})$ continue dans l'espace et obéissant à une loi spatiale. Cette hypothèse implique que l'on se place volontairement dans une optique probabiliste où, d'une manière générale, le résultat quantitatif n'a de valeur que s'il est calculé par prise de moyenne sur un échantillon suffisamment grand.

1.3.1 Propriétés des réalisations de la variable régionalisée

De par sa définition de fonction aléatoire, $Z(\underline{x})$ contient une infinité de réalisations $(z_k(\underline{x}))_{k \in \mathbb{N}}$ admettant toutes les propriétés fondamentales suivantes :

- Une réalisation $z_k(\underline{x})$ de la fonction aléatoire $Z(\underline{x})$ est une variable régionalisée admettant la même loi spatiale que $Z(\underline{x})$. Les réalisations sont donc statistiquement équivalentes, à la fois entre elles et à la fonction aléatoire $Z(\underline{x})$: même espérance, même covariance (et donc même variance) et même histogramme.
- Les réalisations sont statistiquement indépendantes entre elles, mais également indépendantes de la fonction aléatoire $Z(\underline{x})$. S'agissant de fonctions définies dans l'espace, l'indépendance statistique est équivalente à l'indépendance spatiale, ce qui implique la nullité des covariances croisées :

$$E[z_k(\underline{x}) \cdot z_j(\underline{y})] = 0 \quad \forall (\underline{x}, \underline{y}), \quad \forall k \neq j, \quad k \text{ et } j \in \mathbb{N} \quad (1.47)$$

$$E[z_k(\underline{x}) \cdot Z(\underline{y})] = 0 \quad \forall (\underline{x}, \underline{y}), \quad \forall k \in \mathbb{N} \quad (1.48)$$

En langage statistique, chaque réalisation $z_k(\underline{x})$ est assimilée à un tirage de $Z(\underline{x})$ relatif à un nombre aléatoire ξ_k . On fait l'hypothèse que la variable régionalisée $z(\underline{x})$ est une réalisation particulière de la fonction aléatoire $Z(\underline{x})$ relative à un certain nombre aléatoire ξ (figure 1.3).

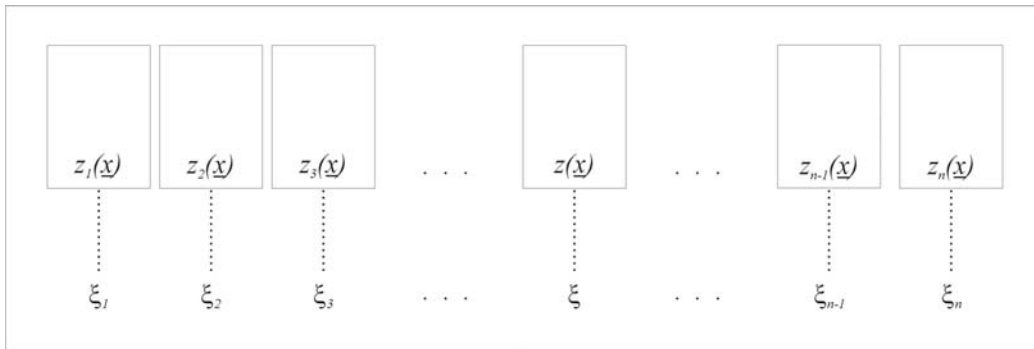


Figure 1.3 : Les réalisations sont autant de tirages aléatoires de la fonction aléatoire $Z(\underline{x})$

1.3.2 Principe des simulations stochastiques

On va chercher à construire des simulations dites **stochastiques** de la variable régionalisée. Ces simulations sont considérées comme autant de réalisations $(z_k(\underline{x}))_{k \in \mathbb{N}}$ ayant les propriétés des réalisations au sens statistique du terme, c'est-à-dire statistiquement équivalentes (de même loi spatiale) et statistiquement indépendantes.

En construisant ces simulations stochastiques, l'objectif n'est pas de "tirer le bon numéro", c'est-à-dire de simuler exactement LA réalisation qui correspond à la variable régionalisée réelle, mais de générer une famille de réalisations ayant les mêmes propriétés statistiques afin de pouvoir donner des résultats probabilistes par prise de moyenne sur un grand nombre. C'est la loi forte des grands nombres, qui mathématiquement peut se traduire par la relation suivante :

$$E[g(Z(\underline{x}))] = \lim_{n \rightarrow \infty} \sum_{k=1}^n g(z_k(\underline{x})) \quad (1.49)$$

avec g , une fonction analytique quelconque, pourvu que l'espérance existe.

Le processus de génération de simulations stochastiques permet, non pas d'obtenir la meilleure estimation (comme dans le processus de krigeage), mais de reproduire la variabilité spatiale de la propriété étudiée. Cependant, aucune réalisation ne peut prétendre être plus proche de la réalité qu'une autre : elles sont toutes **équiprobables**. Les simulations stochastiques fournissent en quelque sorte des exemples de ce que pourrait être la réalité (inconnue) à l'intérieur de la fourchette d'incertitude donnée par l'estimation.

1.3.2.1 Simulations stochastiques non-conditionnelles

Les simulations stochastiques non-conditionnelles, notées $(z_k^{SNC}(\underline{x}))_{k \in \mathbb{N}}$, en tant que réalisations de la même fonction aléatoire Z , en ont les propriétés (énoncées précédemment au point 1.3.1). La similitude entre toutes ces simulations stochastiques est uniquement statistique. La ressemblance avec le milieu réel – qui n'est qu'une réalisation parmi d'autres – n'est que statistique : aux points de mesure $(z^{\text{exp}}(\underline{x}_i))_{i=1,p}$, les valeurs d'une réalisation peuvent différer de celles des données expérimentales. Elles ne sont pas conditionnées aux données expérimentales elles-mêmes mais à leurs statistiques, d'où leur nom : simulations stochastiques non-conditionnelles.

$$z_k^{SNC}(\underline{x}_i) \neq z^{\text{exp}}(\underline{x}_i) \quad , \quad \forall i = 1, p \quad (1.50)$$

1.3.2.2 Simulations stochastiques conditionnelles

Les simulations stochastiques conditionnelles, notées $(z_k^{SC}(\underline{x}))_{k \in \mathbb{N}}$, possèdent les mêmes propriétés que les simulations stochastiques non-conditionnelles, mais diffèrent de celles-ci par le fait qu'elles vérifient les valeurs observées aux points de mesure $(z^{\text{exp}}(\underline{x}_i))_{i=1,p}$.

$$z_k^{SC}(\underline{x}_i) = z^{\text{exp}}(\underline{x}_i) \quad , \quad \forall i = 1, p \quad (1.51)$$

La figure 1.4 illustre la différence entre le krigeage, les simulations stochastiques non-conditionnelles et les simulations stochastiques conditionnelles.

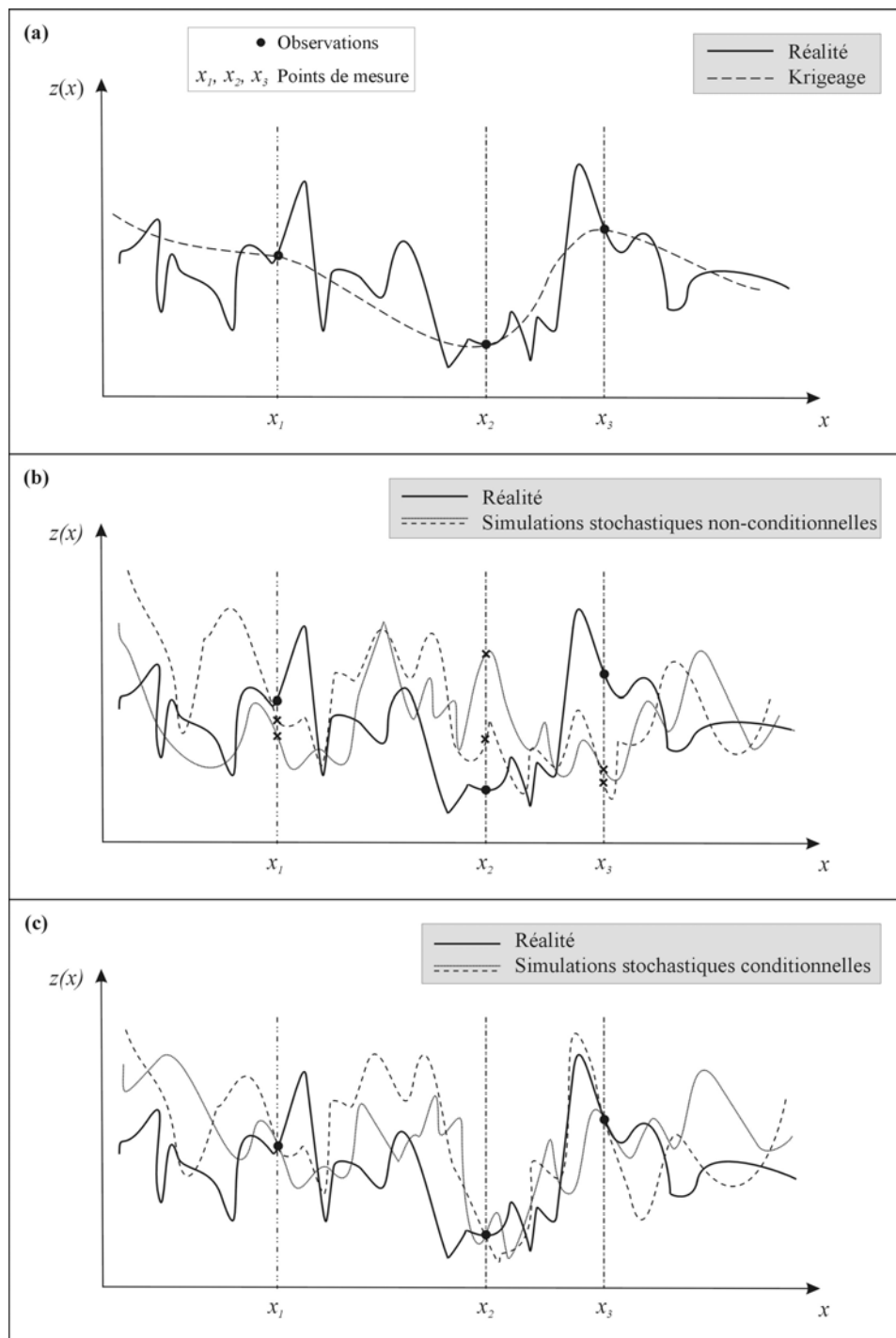


Figure 1.4 : Comparaison 1-D des valeurs réelles de la propriété étudiée et des valeurs obtenues (a) par krigage, (b) par simulations stochastiques non-conditionnelles et (c) par simulations stochastiques conditionnelles

1.3.2.3 Conditionnement d'une simulation stochastique

Simulation stochastique monovariante

De nombreux générateurs de champs aléatoires produisent des simulations stochastiques non-conditionnelles. Pour conditionner ces simulations stochastiques par les valeurs expérimentales de la propriété aux points de mesure, une étape supplémentaire introduisant un double krigeage est nécessaire [DELHOMME, 1979], illustré à la figure 1.5. Pour rappel, le krigeage fournit, sur base des données expérimentales $(z^{\text{exp}}(\underline{x}_i))_{i=1,p}$ aux points de mesure $(\underline{x}_i)_{i=1,p}$ et de sa structure de corrélation spatiale γ , une estimation $z^*(\underline{x})$ de la valeur réelle $z(\underline{x})$ en tout point \underline{x} du domaine. Cependant, puisque $z^*(\underline{x})$ est une estimation, une erreur de krigeage $[z(\underline{x}) - z^*(\underline{x})]$ existe. La relation entre la valeur réelle et l'estimation s'écrit alors :

$$\begin{aligned} \text{valeur réelle} &= \text{estimation par krigeage} + \text{erreur de krigeage} \\ z(\underline{x}) &= z^*(\underline{x}) + [z(\underline{x}) - z^*(\underline{x})] \end{aligned} \quad (1.52)$$

La valeur réelle étant inconnue, l'erreur de krigeage reste inconnue. Cette erreur peut cependant être simulée par un processus aléatoire : une simulation stochastique non-conditionnelle $z_k^{\text{SNC}}(\underline{x})$ est générée à partir de la structure de corrélation spatiale de la propriété. Une estimation par krigeage $z_k^{\text{SNC}*}(\underline{x})$ du champ généré est alors construite sur base des valeurs prises par la simulation stochastique non-conditionnelle aux points de mesure $(\underline{x}_i)_{i=1,p}$ et de la structure de corrélation spatiale γ de la propriété. L'erreur de krigeage, estimée par la différence $[z_k^{\text{SNC}}(\underline{x}) - z_k^{\text{SNC}*}(\underline{x})]$ est alors substituée à $[z(\underline{x}) - z^*(\underline{x})]$. Une simulation stochastique conditionnelle $z_k^{\text{SC}}(\underline{x})$ s'écrit alors

$$z_k^{\text{SC}}(\underline{x}) = z^*(\underline{x}) + [z_k^{\text{SNC}}(\underline{x}) - z_k^{\text{SNC}*}(\underline{x})] \quad (1.53)$$

Le krigeage étant un interpolateur exact, la variance d'estimation aux points de mesure est nulle. Le résidu $[z_k^{\text{SNC}}(\underline{x}) - z_k^{\text{SNC}*}(\underline{x})]$ en tout point de mesure $(\underline{x}_i)_{i=1,p}$, est donc nul et les valeurs mesurées y sont donc bien préservées.

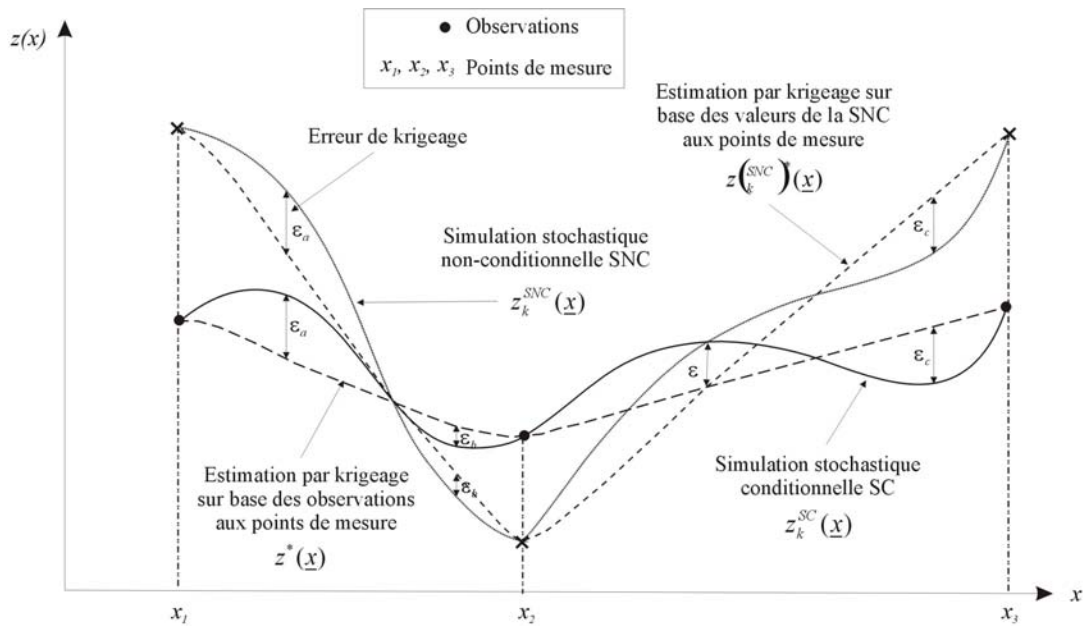


Figure 1.5 : Principe du conditionnement de simulations stochastiques (d'après LAVENUE [1998])

Il est important de noter que la moyenne des valeurs des k simulations stochastiques conditionnelles en un point \underline{x} converge vers l'estimation de krigeage et leur variance converge vers la variance de krigeage [DELHOMME, 1979]. Les fluctuations des simulations stochastiques conditionnelles autour de l'estimation de krigeage illustre l'incertitude existante lorsque l'on s'éloigne des points de mesure. Dès lors, une augmentation du nombre de points de mesure tend à réduire cette incertitude.

Simulation stochastique multivariable

Les simulations stochastiques conditionnelles multivariées, appelées également **simulations stochastiques co-conditionnelles** (ou co-simulations), sont au cokrigeage ce que les simulations stochastiques conditionnelles monovariées sont au krigeage. Les simulations stochastiques co-conditionnelles permettent d'améliorer les réalisations de la variable étudiée par l'apport d'une (ou plusieurs) variable(s) auxiliaire(s) mieux échantillonnée(s).

La génération des simulations stochastiques non-conditionnelles reste identique au cas monovarié. Ce n'est que lors du processus de conditionnement que la (ou les) variable(s) auxiliaire(s) intervient. Le conditionnement est réalisé, non plus sur base du krigeage, mais du cokrigeage et est décrit par l'équation (1.53), en considérant $z^*(\underline{x})$ et $z_k^{(SNC)*}(\underline{x})$ comme des estimations de $Z(\underline{x})$ par cokrigeage et non plus par krigeage.

1.3.3 Méthodes de construction des simulations stochastiques

De nombreuses méthodes géostatistiques permettent de générer des fonctions aléatoires $Z(\underline{x})$. On peut se référer à DEUTSCH & JOURNAL [1992], SRIVASTAVA [1994], KOLTERMANN & GORELICK [1996], ANDERSON [1997], GOOVAERTS [1997] et DE MARSILY & *al.* [1998] pour des synthèses bibliographiques récentes. Parmi ces techniques, on distingue les méthodes gaussiennes et les méthodes non-gaussiennes.

Les approches gaussiennes produisent des images d'une variable continue de loi de distribution gaussienne, possédant toutes la même espérance, la même variance et la même fonction de variogramme (ou de covariance). Les algorithmes les plus utilisés sont ceux de la méthode des bandes tournantes [MATHERON, 1973 ; MANTOGLU & WILSON, 1982], des méthodes spectrales [GUTJAHR, 1989], de la méthode de décomposition LU de Cholesky [CLIFTON & NEUMAN, 1982] et de la méthode des simulations gaussiennes séquentielles [GÓMEZ-HERNÁNDEZ & JOURNAL, 1993]. Les approches non-gaussiennes produisent, elles, des images dont la variable ne suit pas une loi de distribution gaussienne. Ces approches regroupent entre autre les méthodes de simulation des indicatrices (SIS) [JOURNAL & ALABERT, 1989, 1990 ; GÓMEZ-HERNÁNDEZ & SRIVASTAVA, 1990 ; JOURNAL & GÓMEZ-HERNÁNDEZ, 1993], l'algorithme du recuit simulé [DEUTSCH & JOURNAL, 1992], l'approche booléenne [HALDORSEN & CHANG, 1986] et la méthode des chaînes de Markov [DOVETON, 1994].

Seules les méthodes gaussiennes seront décrites. En effet, de nombreux auteurs tels que FREEZE [1975] et HOEKSEMA & KITANIDIS [1985] ont montré que le logarithme népérien de la conductivité hydraulique peut le plus souvent être considéré comme normalement distribué. Nous verrons, au chapitre 4, que les analyses statistiques des données de conductivité hydraulique utilisées dans ce travail vont dans le même sens. La génération des champs de conductivité hydraulique sera réalisée par une approche gaussienne et en particulier, par la méthode des bandes tournantes.

Il faut néanmoins attirer l'attention sur le fait qu'il existe d'autres approches permettant de modéliser la variabilité spatiale, comme l'approche de VANMARCKE & *al.* [1986] (que nous ne développerons pas ici) qui ne se base pas sur la théorie des fonctions spatiales aléatoires mais repose sur le concept de champs aléatoires décrivant les fluctuations des paramètres dans l'espace par leur moyenne, leur variance et leur auto-corrélation [BOLLE, 2000].

Méthode des bandes tournantes [MATHERON, 1973 ; DELHOMME, 1979 ; MANTOGLOU & WILSON, 1982 ; TOMPSON & *al.*, 1989 ; RIVOIRARD, 1995]

Cette méthode génère des simulations stochastiques non-conditionnelles, bi- ou tri-dimensionnelles, de la fonction aléatoire $Z(\underline{x})$ à partir d'un grand nombre de simulations unidimensionnelles, appelées "bandes". Toute simulation de covariance isotrope $C(h)$ dans \mathfrak{R}^n ($n = 2$ ou 3) est donc ramenée à une simulation de fonctions aléatoires indépendantes, de covariance unidimensionnelle $C_1(\zeta)$ équivalente, définies dans \mathfrak{R}^1 (voir MANTOGLOU & WILSON [1982] pour les équations de transformation de C_n en C_1). La figure 1.6 schématise le principe de la méthode des bandes tournantes.

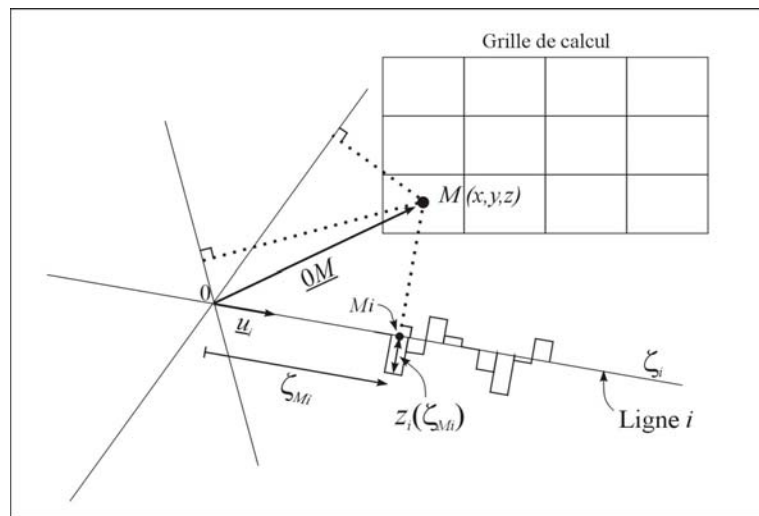


Figure 1.6 : Représentation schématique de la méthode des bandes tournantes (inspiré de MANTOGLOU & WILSON [1982])

Une origine 0 est arbitrairement choisie dans \mathfrak{R}^n . Un nombre L de lignes, passant par cet origine 0 et de direction unitaire $(\underline{u}_i)_{i=1,L}$, sont distribuées de manière uniforme autour de l'origine 0 (suivant un cercle en 2D ou une sphère en 3D). Le long de chaque ligne i , une simulation unidimensionnelle indépendante de covariance $C_1(\zeta_i)$ est générée, ζ_i étant la coordonnée sur la ligne i . Soit un point M de la grille de calcul pour lequel une valeur de la propriété étudiée doit être générée. Le vecteur \underline{OM} est projeté orthogonalement sur chaque ligne i . Au point de projection $\zeta_{Mi} = \underline{OM} \cdot \underline{u}_i$ correspond une valeur $z_i(\zeta_{Mi})$ de la simulation unidimensionnelle sur la ligne i . La valeur simulée $z_k(\underline{OM})$ attribuée au point M est donnée par,

$$z_k(\underline{OM}) = \frac{1}{\sqrt{L}} \sum_{i=1}^L z_i(\underline{OM} \cdot \underline{u}_i) \quad (1.54)$$

Il est bon de noter que la méthode des bandes tournantes peut se généraliser au cas anisotrope.

Méthodes spectrales [GUTJAHR, 1989 ; ROBIN & al., 1993 ; GUTJAHR & al., 1994]

Contrairement au domaine spatial pour lequel la corrélation spatiale est caractérisée par une fonction de covariance ou de variogramme, le domaine fréquentiel reproduit la corrélation spatiale par une fonction de densité spectrale. Ces algorithmes ajustent donc la périodicité de la transformée de Fourier à la périodicité des données observées à partir des fonctions de covariance et de variogramme.

Contrairement à la méthode des bandes tournantes et aux méthodes spectrales qui engendrent des simulations stochastiques non-conditionnelles et nécessitent une étape de double krigeage supplémentaire pour les conditionner, les algorithmes de décomposition LU et de simulation gaussienne séquentielle génèrent directement, par construction, des simulations stochastiques conditionnelles.

Méthode de décomposition LU [CLIFTON & NEUMAN, 1982; ALABERT, 1987; DAVIS, 1987]

Un champ aléatoire est obtenu en multipliant un vecteur de nombres aléatoires w par une matrice L , calculée à partir des informations connues sur la continuité spatiale. Cette matrice L provient de la décomposition de la matrice de covariance (ou de variogramme) de la fonction aléatoire $Z(x)$ de loi de distribution gaussienne, en un produit de matrices triangulaires inférieure L (*lower*) et supérieure U (*upper*). Des réalisations multiples $(z_k(x))_{k=1,n}$ sont obtenues simplement en générant n vecteurs de nombres aléatoires $(w^{(k)})_{k=1,n}$ différents. Cette méthode est simple mais très coûteuse en temps de calcul et en espace mémoire. Elle est limitée à la génération de champs de taille relativement réduite puisque l'algorithme considère tous les nœuds et les données simultanément dans une seule matrice de covariance. L'avantage de cette méthode est que la décomposition de la matrice de covariance ne se fait qu'une seule fois. C'est donc la méthode la plus rapide lorsque le nombre total de données conditionnelles et de nœuds devant être simulés est petit – moins de quelques centaines.

Méthode de simulation Gaussienne séquentielle (sGs) [DEUTSCH & JOURNAL, 1992 ;

GÓMEZ-HERNÁNDEZ & JOURNAL, 1993 ; GÓMEZ-HERNÁNDEZ & CASSIRAGA, 1994]

Dans cet algorithme, un nombre aléatoire est attribué à chaque cellule pour laquelle aucune donnée expérimentale n'existe, définissant ainsi un "ordre" aléatoire parmi toutes les cellules de la grille. Pour chaque cellule, la fonction de densité de probabilité (fdp) est estimée sur base d'un nombre spécifié de données conditionnantes voisines (comprenant les données initiales ainsi que les valeurs simulées précédemment durant la procédure de génération et se trouvant dans le voisinage du point généré). Une valeur est tirée aléatoirement de cette fdp sur base de la description de la continuité spatiale. Le sGs est flexible en ce sens que de nombreuses techniques peuvent être utilisées pour estimer la fdp. Le sGs utilise le krigeage

pour estimer la moyenne et l'écart-type de la fdp en supposant que la distribution est gaussienne [SRIVASTAVA, 1994]. Si des réalisations multiples $(z_k(\underline{x}))_{k=1,n}$ sont désirées, l'algorithme précédent est répété n fois en considérant un ordre aléatoire différent pour chaque réalisation. La configuration des données varie donc d'une réalisation à l'autre et différents systèmes de krigeage doivent être résolus.

1.3.4 Seuillage des simulations stochastiques

Toutes les méthodes gaussiennes citées ci-dessus génèrent directement des champs de conductivité hydraulique sans passer par la création d'une image de la géologie du sous-sol. Ces méthodes fournissent des variations de la valeur du paramètre étudié autour d'une moyenne supposée constante. Elles reproduisent donc la variabilité (continue) de la propriété étudiée.

La méthode des gaussiennes seuillées [MATHERON & *al.*, 1987 ; GALLI & *al.*, 1990 ; DE MARSILY & *al.*, 1998] tente de représenter au mieux la distribution des "faciès" dans l'espace (et non celle d'une variabilité continue de la propriété étudiée). Ces faciès correspondent à des types de roches aux propriétés bien définies (par exemple : un sable, une argile sableuse, une argile) pouvant appartenir à une même unité géologique (à un même aquifère) et dont on recherche la répartition géométrique dans l'espace.

Dans cette méthode, l'intervalle de variation de la propriété étudiée est découpé en une série de classes – correspondant à des faciès géologiques – sur base de seuils choisis. La probabilité cumulée de la loi de Gauss pour chaque classe doit correspondre au pourcentage du faciès dans la formation. De cette manière, tout tirage dans le domaine simulé d'une fonction aléatoire $Z(\underline{x})$ à distribution gaussienne donne un pourcentage de chaque faciès identique à celui observé sur le terrain. Pour un champ stationnaire, les proportions de chaque type de faciès sont donc contrôlées par le choix des seuils. Sans autre contrainte, les faciès ainsi engendrés sont discontinus et irréguliers, avec les bonnes proportions de chaque type de sédiments, mais sans reproduire la structure spatiale réelle. Pour faire apparaître cette structure spatiale, il suffit d'imposer à la fonction $Z(\underline{x})$ d'avoir une certaine covariance spatiale, engendrant une certaine continuité des valeurs de conductivité hydraulique en des points voisins. En sélectionnant la fonction de covariance appropriée, on crée alors des faciès qui ont une très bonne ressemblance avec les distributions réelles de faciès dans une coupe géologique. Une fois les faciès définis, on peut affecter à chacun de ceux-ci soit une valeur unique de la propriété étudiée (ici en l'occurrence, la conductivité hydraulique), soit une valeur continue de la propriété tirée dans une loi de distribution choisie. La méthode des

gaussiennes seuillées permet aisément de conditionner les réalisations par des mesures de la propriété.

Des techniques de seuillage, dérivées de la méthode des gaussiennes seuillées, seront utilisées dans ce travail. Pour faciliter la compréhension, elles seront décrites ultérieurement, au chapitre 5.