

---

---

# Géostatistique non linéaire

---

*Les méthodes géostatistiques dites «non linéaires» telles que le conditionnement uniforme et les simulations conditionnelles, ont été développées non pas dans le but d'estimer une valeur vraie inconnue mais afin de calculer la probabilité avec laquelle cette valeur inconnue est supérieure ou inférieure à un seuil déterminé.*

*Dans ce chapitre, nous présentons le cadre théorique des ces méthodes et leurs applications dans le contexte de notre étude. L'objectif recherché est d'obtenir une cartographie de cette probabilité. La première section est réservée à l'introduction des méthodes de conditionnement uniforme et la simulation conditionnelle. Dans la seconde section, nous présentons les traitements et applications de ces méthodes sur les mesures de la densité de la puissance pour le service GSM 900. Enfin, nous terminons par des discussions et conclusions.*

## 1) Anamorphose gaussienne

### 1.1) Définition

Les méthodes de la géostatistique non linéaire exposées dans ce chapitre, supposent que la distribution de fréquences de la variable étudiée peut être modélisée par un modèle gaussien [32]. Il convient de rappeler que la fonction de répartition d'une variable gaussienne centrée réduite  $Z$  (moyenne nulle et variance égale à 1) est donnée par l'équation 6.1 :

$$G(x) = P[Z \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz \quad (\text{Équation 6.1})$$

Cette fonction n'a pas d'expression analytique simple. En pratique on se sert soit de tables, soit de formules d'approximation, calculées numériquement [36]. Dans le contexte de notre étude, des mesures in situ de champ électrique ont été réalisées dans un contexte urbain, le plus souvent l'histogramme de ces mesures est asymétrique avec quelques valeurs élevées et il est très rare de rencontrer des distributions parfaitement

gaussiennes [58]. Ces valeurs représentent l'exposition dans certains endroits du domaine où l'exposition est plus forte.

Par conséquent, il est intéressant de procéder à une transformation de la variable réelle en variable gaussienne. Le processus consiste à déformer l'histogramme de données réelles en un histogramme gaussien réduit ; sur les histogrammes cumulés (fonctions de répartition  $F(z)$  et  $G(y)$ ) elle consiste à associer à chaque valeur de « z », la valeur « y » (gaussienne) correspondant à la même fréquence cumulée.

## 1.2) Polynômes d'Hermite

La fonction reliant « y » et « z » est appelée fonction d'anamorphose ou tout simplement anamorphose  $\varphi$ . En réalité, on a seulement accès à des observations de Z, on peut donc estimer cette fonction par une anamorphose empirique :

$$\hat{\varphi}(y) = \sum_{k=1}^N z_k I_{G^{-1}(\frac{(k-1)}{N}) \leq y \leq G^{-1}(\frac{k}{N})} \quad (\text{Équation 6.2})$$

Avec  $I$  la fonction indicatrice 
$$I(y) = \begin{cases} 1 & G^{-1}(\frac{(k-1)}{N}) \leq y \leq G^{-1}(\frac{k}{N}) \\ 0 & \text{sinon} \end{cases}$$

La fonction  $\varphi$  enregistre toutes les irrégularités que peut présenter la distribution expérimentale d'exposition; pour une variable gaussienne, l'anamorphose est linéaire [36]. Les fonctions utilisées en géostatistique pour interpoler l'anamorphose empirique par une fonction continue sont les «polynômes d'Hermite». Ces polynômes sont définis à partir de la densité de probabilité gaussienne réduite [32]:

$$H_k(y) = \frac{1}{\sqrt{k!g(y)}} \frac{d^k g(y)}{dy^k} \quad k \geq 0 \quad (\text{Équation 6.3})$$

Où  $g$  c'est la distribution d'une variable gaussienne centrée et réduite :

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \quad (\text{Équation 6.4})$$

A partir de cette équation, les trois premiers polynômes sont donnés par les fonctions suivantes :

$$H_0(y) = 1; H_1(y) = -y; H_2(y) = \frac{(y^2 - 1)}{\sqrt{2}} \quad (\text{Équation 6.5})$$

Ensuite ces polynômes son calculables par une relation de récurrence :

$$H_{k+1}(y) = -\frac{1}{\sqrt{k+1}} y H_k(y) - \sqrt{\frac{k}{k+1}} H_{k-1}(y) \text{ où } k > 0 \quad (\text{Équation 6.6})$$

## Chapitre 6: Géostatistique non linéaire

Le principal avantage de ce type de polynômes est que n'importe quelle fonction de carré intégrable peut être développée en polynômes d'Hermite [36], par exemple, pour l'anamorphose gaussienne on obtient :

$$z = \varphi(y) = \varphi_0 + \varphi_1 H_1(y) + \varphi_2 H_2(y) + \dots \quad (\text{Équation 6.7})$$

$$z = \sum_{k=0}^{+\infty} \varphi_k H_k(y)$$

$y$  : valeur gaussienne centrée réduite

$\varphi_k$  :  $k$ ème coefficient

$H_k(y)$  :  $k$ ème polynôme d'Hermite

Le nombre  $n$  de coefficients  $\varphi_n$  calculés indique l'ordre de troncature. Ces coefficients restent constants et ne dépendent pas de l'ordre de troncature. En pratique, on calcule un nombre déterminé de polynômes à partir des valeurs gaussiennes ( $Y$ ) d'exposition, ensuite on calcule les coefficients de chaque polynôme à l'aide de la formule [36] :

$$\varphi_k = \int z H_k(y) g(y) dy \quad (\text{Équation 6.8})$$

$Y$  : valeur gaussienne centrée réduite

$z$  : valeur réelle

$\varphi_k$  :  $k$ ème coefficient

$H_k(Y)$  :  $k$ ème polynôme d'Hermite

$g(y)$  : fonction de densité d'une loi gaussienne centrée réduite

On peut valider ce calcul par les deux expressions ci-après [36]:

$$E(Z) = \varphi_0 \quad \text{et} \quad \text{Var}(Z) = \sum_{k=1}^{+\infty} \varphi_k^2 \quad (\text{Équation 6.9})$$

On vérifie donc que le premier coefficient est égal à la moyenne expérimentale des données et que la somme du carré de tous les coefficients (sauf le premier) est égale à la variance expérimentale de la variable réelle.

## **2) Conditionnement uniforme**

On s'intéressera souvent non seulement à des valeurs définies sur des points, mais également à des valeurs moyennes sur des éléments de surface. Un support ponctuel n'a pas beaucoup de sens car l'exposition peut varier fortement d'un point à un autre, immédiatement voisin. Le problème que nous étudions comporte donc trois supports : le support ponctuel des points de mesure, le support des cellules et un support de bloc.

Le modèle Gaussien discret, va nous servir à aborder un problème à trois supports tel qu'il est représenté schématiquement sur la Figure 6.1, où nous avons des mesures en des points, des valeurs de modèle numérique pour des cellules elles-mêmes partitionnées en des blocs.

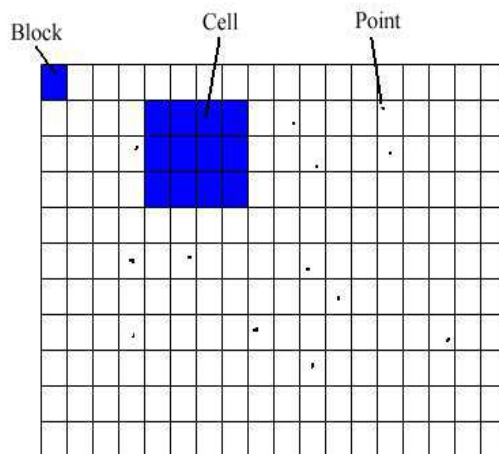


Figure 6.1 : Trois supports : points de mesure à l'intérieur de cellules partitionnées en blocs.

Soit  $Z$  une fonction aléatoire, et  $z(x)$  une réalisation de cette fonction connue dans un bloc  $v$ . La valeur moyenne de  $z$  sur  $v$  :

$$z_v = \frac{1}{|v|} \int_v z(x) dx \text{ est une réalisation de la variable aléatoire } Z_v = \frac{1}{|v|} \int_v Z(x) dx.$$

Comme nous connaissons les observations de  $Z$ , alors il y a équivalence entre l'information portée par  $Z(x)$  et celle portée par  $Y(x)$ . D'où :

$$E[Z(x) | Z_v] = E[\varphi(Y(x)) | Y_v] \quad (\text{Équation 6.10})$$

On considère que  $x$  est un point aléatoire dans le bloc  $v$ . D'après la relation de *Cartier* [6] l'espérance de  $Z(x)$  conditionnellement à la valeur du bloc  $v$  est égale à la moyenne de  $Z$  sur le bloc :

$$\begin{aligned} E[Z(x) | Z_v] &= \frac{1}{|v|} \int_v E[Z(x) | Z_v] dx && (\text{Équation 6.11}) \\ &= E\left[\frac{1}{|v|} \int_v Z(x) | Z_v dx\right] = E[Z_v | Z_v] = Z_v \end{aligned}$$

Avec l'anamorphose de bloc  $\varphi_v(Y(v))$  correspondante, la relation se réécrit

$$E[\varphi(Y(x)) | Z(v)] = E[\varphi(Y(x)) | \varphi_v(Y(v))] = \varphi_v(Y(v)) \quad (\text{Équation 6.12})$$

L'anamorphose Gaussienne des blocs peut être exprimée avec les coefficients  $\varphi_k$  (ajustés aux données) de l'anamorphose ponctuelle en introduisant un coefficient positif  $\rho$  de corrélation point-bloc [39]:

$$\begin{aligned}\varphi_v(Y(v)) &= E[\varphi(Y(x))|Y(v)] \\ &= \sum_{k=0}^{\infty} \frac{\varphi_k}{k!} E(H_k(Y(v))|Y(v)) = \sum_{k=0}^{\infty} \frac{\varphi_k}{k!} \rho^k H_k(Y(v))\end{aligned}\quad (\text{Équation 6.13})$$

La corrélation point-bloc est déterminée à partir de la variance de dispersion des blocs, qui est elle-même calculée avec le modèle de variogramme [20].

$$\text{var}(Z(v)) = \text{var}(\varphi_v(Y(v))) = \sum_{k=0}^{\infty} \frac{\varphi_k^2}{k!} \rho^{2k}\quad (\text{Équation 6.14})$$

De la même manière on peut calculer un coefficient  $\rho'$  de corrélation point-cellule et l'on obtient également un coefficient bloc-cellule en formant le rapport  $\rho_{vV} = \rho'/\rho$ .

### 3) Simulation conditionnelles

Les techniques de krigeage [20], permettent d'estimer localement les valeurs prises par une variable régionalisée à partir d'un échantillonnage de cette dernière.

Cependant, la propriété de lissage induite par le krigeage empêche d'apprécier la variabilité des valeurs inconnues (non mesurées) : ainsi, l'histogramme des valeurs estimées est moins dispersé que celui des valeurs réelles ; le variogramme des valeurs estimées n'est pas non plus représentatif de la réalité.

Le but des simulations est de reproduire la variabilité spatiale de la variable régionalisée tout en respectant ses propriétés statistiques (histogramme, variogramme, etc.), chaque simulation est alors considérée comme une réalisation possible de la réalité [41]. De même une simulation conditionnelle est une simulation restituant aux points de mesures les valeurs qui y sont connues.

La construction de simulations est particulièrement simple dans le cadre des fonctions aléatoires de loi spatiale multigaussienne, c'est-à-dire telles que toute combinaison linéaire de valeurs suive une distribution gaussienne.

#### 3.1) Post-conditionnement

De nombreux algorithmes existent pour obtenir des simulations non conditionnelles des fonctions aléatoires de loi spatiale multigaussienne: bandes tournantes, décomposition

matricielle, moyennes mobiles discrètes, méthodes spectrales, méthode séquentielle, etc [6].

Le conditionnement à un ensemble de données se résout à l'aide d'un krigeage; en effet, on a en tout point  $x$  :

$$Y(x) = Y^k(x) + [Y(x) - Y^k(x)], \quad Y^k = \text{Krigeage de la gaussienne} \quad (\text{Équation 6.15})$$

Dans le cas d'une fonction aléatoire multigaussienne, le résidu en tout point  $[Y(x) - Y^k(x)]$  est indépendant des valeurs aux points de données [6].

L'idée est alors de substituer à ce résidu indépendant mais inconnu un résidu simulé ayant exactement la même structure. Pour ce faire, on fabrique une simulation non conditionnelle de la variable, soit  $Y_s(x)$ , sur le domaine considéré, puis on calcule en tout point  $x$ , le résidu de son krigeage à partir des valeurs prises par  $Y_s$  aux points de données [6] :

$$[Y(x) - Y^k(x)]$$

La recombinaison:

$$Y_s(x) = Y^k(x) + [Y_s(x) - Y_s^k(x)] \quad (\text{Équation 6.16})$$

Donne alors une autre simulation de la fonction aléatoire, mais qui est maintenant conditionnelle, en un point de donnée on retrouve bien la valeur continue :

$$Y_s(x_i)_{SC} = Y^k(x_i) + [Y_s(x_i) - Y_s^k(x_i)] = Y_s(x_i) \quad (\text{Équation 6.17})$$

- 
1. Effectuer une simulation non conditionnelle de la variable gaussienne, aux points désirés pour la simulation et aux points échantillons.
  2. Effectuer deux krigeages: le premier sur la variable gaussienne, le second sur les valeurs de la simulation non conditionnelle aux points des données.
  3. Former:  $Y_s(x_i)_{SC} = Y^k(x_i) + [Y_s(x_i) - Y_s^k(x_i)]$
- 

### Algorithme de Simulation conditionnelle

#### 3.2) Méthode de bandes tournantes

La technique des bandes tournantes est une des méthodes les plus utilisées et a prouvé son efficacité dans le domaine minier. La simulation non conditionnelle par bandes tournantes consiste à simuler une variable aléatoire  $y$  de covariance  $c(s)$  le long d'un

nombre théoriquement infini de lignes passant par une origine  $O$  et uniformément orientées dans toutes les directions de l'espace. La variable aléatoire  $Y$  définie en tout point  $x$  de l'espace  $IR^2$  comme la somme des projetés  $y$  de  $Y$  sur chacune des lignes a une covariance  $C(h)$  liée à la covariance  $c(s)$  par la relation suivante:

$$C(h) = \frac{1}{h} \int_0^h c(s) ds \quad (\text{Équation 6.18})$$

Le calcul numérique de  $Y$  passe par une approximation-discrète et le nombre de lignes réellement simulées est limité à  $N$  lignes. La variable aléatoire  $Y$  définie sur l'ensemble de l'espace devient la somme des valeurs  $y$ , des projetés de  $Y$  sur chacune des lignes, divisée par  $\sqrt{N}$  :

$$Y(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^N y_i \quad (\text{Équation 6.19})$$

### **4) Conclusion**

Dans ce chapitre, nous avons exposé des méthodes pour calculer les probabilités de dépassement des seuils par deux méthodes de la géostatistique non-linéaire. Le conditionnement uniforme offre une mise en œuvre rapide. Cette méthode permet de plus la prise en compte de variables auxiliaires grâce au krigeage avec la dérive externe (il est possible donc de prendre en compte une composante non stationnaire).

Les simulations conditionnelles par bandes tournantes permettent aussi la prise en compte des variables auxiliaires et la modélisation de non stationnarités, mais leur mise en œuvre est plus lourde et nécessite un plus grand temps de calcul.