Interpolation et Géostatistique (III)
Krigeage

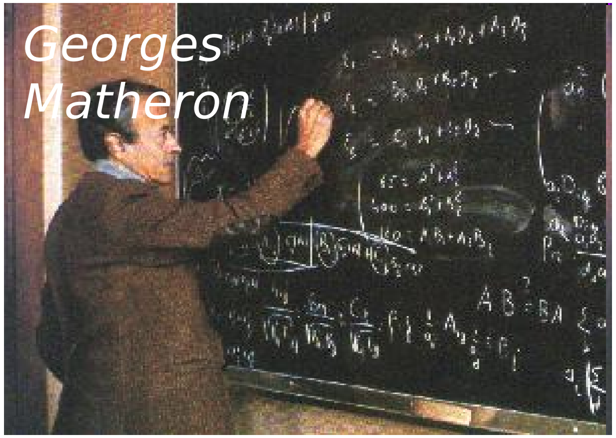
http://dpt-info.u-strasbg.fr/~gancars

Pierre Gançarski
IUT Robert Schuman

5 – Géostatistique

- Le krigeage est nommé d'après Gerhardus Danie Krige, un ingénieur minier sud-africain qui a présenté les idées dans sa thèse de maîtrise en 1951.
  - "A statistical approach to some basic mine valuation problems on the Witwatersrand". J. of the Chem., Metal. and Mining Soc. of South Africa 52 (6): 119–139.

  Ces idées ont ensuite été formalisée par un éminent mathématicien français Georges Matheron en 1962.
  - Traité de Géostatistique Appliquée, Editions Technip, France



- Krigeage a deux aspects :
  - la quantification de la structure spatiale des données (appelée variographie)
  - la prévision des valeurs à des points inconnus

5 – Géostatistique

- Idée de base :
  - A priori, la valeur attendue pour une variable Z ne devrait pas varier avec la latitude et la longitude : la valeur de Z est constante dans toute la région mais cette valeur n'est pas nécessairement connue.
  - Or, on constate en général, une certaine variabilité dans les valeurs pour Z aux différents points : cela peut être considéré cela comme une déviation locale de la structure globale

    → structure locale ou résiduelles ou encore erreur.

➔ Les géostatisticiens décomposent la valeur z_p de Z en un point *p* en un terme lié à la structure globale α et un terme lié à la structure locale ε_p.

$$z_p = \alpha + \varepsilon_p$$

5 – Géostatistique

- Comme pour IDW l'influence d'un point *i* (de valeur z_i mesurée) sur la calcul de la valeur locale estimée ε_p au point P, diminue en fonction de sa distance

$$\varepsilon_p = \sum_{i=1,\dots,N} w_i z_i$$

Le problème consiste à trouver les w_i

$$z'_p = \alpha + \varepsilon_p = \alpha + \sum_{i=1,\dots,N} w_i z_i$$

5 – Géostatistique

- Propriétés de l’estimateur
  - Contrainte de non-biais : l’espérance des valeurs estimées doit être la même que l’espérance des valeurs réelles
  - Contrainte d’optimalité : la variance de l’erreur d’estimation doit être la plus petite possible
    - l’erreur d’estimation est la différence entre la valeur estimée et la valeur réelle.

5 – Géostatistique : Krigeage simple

- Cas stationnaire : Krigeage simple
  - Contrainte de non-biais : l'espérance des valeurs estimées doit être la même que l'espérance des valeurs réelles

$$E\left[z'_p\right] = E\left[z_p\right] \Rightarrow \alpha + \sum_{i=1,\dots,n} w_i E\left[z_i\right] - E\left[z_p\right] = 0$$

p appartenant à l'échantillon, z_p sa valeur réelle, *n* le nbre points dans son voisinage

  - d'où $$\alpha + \left(\sum_{i=1,\dots,n} w_i - 1\right).\mu = 0$$ où µ est la moyenne

d'où

$$\alpha = \mu\left(1 - \sum_{i=1,\dots,n} w_i\right)$$

5 – Géostatistique

- Plusieurs cas
  - Champs stationnaires : la variation de la mesure entre deux points distants de *h* ne dépend que de *h*, et non de la position des points
    - Krigeage simple : on connaît μ (en général, la moyenne du champ à estimer) ou on sait l'estimer → on peut calculer α
    - Krigeage ordinaire : on ne connaît pas μ
  - Champs non-stationnaires
    - Krigeage universel

5 – Géostatistique : Krigeage simple

- La variance de l'erreur doit être minimale

$$Var[z'_p - z_p] = Var[z'_p] + Var[z_p] - 2Cov[z'_p, z_p]$$

or $$Var[z'_p] = Var\left[\sum_{i=1}^{n} w_i z_i\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov[z_i, z_j]$$

et $$Var[z_p] = \sigma^2$$

où σ est la variance de la mesure

5 – Géostatistique : Krigeage simple

- D'où

$$Var\left[z'_p - z_p\right] = Var\left[z'_p\right] + Var\left[z_p\right] - 2Cov\left[z'_p , z_p\right]$$

se ré-écrit en :

$$Var\left[z'_p - z_p\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov\left[z_i, z_j\right] + \sigma^2 - 2Cov\left[z'_p, z_p\right]$$

Reste à calculer la covariance entre z'_p et z_p

5 – Géostatistique : Krigeage simple

- Il faut encore calculer la covariance entre z'_p et z_p

$$Cov\left[z'_p, z_p\right] = Cov\left(\sum_{i=1}^{n} w_i z_i,\ z_p\right)$$

$$= E\left(\sum_{i=1}^{n} w_i z_i \cdot z_p\right) - E\left(\sum_{i=1}^{n} w_i z_i\right) \cdot E(z_p) \quad \textit{par définition}$$

$$= \sum_{i=1}^{n} w_i E(z_i . z_p) - \sum_{i=1}^{n} w_i E(z_i) \cdot E(z_p)$$

$$= \sum_{i=1}^{n} w_i Cov\left[z_i, z_p\right]$$

On se ramène à un calcul de covariance entre les points connus et le point à estimer.

5 – Géostatistique : Krigeage simple

- D'où

$$Var\left[z'_p - z_p\right] = Var\left[z'_p\right] + Var\left[z_p\right] - 2Cov\left[z'_p, z_p\right]$$

se ré-écrit en :

$$Var\left[z'_p - z_p\right] = \sigma^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov\left[z_i, z_j\right] - 2\sum_{i=1}^{n} w_i Cov\left[z_i, z_p\right]$$

5 – Géostatistique : Krigeage simple

La variance doit être minimale : il faut donc trouver $w=\{w_1, \ldots, w_n\}$ tels que la dérivée soit nulle

$$\frac{d\left(Var\left[z'_p - z_p\right]\right)}{dw} = \frac{d\left(\sigma^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov\left[z_i, z_j\right] - 2\sum_{i=1}^{n} w_i Cov\left[z_i, z_p\right]\right)}{dw} = 0$$

c'est-à-dire $\sum_{j=1}^{n} w_j Cov\left[z_i, z_j\right] = Cov\left[z_i, z_p\right]$ et ce pour tout i dans le voisinage de p

➔ On obtient un système linéaire à n équations à résoudre

- Le système d'équation $\begin{cases} w_1C_{1,1} + w_2C_{1,2} + w_3C_{1,3} + \cdots + w_nC_{1,n} & = & C_{1,p} \\ w_1C_{2,1} + w_2C_{2,2} + w_3C_{2,3} + \cdots + w_nC_{2,n} & = & C_{2,p} \\ \cdots & & \cdots \\ w_1C_{n,1} + w_2C_{n,2} + w_3C_{n,3} + \cdots + w_nC_{n,n} & = & C_{n,p} \end{cases}$

peut s'écrire
$$\begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,n} \\ & \ddots & & \\ C_{n,1} & C_{n,1} & \cdots & C_{n,n} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} C_{1,p} \\ \vdots \\ C_{n,p} \end{bmatrix}$$

c'est-à-dire
$$\begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,n} \\ & \ddots & & \\ C_{n,1} & C_{n,1} & \cdots & C_{n,n} \end{bmatrix}^{-1} \begin{bmatrix} C_{1,p} \\ \vdots \\ C_{n,p} \end{bmatrix}$$

5 – Géostatistique : Krigeage ordinaire

- Cas stationnaire : Krigeage ordinaire
  - En général, on ne connaît pas la moyenne : Le seul moyen de garantir le non-biais est que dans la formule

$$E\left[z'_p - z_p\right] = \alpha + \left(\sum_{i=1,\dots,n} w_i - 1\right).\mu = 0$$

    - α soit nul
    - et que

$$\sum_{i=1,\dots,n} w_i = 1$$

  - La propriété sur la variance étant la même que précédemment, nous obtenons donc une équation de plus.

5 – Géostatistique : Krigeage cas stationnaire

- Reste à résoudre le système linéaire

1. Comment calculer les valeurs de covariance : utiliser le variogramme *y*
   - Une estimation de la *Cov(x',x)* peut être donnée par

$$Cov(x',x) = \sigma - y(h)$$

où σ est le palier du variogramme et *h* la distance de x à x'

- Le système peut s'écrire sous forme matricielle (exemple pour 4 points)

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ \infty \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & 1 \\ C_{21} & C_{22} & C_{23} & C_{24} & 1 \\ C_{31} & C_{32} & C_{33} & C_{34} & 1 \\ C_{41} & C_{42} & C_{43} & C_{44} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{p,1} \\ C_{p,2} \\ C_{p,3} \\ C_{p,4} \\ 1 \end{bmatrix}$$

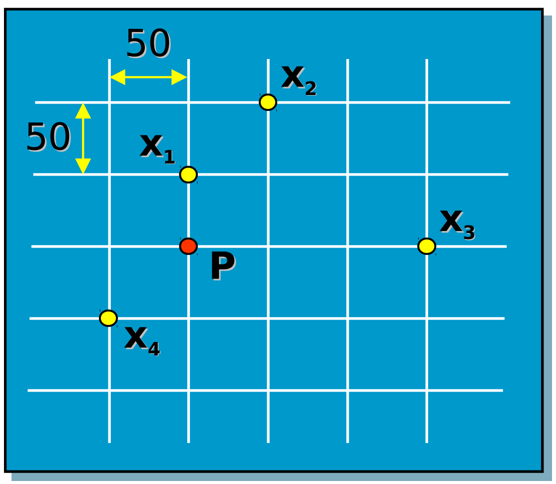
- Il « suffit » donc d'inverser la matrice

5 – Géostatistique : exemple de krigeage

- Un exemple complet

5 – Géostatistique : exemple de krigeage

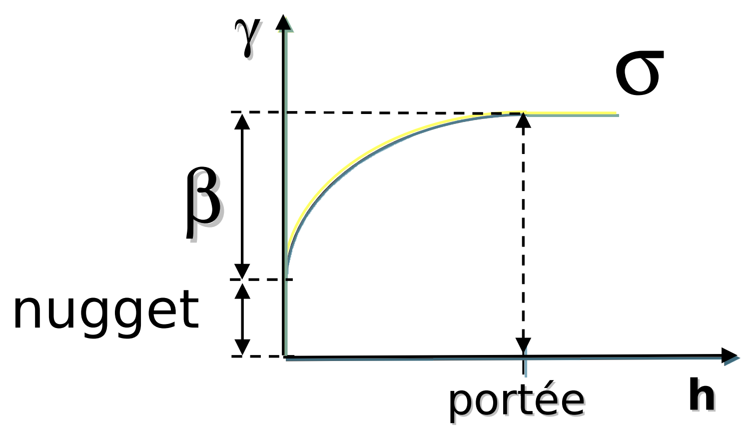
- Un exemple complet
  - Les données



5 – Géostatistique : exemple de krigeage

- Un exemple complet
  - Le variogramme théorique

$$\gamma(\mathbf{h}) = \begin{cases} 0 & pour\ |\mathbf{h}| = 0 \\ nugget + \lambda\left[\frac{3}{2}\left(\frac{|\mathbf{h}|}{a}\right) - \frac{1}{2}\left(\frac{|\mathbf{h}|}{a}\right)^3\right], & pour\ \ 0 < |\mathbf{h}| \le portée \\ \sigma & pour\ |\mathbf{h}| > portée \end{cases}$$



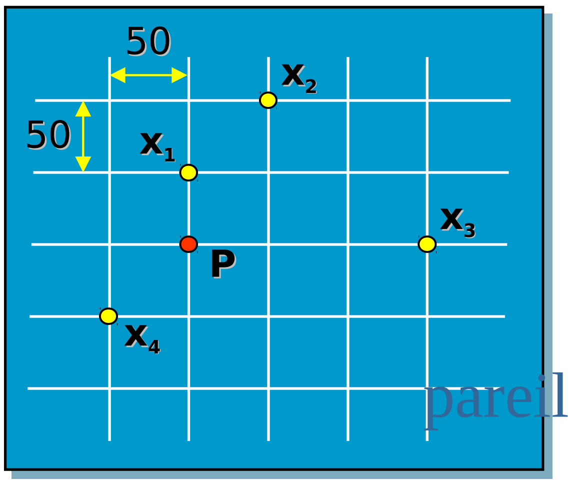
5 – Géostatistique : exemple de krigeage

- Un exemple complet
  - Calcul de la matrice

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ 0 \end{bmatrix} = \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & 1 \\ C_{2,1} & C_{2,2} & C_{2,3} & C_{2,4} & 1 \\ C_{3,1} & C_{3,2} & C_{3,3} & C_{3,4} & 1 \\ C_{4,1} & C_{4,2} & C_{4,3} & C_{4,4} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{1,p} \\ C_{2,p} \\ C_{3,p} \\ C_{4,p} \\ 1 \end{bmatrix}$$

5 – Géostatistique : exemple de krigeage

- Un exemple complet
  - Calcul de la matrice



$$C_{1,2} = C_{2,1} = C_0 + C_1 - \gamma(50\sqrt{2})$$

$$= (2+20) - \left[2+20\left(1,5\,\frac{50\sqrt{2}}{200} - 0,5\,\frac{(50\sqrt{2})^3}{(200)^3}\right)\right] = 9,84$$

pareil pour :

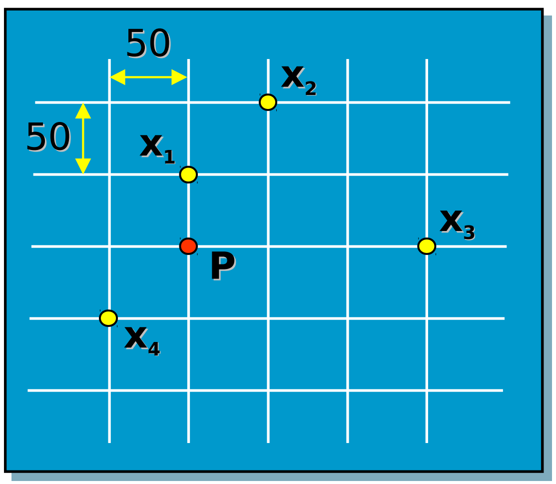
$$C_{4,p} = C_0 + C_1 - \gamma(50\sqrt{2}) = 9,84$$

On continue ...

5 – Géostatistique : exemple de krigeage

Un exemple complet

- Calcul de la matrice



$$C_{1,3} = C_{3,1} = (C_0 + C_1) - \gamma[\sqrt{(150)^2 + (50)^2}\,] = 1,23$$

$$C_{1,4} = C_{4,1} = (C_0 + C_1) - \gamma[\sqrt{(100)^2 + (50)^2}\,] = 4,98$$

$$C_{2,3} = C_{3,2} = (C_0 + C_1) - \gamma[\sqrt{(100)^2 + (100)^2}\,] = 2,33$$

$$C_{2,4} = C_{4,2} = (C_0 + C_1) - \gamma[\sqrt{(100)^2 + (150)^2}\,] = 0,29$$

$$C_{3,4} = C_{4,3} = (C_0 + C_1) - \gamma[\sqrt{(200)^2 + (50)^2}\,] = 0$$

$$C_{1,p} = (C_0 + C_1) - \gamma(50) = 12,66$$

$$C_{2,p} = (C_0 + C_1) - \gamma[\sqrt{(100)^2 + (50)^2}\,] = 4,98$$

$$C_{3,p} = (C_0 + C_1) - \gamma(150) = 1,72$$

$$C_{1,1} = C_{2,2} = C_{3,3} = C_{4,4} = (C_0 + C_1) - \gamma(0) = 22$$

5 – Géostatistique : exemple de krigeage

- Un exemple complet
  - Inversion de la matrice (laissons faire la machine …)
    - w1 = 0,518 w2 = 0,089 w3 = 0,022 w4 = 0,371
  - Calcul de l'estimation
    - $z'_p = 0{,}518\, \mathbf{z}_1 + 0{,}022\, \mathbf{z}_2 + 0{,}089\, \mathbf{z}_3 + 0{,}371\, \mathbf{z}_4$
  - où $\mathbf{z}_1$ est la valeur relevée en $\mathbf{x}_1$, $\mathbf{z}_2$ est la valeur relevée en $\mathbf{x}_2$ etc.

5 – Géostatistique : Krigeage cas non stat.

- Soit les données sur la température de 100 stations météorologiques s_1 .. s_{100} sur une vaste étendue orientée Nord-Sud
  - Comment prédire les valeurs de la température T(s) à chaque point *s* en utilisant ces données ?
  - On sait que que la température à des latitudes plus petites (plus au sud) sont plus élevées.
    - il faut que T(s) tienne compte la latitude
- Krigeage ordinaire n'est pas approprié ici, car il suppose que la structure globale est la même partout.

→ La méthode du krigeage universel permet une structure globale non constante

5 – Géostatistique : Krigeage universel

- La méthode du krigeage universel consiste à modéliser la structure globale.
  - α doit être exprimé en fonction de s → $\alpha(s)$
    - par exemple :
      - soit t_{max} est la température moyenne au sud de la zone de latitude s_{max} et t_{min} est la température moyenne au nord de latitude s_{min}
      - on suppose que la température diminue linéairement avec la latitude

    on peut alors poser $\alpha(s) = t_{max} - (t_{max} - t_{min}).\dfrac{s - s_{max}}{s_{max} - s_{min}}$

- Les autres étapes sont sensiblement les mêmes (étude du variogramme)
  - Néanmoins, risque de biais dans le calcul des corrélations
  - On peut essayer de supprimer cette « dérive » *a priori* avant de calculer les corrélations

5 – Géostatistique : Krigeage

- Les étapes du krigeage
  - Examiner les données
    - Trouver les tendances spatiales
    - Vérifier la normalité de la loi
    - Transformer les variables, si nécessaire
    - Déterminer la zone d'influence des points (anisotropie)
  - Etude du variogramme (variographie)
    - Calculer le variogramme empirique
    - Sélectionner du modèle qui épouse le mieux la répartition des points dans le semi-variogramme
    - Vérifier le modèle par validation croisée : examiner l'écart quadratique moyen
  - Krigeage
    - Choisir de la méthode de krigeage à utiliser
    - Générer une grille de données à interpoler
    - Réaliser l'interpolation en utilisant le modèle choisi

Exercices

- Exercices :

1. Sur les données Ozone de la région de Los Angeles, essayer les différentes méthodes de krigeage

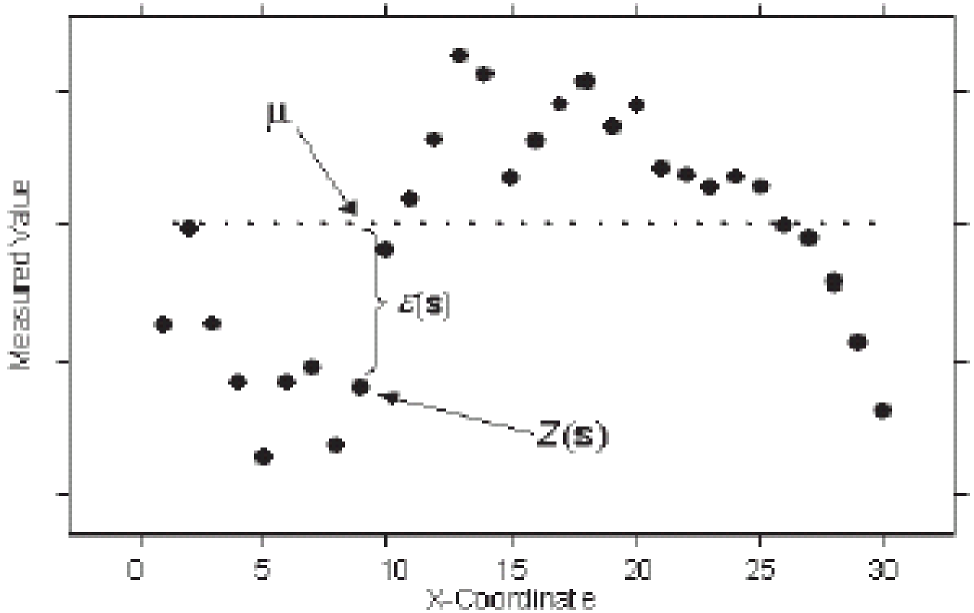
5 – Géostatistique : Propriétés du Krigeage

- Interpolation exacte
  - Il est facile de vérifier que pour un point de l’échantillon, la résolution du système fournit un poids égal à 1 pour ce point et un poids nul pour tous les autres
- Propriété de lissage
  - Les valeurs estimées présentent une dispersion moindre que les valeurs vraies
- Biais conditionnel
  - La moyenne des estimations est à peu égale à la moyenne réelle sauf si on ne prend que une zone où l’estimation est supérieure à un certain seuil → on ne trouve pas nécessairement la bonne moyenne sur cette zone

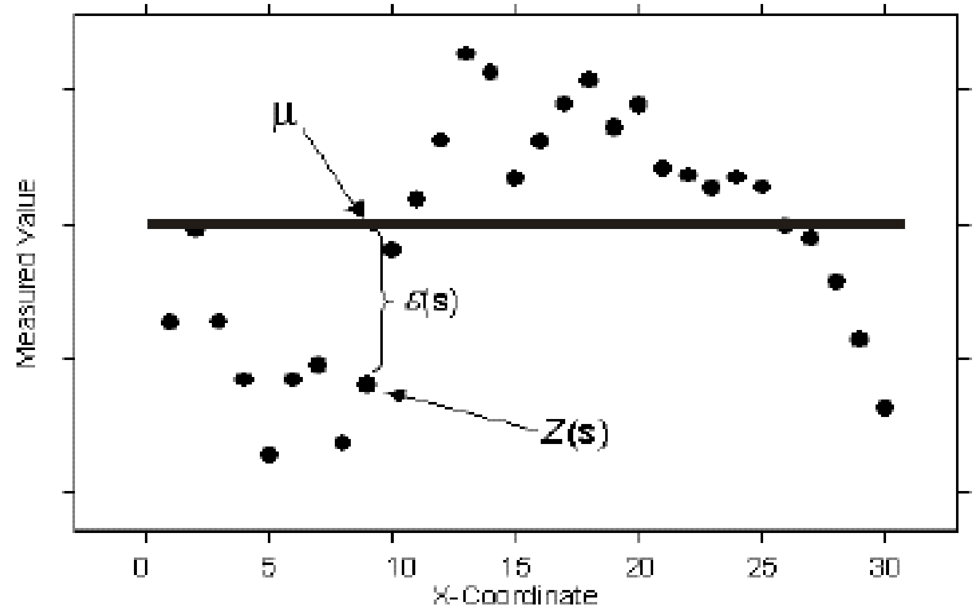
5 – Géostatistique : méthodes avancées

- Krigeage d'indicateurs :
  - méthode d'interpolation géostatistique n'exigeant pas que les données soient distribuées normalement.
- Co-krigeage
  - technique d'interpolation utilisée lorsque il existe une ou plusieurs variables fortement corrélées avec la variable à estimer.
    - Attention : il faut que ces variables soient a minima échantillonnées à la même série de lieux que la variable d'intérêt.

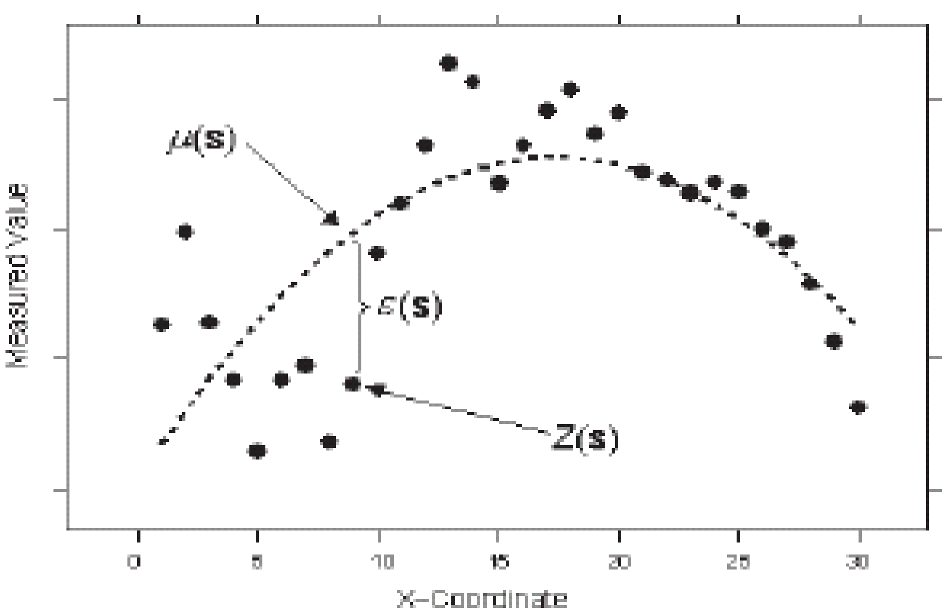
Ordinary Kriging



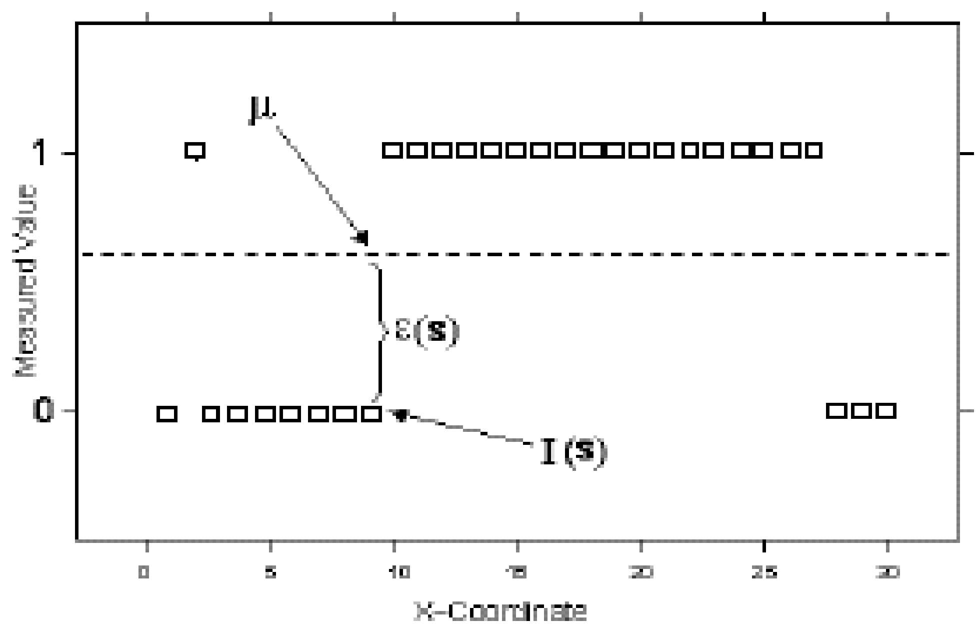
Simple Kriging



Universal Kriging



Indicator Kriging



Conclusion sur le Krigeage

- Conclusion générale
  - Il existe plusieurs solutions d'interpolation pour un même ensemble de données.
  - La qualité des résultats d'interpolation spatiale dépend de :
    - L'exactitude, la quantité et la distribution spatiale des valeurs utilisées pour l'interpolation.
    - L'efficacité du modèle d'interpolation utilisé à modéliser correctement le phénomène à l'étude.
  - Il faut associer les méthodes disponibles avec une bonne connaissance du terrain.

Principales références

- Sources
  - ESRI book “using ArcGIS Geostatistical Analyst”
  - Spatial Interpolation: A Brief Introduction - Eugene Brusilovskiy
  - www.spatialanalysisonline.com
  - Geo-statistical Analysis - Dr. A.K.M. Saiful Islam
  - Hétérogénéité spatiale à différentes échelles, S.GARRIGUES, D.ALLARD, F.BARET,S.MARNI,H.JEANJEAN
  - Représentation cartographique de la qualité de l’air à l’échelle d’une agglomération ou d’une région, Giovanni CARDENAS
  - Geospatial Analysis and Modeling, Helena Mitasova
  - Estimation et interpolation spatiale, Michel Arnaud et Xavier Emery (Hermès)
  - Statistique spatiale, méthodes et applications géomatique, Jean-Marc Zaninetti (Hermès)