# LE KRIGEAGE : LA MÉTHODE OPTIMALE D'INTERPOLATION SPATIALE


Yves Gratton

INSTITUT NATIONAL DE LA RECHERCHE SCIENTIFIQUE
INRS-Eau-Terre-Environnement
B.P. 7500, Québec, Qc, Canada, G1V4C7
Yves_Gratton@inrs-ete.uquebec.ca



Nous présentons une introduction informelle à la méthode d'estimation (interpolation) spatiale connue sous le nom de Krigeage. La notion de variogramme est introduite et discutée. La supériorité du Krigeage est illustrée à l'aide d'un exemple simple.



***Mots clés*** **:** Krigeage, géostatistiques, analyse objective, interpolation spatiale, interpolation optimale, interpolation de Gauss-Markov.


## 1. INTRODUCTION

L'objectif de cet article est de présenter informellement les principes de la méthode d'interpolation spatiale connue sous le nom de « Krigeage ». Le Krigeage est la méthode optimale, au sens statistique du terme, d'estimation. On peut l'utiliser autant pour l'interpolation que l'extrapolation. Ici nous nous restreindrons à l'interpolation et l'extrapolation spatiales en deux dimensions. Le Krigeage porte le nom de son précurseur, l'ingénieur minier sud-africain D.G. Krige. Dans les années 50, Krige[1] a développé une série de méthodes statistiques empiriques afin de déterminer la distribution spatiale de minerais à partir d'un ensemble de forages. C'est cependant le français Matheron[2] qui a formalisé l'approche en utilisant les corrélations entre les forages pour en estimer la répartition spatiale. C'est lui qui a baptisé la méthode « Krigeage ». Il a aussi été le premier a utiliser le terme « géostatistiques » pour désigner la modélisation statistique de données spatiales. Les mêmes idées ont été développées parallèlement en URSS par L.S. Gandin[3]. Gandin a baptisé sa méthode « interpolation optimale ». Il a introduit la notion d'« analyse objective » pour décrire cette approche basée sur les corrélations. C'est le nom sous lequel la méthode est connue en météorologie. En océanologie, la méthode a été introduite par Bretherton et al.[4] et elle est connue sous le nom de « méthode d'interpolation de Gauss-Markov », d'après le nom qu'on lui donne formellement dans les livres de statistiques (voir Liebelt[5], par exemple). Dans les prochaines sections, nous présenterons une description du variogramme, le cœur du Krigeage. Suivront une présentation du Krigeage ordinaire, un exemple illustrant la supériorité de la méthode, ainsi que la conclusion.

## 2. LE VARIOGRAMME

L'interpolation spatiale est un problème classique d'estimation d'une fonction F(**x**), où **x** = (x,y), en un point $\mathbf{x}_p$ du plan à partir de valeurs connues de F en un certain nombre, m, de points environnants $x_i$:

$$F(x_p) = \sum_{i=1}^{m} W_i \cdot F(x_i) \qquad (1)$$

Le problème consiste à déterminer la pondération, i.e. les $W_i$, de chacun des points environnants. Il existe plusieurs façons de choisir ces poids. Les deux méthodes les plus connues sont l'interpolation linéaire (en fonction de l'inverse de la distance) et la méthode des splines cubiques (ajustement de polynômes cubiques). Le Krigeage choisit plutôt les poids à partir du degré de similarité entre les valeurs de F, i.e. à partir de la covariance entre les points en fonction de la distance entre ces points.

Un utilisateur sérieux du Krigeage se doit de bien connaître les conditions d'utilisation de la méthode. Ici, nous dirons simplement que la seule condition indispensable pour utiliser le Krigeage est que la moyenne et la variance (voir l'Annexe) de la fonction F soient stationnaires, c'est-à-dire qu'elles ne dépendent pas de la position des points, seulement de la distance entre les points. Le variogramme est alors simplement la variance totale moins la covariance, en fonction de la distance entre les points. Le livre de Journel et Huijbregts[6] présente une excellente description formelle du variogramme (et aussi du Krigeage), tout en maintenant le niveau mathématique à son minimum. Le Krigeage utilisera alors le semi-variogramme (la moitié du variogramme) pour déterminer les poids dans l'équation (1). Le semi-variogramme est calculé à l'aide de l'équation (2) pour les n(h) points $x_i$ et $y_i$ séparés par une distance $h = | x_i - y_i |$ :

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (x_i - y_i)^2 \qquad (2)$$

La figure 1 présente un exemple de semi-variogramme $\gamma(h)$. Il suffit d'ajuster une fonction analytique à tous ces points à l'aide de la méthode des moindres carrés et nous obtenons alors une fonction continue caractérisant complètement la semi-variance en fonction de la distance entre les points.

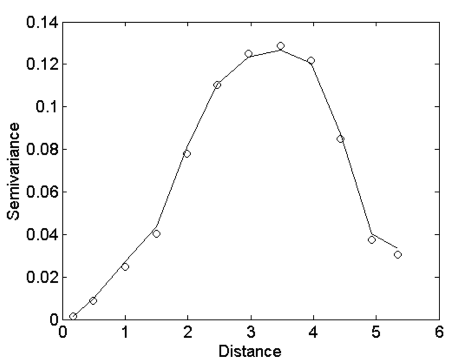


**Figure 1.** Exemple de semi-variogramme. Les points (cercles) sont obtenus à l'aide de l'éq. (2) pour l'ensemble des distances h possibles, où $h = |x_i - y_i|$. Une fonction continue (la ligne continue) a ensuite été ajustée à l'aide de la méthode des moindres carrés.

Le choix et l'ajustement d'une fonction au semi-variogramme est la partie la plus délicate du Krigeage : c'est presqu'un art plutôt qu'une science. Le lecteur est une fois de plus renvoyé au livre de Journel et Huijbregts[6] qui présente, à mon avis, la meilleure discussion des types de variogramme et de leurs comportements. Le choix du semi-variogramme doit ensuite être validé par des tests statistiques.

## 3. LE KRIGEAGE ORDINAIRE

Le Krigeage consiste ensuite à calculer les $W_i$ de l'équation (1) à l'aide des valeurs de la fonction $\gamma(h)$ correspondant aux m points choisis. Il existe trois types de Krigeage univarié (i.e. à une seule variable) : le Krigeage simple, le Krigeage ordinaire et le Krigeage universel. La différence entre ces types d'estimation réside dans la connaissance de la statistique de la variable à interpoler :

(1) Krigeage simple : variable stationnaire de moyenne connue;

(2) Krigeage ordinaire : variable stationnaire de moyenne inconnue et

(3) Krigeage universel : variable non-stationnaire (qui contient une tendance).

Ici, nous nous restreindrons au Krigeage ordinaire, aussi appelé Krigeage ponctuel par certains auteurs, car il est le plus fréquemment utilisé. La méthode consiste à déterminer la combinaison de poids, i.e. la combinaison des $W_i$ de l'éq. (1), qui garantit que les semi-variances calculées à l'aide du point cible $\mathbf{x}_p$ se retrouveront sur la courbe de la figure 1. Les poids sont obtenus en multipliant les $W_i$, pour chacun des m points, par chacune des m semi-variances associées à ce point (les lignes de la matrice A de l'éq. (3)). Le problème s'exprime finalement sous la forme du système de m+1 équations linéaires à m+1 inconnues suivant (Davis 1986)

$$A \cdot W = B \qquad (3)$$

$$\text{où} \quad A = \begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \dots & \gamma(h_{1m}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \dots & \gamma(h_{2m}) & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \gamma(h_{m1}) & \gamma(h_{m2}) & \dots & \gamma(h_{mm}) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix},$$

$$W = \begin{bmatrix} W_1 \\ W_2 \\ \dots \\ W_m \\ \lambda \end{bmatrix} \quad \text{et} \quad B = \begin{bmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \dots \\ \gamma(h_{mp}) \\ 1 \end{bmatrix}.$$

Les $\gamma(h_{ij})$ sont les valeurs du semi-variogramme qui correspondent à la distance $h_{ij}$ entre les points $x_i$ et $x_j$. Les $\gamma(h_{ij})$ ont déjà été calculés à partir des données à l'aide de l'éq. (2), tandis que les $\gamma(h_{ip})$ sont calculés à l'aide de la fonction analytique qui a été ajustée aux points au semi-variogramme de la fig. 1. Pour que la solution soit non-biaisée, la somme des poids, les $W_i$, doit être égale à 1. Cette dernière contrainte introduit un degré de liberté supplémentaire dans le problème. Ce degré supplémentaire est utilisé en ajoutant une variable libre, $\lambda$ (un multiplicateur de Lagrange), dans le but de minimiser l'erreur d'estimation. Le vecteur *W* est obtenu en multipliant les deux côtés de l'équation (3) par l'inverse de la matrice *A*. La valeur recherchée au point $\mathbf{x}_p$ est ensuite calculée en utilisant les valeurs connues de F, les $F(x_i)$, à l'aide de l'équation (1).

La variance de l'estimation $s_p^2$, c'est-à-dire le carré de l'erreur standard en chaque point, est obtenue par la relation

$$s_p^2 = W^T \cdot B. \qquad (4)$$

où le T indique qu'il faut utiliser la transposée du vecteur W. Si nous supposons que les erreurs d'estimation sont normalement distribuées autour de la vraie valeur, alors la probabilité que la vraie valeur soit $F(x_p) \pm s_p$ est de 68%, tandis que la probabilité que la vraie valeur soit $F(x_p) \pm 2s_p$ est de 95% (Davis 1986).

## 4. EXEMPLE DE KRIGEAGE ORDINAIRE

La figure 2 propose les contours de la fonction

$$F(R) = \frac{\sin R}{R}, \text{ où } R = (x^2 + y^2)^{1/2} \qquad (5)$$

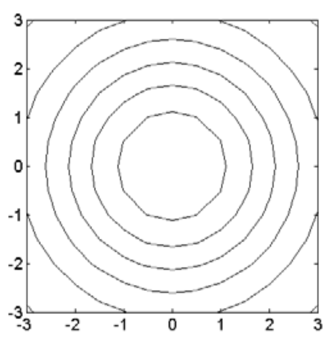

**Figure 2.** Contours de F(R), où F(R) est défini par l'éq. (5)

Dans le but de vérifier l'efficacité de la méthode, nous (Gratton et al.[8]) avons échantillonné cette fonction à l'aide de la grille irrégulière que nous avions utilisée lors d'une campagne océanographique en mer d'Alborán (en Méditerranée occidentale). Les figures 3a et 3b présentent deux estimations des contours de la fonction F(R) obtenues en calculant les pondérations à l'aide de la méthode de la distance inverse et par Krigeage. La grille irrégulière d'échantillonnage est aussi reproduite sur ces figures. Autrement dit, nous tentons de reproduire, par interpolation et extrapolation, les contours parfaitement circulaires de la figure 2, à partir des valeurs de F(R) calculées aux points identifiés sur les figures 3a,b.

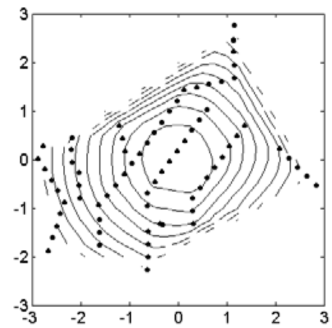

**Figure 3a.** Estimation de la fonction F(R) à l'aide d'une interpolation linéaire basée sur l'inverse de la distance. La grille irrégulière d'échantillonnage est aussi présentée.

On observe que l'interpolation à l'aide du Krigeage est de beaucoup supérieure à l'autre. De plus, le Krigeage nous permet d'obtenir une estimation « raisonnable » de la fonction à l'extérieur (extrapolation) de la grille d'échantillonnage. Il est aussi possible (et souhaitable) de tracer les contours de l'erreur d'estimation donnée par l'éq. (4).

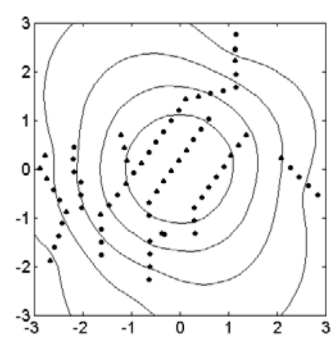

**Figure 3b.** Même chose que la figure 3a, mais à l'aide du Krigeage ordinaire.

## 5. DISCUSSION ET CONCLUSIONS

Nous avons présenté une introduction informelle au Krigeage et à son outil privilégié, le semi-variogramme. Nous avons eu recours à un exemple en deux dimensions mais le Krigeage s'applique tout aussi bien en trois dimensions. Lors du calcul du variogramme, nous avons utilisé le fait, sans l'énoncer, que le variogramme était isotrope, c'est-à-dire que la variabilité était la même dans les toutes les directions du plan. Ce n'est généralement pas le cas. Lorsque cela se produit, nous devons définir un variogramme pour chacun des axes du plan. De plus, nous avons utilisé tous les points disponibles pour calculer les $W_i$ : la plupart des logiciels de *contouring* n'en utilise qu'un sous-ensemble. Davis[7] présente une description des différentes méthodes de *contouring*. Finalement, il existe une version multivariée (i.e. à plusieurs variables distinctes) du Krigeage. Le lecteur intéressé est invité à consulter l'article de Marcotte[9] et le livre de Deutsch and Journel[10] qui nous proposent des logiciels écrits en Matlab® et en Fortran®, respectivement. Nous (Gratton et Lafleur[11]) les avons réunis en une seule boîte à outils en Matlab®. Il existe cependant plusieurs autres logiciels gratuits aussi disponibles sur Internet. Le site suivant de l'Université de Lausanne, entre autres, est un excellent point de départ pour obtenir plus d'informations sur les logiciels disponibles :

http://www.ai-geostats.org/.

Le Krigeage est la méthode optimale, au sens statistique, d'interpolation et d'extrapolation. C'est la méthode d'estimation la plus précise. Contrairement à toutes les autres méthodes, elle nous permet aussi de calculer l'erreur d'estimation. Il faut cependant ajouter que dans plusieurs cas, et spécialement dans le cas d'une grille régulière d'échantillonnage, la méthode des splines cubiques produit des résultats à peu près « équivalents » à ceux obtenus par Krigeage[12], tout en étant plus rapide et plus simple à utiliser. Si la précision des résultats est importante, le Krigeage demeure, même dans ce dernier cas, la méthode de prédilection.

## ANNEXE : Moments d'ordre un et deux

Une variable aléatoire, $Z(\mathbf{x})$ où $\mathbf{x}$ est le vecteur (x,y), est une fonction qui prend un ensemble de valeurs ou de réalisations selon une distribution de probabilités quelconque. Ces réalisations peuvent être des températures, des concentrations de minerai, l'abondance de zooplancton, etc. Une fonction aléatoire est un ensemble de variables aléatoires définies sur une région d'intérêt : $\{Z(\mathbf{x}),\ \mathbf{x} \in$ une zone d'étude$\}$. Les deux premiers moments d'une fonction aléatoire sont

*Moment d'ordre un ou l'espérance mathématique.*

La moyenne: $m(\mathbf{x}) = E\{Z(\mathbf{x})\}$

*Moments du second ordre.*

a) La variance: $\mathrm{Var}\{Z(\mathbf{x})\} = E\{\ [Z(\mathbf{x}) - m(\mathbf{x})]^2\ \}$

b) La covariance: $C(\mathbf{x}_1,\mathbf{x}_2) = E\{\ [Z(\mathbf{x}_1)-m(\mathbf{x}_1)] \bullet [Z(\mathbf{x}_2) - m(\mathbf{x}_2)]\ \}$

c) Le variogramme: $2\gamma(\mathbf{x}_1,\mathbf{x}_2) = \mathrm{Var}\{\ Z(\mathbf{x}_1) - Z(\mathbf{x}_2)\ \}$

Si la fonction aléatoire est stationnaire, alors

$\mathrm{Var}\{Z(\mathbf{x})\} = E\{Z(\mathbf{x}) - m^2\} = C(0)$ et

$\gamma(h) = ½\, E\{[Z(\mathbf{x}+h) - Z(\mathbf{x})]^2\} = C(0) - C(h)$.

## REFERENCES


1) Krige, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. 1951. J. of Chem., Metal. and Mining Soc. of South Africa, 52, 119-139.
2) Matheron, G. 1963. Principles of Geostatistics. Economic Geol., 58, 1246-1268.
3) Gandin, L.S. 1965. Objective Analysis of Meteorological fields. Israël Program for Scientific Translations, No. 1373, 242 p.
4) Bretherton, F.B., R.E. Davis and C.B. Fandry. 1976. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. Deep-Sea Res., 23, 559-582.
5) Liebelt, P.B. 1967. *An introduction to Optimal Estimation*, Addison-Wesley, 267 p.
6) Journel, A.G. and C.J. Huijbregts. 1978. *Mining Geostatistics*. Academic Press, 600 p.
7) Davis, J. C. 1986. *Statistics and Data Analysis in Geology*. 2nd ed, John Wiley & Sons. New York, 289 p.
8) Gratton, Y., L. Prieur, R.G. Ingram, et C. Lafleur. 2002. *Les courants en mer d'Alborán Est pendant la campagne Almofront-I*. Rapport interne, INRS-ETE.
9) Marcotte, D. 1991. Cokriging with MATLAB. Computers & Geosciences, 17(9): 1265-1280.
10) Deutsch, C.V. et A.G. Journel. 1992. *GSLIB - Geostatistical Software Library*. Oxford Univ. Press, New York, 340 p.
11) Gratton, Y. et C. Lafleur . 2001. *Le Matlab Kriging Toolbox*. Version 4.0. Manuel de référence, INRS-ETE. Le logiciel est disponible gratuitement à l'adresse suivante : http://www.inrs-ete.uquebec.ca/activites/repertoire/yves_gratton/krig.htm
12) Dubrule, O. 1984. Comparing Splines and Kriging. Computers & Geosciences, 10(2-3): 327-33.