

TECHNIQUES GÉOSTATISTIQUES POUR L'INTERPOLATION SPATIALE À PARTIR D'OBSERVATIONS ET DE SIMULATIONS NUMÉRIQUES

Geostatistical techniques for spatial interpolation from
observations and numerical simulation output

Hans WACKERNAGEL * - Walid TABBARA **

(*): MINES-ParisTech, Equipe de Géostatistique, 35 rue St Honoré
F-77305 Fontainebleau, hans.wackernagel@mines-paristech.fr

(**): Université Pierre et Marie Curie-Paris 6; Supélec, L2S, 3 rue Joliot Curie,
Plateau du Moulon
F-91192 Gif sur Yvette, tabbara@lss.supelec.fr

Résumé. Nous proposons une introduction à quelques concepts et méthodes de la géostatistique, qui regroupe dans un formalisme de fonctions aléatoires des méthodes pour l'estimation de phénomènes déployés dans l'espace géographique. Cette estimation est faite sur la base d'observations, qui peuvent inclure des prévisions effectuées avec des modèles numériques. Nous illustrons le propos au moyen de deux exemples d'application. Le premier est relatif à l'estimation du risque de dépassement d'un seuil environnemental en pollution de l'air. Le second traite du couplage d'une onde électromagnétique sur un déploiement de structures.

Synopsis. We provide an introduction to some concepts and methods of geostatistics, which proposes in the framework of a random functions formalism a number of methods for the estimation of phenomena in a geographical space. The estimation is performed on the basis of observations which may include forecasts performed with numerical models. Illustrations come from two areas. The first is about estimating the risk of threshold trespassing in air pollution. The second deals with the coupling of an electromagnetic wave on a deployment of structures.

Mots-Clés: Géostatistique, interpolation spatiale, mesures, couplage, électromagnétisme.

Keywords: Geostatistics, spatial interpolation, measurements, coupling, electromagnetism.

I. LA GÉOSTATISTIQUE

La géostatistique [1,2] est issue des sciences de l'ingénieur, née de problèmes d'estimation de ressources minières, et elle a trouvé au cours des dernières décennies des

applications dans de nombreux domaines, notamment en environnement, comme, par exemple, celle d'estimer l'exposition acoustique en milieu ouvert [3] ou encore l'exposition radioélectrique en milieu urbain [4].

La géostatistique est basée sur la théorie des variables régionalisées. Une variable régionalisée, notée $z(x)$, décrit une quantité z de coordonnées x dans un domaine spatial D . On ne connaît généralement pas la quantité $\{z(x), D\}$ dans l'ensemble du domaine D , d'où l'intérêt d'avoir recours à un modèle probabiliste. L'avantage de l'introduction d'un cadre probabiliste est qu'il permettra non seulement de résoudre des problèmes d'estimation en des endroits non échantillonnés, mais aussi de simuler des réalisations conditionnellement aux valeurs échantillonnées [5].

I.1 Le variogramme

On considérera ainsi $\{z(x), D\}$ comme une réalisation d'une fonction aléatoire $\{Z(x), D\}$, dont on supposera souvent que les incréments sont stationnaires d'ordre deux:

$$\begin{aligned} E[Z(x+h) - Z(x)] &= 0 \\ \text{var}[Z(x+h) - Z(x)] &= 2\gamma(h) \end{aligned}$$

c'est-à-dire qu'au niveau du modèle de fonction aléatoire la différence des valeurs en deux points espacés de h est nulle en moyenne, quelle que soit la localisation du point x et quelle que soit la longueur et l'orientation du vecteur h ; la variance de cette différence, quant à elle, ne dépend que du vecteur h et non de la position de la paire de points dans le domaine.

La quantité $\gamma(h)$ est le *variogramme*, que l'on peut donc écrire:

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

I.2 Le krigeage

Pour l'estimation de valeurs en des points non-échantillonnés de l'espace géographique, la géostatistique se sert d'une forme particulière de régression linéaire, qui été qualifiée de *krigeage*, d'après l'ingénieur D.G. Krige qui en a eu l'idée en 1950. La version la plus courante en est le *krigeage ordinaire*. Pour estimer une valeur $Z^*(x_0)$ en un point quelconque du domaine on a recours à une combinaison linéaire des n observations $Z(x)$:

$$Z^*(x_0) = \sum_{\alpha=1}^n w_{\alpha} Z(x_{\alpha})$$

pour laquelle on imposera des pondérateurs w de somme unité (pour que l'estimateur soit sans biais). En minimisant la variance d'estimation avec cette contrainte (laquelle mène à l'introduction d'un multiplicateur de Lagrange), on obtient le système de $n+1$ équations du krigeage ordinaire:

$$\sum_{\beta=1}^n w_{\beta} \gamma(x_{\alpha} - x_{\beta}) + \mu = \gamma(x_{\alpha} - x_0), \quad \forall \alpha$$

$$\sum_{\beta=1}^n w_{\beta} = 1$$

dont la résolution nous livre les pondérateurs optimaux w de notre estimateur.

I.3 Le krigeage avec dérive externe

On peut disposer dans certains cas d'un facteur explicatif défini sur tout le domaine, comme, par exemple, dans l'étude de températures de surface, l'altitude fournie par un modèle numérique de terrain. Un problème analogue est celui d'une prévision de la quantité d'intérêt par un modèle numérique, comme par exemple, pour l'estimation de l'ozone, la prévision d'un modèle transport atmosphérique. On supposera qu'une telle variable régionalisée $s(x)$, connue sur tout le champ et appelée *dérive externe*, est équivalente en moyenne à $Z(x)$ à des constantes additive a et multiplicative b près,

$$E[Z(x)] = a + bs(x)$$

et cela mène, selon les mêmes principes que précédemment, à établir le système à $n+2$ équations du krigeage avec dérive externe:

$$\sum_{\beta=1}^n w_{\beta} \gamma(x_{\alpha} - x_{\beta}) + \mu_1 + \mu_2 s(x_{\alpha}) = \gamma(x_{\alpha} - x_0), \quad \forall \alpha$$

$$\sum_{\beta=1}^n w_{\beta} = 1$$

$$\sum_{\beta=1}^n w_{\beta} s(x_{\beta}) = s(x_0)$$

I.4 Le changement de support

Les mesures que l'on effectue sur le terrain ont souvent un caractère ponctuel (ce que l'on considérera dans la suite, sans perte de généralité). Un problème fréquemment posé dans la pratique est celui de savoir si une valeur réglementaire environnementale est dépassée à tel ou tel endroit sur un volume d'intégration spatial donné; on appellera ce volume le *support*. Le problème du dépassement de seuil a deux implications: 1) l'estimation du dépassement de seuil fait appel à un estimateur qui n'est plus une simple combinaison linéaire des données, 2) pour calculer cet estimateur il faut faire référence à la loi de distribution pour le support spécifié. En général le support donné dans la réglementation est différent du support de mesure, car le premier est établi pour être représentatif de l'exposition, tandis que le second est défini sur la base des possibilités techniques de prise de mesures. Pour résoudre le problème de l'estimation de dépassement de seuil la géostatistique non-linéaire dispose de modèles décrivant la déformation que subit une loi de distribution lorsque l'on passe d'un support à un autre. Nous donnons dans la suite un exemple d'un tel modèle de changement de support issu de la pratique minière et indiquons un estimateur non-linéaire basé sur ce modèle.

Le point de départ du *modèle Gaussien discret* est la relation de Cartier qui stipule que pour un point de position aléatoire dans un volume, $\underline{x} \in V$, l'espérance conditionnelle de $Z(\underline{x})$ est égale à la valeur moyenne de Z pour ce volume:

$$E[Z(\underline{x}) | Z(V)] = Z(V)$$

En général, les observations ne sont pas Gaussiennes, on définira donc une transformation $\varphi(Y)$ d'une Gaussienne Y sur la base d'un développement en polynômes d'Hermite H_k (qui sont associés à la loi Gaussienne) avec des coefficients φ_k . C'est ce que l'on appelle une *anamorphose Gaussienne*:

$$Z(x) = \varphi(Y) = \sum_{k=0}^{\infty} \frac{\varphi_k}{k!} H_k(Y)$$

Pour une variable définie sur un support de blocs, $Z(V)$, on peut définir une fonction d'anamorphose de blocs pour laquelle la relation de Cartier s'écrit:

$$E[\varphi(Y(\underline{x})) | Z(V)] = \varphi_v(Y(V))$$

Dans ce contexte Gaussien, les coefficients de l'anamorphose de blocs sont reliés à ceux de l'anamorphose ponctuelle par un coefficient r :

$$\varphi_{vk} = \frac{\varphi_k}{k!} r^k$$

où le coefficient de changement de support point-bloc r est déduit de la variance de dispersion des blocs, qui peut être calculée à l'aide du modèle de variogramme.

De la même manière pour des cellules V , sachant que V est partitionné en blocs v , on peut définir, puis calculer, la valeur d'un coefficient point-cellule r' , qui servira au calcul du coefficient bloc-cellule r_{vV} , qu'on obtient simplement en formant le rapport:

$$r_{vV} = r' / r$$

La quantité d'intérêt, la proportion de blocs v dans une cellule particulière V pour lesquelles une valeur limite réglementaire z_c est dépassée, est obtenue par *conditionnement uniforme*, c'est-à-dire en formant l'espérance conditionnelle de l'indicatrice I testant le dépassement:

$$E [I_{Z(\underline{y}) > z_c} | Z(V)] = 1 - G \left(\frac{y_c - r_{vV} y(V_0)}{\sqrt{1 - r_{vV}^2}} \right)$$

où G est la loi Gaussienne et où la variable conditionnelle $Y(\underline{y}) = y(V_0)$ est Gaussienne de moyenne $r_{vV} y(V_0)$ et de variance $1 - r_{vV}^2$.

II EXEMPLE: Estimation de probabilité de dépassement de seuil

Cet exemple est tiré de la publication [6]. Sur la Fig. 1 on voit (à gauche) la localisation de 19 stations du réseau de mesures d'ozone d'Airparif pour l'Ile de France le 17 juillet 1999; les centres des cellules du modèle de transport chimique CHIMÈRE sont également représentés. La carte (Fig. 1, à droite) représente une prévision du modèle numérique, corrigée par les valeurs des stations au moyen d'un krigeage par dérive externe.

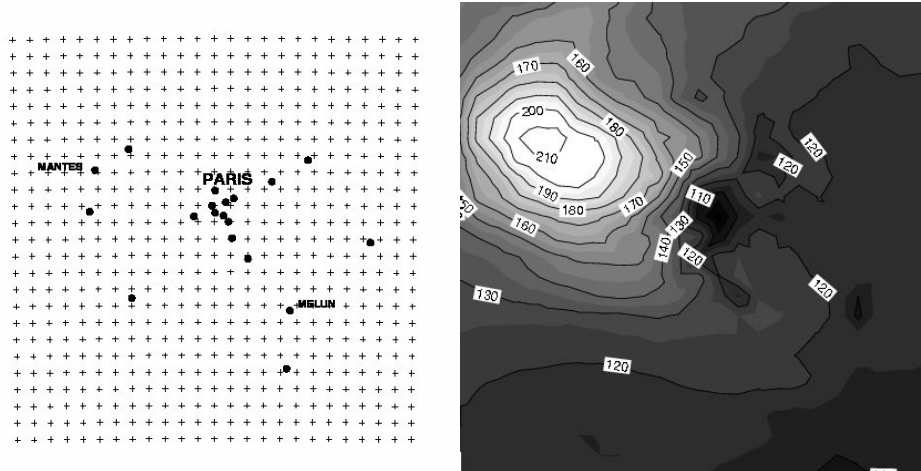


Figure 1: Localisation des 19 stations et des centres des cellules 6x6 km² (à gauche).
Prévission du modèle CHIMÈRE : valeurs d'ozone en g/m³ (à droite).

La Fig. 2 exhibe les fonctions d'anamorphose pour les blocs (à gauche, $r=.97$) et les cellules (à droite, $r'=.72$) calculées à partir des 19 valeurs d'ozone, en faisant intervenir le modèle de variogramme ajusté à ces valeurs.

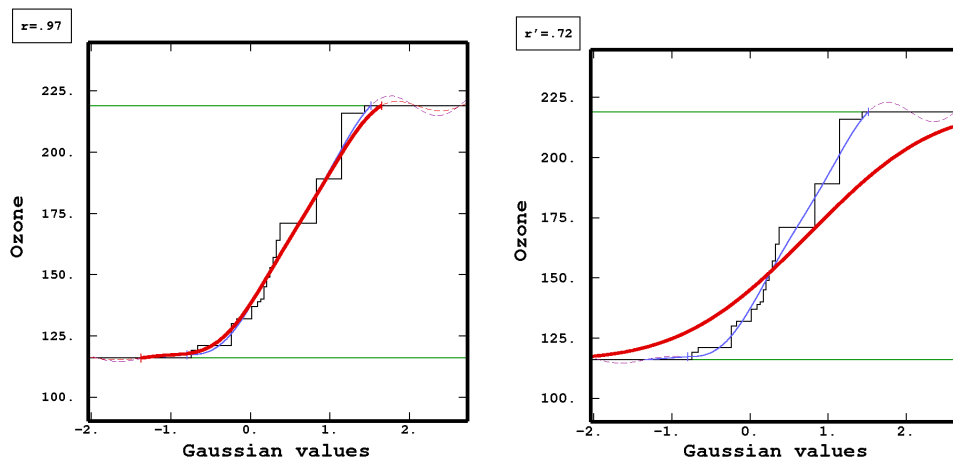


Figure 2: Fonctions d'anamorphose pour des supports de blocs 1x1 km² (à gauche) et de cellules 6x6 km² (à droite).

Sur la Fig. 3 on peut voir (à gauche) les histogrammes des valeurs moyennées sur des blocs et des cellules, calculés sur la base du modèle de changement de support. On voit également (à droite) les courbes de proportion de blocs dont les valeurs dépassent un seuil environnemental en fonction de différentes valeurs possibles de seuils environnementaux données en abscisse.

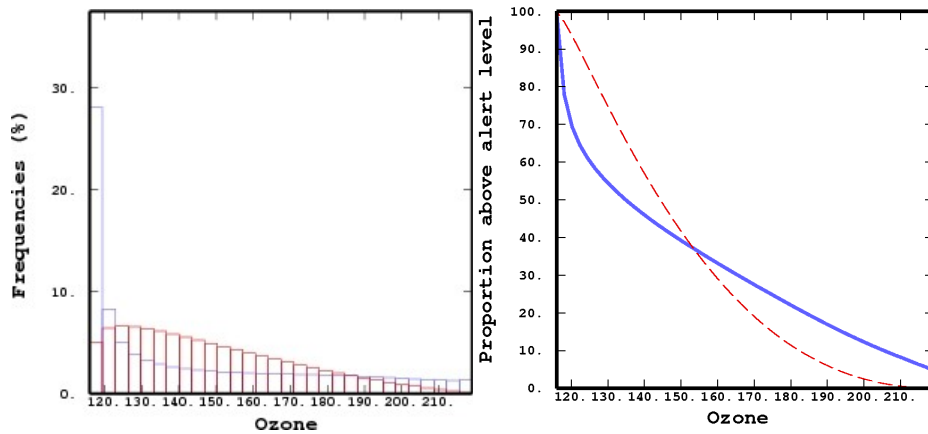


Figure 3: Histogrammes (à gauche), courbes de proportion de dépassement de seuils (à droite), pour des unités de $1 \times 1 \text{ km}^2$ [courbe continue] et de $6 \times 6 \text{ km}^2$ [courbe discontinue] .

La Fig. 4 représente des cartes de dépassement de seuil obtenues par conditionnement uniforme en considérant (à gauche) un seuil environnemental Suisse et (à droite) un seuil Européen. Ces cartes de probabilité de dépassement de seuil peuvent alors servir à informer la population sur le risque de pollution dans une partie spécifique de l'Ile de France.

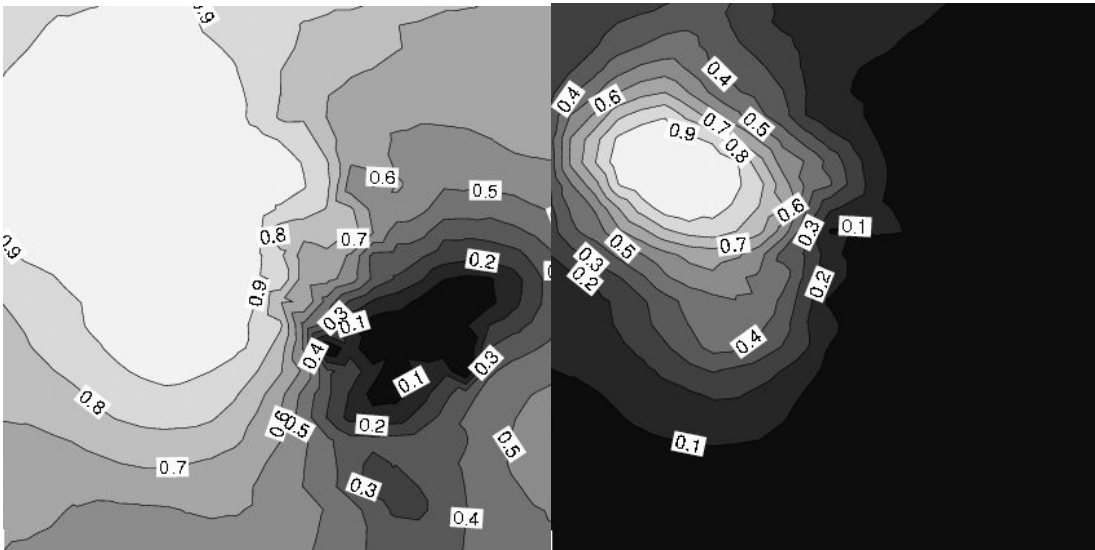


Figure 4: Probabilités de dépassement de seuils de 120 g/m^3 (à gauche) et de 180 g/m^3 (à droite), sur un support de $1 \times 1 \text{ km}^3$.

III EXEMPLE: Couplage d'une onde électromagnétique sur un déploiement de structures [7]

L'étude et l'analyse du couplage d'une onde électromagnétique sur une structure complexe, comme le déploiement d'abris (shelters) considéré ici, fait appel à une modélisation aussi réaliste que possible mais nécessairement simplifiée. Il n'en demeure pas moins que le coût informatique, lié au nombre de configurations possibles de la structure et des directions d'incidence de l'onde, nécessaire à la réalisation d'une analyse fine du problème est prohibitif.

Un échantillon de taille raisonnable de configurations est considéré pour lesquelles les valeurs d'un paramètre caractéristique du couplage sont calculées. Le krigeage ordinaire permet, à partir de ces valeurs, d'estimer celles relatives à de nouvelles configurations et d'évaluer la qualité de l'estimation.

Un ensemble de 11 abris est considéré, le paramètre caractéristique du couplage (observable) est le maximum temporel d'un champ calculé dans une zone d'un abris connecté aux dix autres. Pour un déploiement fixé seule la direction d'incidence varie de façon aléatoire, caractérisée par les deux facteurs : angle d'élévation (facteur f1) qui varie de 10° à 90° par pas de 20° , angle d'azimuth (facteur f2) qui varie de 0° à 135° par pas de 45° . Le modèle du krigeage retenu est à dérive externe représentée par une constante, les valeurs normalisées des facteurs varient de -1 à +1.

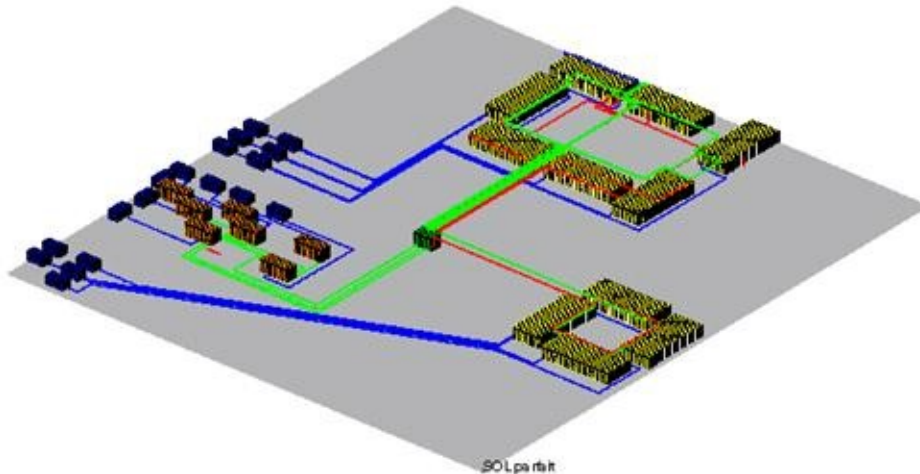


Figure 5: Le modèle d'un déploiement comprenant 11 abris et des équipements annexes.

La Fig. 5 représente le modèle d'un déploiement comprenant 11 abris et des équipements annexes. La Fig. 6 représente une coupe pour une élévation de 40° du résultat de l'estimation (courbe continue) et des bornes de l'écart-type à ± 2 .

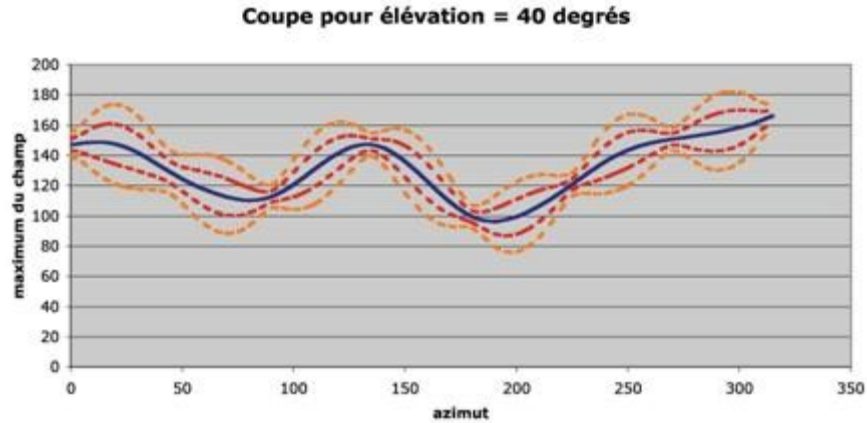


Figure 6: Une coupe pour une élévation de 40° du résultat de l'estimation (courbe continue) et des bornes de l'écart-type à $\pm 2\sigma$.

La Fig. 7 représente le résultat de la validation croisée, elle permet d'apprécier la précision de l'estimation.

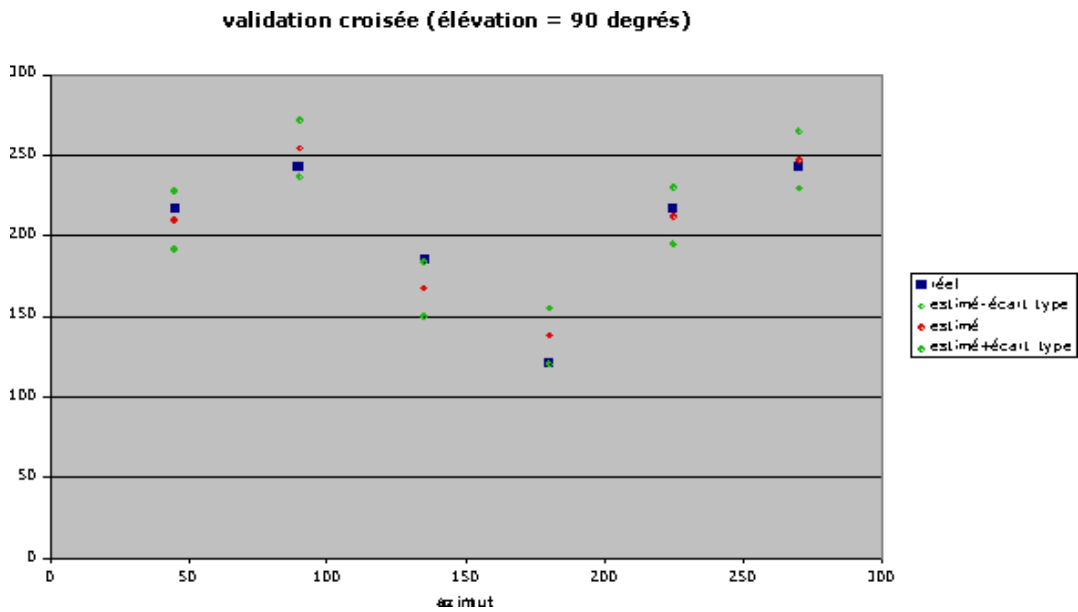


Figure 7: Le résultat de la validation croisée.

IV. CONCLUSION

La géostatistique n'a été introduite que récemment en électromagnétisme [8]. Pour ce qui est de son application à l'estimation de l'exposition radioélectrique en milieu urbain, elle est actuellement une composante essentielle du projet ANR SAMPER (2008-2010) visant à élaborer un système d'acquisition et de modélisation pour la prévision de l'exposition radioélectrique et impliquant le CSTB, l'Ecole des Mines, Orange Labs, Satimo et la Ville de Grenoble.

REFERENCES

- [1] J. Chilès and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*. New York: Wiley, 1999.
- [2] H. Wackernagel, *Multivariate Geostatistics: an Introduction with Applications*. Berlin: Springer-Verlag, 3rd ed., 2003.
- [3] O. Baume, B. Gauvreau, M. Bérangier, F. Junker, and H. Wackernagel, "Apport des méthodes géostatistiques à l'estimation de l'impact acoustique d'une source en environnement ouvert," *Acoustique & Techniques*, vol. 45, pp. 29-33, 2006.
- [4] Y. Ould Isselmou, H. Wackernagel, W. Tabbara, and J. Wiart, "Geostatistical estimation of electromagnetic exposure," in *geoENV VI - Geostatistics for Environmental Applications* (A. Soares, M. Pereira, and R. Dimitrakopoulos, eds.), pp. 59-69, Springer, 2008.
- [5] C. Lantuéjoul, *Geostatistical Simulation: Models and Algorithms*. Berlin: Springer-Verlag, 2002.
- [6] H. Wackernagel, C. Lajaunie, N. Blond, C. Roth, and R. Vautard, "Geostatistical risk mapping with chemical transport model output and ozone station data," *Ecological Modelling*, vol. 179, pp. 177-185, 2004.
- [7] J. Geiswiller, Y. Beniguel, J. C. Joly, B. Pecqueux, W. Tabbara, M. Hélier, and D. Lecointe, "Electromagnetic susceptibility of a structure composed of several shelters networked in a complex manner and constituting a command control unit," in *EUROEM 2004*, (Magdeburg), 2004.
- [8] P. De Doncker, J. M. Dricot, R. Meys, M. Hélier, and W. Tabbara, "Electromagnetic fields estimation using spatial statistics," *Electromagnetics*, vol. 26, pp. 111-122, 2006.

Hans WACKERNAGEL. Maître de Recherche à MINES-ParisTech au sein de l'équipe de Géostatistique du Centre de Géosciences. Il est docteur en sciences et techniques minières (Ecole des Mines de Paris, 1985) et habilité à diriger des recherches (Université Pierre-et-Marie Curie Paris VI, 2004). Auteur d'une monographie sur la géostatistique multivariable (en anglais et en japonais) il a été impliqué dans de nombreux projets nationaux et internationaux en modélisation statistique et en assimilation de données dans des domaines aussi divers que l'océanographie, l'étude du climat, la pollution de l'air, l'évaluation de l'exposition acoustique et radio-électrique, l'épidémiologie.

Walid TABBARA. Né à Beyrouth (Liban) en 1946, Docteur ès Sciences de l'Université Paris VI (1976), Professeur à cette même université. Il effectue ses recherches à Supélec au Département de Recherche en Électromagnétisme, ainsi qu'au laboratoire Franco-Singapourien SONDRRA où il occupe le poste de Technical Manager. Ses domaines de recherche actuels sont la compatibilité électromagnétique et la télédétection des milieux naturels.