

Contributions pratiques d'une géostatistique raisonnée en environnement : méthodes et application à la cartographie nationale de la pollution par l'ozone en France

Nicolas Jeannée,
GEOVARIANCES
49bis av. Franklin Roosevelt, BP 91 - 77212 AVON
Tel : 01.60.74.91.04 – Fax : 01.64.22.87.28,
mail : jeannee@geovariances.fr

Erwann Fangeat,
ADEME
27 rue Louis Vicat, 75015 PARIS
Tel : 01.47.65.20.82,
mail : erwann.fangeat@ademe.fr

Mohamedou Ba,
IFEN
61 Bd. Alexandre Martin, 45058 ORLEANS
Chemin de la Censive du Tertre BP 81227 – 44312 Nantes CEDEX 3
Tel : 02.38.79.78.57,
mail : mohamedou.ba@ifen.fr

Session de rattachement : comment faire parler vos données géographiques ?

Résumé :

L'utilisation de la géostatistique, comme outil d'aide à la décision lors de l'analyse de données spatialisées, s'est répandue ces dernières années dans le domaine de l'environnement grâce à sa capacité à :

- affiner la cartographie d'un phénomène (concentration en polluant par exemple) en prenant en compte de façon cohérente les variables auxiliaires (cofacteurs) qui lui sont liées,
- quantifier l'erreur associée à cette estimation – qui, en raison de la connaissance fragmentaire du phénomène, existe toujours, quelque soit la méthode d'estimation,
- fournir des outils d'analyse de risque : dépassement de seuils de pollution, exposition, ...

Les principes ainsi que la valeur ajoutée des outils géostatistiques - mais aussi le risque d'une utilisation « boîte noire » - sont illustrés à travers une cartographie au niveau national de la pollution par l'ozone. Priorité du programme CAFE (Clean Air For Europe) sur la préservation de la qualité de l'air en Europe, l'ozone fait partie des polluants surveillés en France par des Associations Agréées de Surveillance de la Qualité de l'Air (AASQA). Les données sont regroupées au sein de la base de données nationale sur la qualité de l'air (BDQA), gérée par l'ADEME. La géostatistique permet l'obtention d'une cartographie pertinente de ce polluant complexe, enrichie par la prise en compte de cofacteurs tels que l'altitude et l'occupation des sols issue de la base de données européenne CORINE land cover gérée par l'IFEN.

Mot-clés : géostatistique, cokrigeage, cofacteur, cartographie, surveillance qualité de l'air, pollution photochimique, ozone, réseau ATMO, exposition.

Introduction

La BDQA, Base de Données de la Qualité de l'Air, fait partie du système national d'information sur la qualité de l'air gérée par l'ADEME et est alimentée par les quarante associations agréées de surveillance de la qualité de l'air (AASQA) réparties sur l'ensemble du territoire français (métropolitain et DOM-TOM).

Constituant tout d'abord un lieu de regroupement d'informations validées, régulièrement actualisées, cette base est exploitée par différentes équipes. Du point de vue de son utilisation scientifique, elle est mise à disposition de chercheurs. Elle sert aussi de base de connaissance commune et permet la réalisation de rapports et d'études. Dans ce contexte, l'ADEME a sollicité Géovariances afin de réaliser une cartographie nationale de la pollution par l'ozone à l'aide des moyennes annuelles des stations couvrant le territoire (Jeannée, 2003).

La géostatistique est aujourd'hui régulièrement utilisée pour cartographier des pollutions atmosphériques au niveau local et régional, notamment au sein des AASQA (voir par exemple Chaxel & Chappaz, 2001, ou ASPA, 2002). L'objectif est de montrer l'intérêt des outils géostatistiques de cartographie et d'analyse de risque, à partir d'une étude de la pollution par l'ozone en France, à partir des concentrations moyennes annuelles (afin de ne pas alourdir le texte, nous sous-entendons fréquemment dans la suite « moyennes annuelles », pour ne parler que de concentrations) obtenues en 2001 sur les stations de mesures du réseau ATMO. Cette cartographie intégrera plusieurs cofacteurs corrélés à la pollution en ozone : classification des stations, altitude, type d'occupation du sol. La base de données Corine Land Cover, gérée par l'IFEN, a rendu possible l'utilisation des données d'altitude et d'occupation des sols connues sur une maille de 4 km couvrant l'ensemble du territoire.

L'ensemble des modélisations et estimations géostatistiques a été réalisé avec le logiciel Isatis, version 4.1, développé par Géovariances (GEOVARIANCES, 2002).

Analyse de données

Suite à l'analyse exploratoire des concentrations en ozone et de leurs corrélations avec les cofacteurs disponibles, plusieurs combinaisons de ces cofacteurs se sont avérées intéressantes :

- une analyse en composantes principales menée à partir de l'altitude moyenne et des postes Corine Land Cover a permis d'identifier une combinaison linéaire des cofacteurs, explicable physiquement, qui présente une corrélation non négligeable avec les concentrations en ozone,
- la combinaison linéaire correspondant au premier facteur obtenu par ACP a ensuite été modifiée de façon à optimiser la corrélation avec les concentrations en ozone.

La Figure 1 illustre à la fois la localisation des stations de fond du réseau ATMO, avec des croix proportionnelles à la concentration en ozone, et la combinaison linéaire des cofacteurs qui, en raison de sa corrélation avec les concentrations en ozone, est utilisée pour en améliorer l'estimation. Cette combinaison linéaire met en évidence plusieurs caractéristiques cohérentes avec les mécanismes de production / transformation de l'ozone :

- les valeurs de la combinaison ont tendance à augmenter avec l'altitude, tout comme les concentrations en ozone,
- on constate bien des valeurs fortes du cofacteur dans les zones de conifères, en particulier dans la région des Landes, ces derniers favorisant la production d'ozone,

- les zones urbaines, les zones industrielles et les réseaux routiers, qui ont tendance à générer des déficits en ozone, présentent bien des valeurs faibles du cofacteur.
- Afin d'éviter des résultats qui, bien qu'étant statistiquement défendables, seraient complètement irréalistes physiquement, nous discuterons l'importance de travailler à partir de tels cofacteurs qui combinent simplicité et pertinence d'un point de vue physico-chimique.

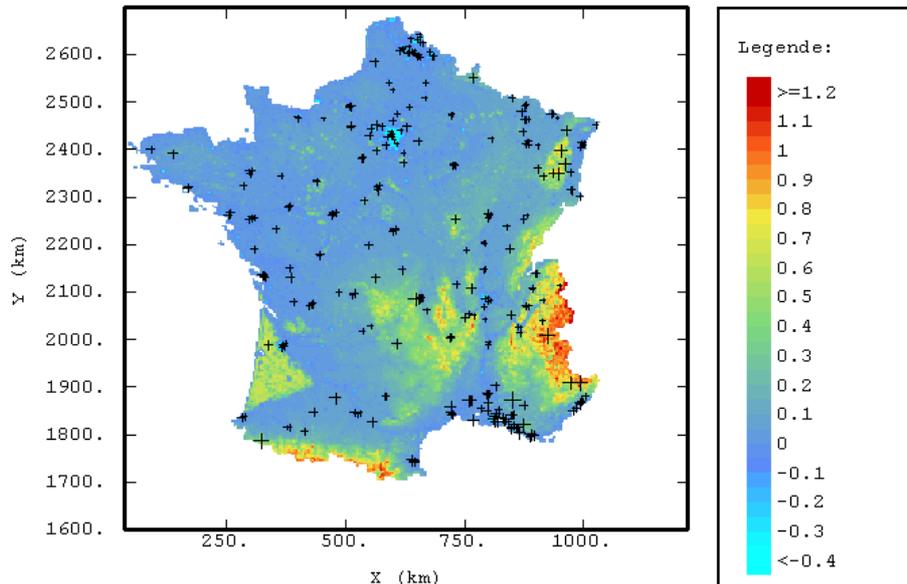


Figure 1 : Carte de la combinaison linéaire des cofacteurs, représentation proportionnelle des concentrations en ozone aux stations de fond du réseau ATMO.

Cartographie, incertitude et analyse de risque

Estimations monovariabiles

Plusieurs techniques de cartographie sont proposées, tout d'abord à partir des concentrations en ozone seules. Une **interpolation par inverse des distances quadratiques (IDQ)**, technique classique d'interpolation déterministe, servira de référence par rapport aux estimations géostatistiques.

Un **krigeage ordinaire** est ensuite réalisé grâce à une modélisation de la structure spatiale des concentrations en ozone. Tout comme pour les interpolations classiques, l'estimation de la concentration en ozone au point x_0 , notée \hat{z}_0 , est obtenue par combinaison linéaire des n concentrations aux points de mesure x_i :

$$\hat{z}_0 = \sum_{i=1}^n \lambda_i z(x_i) .$$

Le krigage se différencie uniquement dans le choix des coefficients λ_i . Appelés poids de krigage, ils dépendent :

- des distances entre les données et le point cible x_0 (tout comme les interpolateurs classiques),
- des distances séparant les données entre elles (clusterisation, existence de regroupements),
- de la structure spatiale du phénomène.

La configuration des données et la position du point cible étant connues, il suffit pour réaliser l'estimation de calculer expérimentalement la structure spatiale puis de la modéliser. C'est l'objectif de l'analyse variographique, qui constitue le cœur de l'analyse géostatistique. Le variogramme (voir Figure 2) représente l'évolution de la variabilité du phénomène en fonction de l'échelle d'observation. Nous discutons dans la présentation l'intérêt d'une modélisation de la structure spatiale qui soit cohérente avec le phénomène, et interprétable physiquement.

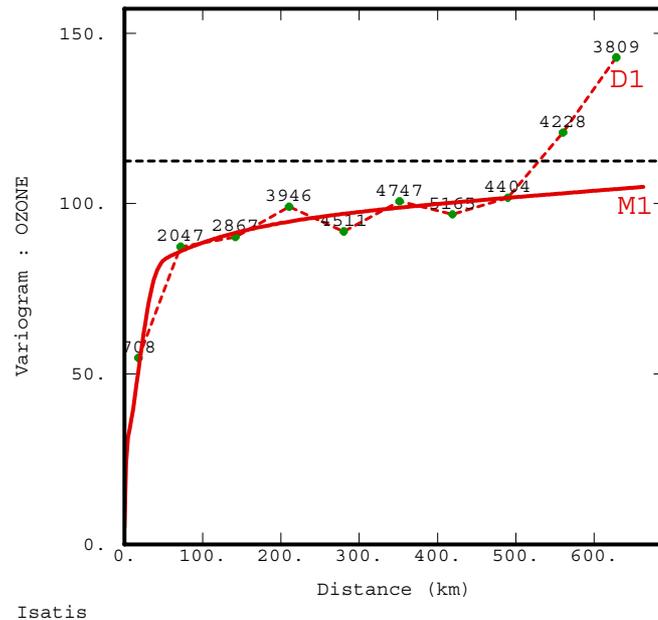


Figure 2 : Variogramme expérimental (courbe en tirets) et modèle de variogramme ajusté (courbe continue) pour la concentration en ozone.

La Figure 3 illustre le résultat de l'estimation des concentrations en ozone par krigeage ordinaire - toutes les estimations des concentrations en ozone sont données en $\mu\text{g}/\text{m}^3$. Cette carte laisse apparaître les grandes tendances des concentrations sur l'ensemble du territoire, mais son allure très lisse, due à un échantillonnage localement pauvre, en limite le réalisme, et donc l'intérêt.

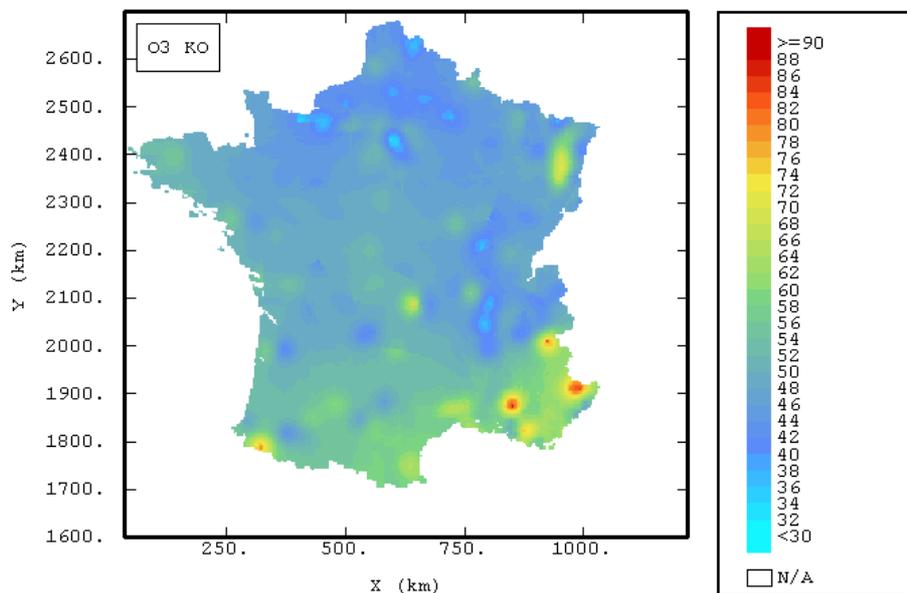


Figure 3 : Estimation de la concentration en ozone par krigeage ordinaire.

Un avantage essentiel du krigeage ordinaire (et des techniques géostatistiques présentées dans la suite de ce chapitre) par rapport aux interpolateurs classiques réside dans la quantification de l'incertitude associée à l'estimation (qui existe toujours), rendue possible par la modélisation de la structure spatiale. Cette incertitude est usuellement représentée par la carte d'écart-types (de l'erreur) de krigeage.

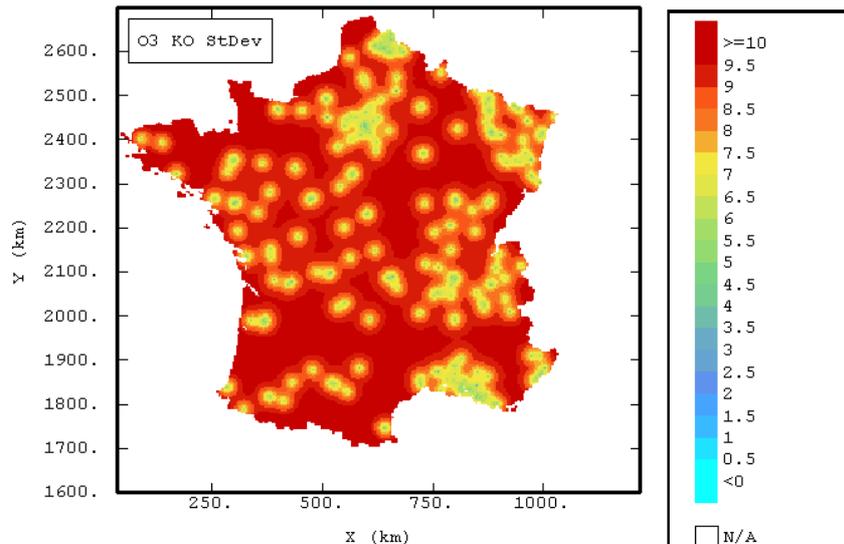


Figure 4 : Ecart-types de l'erreur de krigeage ordinaire de la concentration en ozone.

L'écart-type de krigeage :

- prend des valeurs minimales à proximité des points de données¹, la confiance que nous pouvons avoir en l'estimation étant alors grande,
- augmente lorsque l'on s'écarte des données ; cette augmentation est fonction de la structure spatiale choisie.

La carte d'écart-types de krigeage est un indicateur qualitatif de la qualité de l'estimation. De ce point de vue, la Figure 4 permet de bien identifier les zones mal reconnues, où la qualité de l'estimation est la moins bonne. L'utilisation quantitative des écart-types de krigeage comme un outil permettant d'obtenir un intervalle de confiance pour les concentrations en ozone nécessite des hypothèses supplémentaires et doit se faire prudemment.

Intégration de cofacteurs

Nous illustrons l'intégration des cofacteurs par un **cokrigeage colocalisé** des concentrations en ozone avec la combinaison linéaire des cofacteurs conduisant à la meilleure corrélation avec les concentrations en ozone. Supposons qu'en plus des valeurs de concentration en ozone aux stations de mesure z_{1i} , nous connaissions une variable auxiliaire z_{2j} en certains points potentiellement différents des stations de mesure. L'estimation $\hat{z}_1(x_0)$ par cokrigeage de la concentration en ozone au point x_0 en utilisant un cofacteur z_2 est obtenue par combinaison linéaire des n concentrations aux stations de mesure x_i et des p valeurs de cofacteur aux points x_j :

$$\hat{z}_1(x_0) = \sum_{i=1}^n \lambda_{1i} z_{1i} + \sum_{j=1}^p \lambda_{2j} z_{2j} .$$

Dans notre cas, la variable auxiliaire est connue en chaque nœud de la grille d'estimation. Prendre en compte l'ensemble des informations disponibles (291 points de mesure d'ozone, 34800 nœuds de grille) conduirait à résoudre un système d'équations très lourd, alors qu'il apparaît intuitivement inutile d'utiliser toutes les valeurs du cofacteur pour estimer la concentration en ozone en un nœud donné. Le cokrigeage colocalisé est une « simplification »

¹ L'écart-type de krigeage est même nul si l'estimation est réalisée exactement en un point de mesure : la confiance dans l'estimation est alors complète, vu que nous connaissons la valeur mesurée (en cas d'absence d'erreurs de mesure).

du cokrigage – rigoureuse sous certaines hypothèses – qui consiste à ne retenir dans le système de cokrigage que les points de mesure en ozone, les valeurs du cofacteur en ces points de mesure, plus la valeur du cofacteur au point où l'on procède à l'estimation (donc, $291 * 2 + 1$ données en voisinage unique, et en voisinage glissant seulement les données situées dans le voisinage).

La Figure 5 montre clairement l'influence du cofacteur sur l'estimation des concentrations en ozone, rendant cette dernière nettement plus réaliste que les résultats précédents : concentrations élevées en altitude (sauf au voisinage d'agglomérations telles que Grenoble), déficits en ozone au niveau des agglomérations, production d'ozone au niveau des Landes.

Cette estimation multivariable conduit à une diminution de l'incertitude associée à l'estimation, les écart-types obtenus étant inférieurs à ceux obtenus par krigeage ordinaire.

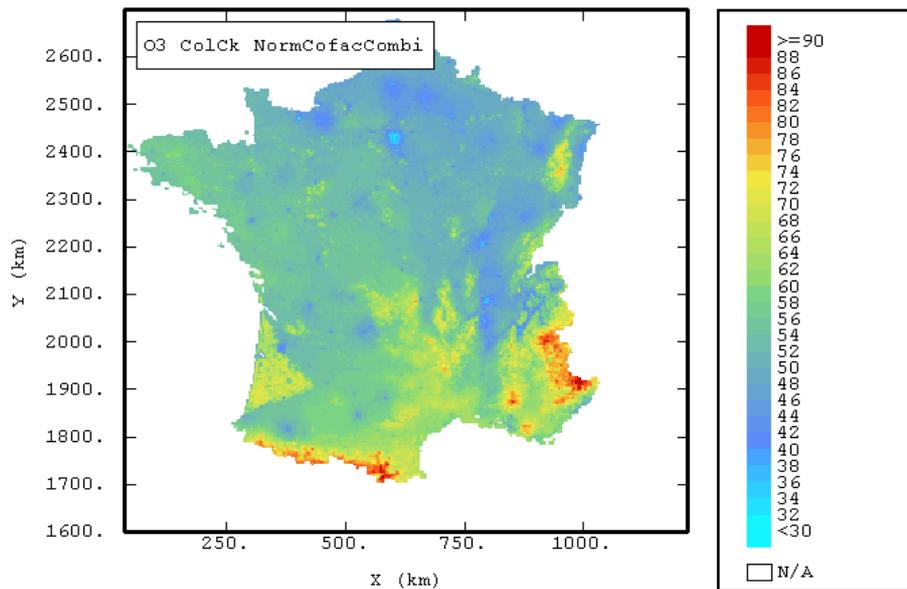


Figure 5 : Estimation de la concentration en ozone par cokrigage colocalisé avec la meilleure combinaison linéaire de l'altitude moyenne et des postes Corine Land Cover.

Face à des estimations obtenues par différentes méthodes, il peut être intéressant d'aider à l'évaluation qualitative du modèle de cartographie le plus approprié en comparant leur qualité prédictive. Cinq jeux de validation, contenant chacun approximativement 10 % des données, ont été choisis aléatoirement dans ce but. Pour chaque jeu de validation et chaque technique, l'évaluation a porté sur le calcul des erreurs quadratiques moyennes entre concentrations vraies (aux points de validation, non utilisés) et estimées, et l'observation des nuages de corrélation correspondants. A titre d'illustration, la Figure 6 donne les erreurs quadratiques moyennes obtenues ; elles montrent la valeur ajoutée tout d'abord de l'estimation géostatistique par rapport à une estimation déterministe, mais surtout de la prise en compte rigoureuse de cofacteurs liés aux concentrations en ozone, avec une diminution de l'erreur allant jusqu'à 50 %.

Dépassement de seuils de pollution

La méthode la plus pertinente pourra ensuite être utilisée dans une optique d'analyse de risque. La présentation illustre comment la géostatistique permet l'obtention de cartes de probabilités de dépassement d'un seuil de pollution en ozone, ainsi que l'optimisation du réseau d'échantillonnage.

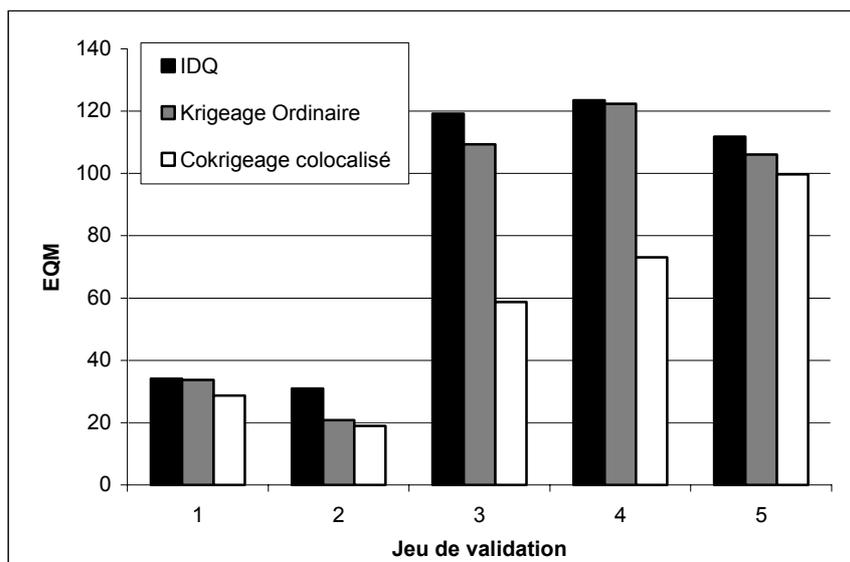


Figure 6 : Erreurs Quadratiques Moyennes (EQM) obtenues pour chaque méthode d'estimation sur les différents jeux de validation.

Conclusion

Le recours à des outils de modélisation et d'estimation automatiques (boîtes noires) est selon nous à proscrire. En effet, les résultats peuvent s'avérer désastreux, notamment en raison de l'absence de tout moyen de contrôle sur le modèle construit. Il faut accepter que toute estimation présente un risque inévitable d'erreur radicale dans les zones non échantillonnées. Il importe donc de tenir compte, non seulement des données numériques, mais également de toutes les autres sources d'information dont on dispose : connaissances générales sur le phénomène, expérience des praticiens, (Matheron, 1978). Toute estimation doit être raisonnée : elle doit, autant que faire se peut, chercher à assurer que le résultat sera physiquement plausible et réaliste.

Lorsqu'ils sont mis en œuvre dans cet esprit, les outils géostatistiques apportent une valeur ajoutée indéniable pour la cartographie et l'analyse de risque de variables environnementales. Cette valeur ajoutée réside dans :

- l'utilisation de la structure spatiale intrinsèque du phénomène pour son estimation,
- l'intégration rigoureuse de variables auxiliaires liées au phénomène d'intérêt, en améliorant ainsi l'estimation,
- la quantification de l'erreur associée à chaque estimation.

Remerciements

L'étude présentée fait l'objet d'un financement par l'ADEME, contrat n° 02 62 031.

Bibliographie

ASPAS, 2002, *Diagnostic de la Qualité de l'Air sur l'Agglomération de Mulhouse - Annexe au rapport final sur la répartition spatiale de la pollution atmosphérique*. ASPA 02031901-I-D, ASPA.

CHAXEL E., CHAPPAZ C., 2001, *Mise en place d'une méthodologie pour la cartographie de l'ozone à l'échelle du département de l'Isère*. ASCOPARG.

GEOVARIANCES, 2002. *ISATIS Software Manual, 4th Edition*. Géovariations Fontainebleau, 585 pp.

JEANNEE N., 2003, *Cartographie nationale de la pollution par l'ozone en France*. Rapport d'avancement du contrat ADEME n° 02 62 031.

MATHERON G., 1978, *Estimer et choisir*. Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fasc. 7, Ecole des Mines de Paris.