

ARTICLES

L'intégration des informations indirectes à la cartographie géostatistique des polluants

Integrating indirect information when mapping pollutants

Michel BOBBIA*, Valérie PERNELET**, Chris ROTH**

Résumé

La demande d'information du public ainsi que la législation existante impliquent que les organismes surveillant la qualité de l'air soient en mesure de produire rapidement des cartes précises des polluants. Les techniques classiques de cartographie ne suffisent plus. Les cartes qui en résultent sont souvent irréalistes et le choix de la méthode retenue pour les établir est fondé sur des critères subjectifs. L'algorithme géostatistique de cartographie est mieux adapté au problème car il permet de prendre en compte le comportement spatial spécifique du polluant grâce à une fonction de corrélation spatiale calculée à partir des concentrations observées aux stations. De plus, la corrélation entre le polluant et les variables auxiliaires (comme les émissions ou le trafic) peut être quantifiée et utilisée afin d'obtenir une carte plus représentative de la réalité. Après un bref rappel des concepts de base de la géostatistique, les deux applications pratiques présentées permettent d'apprécier l'intérêt de la géostatistique. Cette méthode est bien adaptée au problème de la surveillance de la qualité de l'air car elle permet d'obtenir des cartes représentatives du comportement spatial spécifique des polluants traités.

Abstract

Being able to provide a quick and accurate pollutant map from readings at isolated measurement stations is becoming more important today in light of the European norms on air quality and the public's demand to be informed. Commonly used algorithms, like spline or inverse distance based techniques, are quick but their accuracy remains to be determined for two reasons. Firstly, the choice of method is arbitrary. The method retained will be the one that produces the most satisfactory result according to the user's subjective perception. Secondly, none of these methods can take relevant auxiliary information (like meteorological data, traffic density or industrial emissions) into account, thus limiting the realism of the final map. The geostatistical mapping algorithm, called kriging, provides an answer to both these problems. This algorithm is based on the specific spatial behaviour of the mapped pollutant via a spatial correlation function calculated from the sample measurements. Because the form of this correlation function determines the shape of the final map, the geostatistical algorithm adapts itself directly to the spatial characteristics of the data, thus eliminating the need for a subjective evaluation. Doubly important, auxiliary information can also be incorporated in the final map by integrating any relationship (linear or non-linear) with the mapped pollutant into the kriging algorithm. The paper begins with a presentation of the concepts behind the geostatistical mapping algorithm. Two practical applications, mapping nitrogen dioxide over Rouen and benzene over Paris, are used to highlight the capacity of the geostatistical approach to produce quick and realistic maps that take traffic information and emissions into account when monitoring air quality in real time.

* AIR NORMAND, 21 avenue de la Porte des Champs, 76000 Rouen.

** AIRPARIF, 7 rue Crillon, 75004 Paris.

Introduction

Aujourd'hui, la législation existante, le besoin d'affinement des connaissances et la demande du public imposent aux organismes responsables de la surveillance la qualité de l'air de produire des cartes de polluants à partir des observations aux stations de mesures isolées. L'information apportée par une carte est plus complète puisqu'elle couvre tout l'espace. La mesure de l'exposition des populations aux polluants atmosphériques s'en voit améliorée ; en terme de santé publique, les résultats sont plus réalistes. L'effet visuel est enrichi par l'émergence de zones de meilleure ou moins bonne qualité de l'air. L'effort de compréhension est largement facilité par des couleurs plutôt que des valeurs. Par ailleurs, il faut pouvoir produire ces cartes dans un délai acceptable, ce qui n'est pas forcément le cas d'une modélisation numérique des processus physico-chimiques. Un algorithme de cartographie aussi précis que possible mais moins lourd à mettre en œuvre apparaît donc nécessaire.

Pour cartographier le polluant il faut d'abord interpoler entre les stations de mesures, aux nœuds d'une grille régulière qui couvre le champ d'étude. Les valeurs interpolées sont ensuite présentées graphiquement sous forme d'une carte. Interpoler entre les stations de mesure signifie affecter une concentration estimée à chaque nœud de la grille. Pour un nœud donné, cette concentration obtenue est une moyenne pondérée des concentrations observées aux stations de mesure voisines du nœud. Il existe cependant plusieurs algorithmes de pondération possibles. (Voir par exemple : Arnaud et Emery [1] ou Ionescu [2] pour une présentation du problème d'interpolation). Chaque algorithme mène à une concentration estimée différente et donc à une carte du polluant différente. Le choix de l'algorithme utilisé doit donc être dicté par sa capacité à combiner les données expérimentales d'une façon aussi cohérente que possible. Ce choix ne peut être fondé que sur la compréhension des différents algorithmes.

L'objectif du présent article n'est pas de présenter formellement les différentes méthodes. En ce sens, aucune formule mathématique n'est fournie – la littérature est suffisamment abondante sur ce thème – et nous avons fait le choix de vulgariser complètement les théories des méthodes abordées tout au long de l'article. Nous nous attachons donc à montrer que différentes techniques d'interpolation conduisent à des résultats différents, le but ultime étant d'améliorer les cartes finales par l'intégration d'informations supplémentaires aux mesures. Seul le cadre géostatistique permet cela. C'est la comparaison des différentes méthodes qui est en jeu ici, et non pas la présentation détaillée de celles-ci ; notamment toutes les étapes d'ajustement des modèles (optimisation, variographie...) ne sont pas abordées.

Quelques techniques classiques d'interpolation sont d'abord présentées afin d'apprécier leurs limita-

tions pratiques. On se limite aux techniques fondées sur la pondération des concentrations observées afin de comparer ces méthodes à l'algorithme géostatistique, le krigeage [3], qui applique également une règle de pondération aux concentrations observées. Les interpolateurs fondés sur l'ajustement d'une fonction analytique à travers des concentrations observées comme la triangulation, les splines ou les polynômes [1, 2], ne sont pas considérés ici. Ces méthodes ne sont pas réellement applicables lorsque le polluant présente des fluctuations spatiales trop importantes à petite échelle pour être modélisé analytiquement d'une façon réaliste.

Les idées génératrices du krigeage sont ensuite présentées pour montrer que, de par sa définition, cette méthode s'adapte aux caractéristiques spatiales spécifiques du polluant considéré pour utiliser d'une façon optimale des concentrations observées aux sites de mesure lors de la production de la carte. L'application du krigeage à la pollution de l'air se fait depuis plusieurs années (par exemple : [4-6]), mais les capacités totales de la géostatistique sont utilisées depuis peu en cartographie. L'interpolation géostatistique permet de prendre en compte des relations éventuelles entre le polluant et d'autres variables dites auxiliaires, comme par exemple pour la pollution urbaine la densité du trafic, les conditions météorologiques ou la texture urbaine. Figuiera et al. [7] cartographient la déposition du chlore provenant du sel de mer en utilisant la distance jusqu'à la mer comme variable auxiliaire. Nunes et al. [8] cartographient le dioxyde de soufre en utilisant le résultat d'un modèle déterministe très simple comme variable auxiliaire. Dans cet article, deux nouvelles applications sont présentées : la cartographie de la pollution urbaine en prenant en compte sa relation avec les facteurs d'urbanisation d'une part via la population, et d'autre part via la texture urbaine.

Les études sont entièrement réalisées à l'aide du logiciel ISATIS.

L'interpolation par des techniques classiques

Afin de comprendre les limites des méthodes d'interpolations classiques, trois d'entre elles sont présentées. La méthode polygonale (autrement appelée la méthode du plus proche voisin) implique que la donnée la plus près du nœud reçoit tout le poids et les autres données aucun poids. Cette méthode est trop simpliste pour fournir une carte satisfaisante. Elle n'apporte pas vraiment plus d'informations que celles fournies par les mesures elles-mêmes. La méthode de la moyenne mobile affecte le même poids à toutes les données voisines du nœud à estimer. Il faut cependant définir le nombre de données voisines à utiliser pour interpoler, un choix qui change automatiquement la carte résultante. Selon la méthode de l'inverse des distances, le poids affecté aux données voisines est proportionnel à l'inverse de la distance entre le nœud et la donnée

considérée. Logiquement, plus la donnée est proche du nœud, plus elle reçoit un poids important. Cependant on pourrait également affecter un poids qui est proportionnel à l'inverse des carrés des distances afin de privilégier encore davantage les données les plus proches du nœud. Ceci devient un choix à définir par l'utilisateur, avec sa part d'arbitraire. Couramment utilisée par ailleurs, cette méthode a pour effet dans notre cas de considérer les points de mesure comme des « puits » ou des « sources » de pollution.

Pour illustrer l'influence de la méthodologie sur la carte résultante, la figure 1 présente trois cartes possibles obtenues à l'aide des méthodes classiques. Les données utilisées sont des concentrations en benzène (C_6H_6) obtenues par 71 tubes à diffusion en

Ile-de-France [9]. La localisation des tubes est représentée sur les cartes par un point et les limites des départements franciliens par une ligne noire. Le champ d'étude, la région Ile-de-France couvre approximativement $165 \text{ km} \times 130 \text{ km}$ autour de Paris. La forme des cartes reflète la méthode utilisée qui ne dépend que de la localisation des stations de mesures par rapport au nœud considéré. Tandis que l'utilisateur retient la carte qui lui paraît la plus représentative de la réalité, ce choix est nécessairement fondé sur des critères subjectifs, c'est-à-dire les idées a priori de l'utilisateur. Afin de toujours obtenir la carte la plus représentative, il faut s'affranchir du choix de la carte qui est basé sur des critères visuels subjectifs. La méthode utilisée doit alors prendre en compte les caractéristiques spatiales spécifiques du polluant.

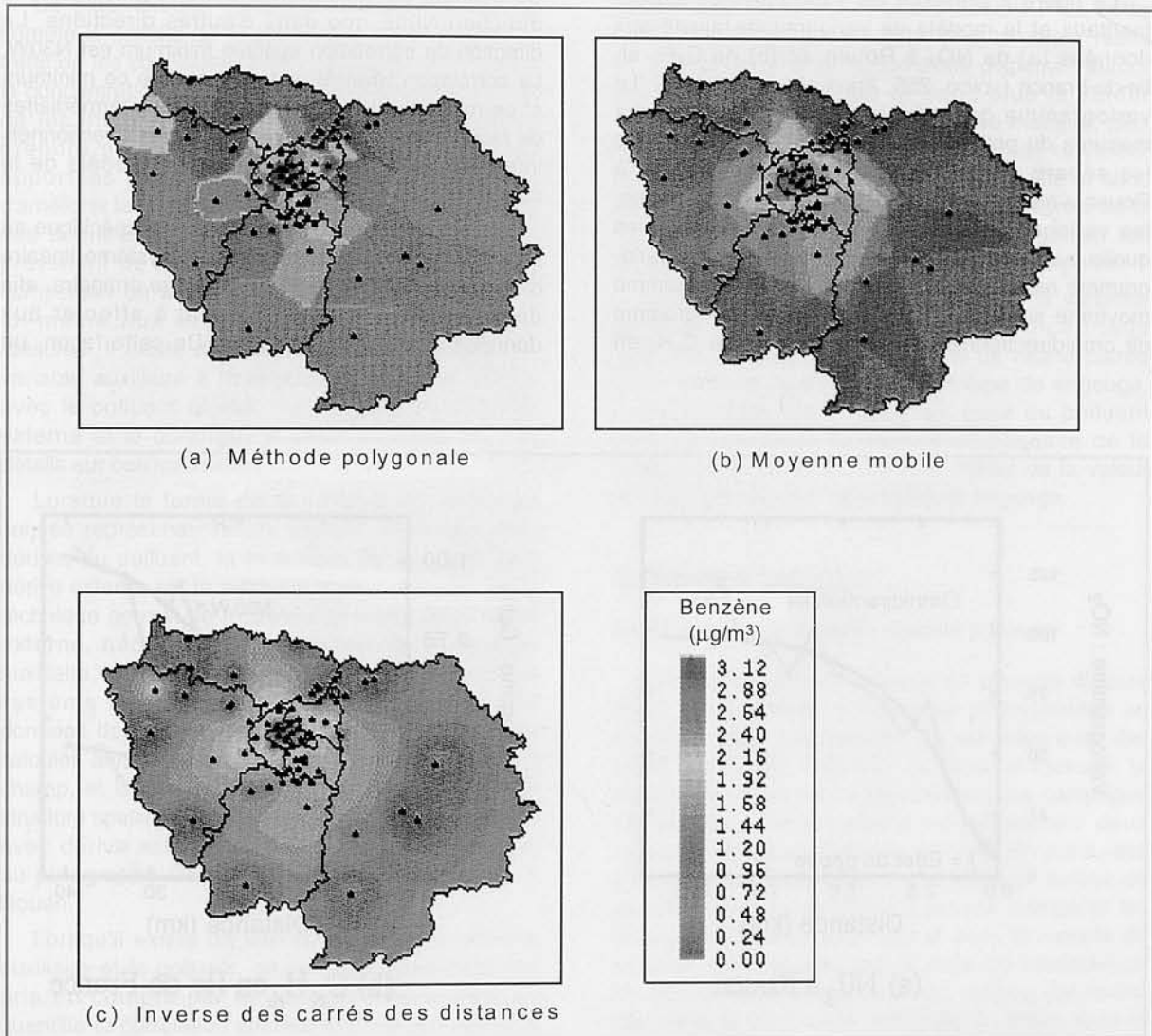


Figure 1.

Les cartes obtenues par (a) la méthode polygonale, (b) la moyenne mobile des cinq données les plus proches, et (c) l'inverse des carrés des distances.

The maps obtained by (a) the polygonal method, (b) a moving average of the 5 nearest points, and (c) the inverse distance squared method.

L'approche géostatistique

Interpolation par krigeage ordinaire

Selon l'algorithme d'interpolation géostatistique, le krigeage, la règle de pondération et donc la carte qui en résulte sont directement déterminées par le comportement spatial des données du polluant. Ledit variogramme [10, 11], calculé à partir des valeurs observées aux stations de mesure, permet de quantifier la continuité spatiale du polluant. La forme du variogramme peut varier avec la direction considérée, par exemple, lorsque les conditions météorologiques régissent la diffusion du polluant. Une fonction mathématique est ajustée aux variogrammes directionnels afin d'obtenir un modèle de variogramme qui caractérise la variabilité spatiale pour toute distance et toute direction dans le plan à 2D.

La figure 2 présente les variogrammes expérimentaux et le modèle de variogramme ajusté aux données (a) de NO₂ à Rouen, et (b) de C₆H₆ en Ile-de-France (voir p. 255, Applications pratiques). Le variogramme quantifie la variabilité entre deux mesures du polluant en fonction de la distance qui les sépare. L'analyse variographique du NO₂ à Rouen a montré que ce polluant est en effet isotrope, les variogrammes directionnels étant semblables quelle que soit la direction considérée. Ainsi le variogramme retenu pour le krigeage est le variogramme moyenné sur toutes les directions, le variogramme dit omnidirectionnel. Le comportement du C₆H₆ en

Ile-de-France varie selon la direction considérée. On présente ici la direction de variabilité maximale (N30W) et celle de variabilité minimale (N60E). Le modèle de variogramme, en ajustant tous les variogrammes directionnels, reflète le comportement spatial du polluant quelle soit la direction dans le plan à 2D.

Le variogramme peut être interprété en termes de continuité ou de corrélation spatiale si on l'inverse graphiquement pour obtenir une fonction décroissante : plus la distance entre deux mesures est importante moins les mesures sont corrélées. Le polluant est non corrélé au-delà de la distance où le modèle de variogramme se stabilise au niveau d'un palier. De cette façon, pour toute direction, le NO₂ à Rouen est corrélé dans l'espace jusqu'à une distance d'environ 8 km. La corrélation entre deux mesures de NO₂ séparées de plus de 8 km est nulle. Le comportement du C₆H₆ en Ile-de-France est plus complexe. Ce polluant est plus corrélé dans l'espace dans la direction N60E que dans d'autres directions. La direction de corrélation spatiale minimum est N30W. La corrélation spatiale varie donc entre ce minimum et ce maximum le long des directions intermédiaires de façon à respecter les variogrammes directionnels intermédiaires. Ainsi on obtient un modèle de la corrélation ou continuité spatiale à 2D.

Le modèle de variogramme, qui est spécifique au polluant, est ensuite entré dans un système linéaire d'équations, ledit système de krigeage ordinaire, afin de déterminer le poids optimal à affecter aux données lors de l'interpolation. De cette façon, un

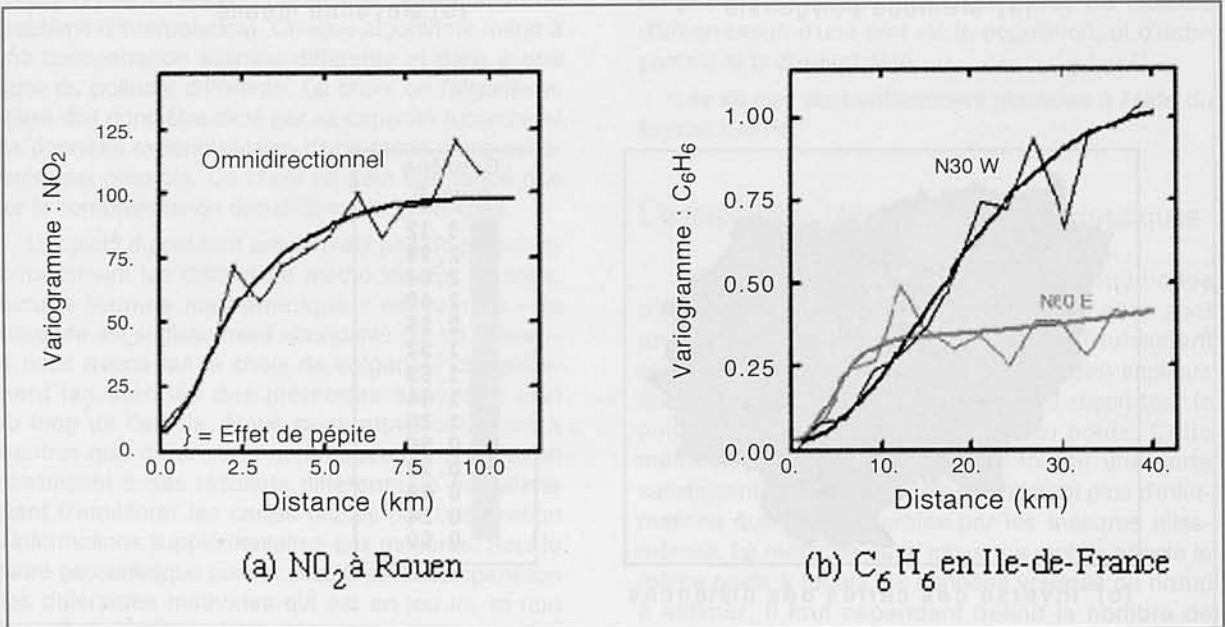


Figure 2.
Les variogrammes expérimentaux (ligne mince) et les modèles de variogramme (ligne épaisse) ajustés aux données (a) de NO₂ à Rouen, et (b) de C₆H₆ en Ile-de-France.
The experimental variograms (thin line) and the variogram model (thick line) fitted to (a) NO₂ data from Rouen, and (b) C₆H₆ data from Ile-de-France.

poids plus grand est affecté aux données qui sont mieux corrélées avec la concentration au nœud considéré. Parallèlement, un poids moins important est affecté aux données qui sont moins corrélées avec la concentration au nœud. Donc l'interpolation géostatistique prend en compte non pas la distance entre les stations de mesures et le nœud mais la corrélation spatiale du polluant entre ces stations et la concentration au nœud considéré. De cette façon le krigeage s'adapte aux caractéristiques spatiales du polluant via le modèle de variogramme. La valeur estimée en un nœud n'est plus le résultat d'une fonction mathématique choisie arbitrairement (polygone, moyenne, fonction inversement proportionnelle à une distance...), mais la valeur la plus probable comme résultant d'un modèle statistique adapté et optimisé pour les données expérimentales.

Intégration des informations indirectes

De plus, lorsque certaines variables auxiliaires, comme la densité de la population, les émissions du trafic ou les conditions météorologiques, sont liées à la concentration du polluant, la géostatistique permet de les intégrer dans l'interpolation. L'intérêt de prendre en compte les informations indirectes apportées par la ou les variables externes est d'améliorer la carte surtout dans les zones où la densité des données est la plus faible. Ainsi l'information provenant de la variable auxiliaire permettrait de compenser un manque d'information sur le polluant lui-même aux endroits non informés par des mesures. Il existe deux méthodes pour intégrer une variable auxiliaire à l'interpolation selon sa relation avec le polluant étudié : le krigeage avec dérive externe et le co-krigeage (*voir* [12] pour plus de détails sur ces méthodes).

Lorsque la forme de la variable auxiliaire est censée représenter l'allure globale, autrement dit la dérive, du polluant, la technique du krigeage avec dérive externe est la méthode la plus adaptée. Cette technique permet de modéliser la forme de la dérive externe, nécessairement fondée sur la variable auxiliaire, et d'incorporer cette dérive externe dans le système de krigeage. Les poids affectés aux données de polluant lors de l'interpolation sont donc calculés afin de respecter l'allure de la dérive sur le champ, et la carte qui en résulte reproduit mieux la structure spatiale du polluant. L'application du krigeage avec dérive externe au NO_2 à Rouen est décrite au paragraphe Cartographie du dioxyde d'azote à Rouen.

Lorsqu'il existe un lien direct entre la variable auxiliaire et le polluant, ce lien peut également être pris en compte par le variogramme croisé qui quantifie la corrélation spatiale entre le polluant et la variable auxiliaire. La technique du co-krigeage prend en compte le comportement spatial du polluant et celui de la variable auxiliaire ainsi que la corrélation croisée entre eux afin d'estimer le polluant à partir des mesures de concentration et des données de la

variable auxiliaire. Pour mettre en œuvre le co-krigeage il faut donc modéliser les variogrammes simples et croisés entre la variable auxiliaire et le polluant. Ces trois fonctions de corrélation spatiale sont ensuite entrées dans le système de co-krigeage, afin de déterminer le poids optimal à affecter aux données lors de l'interpolation. Ainsi on affecte un poids aux données du polluant et aux données de la variable auxiliaire lorsqu'on estime le polluant en chaque nœud de la grille. L'application du co-krigeage à C_6H_6 en Ile-de-France est décrite p. 258, au paragraphe Cartographie du benzène hivernal à Paris.

Quantification de l'incertitude

Le cadre probabiliste de la géostatistique permet de quantifier l'incertitude associée à la valeur interpolée à l'aide de la variance de krigeage [12]. Cette variance représente la variabilité ou la dispersion possible de l'erreur de l'interpolation. Plus cette variance de krigeage est forte, plus la valeur réelle du polluant en un point risque d'être dispersée autour de la valeur interpolée et donc plus la valeur interpolée est incertaine. De la même manière, plus la variance de krigeage est faible plus la valeur interpolée est en moyenne près de la réalité et donc plus la carte est précise. Ainsi les fortes valeurs de la variance de krigeage permettent de localiser les zones sous-échantillonnées sur la carte interpolée.

Cependant au lieu de travailler avec la variance de krigeage on exprime souvent l'incertitude associée à l'interpolation en terme de racine carrée de la variance, c'est-à-dire l'écart-type de krigeage. Cette variable, dont l'unité est celle du polluant même, représente la dispersion possible de la concentration réelle et inconnue autour de la valeur obtenue par une des techniques de krigeage.

Applications pratiques

Cartographie du dioxyde d'azote à Rouen

Une campagne de mesure du dioxyde d'azote (NO_2) a été réalisée à Rouen du 20 septembre au 4 octobre 1999. Les mesures ont été faites avec des tubes à diffusion localisés de façon à mesurer la pollution de fond sur l'agglomération. La campagne fournit des concentrations moyennes sur deux semaines pour une centaine de sites. En outre, des séries de trois tubes ont été disposées autour de certains capteurs de façon à pouvoir comparer les mesures des tubes entre eux et avec la mesure du capteur. Ceci permet par la suite de caractériser l'erreur de mesure associée aux valeurs provenant des tubes et de l'inclure dans l'étude. Enfin, dans le but d'améliorer la carte obtenue à partir des échantillons de dioxyde d'azote, la densité de population a été prise en considération comme variable auxiliaire. A priori la concentration en NO_2 n'est pas directement liée à la densité de population, mais il pourrait

y avoir une relation indirecte en appliquant le raisonnement suivant : plus la population est dense, plus il y a de voitures et donc plus la concentration en NO_2 risque d'être importante. Concernant les émissions routières, cette information a été écartée pour deux raisons : les émissions de NO_2 industrielles sont légèrement majoritaires par rapport à celles du trafic routier sur l'agglomération rouennaise, et ces émissions sont représentatives d'une pollution de proximité qui ne devient une pollution de fond qu'après diffusion. Il faudrait disposer d'un cadastre d'émissions spatialisé et suffisamment complet et précis, et de la météo intervenant entre émissions et immissions, y appliquer une diffusion physico-chimique... ce qui sort du cadre de cet exposé.

Les données des tubes et celles des capteurs sont entachées d'erreur. Afin d'étudier l'incertitude associée aux données des tubes à diffusion, des triplets de tubes sont disposés à très petite distance (quelques centimètres) autour d'un capteur. Cette configuration permet d'abord de comparer les données des différents tubes entre elles et ensuite de comparer ces valeurs à la valeur du capteur correspondant. Ainsi, nous avons pu quantifier l'incertitude, définie en terme d'une *variance de l'erreur de mesure*, associée aux données des tubes afin d'incorporer celle-ci au stade de l'analyse variographique et de la cartographie par krigeage. Le krigeage ordinaire se fait en prenant en compte la variance de l'erreur de mesure. Cette variance de l'erreur de mesure est intégrée dans le modèle de variogramme ajusté aux données expérimentales comme un *effet de pépité* (voir Figure 2, p. 254).

Pour intégrer la variable auxiliaire dans la cartographie, nous construisons une fonction de la population qui indique la tendance générale, dite *dérive externe* (voir par exemple [13]), de la concentration en NO_2 et qui soit représentative du comportement global du polluant (voir Figure 3 (a) ; cette figure permet en outre de se représenter la forme de l'agglomération rouennaise). La relation entre la concentration en NO_2 et la dérive externe est montrée dans la figure 3 (b).

Afin d'évaluer l'influence de la dérive, les cartes obtenues par krigeage sans et avec prise en compte de la dérive externe sont présentées respectivement dans la figure 4 ci-contre. L'intégration de la dérive au NO_2 améliore nettement l'estimation. La carte est plus détaillée, et les zones en extrapolation ou à faible densité de points de mesure sont mieux renseignées.

Les écarts-types de krigeage [voir Figures 4 (c) et 4 (d)] permettent de quantifier l'incertitude associée à chaque carte en fonction de la méthode d'estimation. Essentiellement, l'écart-type de krigeage ne dépend que de la forme du variogramme et de la géométrie des données ; plus on s'éloigne d'un point de mesure, plus l'incertitude associée à la carte augmente. En krigeage avec dérive externe, en revanche, l'écart-type dépend également de la forme de la dérive ; moins cette dérive reflète le comportement local des données de NO_2 , plus l'incertitude associée à la carte est forte. Ceci est mis en évidence vers le nord-est du champ où la carte de l'écart-type reflète la différence entre la valeur modérée de NO_2 capturée par les tubes et la forte densité de population dans cette partie de Rouen.

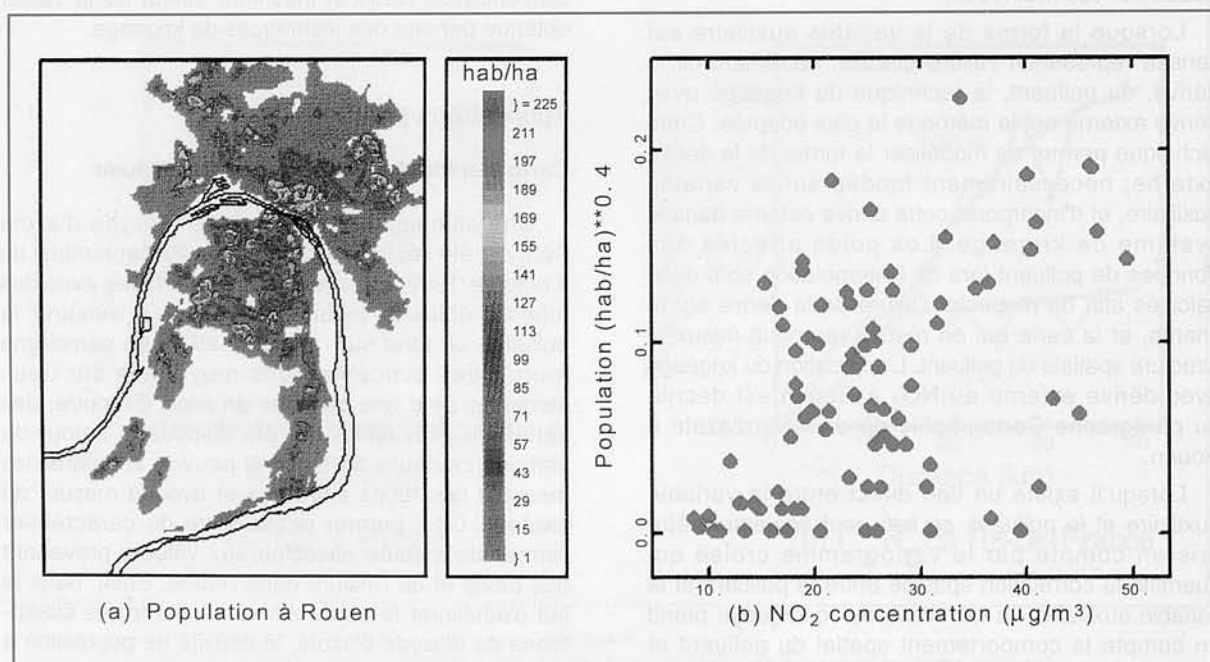


Figure 3. Relation entre la population et la concentration en NO_2 aux sites de mesures.
The relationship between the population and the NO_2 concentration at the sampling points.

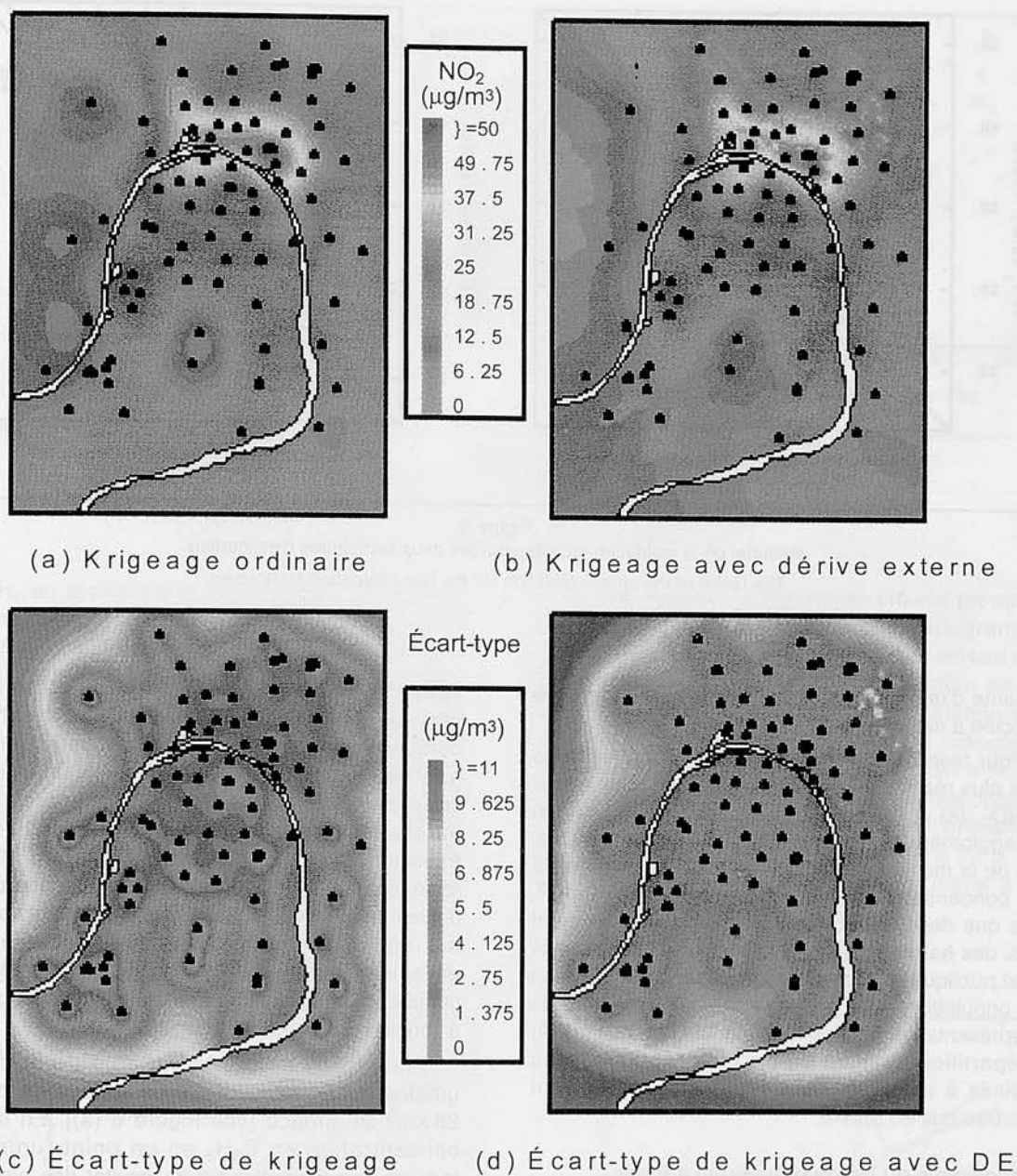


Figure 4.
Cartes du NO₂ krigé (a) à partir des données et (b) en y intégrant la densité de population, accompagnées de leur écart-type de krigage.

The kriged NO₂ maps obtained from (a) only the NO₂ samples, and (b) after integrating the population density, together with the respective kriging standard deviation maps (c) and (d).

Outre la variance de krigage, la validation croisée permet de juger de la fiabilité d'une carte par rapport à une autre. Elle consiste à estimer un point dont on connaît la valeur en l'éliminant des données d'entrée. En faisant cela successivement sur tous les sites de mesure, on obtient un nuage de points (valeurs vraies contre valeurs estimées) qui est d'autant plus proche de la bissectrice que la carte est bien estimée. La figure 5 p. 258, présente les

résultats comparatifs de la validation croisée pour les deux estimations (avec et sans dérive externe).

Les résultats de la validation croisée montrent que la prise en compte de la dérive, c'est-à-dire des informations indirectes, améliore la précision de la carte obtenue à l'intérieur de l'enveloppe convexe des sites de mesure. Dans les zones en extrapolation, l'écart-type du krigage avec dérive externe

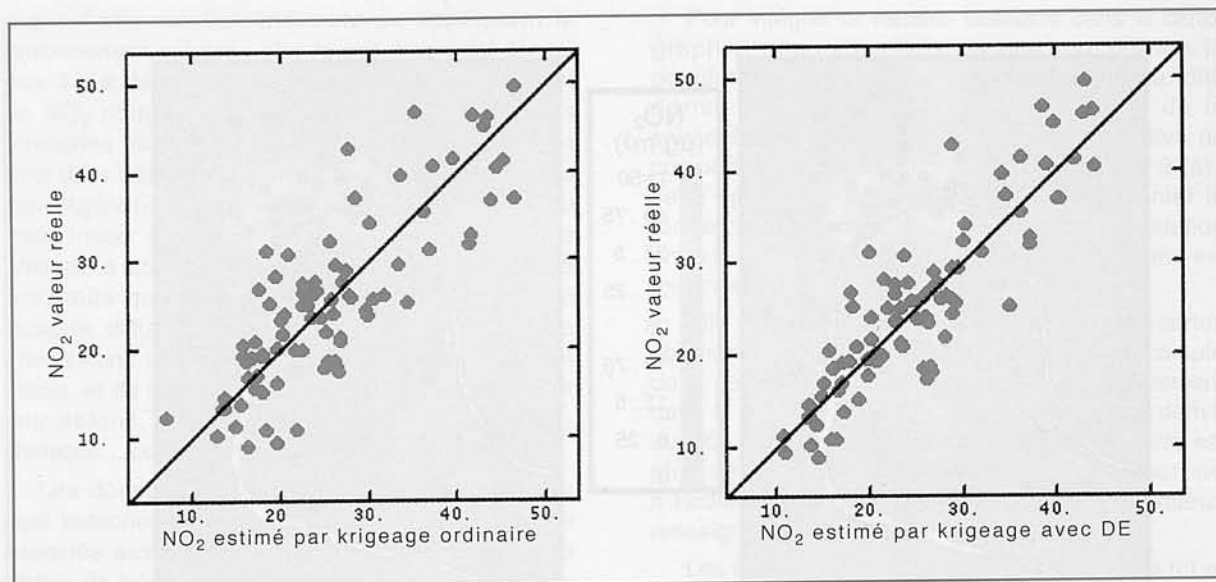


Figure 5.
Résultat de la validation croisée pour les deux techniques d'estimation.
The result of the cross validation for the two estimation techniques.

présente d'une façon plus réaliste la forte incertitude associée à ces zones.

Pour bien comprendre la nécessité d'obtenir une carte plus réaliste, nous avons croisé les deux cartes du NO_2 , (a) et (b) de la figure 4, avec la population de l'agglomération rouennaise. Dans le premier cas, plus de la moitié des habitants (51 %) est exposée à une concentration moyenne inférieure à $25 \mu\text{g}/\text{m}^3$, alors que dans le deuxième cas, ce sont seulement 40 % des habitants qui y sont exposés. En terme de santé publique, il apparaît clairement que l'exposition des populations à la pollution dépend largement de la représentation plus ou moins fidèle que l'on fait de la répartition spatiale du polluant. Tous moyens destinés à améliorer cette représentation doivent donc être mis en œuvre.

Cartographie du benzène hivernal à Paris

Il s'agit d'une étude cartographique des données de benzène (C_6H_6) en Ile-de-France. Les données proviennent de trois campagnes consécutives de mesures par tube à diffusion [9] du 4 janvier au 16 février 1999 sur 71 sites localisés pour mesurer la pollution de fond. Chaque campagne fournit des concentrations moyennes de C_6H_6 sur la période d'exposition de deux semaines. Nous cartographions la concentration moyenne des trois campagnes consécutives, c'est-à-dire la concentration moyenne de C_6H_6 sur la période de six semaines.

Une première approche consiste à utiliser le krigeage ordinaire pour obtenir une carte de C_6H_6 à partir des seules mesures fournies par les 71 tubes C_6H_6 . Nous étudions ensuite l'utilisation d'une variable auxiliaire afin d'améliorer cette première carte. Nous retenons la densité urbaine comme

variable auxiliaire. La décision de retenir cette variable est justifiée tant elle s'inscrit dans une certaine logique. Certes la concentration en C_6H_6 n'est pas directement liée à la densité urbaine, mais cette dernière intègre de manière indirecte, tout un ensemble de variables explicatives liées aux émissions de C_6H_6 dans l'air ambiant à l'échelle régionale. Ainsi, les variations de la densité urbaine doivent-elles être logiquement positivement corrélées aux variations de la densité de population, des émissions surfaciques (comme, par exemple, le chauffage urbain) et des émissions du trafic, qui sont à l'origine des niveaux de fond de C_6H_6 .

Plus précisément, la variable de la densité urbaine est le pourcentage de surface de bâti par 25 km^2 de surface [voir Figure 6 (a)]. En effet, la concentration en C_6H_6 en un point (un site de mesure, par exemple) est le résultat des processus de transport et de diffusion atmosphériques des émissions de C_6H_6 situées non seulement au point considéré mais dans un voisinage de quelques kilomètres autour de celui-ci. La figure 6 (a) montre également le champ d'étude, la région Ile-de-France qui couvre approximativement $165 \text{ km} \times 130 \text{ km}$ autour de Paris. Les limites des départements franciliens sont représentées en noir, et la localisation des tubes par un point.

Le nuage de corrélation obtenu entre les concentrations de C_6H_6 et la densité urbaine (voir Figure 6 (b)) montre une relation linéaire entre ces deux variables. La corrélation positive implique que globalement plus la densité urbaine est importante, plus les concentrations en C_6H_6 sont élevées.

La relation linéaire directe entre C_6H_6 et la densité urbaine peut être prise en compte dans la cartographie

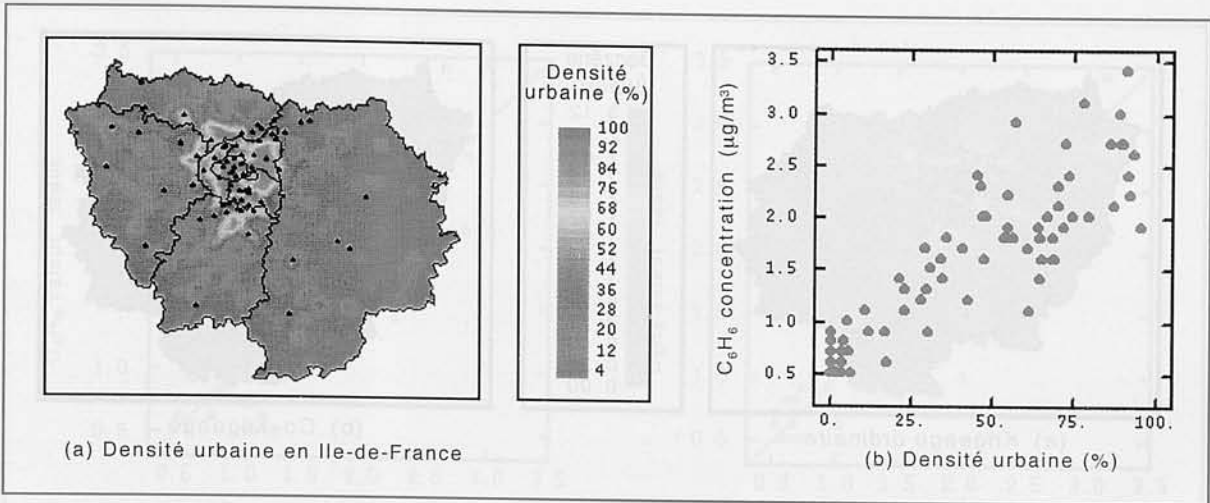


Figure 6.
Relation entre la densité urbaine et la concentration en C_6H_6 aux sites de mesures.
The relationship between the urban texture and the C_6H_6 concentration at the sampling locations.

de C_6H_6 en appliquant la technique du co-krigeage. Pour ce faire, il faut modéliser non seulement les variogrammes directionnels de C_6H_6 (voir Figure 2, p. 254) mais également ceux de la densité urbaine ainsi que les variogrammes directionnels croisés entre C_6H_6 et la densité urbaine [voir respectivement les figures 7 (a) et 7 (b)]. Les variogrammes expérimentaux sont représentés par une ligne mince et le modèle de variogramme par une ligne épaisse continue.

La forte relation entre C_6H_6 et la densité urbaine implique que les variogrammes expérimentaux de

C_6H_6 ainsi que les variogrammes croisés présentent une allure semblable à celle des variogrammes directionnels de C_6H_6 . La connaissance en tout point de la densité urbaine et la modélisation de son comportement spatial, ainsi que sa relation avec C_6H_6 permettraient a priori d'améliorer la cartographie de C_6H_6 entre les sites de mesure. Afin d'évaluer l'influence de la variable auxiliaire, les cartes obtenues par krigage et par co-krigeage sont présentées respectivement dans les figures 8 (a) et 8 (b) p. 260. L'intégration de la variable auxiliaire mène à une

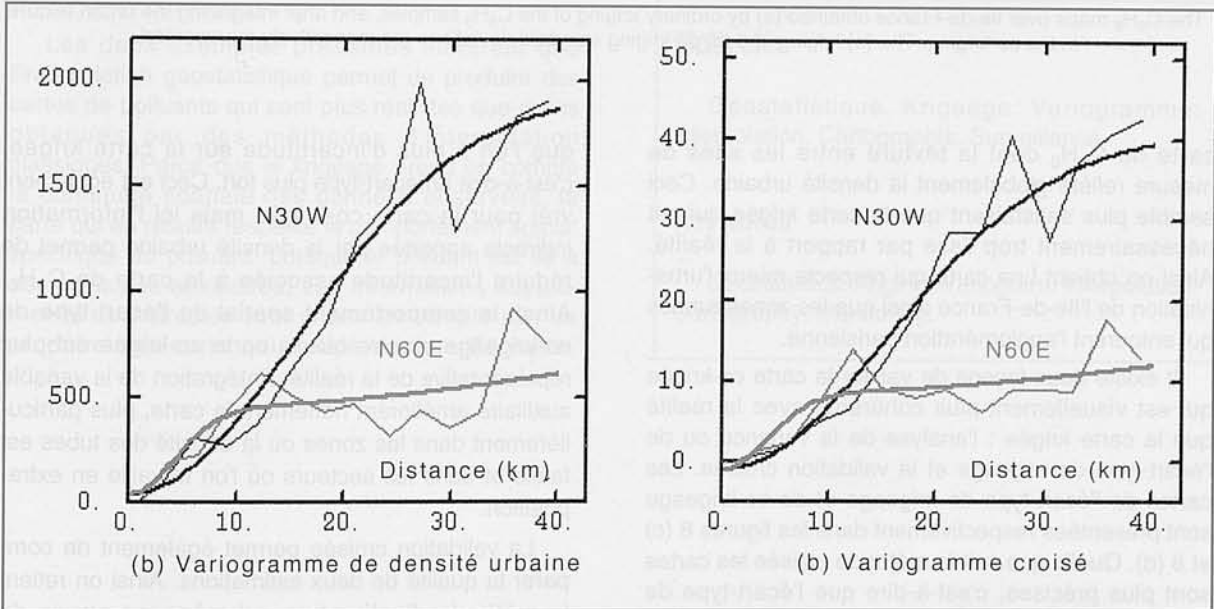


Figure 7.
Les variogrammes directionnels expérimentaux (ligne mince) et le modèle de variogramme (ligne épaisse) pour (a) la densité urbaine et (b) le variogramme croisé entre C_6H_6 et la densité urbaine.
The directional experimental variograms (thin line) and the variogram model (thick line) for (a) the urban texture and (b) the cross variogram between C_6H_6 and the urban texture.

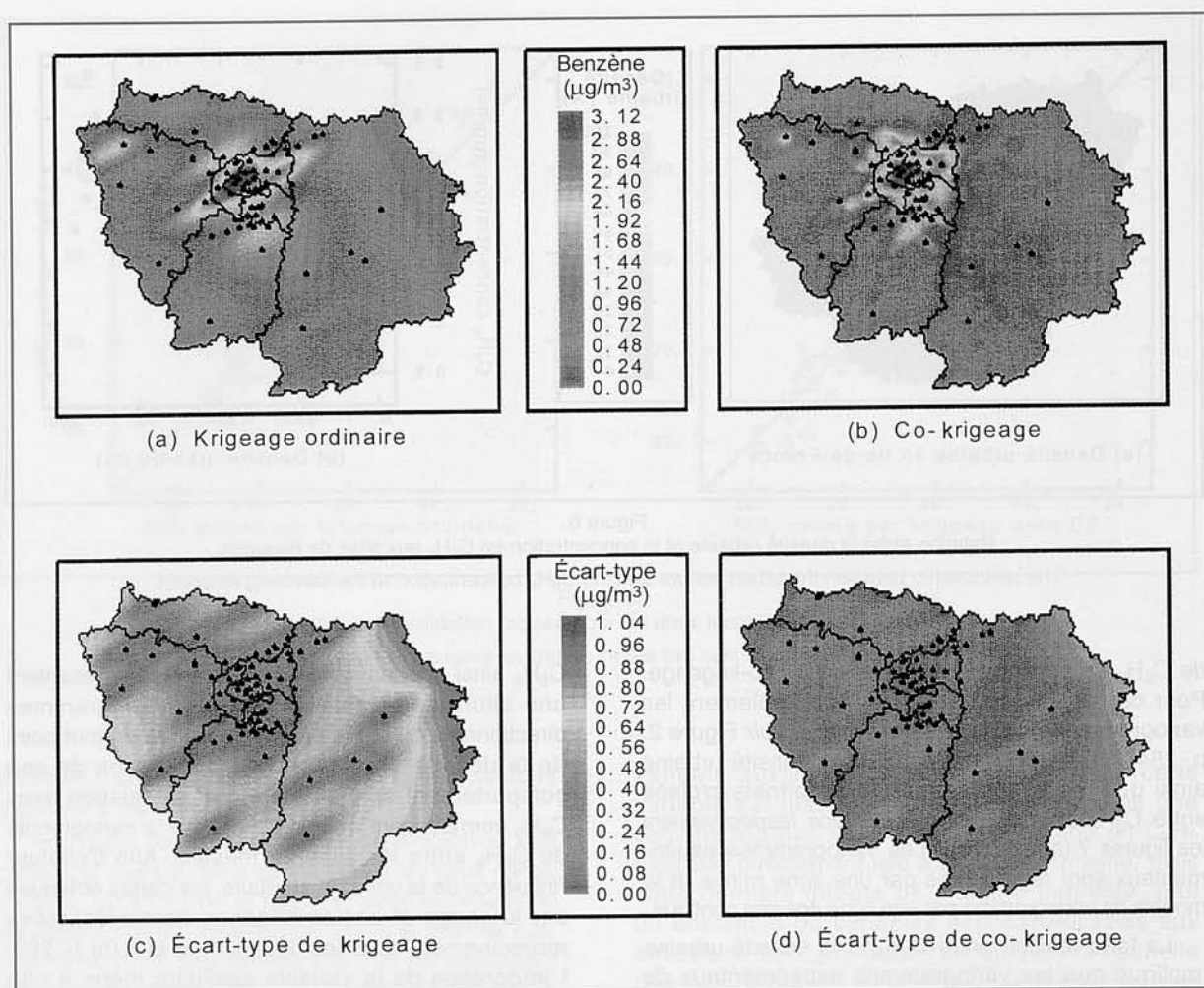


Figure 8.

Cartes de C_6H_6 en Ile-de-France (a) à partir du krigage ordinaire des données de C_6H_6 et (b) en intégrant la densité urbaine au co-krigeage. Cartes de l'écart-type (c) de krigage et (d) de co-krigeage.

The C_6H_6 maps over Ile-de-France obtained (a) by ordinary kriging of the C_6H_6 samples, and after integrating the urban texture in the co-kriging. The (a) kriging and (b) co-kriging standard deviations are respectively represented.

carte de C_6H_6 dont la texture entre les sites de mesure reflète globalement la densité urbaine. Ceci semble plus satisfaisant que la carte krigée qui est nécessairement trop lisse par rapport à la réalité. Ainsi on obtient une carte qui respecte mieux l'urbanisation de l'Ile-de-France ainsi que les zones rurales qui entourent l'agglomération parisienne.

Il existe deux façons de valider la carte co-krigée qui est visuellement plus cohérente avec la réalité que la carte krigée : l'analyse de la variance ou de l'écart-type de krigage et la validation croisée. Les cartes de l'écart-type de krigage et de co-krigeage sont présentées respectivement dans les figures 8 (c) et 8 (d). Quelle que soit la méthode utilisée les cartes sont plus précises, c'est-à-dire que l'écart-type de krigage est plus faible, près des sites de mesures. Loin des sites de mesures, l'écart-type de krigage reflète la continuité des variogrammes directionnels qui définissent la direction N60E comme celle de la plus forte variabilité. C'est donc dans cette direction

que l'on a plus d'incertitude sur la carte krigée, c'est-à-dire un écart-type plus fort. Ceci est également vrai pour la carte co-krigée mais ici l'information indirecte apportée par la densité urbaine permet de réduire l'incertitude associée à la carte de C_6H_6 . Ainsi, le comportement spatial de l'écart-type de co-krigeage montre que la carte co-krigée est plus représentative de la réalité, l'intégration de la variable auxiliaire améliorant nettement la carte, plus particulièrement dans les zones où la densité des tubes est faible et dans les secteurs où l'on travaille en extrapolation.

La validation croisée permet également de comparer la qualité de deux estimations. Ainsi on retient la méthode d'estimation qui mène au nuage de points (valeurs vraies versus valeurs estimées) qui est le plus près de la bissectrice. La figure 9 présente les résultats de la validation croisée pour le krigage et pour le co-krigeage de C_6H_6 en Ile-de-France.

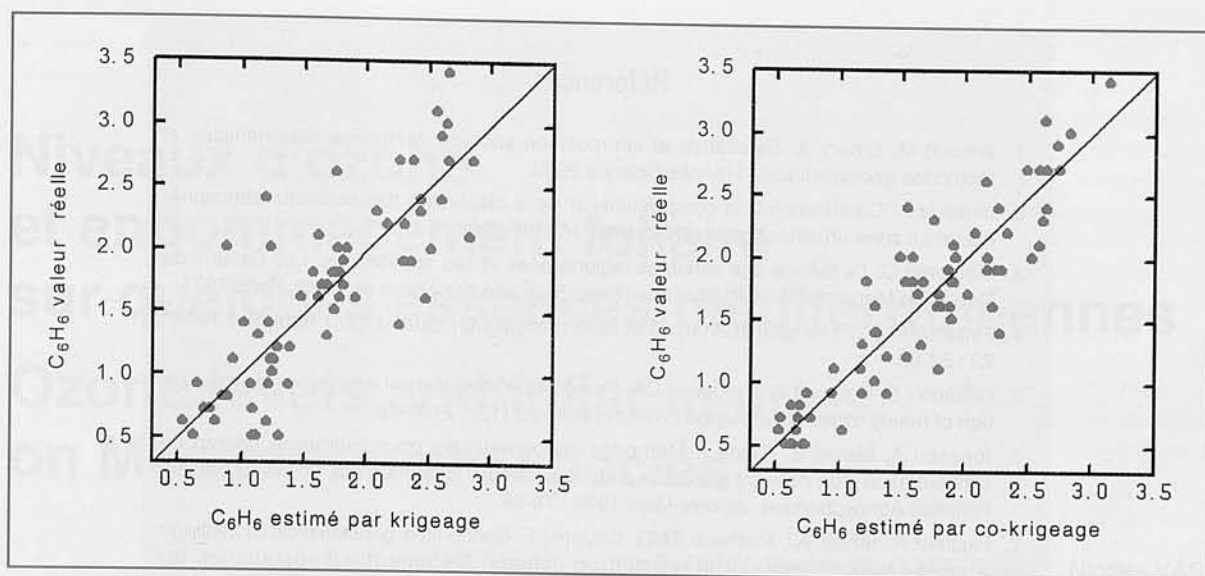


Figure 9.

Résultat de la validation croisée pour (a) le krigeage ordinaire et pour (b) le co-krigeage.

The cross validation result for (a) ordinary kriging and (b) co-kriging.

Le nuage correspondant au co-krigeage est plus centré et plus serré autour de la bissectrice. Ainsi on confirme que la carte co-krigée est plus représentative de la réalité. La prise en compte de la variable auxiliaire, ici la densité urbaine, améliore donc la précision de la carte obtenue, non seulement à l'intérieur de l'enveloppe convexe des sites de mesure mais également dans les zones en extrapolation.

Conclusion

Les deux exemples présentés montrent que l'interpolation géostatistique permet de produire des cartes de polluants qui sont plus réalistes que celles obtenues par des méthodes d'interpolation classiques. Parce que le krigeage prend en compte la continuité spatiale des données observées, la carte qui en résulte respecte le comportement spatial spécifique du polluant. Lorsque le polluant est lié à des variables auxiliaires, ces informations supplémentaires indirectes sont prises en compte afin de produire une carte qui reflète ce lien. En outre,

l'approche géostatistique permet de quantifier l'incertitude associée à la carte, et de définir des zones sous-échantillonnées. Par l'utilisation optimale des mesures, les techniques géostatistiques permettent de répondre au problème de la cartographie précise des polluants pour la surveillance de la qualité de l'air selon les exigences du public, des acteurs du domaine et des législateurs.

Mots clés

Géostatistique. Krigeage. Variogramme. Interpolation. Cartographie. Surveillance.

Keywords

Geostatistics. Kriging. Variogram. Interpolation. Cartography. Monitoring.

Références

1. Arnaud M, Emery X. Estimation et interpolation spatiale. Méthodes déterministes et méthodes géostatistiques. Hermès Science 2000.
2. Ionescu A. Contribution à la compréhension de la dispersion des polluants atmosphériques en zone urbaine. Thèse de Doctorat, IAURIF 1998.
3. Matheron G. La théorie des variables régionalisées et ses applications. Les Cahiers du Centre de Morphologie Mathématique (Fasc. 5), École des Mines de Paris, Paris 1971.
4. Finkelstein PL. The spatial analysis of acid precipitation data. J Clim Appl Met 1984 ; 23 : 52-62.
5. Casado LS, Rouhani S, Cardelino CA, Ferrier AJ. Geostatistical analysis and visualization of hourly ozone data. Atmos Environ 1994 ; 28 (12) : 2105-18.
6. Ionescu A, Mayer E, Colda I. Méthodes mathématiques pour estimer le champ de concentration d'un polluant gazeux à partir des valeurs mesurées aux points dispersés. Pollution Atmosphérique Janvier-Mars 1996 : 78-89.
7. Figuera R, Sousa AJ, Pacheco AMG, Catarino F. Space-time geostatistical modelling : A case-study of sea-salt measured on lichens, Geoenviron II - Geostatistics for Environmental Applications : Proceedings of the Second European Conference on Geostatistics for Environmental Applications. J. Gomez-Hernandez, A. Soares, R. Froidevaux (eds) 1999 : 53-64.
8. Nunes C, Soares A, Ferreira F. Evaluation of environmental costs of SO₂ emissions using stochastic images, Geoenviron II - Geostatistics for Environmental Applications : Proceedings of the Second European Conference on Geostatistics for Environmental Applications. J. Gomez-Hernandez, A. Soares, R. Froidevaux (eds) 1999 : 113-24.
9. AIRPARIF. Évaluation de l'exposition potentielle de la population francilienne à la pollution par le benzène. Rapport PRQA Ile-de-France, Département Études & Communications, AIRPARIF, Paris 2000.
10. Chauvet P. Aide-mémoire de la géostatistique linéaire. Les Presses de l'École des Mines de Paris, Paris 1994.
11. Armstrong M, Carignan J. Géostatistique linéaire : application au domaine minier. Les Presses de l'École des Mines de Paris, Paris 1997.
12. Chilès J-P, Delfiner P. Geostatistics: modelling spatial uncertainty. John Wiley & Sons, New York 1999.
13. Wackernagel H. Multivariate Geostatistics, Springer 1998.