

Surveillance de la qualité de l'air par cartographie : l'approche géostatistique

*(Prix du meilleur poster **INRETS 2000**, 5-8 juin à Avignon France)*

Résumé

Aujourd'hui 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, deux applications pratiques sont présentées. Ainsi la géostatistique se montre très adaptée au problème de la surveillance de la qualité de l'air. Seul le krigeage permet d'obtenir des cartes représentatives du comportement spatial spécifique des polluants traités. Tous les calculs sont faits à l'aide du logiciel géostatistique ISATIS.

Mots-clefs : géostatistique, krigeage, variogramme, interpolation, cartographie, surveillance.

Introduction

Aujourd'hui, la législation existante et la demande du public imposent aux organismes responsables de la surveillance de la qualité de l'air de produire des cartes précises de polluant à partir des observations aux stations de mesures isolées.

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 mesures veut dire affecter une concentration estimée à chaque nœud de la grille. Pour un nœud donné, cette concentration est obtenue comme une moyenne pondérée des concentrations observées aux stations de mesures voisines du nœud. Il existe cependant plusieurs algorithmes de pondération possibles. 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 et à prendre en compte le maximum d'informations (comme la densité du trafic ou les conditions météorologiques).

1 - L'interpolation par des techniques classiques

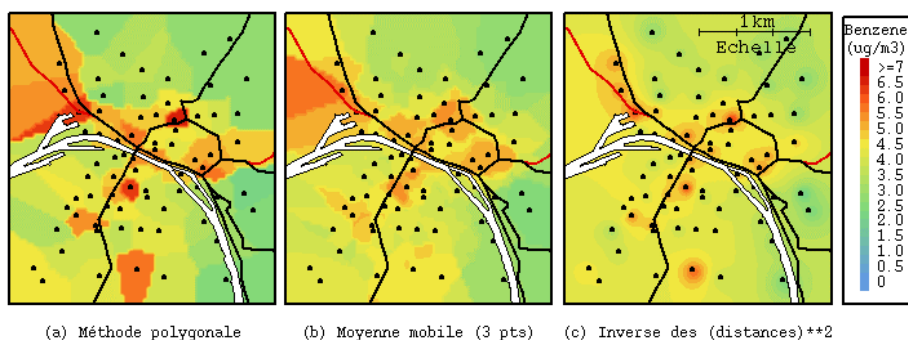


Figure 1 : Les cartes obtenues par (a) la méthode polygonale, (b) la moyenne mobile des 3 données les plus près et par (c) l'inverse des carrés des distances.

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 73 tubes à diffusion sur l'agglomération rouennaise (cf. paragraphe 3 ci-dessous pour une description). La localisation des tubes est présentée sur les cartes par un point noir, la Seine en blanc et les axes routiers principaux en noir et en rouge (pour les autoroutes). 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 que la méthode utilisée prenne également en compte les caractéristiques spatiales spécifiques du polluant.

2 - L'approche géostatistique : interpolation par krigeage

Selon l'algorithme 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. Le dit variogramme, calculé à partir des valeurs observées aux stations de mesure, permet de quantifier la continuité spatiale du polluant.

La Figure 2 présente les variogrammes expérimentaux et le modèle de variogramme ajusté aux données (a) de C6H6 à Rouen, et (b) de NO2 à Paris. Le variogramme quantifie la variabilité entre deux mesures du polluant en fonction de la distance qui les sépare. Le variogramme peut être interprété en terme 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. De cette façon, C6H6 à Rouen est moins corrélé dans l'espace que NO2 à Paris. Le polluant est non corrélé au-delà de la distance où le modèle de variogramme se stabilise autour d'un palier, 3km pour C6H6 et 35-40km pour NO2. La corrélation entre deux mesures séparées d'au moins cette distance est nulle.

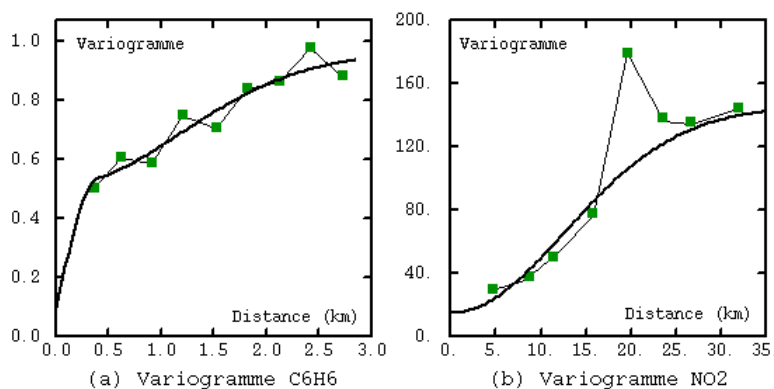


Figure 2 : Les variogrammes expérimentaux (ligne mince) et les modèles de variogramme (ligne épaisse) ajustés aux données (a) de C6H6 à Rouen, et (b) de NO2 à Paris.

Le modèle de variogramme, qui est spécifique au polluant, est ensuite entré dans un système linéaire d'équations, dit système de krigeage, afin de déterminer le poids optimal à affecter aux données lors de l'interpolation. De cette façon, un 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. L'interpolation géostatistique prend donc en compte non pas la distance entre les stations des 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.

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. 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 leur corrélation croisée afin d'estimer le polluant à partir des mesures de concentration et des données de la variable auxiliaire. Lorsque la forme de la variable auxiliaire présente l'allure globale du polluant, la technique du krigeage avec dérive externe permet de modéliser cette allure et de l'incorporer dans la carte interpolée afin de mieux reproduire la structure spatiale du polluant.

Le cadre probabiliste de la géostatistique permet de quantifier l'incertitude associée à la valeur interpolée à l'aide de la variance de krigeage. Cette variance ou sa racine carrée, l'écart-type de krigeage, représente la dispersion possible de la concentration réelle et inconnue autour de la valeur obtenue par krigeage. Plus cette dispersion, c'est-à-dire 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.

Les deux exemples présentés ci-dessous montrent que l'interpolation géostatistique permet de produire des cartes de polluant 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. De plus l'approche géostatistique permet de quantifier l'incertitude associée à la carte et de définir des zones sous-échantillonnées.

3 – Application pratique : Cartographie du benzène à Rouen

Six campagnes de mesures du benzène (C6H6) ont été réalisées à Rouen au cours de l'année 1998. Les mesures ont été faites avec des tubes à diffusion, localisés de façon à mesurer la pollution de fond sur l'agglomération. Chaque campagne fournit des concentrations moyennes sur une semaine. C'est la moyenne des 6 campagnes qui est utilisée pour cartographier ce qui représentera un niveau de fond annuel moyen. Pour ce faire, 73 sites ont pu être exploités. Afin d'améliorer la carte obtenue à partir des échantillons de benzène, deux variables auxiliaires ont été considérées (faute de mieux) : la densité de population et le relief. A priori la concentration en benzène n'est pas directement liée à ces variables, mais il pourrait y avoir une relation indirecte en appliquant les raisonnements suivants : (a) le centre ville et donc les axes routiers principaux se trouvent dans la cuvette créée par la Seine. Ceci mènerait à une relation inverse entre la concentration du polluant et le niveau topographique due plus au hasard de la configuration locale qu'à la physique du phénomène, et (b) plus la population est dense, plus il y a de voitures et donc plus la concentration en benzène risque d'être importante.

Pour intégrer ces variables dans la cartographie, nous construisons une fonction de la population et du relief qui indique la tendance générale, dite dérive externe, de la concentration en C6H6 et qui soit représentative du comportement global du polluant (cf. Figure 3(a)). 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 les Figures 3(b) et 3(c). L'intégration de la dérive au C6H6 améliore nettement l'estimation. La carte est plus détaillée et les zones en extrapolation ou à faible densité de points de mesures sont mieux renseignées.

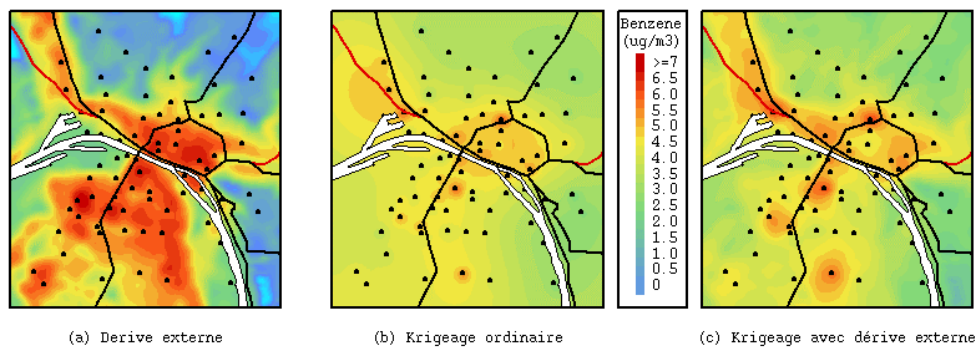


Figure 3 : (a) La forme de la dérive externe, et les cartes krigées de C6H6 à Rouen (b) à partir des données de C6H6 et (c) en y intégrant la densité de la population et le relief.

Deux techniques permettent de juger de la fiabilité d'une carte. Outre la variance de krigeage précédemment évoquée, la validation croisée 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 4 présente les cartes de l'écart-type de krigeage (sans et avec dérive externe), ainsi que le résultat de la validation croisée pour les deux estimations.

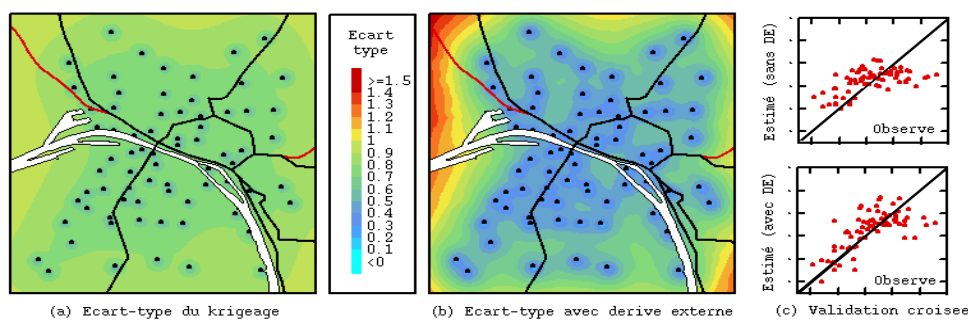


Figure 4 : Les cartes de l'écart-type de krigeage (a) sans et (b) avec prise en compte de la dérive. Le résultat de la validation croisée (c) pour les deux techniques d'estimation.

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 krigeage avec dérive externe présente d'une façon plus réaliste la forte incertitude associée à ces zones.

4 – Application pratique : Cartographie du dioxyde d’azote à Paris

Il s'agit d'une étude de cartographie des données statistiques de percentile 50 (la valeur médiane) de NO₂ enregistrée sur la période 1997-1999 sur 20 capteurs situés en région parisienne. Un cadastre des émissions moyennes de NO_x dans un voisinage de 29 km² pour l'année 1998 a été utilisé comme variable auxiliaire car il existe une relation indirecte entre émissions de NO_x et concentration de NO₂. En effet, la concentration en NO₂ en un point est liée de manière non linéaire du fait des réactions de chimie atmosphérique à la concentration en NO_x en ce point, la concentration en NO_x étant elle-même le résultat des processus de transport et de diffusion atmosphériques des émissions de NO_x situées dans le voisinage du point.

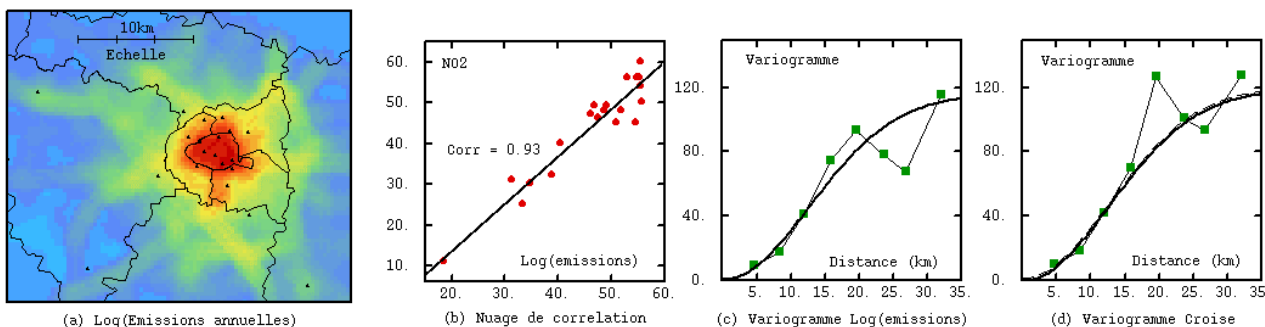


Figure 5 : (a) Le logarithme des émissions en NO_x, (b) le nuage de corrélation entre NO₂ et log(émissions), (c) le variogramme du log(émissions), et (d) le variogramme croisé entre NO₂ et log(émissions).

Si la concentration en NO₂ n'est pas liée linéairement aux émissions de NO_x, on s'aperçoit par contre expérimentalement que celle-ci l'est relativement bien avec le logarithme des émissions, noté désormais log(émissions) (cf. Figure 5(b)). Le logarithme des émissions, présenté dans la Figure 5(a), peut donc être utilisé comme variable auxiliaire pour cartographier le NO₂ en Ile de France. Le champ d'étude couvre une région de 90 x 75 km² autour de Paris. Les limites des départements voisins sont présentées en noir. La localisation des capteurs est représentée par un point noir.

La relation linéaire directe entre NO₂ et log(émissions) peut être prise en compte dans la cartographie de NO₂ en appliquant la technique du co-krigeage. Pour ce faire, il faut également modéliser le variogramme de log(émissions) ainsi que le variogramme croisé entre NO₂ et log(émissions) ; voir respectivement les Figures 5(c) et (d). La connaissance en tout point de la variable log(émissions) et la modélisation de sa relation avec NO₂ permettent d'améliorer la cartographie de NO₂ entre les sites de mesure. Afin d'évaluer l'influence de la variable auxiliaire log(émissions), les cartes obtenues par krigage et par co-krigeage sont présentées respectivement dans les Figures 6(a) et 6(b). On peut constater que l'intégration de la variable auxiliaire log(émissions) améliore nettement la carte, plus particulièrement dans les zones où on a une faible densité de capteurs et dans les secteurs où on travaille en extrapolation.

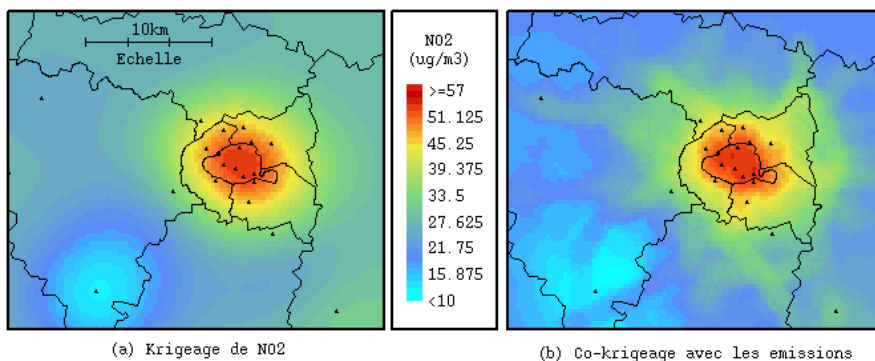


Figure 6 : Les cartes krigées de NO₂ à Paris (a) à partir des données de NO₂ et (b) en y intégrant les émissions en NO_x.

La prise en compte d'une variable auxiliaire, ici le logarithme des émissions en NO_x, améliore la précision de la carte obtenue, non seulement à l'intérieur de l'enveloppe convexe des capteurs mais également dans les zones en extrapolation.