



## Méthodes mathématiques pour estimer le champ de concentration d'un polluant gazeux à partir des valeurs mesurées aux points dispersés

### *Mathematical methods to evaluate the concentration field of a gaseous pollutant from values measured at scattered locations*

Anda IONESCU (\*) (\*\*), Eric MAYER (\*), Iolanda COLDA (\*\*)

#### RÉSUMÉ

Pour estimer le champ de concentration de polluants sur un domaine donné, à partir des points de mesure dispersés, il est nécessaire d'évaluer la concentration aux nœuds d'un maillage rectangulaire, à l'aide d'une méthode d'interpolation ou d'ajustement des données. Les valeurs du maillage rectangulaire nous permettent de tracer les lignes d'isoconcentration, afin d'avoir une image de la variation spatiale de la concentration, plus facile à interpréter. On peut, également, représenter, dans un système tridimensionnel, la surface qui correspond à la distribution spatiale de ce paramètre.

Pour cela, on a analysé deux catégories de méthodes d'approximation : statistiques et analytiques, qui ont été appliquées, ultérieurement, aux données de concentration de  $\text{SO}_2$  mesurées par le réseau Airparif.

Pour conclure, on a remarqué que, dans le cas étudié, la méthode statistique dénommée Krigeage, basée sur les semi-variogrammes ordinaires, ne peut pas être appliquée à cause de la non-corrélation spatiale des valeurs mesurées ; parmi les méthodes analytiques étudiées, celle qui fournit les meilleurs résultats est la méthode des fonctions spline type *plaque mince*.

#### ABSTRACT

In order to evaluate a pollutant concentration field, when only few scattered values of pollutant concentration are available on a given territory, it is necessary to estimate the corresponding values in many other points. An approximation method to quantify the concentrations at the nodes of a rectangular mesh associated to the given territory was applied. These values were then used to generate concentration contours, in order to obtain an image of the concentration spatial distribution.

This paper is concerned with the analysis of two types of approximation methods : statistical and analytical ones, which were applied to the AIRPARIF Network data.

The Kriging statistical method, based on ordinary semivariograms can not be used in the particular case of AIRPARIF data. Among the analytical methods studied here, the thin plate interpolation spline offers the best results.

#### 1. Introduction

Le réseau Airparif a été mis en place en 1982 pour mesurer la pollution atmosphérique et fonctionner en réseau d'alerte dans la région Ile-de-France. Il regroupe une soixantaine de stations qui mesurent différents types de polluants, par exemple  $\text{SO}_2$ ,  $\text{CO}_2$ ,  $\text{NO}_x$ .

Airparif, en définissant ses objectifs [Lamelloise, 1991], propose la mise au point d'un système « permettant une visualisation attractive des résultats de mesure (cartographie vidéo, par exem-

ple) », pour « détecter les situations à risques particuliers », pour « contribuer à mieux connaître l'exposition cumulée des citoyens aux pollutions atmosphériques et, en conséquence, à améliorer la connaissance épidémiologique ».

Cet article se propose de comparer plusieurs méthodes d'interpolation ou d'ajustement et de justifier pourquoi nous avons choisi celle utilisant les fonctions spline type *plaque mince*.

Une validation de cette dernière méthode a été obtenue à partir des données sur la topographie de la ville de Paris. En utilisant un nombre discret de sites d'altitude connue, on a déterminé des courbes de niveau qui ont été comparées aux courbes de niveau fournies par l'Institut Géographique National.

(\*) LETIEF, IUT de Créteil, Université Paris XII, Val de Marne.

(\*\*) Université Technique des Constructions de Bucarest.

Dans le domaine de la pollution de l'air, cette méthode a été également vérifiée en essayant de reconstituer, à partir des points dispersés, la distribution de la pollution autour de la centrale de production thermique de Porcheville, simulée à l'aide du modèle gaussien de Pasquill.

La méthode proposée permet le tracé, sur une base cartographique, des lignes d'isoconcentration et la visualisation en deux ou trois dimensions des niveaux de la pollution atmosphérique mesurée par le réseau Airparif. Ainsi, il est possible d'identifier, presque en temps réel, les situations, les zones critiques et l'origine des émissions.

L'utilisation d'observations de terrain permet de valider et corriger des données fournies par un satellite afin de pouvoir exploiter cette technique dans les zones urbaines non-équipées de réseaux de mesure.

## II. Modèle mathématique

Nous partons d'un territoire  $\mathcal{D}$  où se trouvent des points de mesure dispersés  $P_i$  auxquels sont associées des valeurs  $z_i$  de la concentration en  $\text{SO}_2$  qui correspondent à la somme cumulée des quantités reçues pendant une heure. Les valeurs  $z_i$  sont donc des valeurs horaires indicatrices du niveau de pollution. Ces données horaires ont été fournies par le réseau Airparif pour la période Janvier-Mars 94.

Soit le domaine  $\mathcal{D} \subset \mathbb{R}^2$ . On connaît les coordonnées rectangulaires (coordonnées Lambert II) des points  $P_1, P_2, \dots, P_n \in \mathcal{D}$  et les valeurs mesurées  $z_1, z_2, \dots, z_n$  en  $P_1, P_2, \dots, P_n$ .

On superpose un maillage rectangulaire sur  $\mathcal{D}$  et on se propose d'estimer les valeurs  $z_{ij}$  aux nœuds du maillage.

Les méthodes d'approximation qui ont été étudiées sont les suivantes : Krigeage, coefficients indéterminés, moindres carrés, pondération par rapport aux distances, Shepard, triangulation de Delaunay avec interpolation linéaire, fonctions spline type *plaque mince*.

## III. Méthodes d'estimation

1. *La méthode de Krigeage* [Deutsch, 1992] est une technique géostatistique qui peut être définie comme une méthode d'optimisation de l'estimation d'une grandeur qui est répartie dans l'espace et mesurée sur un réseau de points. Elle est basée sur le concept de variables régionalisées, qui sont des variables typiques d'un phénomène qui se développe dans l'espace et qui possède une certaine structure. Ici, le terme « structure » se réfère à une corrélation spatiale.

En théorie, le Krigeage est considéré comme la méthode qui produit les estimations les plus précises,

mais dans la pratique, son efficacité dépend du choix approprié de différents paramètres.

On commence par l'hypothèse que le champ de concentrations est homogène. Ici, l'homogénéité est prise dans le sens stochastique : la moyenne  $m_z(s)$  est constante et la covariance spatiale  $c_z(s, s')$  dépend seulement de vecteur différence  $h = s - s'$ , où  $s$  et  $s'$  représentent des vecteurs de position des points de mesure.

Pour savoir si l'utilisation de cette méthode est licite, il faut faire des tests de corrélation qui consistent à établir un semi-variogramme  $\gamma(h)$  (la semi-variance de l'accroissement de la variable aléatoire attachée à la variable régionalisée, par rapport aux distances) :

$$\gamma(h) = \text{Var} [Z(s+h) - Z(s)]/2$$

où  $Z$  est une variable aléatoire attachée à la variable régionalisée,  $s$  le vecteur de position d'un point et  $s+h$  la position d'un autre point.

Plus les points sont éloignés, plus la variance de la différence des variables  $Z$  est grande, l'information se dégrade au fur et à mesure qu'on s'éloigne de la source d'information. On considère tous les couples de points, on établit des classes en fonction des distances entre les points et pour chaque classe on calcule la semi-variance de l'accroissement de  $Z$ ,  $\gamma(h)$ .

Les valeurs mesurées par les capteurs placés dans le région Ile-de-France, dans 80 % des cas étudiés, ne présentent pas une corrélation spatiale, le variogramme étant plat ou décroissant. Les tests réalisés (Fig. 1) montrent que, dans ce cas, la méthode n'est pas utilisable.

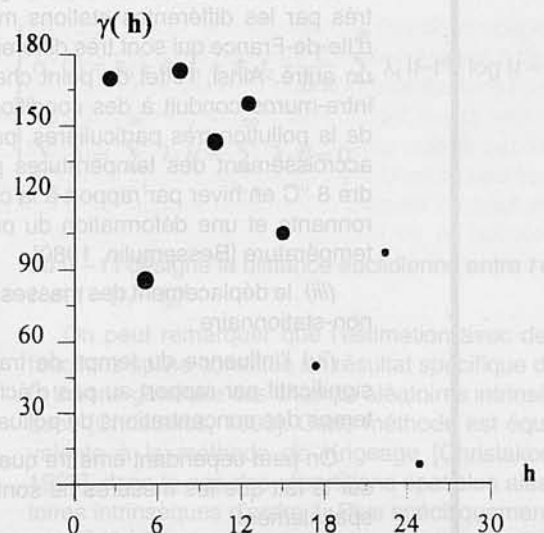


Figure 1.  
Variogramme obtenu à partir des valeurs de concentration mesurées sur la région Ile-de-France.  
Variogram of the concentration measured values in the Ile-de-France region.

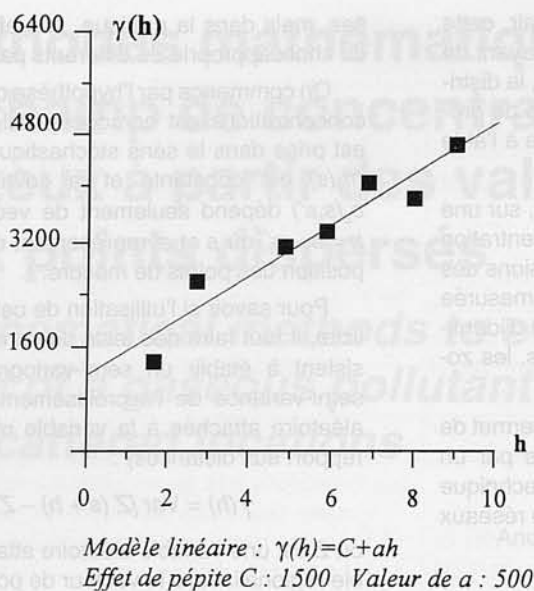


Figure 2.

Variogramme obtenu à partir des valeurs de concentration calculées avec le modèle gaussien de Pasquill.

Variogram of the concentrations calculated with the Pasquill gaussian model.

Par contre, quand on dispose de valeurs calculées correspondant à une seule source, la corrélation spatiale est évidente (Fig. 2).

L'absence de corrélation dans le cas des valeurs de concentration mesurée pourrait être expliquée par :

(i) la présence sur la région Ile-de-France de plusieurs sources de pollution : il y a des sources ponctuelles (i.e. cheminées industrielles, usines d'incinération), mais aussi et essentiellement des sources diffuses (le chauffage, les automobiles) ;

(ii) les paramètres météorologiques enregistrés par les différentes stations météorologiques d'Ile-de-France qui sont très différents d'un point à un autre. Ainsi, l'effet de point chaud dans Paris intra-muros conduit à des conditions de diffusion de la pollution très particulières, par exemple, un accroissement des températures pouvant atteindre 8 °C en hiver par rapport à la campagne environnante et une déformation du profil vertical de température [Bessemulin, 1980] ;

(iii) le déplacement des masses d'air en régime non-stationnaire ;

(iv) l'influence du temps de transport, qui est significatif par rapport au pas d'échantillonnage en temps des concentrations de polluant.

On peut cependant émettre quelques réserves sur le fait que les mesures ne sont pas corrélées spatialement.

Le nombre de capteurs est assez limité (une trentaine de points) ; en outre, quand une tendance non-constante est présente, le semi-variogramme empirique peut être sérieusement biaisé [Cressie, 1991] et conduire à de fausses interprétations vis-

à-vis de la structure spatiale du processus. En effet, on peut montrer [Journel, 1978] que l'estimateur  $\hat{\gamma}(h)$  est un estimateur biaisé de  $\gamma(h)$  si la tendance n'est pas constante :

$$E[\hat{\gamma}(h)] = \gamma \frac{(h+1)}{2N(h)} \sum_{i=1}^{N(h)} \{E[Z(s+h)] - E[Z(s)]\}^2$$

Si on rejette l'hypothèse d'homogénéité, on peut appliquer la méthode de Krigage d'une autre façon. On passe du champ spatial aléatoire  $Z$ , qui est non-homogène, à une répartition spatiale aléatoire intrinsèque d'ordre  $v$ , qui est un champ aléatoire avec des incréments homogènes d'ordre  $v$  et avec une moyenne nulle [Christakos, 1993]. On utilise, ensuite, une procédure automatique pour établir le type de fonction de covariance généralisée et son ordre. Cette transformation élimine de possibles inconvénients, tels que la dichotomie entre la tendance et les résidus.

Dans notre cas, on a obtenu une fonction de covariance d'ordre un, mais les cartes de variances mettent en évidence un degré élevé d'incertitude de l'estimation et l'aspect des distributions de pollution obtenues n'est pas réaliste.

Ces résultats nous conduisent à abandonner la méthode statistique en raison de la non-corrélation des mesures de pollution aux différents emplacements de mesure et de la relativement faible taille de l'échantillon. On recherche, alors, une méthode analytique.

## 2. Méthodes analytiques

2.1. La méthode analytique d'interpolation la plus simple est la méthode des coefficients indéterminés [Bakhvalov, 1976], qui consiste à déterminer une fonction polynomiale qui coïncide en  $P_1, P_2, \dots, P_n$  avec les valeurs mesurées. Plus le nombre de points est grand, plus la fonction d'interpolation contient un nombre élevé de paramètres et plus le degré du polynôme augmente. Alors, la fonction peut s'éloigner notablement en dehors des points où elle coïncide avec les valeurs mesurées. Les résultats obtenus ne sont pas satisfaisants.

2.2. Une meilleure approximation peut être obtenue par la méthode des moindres carrés [Bakhvalov, 1976], mais le problème est, dans ce cas-là, comment choisir le type de fonction d'approximation. Les fonctions polynomiales ne donnent pas de résultats satisfaisants, mais commencent à s'améliorer quand on utilise une fonction exponentielle, qui a comme paramètres l'orientation et la distance de chaque point par rapport au point où la mesure est maximum [Ionescu, 1994].

Pour ce deuxième type de fonction (exponentielle), dans un cas particulier de pollution dans la région Ile-de-France, les coefficients de la surface d'interpolation obtenus à partir de 34 points de mesure étaient affectés d'une incertitude comprise entre 22 % et 55 %.

2.3. Une autre méthode d'interpolation, utilisée par les géologues, est la *méthode de pondération par rapport aux distances*. Pour trouver des valeurs  $z_{ij}$  aux nœuds d'un maillage rectangulaire, on calcule des moyennes pondérées des valeurs mesurées  $z_k$ , les poids étant des fonctions décroissantes par rapport aux distances, le plus fréquemment, l'inverse du carré des distances. La pondération peut être faite, soit sur tout le domaine  $\mathcal{D}$ , soit sur une région autour du nœud de calcul, quand on délimite une zone d'influence. C'est un interpolateur rapide ; comme avantage, on peut signaler que les valeurs se situent toujours dans l'intervalle délimité par le maximum et le minimum, tandis que pour d'autres méthodes, on peut obtenir même des valeurs négatives, quand elle doivent être positives. Comme inconvénient, on remarque que les points isolés sont traités comme des pics ou des puits. Cet effet peut être réduit en ajoutant à la distance un terme appelé paramètre de lissage.

2.4. Une meilleure méthode par rapport à la précédente est la méthode de *Shepard* qui utilise les inverses des distances pondérées par la méthode des moindres carrés. L'algorithme [Franke, 1980] proposé par Shepard est modifié ultérieurement par Franke et Nielson.

2.5. Vu qu'à partir des points dispersés on ne peut pas construire un maillage rectangulaire (même de pas variable) avec toutes les valeurs aux nœuds connues, une autre alternative est de créer un réseau tridimensionnel de triangles, c'est-à-dire de trianguler l'enveloppe convexe  $[P]$  des points  $P_i$  pour que les intérieurs des triangles soient disjoints deux à deux et la réunion des triangles coïncide avec l'enveloppe convexe  $[P]$ . La solution n'est pas unique, mais la triangulation la plus utilisée est celle de Delaunay. Parmi toutes les triangulations associés, elle rend maximum le plus petit des angles des triangles [Berstel, 1991], [Risler, 1991].

A partir de ce réseau, une estimation peut être faite par exemple à partir des facettes de Bézier, en utilisant l'algorithme de De Casteljau [Risler, 1991] ou, plus simplement, en adoptant une interpolation linéaire.

La triangulation donne de bons résultats à partir d'environ 200 points initiaux. Un avantage de cette méthode peut être sa capacité à restituer les fortes variations, quand il y a assez de données, alors que d'autres méthodes d'approximation ont une tendance à lisser.

2.6. La *méthode des fonctions spline type plaque mince* [Carasso, 1991] est utilisée d'habitude pour le lissage des données.

Pour faire passer une courbe lisse par des points, les dessinateurs utilisaient une latte de bois. Cette latte est connue en anglais sous le nom de *spline* et en position d'équilibre elle minimise l'énergie de flexion.

Cette fonction est unique et elle peut être déterminée à partir de la minimisation d'une fonctionnelle qui impose les contraintes suivantes : passer le plus proche possible des points donnés et constituer la surface de courbure minimale.

Soit  $\mathcal{U}$  l'espace des fonctions sur  $\mathcal{R}^2$  dont les dérivées d'ordre 2 sont dans  $\mathcal{L}^2(\mathcal{R}^2)$  et tel que les fonctionnelles  $u \rightarrow u(t)$  soient continues,  $\mathcal{V}$  un espace  $(\mathcal{L}^2(\mathcal{R}^2))^4$  et l'application  $\mathcal{L} : \mathcal{U} \rightarrow \mathcal{V}$  définie par :

$$\mathcal{L}(u) = \left( \frac{\partial^2 u}{\partial t_1^2}, \frac{\partial^2 u}{\partial t_1 \partial t_2}, \frac{\partial^2 u}{\partial t_2 \partial t_1}, \frac{\partial^2 u}{\partial t_2^2} \right)$$

On pose

$$I_p(u) = \| \mathcal{L}(u) \|_{\mathcal{V}}^2 + \sum_{i=1}^m \frac{1}{\rho_i} (u(t^i) - z_i)^2$$

où  $\rho_1, \rho_2, \dots, \rho_m$  sont des scalaires positifs donnés.

Il existe une fonction unique  $\sigma_p(u)$  qui minimise  $I_p(u)$ . Cette minimisation revient à la résolution de l'équation de Laplace :

$$\Delta^2 \sigma_p = \sum_{i=1}^m \lambda_i \delta_i$$

où  $\delta_i$  est la distribution de Dirac en  $t^i$ .

$$\Delta^2 \sigma_p = \sum_{i,j=1}^2 \frac{\partial^4 \sigma_p}{\partial t_i^2 \partial t_j^2}$$

$$\lambda_i = \frac{1}{\rho_i} (z_i - \sigma_p(t^i)) \quad i = 1 \dots m$$

A partir de la solution élémentaire de cette équation, on montre que la fonction spline type *plaque mince* a la forme :

$$\left\{ \begin{array}{l} \sigma_p(t) = a_0 + a_1 t_1 + a_2 t_2 + \frac{1}{8\pi} \sum_{i=1}^m \lambda_i |t - t^i|^2 \log |t - t^i|^2 \\ \sum_{i=1}^m \lambda_i = \sum_{i=1}^m \lambda_i t_1^i = \sum_{i=1}^m \lambda_i t_2^i = 0 \end{array} \right.$$

où  $|t - t^i|$  désigne la distance euclidienne entre  $t$  et  $t^i$  et  $t = (t_1 - t_2)^t$ .

On peut remarquer que l'estimation avec des fonctions spline constitue un résultat spécifique de la théorie générale des champs aléatoires intrinsèques [Christakos, 1993]. Cette méthode est équivalente à la méthode de Krigeage [Christakos, 1993], dans le cas des répartitions spatiales aléatoires intrinsèques d'ordre 1. Plus spécifiquement, en  $\mathcal{R}^2$ , l'approximation avec des fonctions spline type *plaque mince* équivaut au Krigeage intrinsèque avec une covariance spatiale généralisée d'ordre 1,

$$k_x(r) = c_s r^2 \log r.$$

#### IV. Argumentation du choix de la méthode d'estimation

A. Prenons comme cas d'étude le relief de la région parisienne, zone délimitée par les coordonnées Lambert II ( $x_{\min} = 593$  km,  $x_{\max} = 604$  km,  $y_{\min} = 2425$  km,  $y_{\max} = 2433$  km), pour laquelle la carte nous a été fournie par l'IGN.

1. Malgré la subjectivité d'une comparaison visuelle entre les cartes obtenues à partir de 60 points dispersés, avec les différentes méthodes (figures 3b, 3c, 3d, 3e) et la carte IGN (figure 3a), il est évident que celle qui reconstitue le mieux les courbes IGN est obtenue à partir des fonctions splines.

2. La qualité de l'approximation aux points initiaux peut être mise en évidence par un traitement des résidus en ces points.

Notons par  $r_i$  le résidu au point "i", défini par :

$$r_i = z_i^c - z_i^m$$

où  $z_i^c$  est la valeur calculée (interpolée) au point "i"  
 $z_i^m$  est la valeur mesurée ou la valeur initiale au même point "i".

En continuant à analyser le même cas du relief, quelques statistiques des résidus sont données dans le tableau I (pour chaque méthode, les extrêmes des résidus, la moyenne, l'écart type). Il est évident que la méthode des fonctions splines est la plus performante.

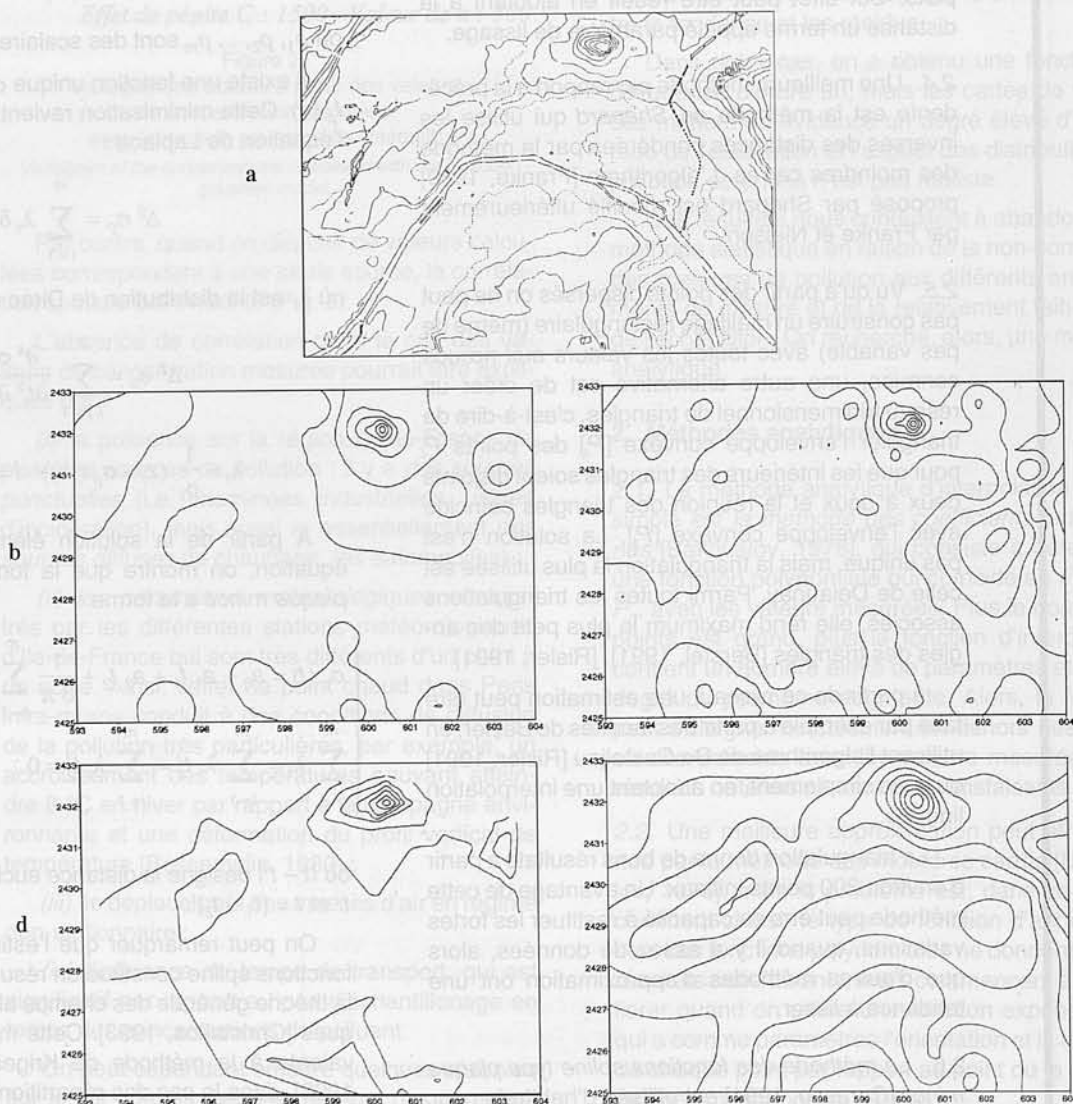


Figure 3.

Cartographie de la ville de Paris a. fournie par l'IGN ; b. en utilisant la méthode de pondération par rapport aux distances ; c. en utilisant la méthode de Shepard ; d. en utilisant la triangulation de Delaunay avec interpolation linéaire ; e. en utilisant les fonctions splines.

Paris's altitude contours a. provided by the National Geographic Institute ; b. obtained by inverse distance ponderation method ; c. by means of Shepard's method ; d. by means of Delaunay triangulation method with linear interpolation ; e. by means of spline method.

Tableau I.  
Traitement des résidus aux points initiaux  
pour les méthodes étudiées

| Méthode     | Moyenne | Ecart type | Min.    | Max.  |
|-------------|---------|------------|---------|-------|
| Pondération | 0,0034  | 1,701      | 3,035   | 9,019 |
| Shepard     | 0,3177  | 1,863      | - 2,978 | 9,918 |
| Delaunay    | 0,5857  | 2,148      | - 2,378 | 9,325 |
| Spline      | 0,0085  | 0,547      | - 1,637 | 1,667 |

3. On a testé le comportement des méthodes par rapport aux points initiaux, maintenant, on va choisir sur la carte une vingtaine de points différents des points de départ et on va comparer les valeurs lues aux valeurs calculées. Les points ont été choisis sur les courbes de niveau pour ne pas introduire des erreurs d'interpolation lors de la lecture sur la carte.

Un traitement des résidus calculés pour la vingtaine des points pour les différentes méthodes, ainsi que les coefficients de corrélation entre les valeurs lues sur la carte et les valeurs calculées sont présentés dans le tableau II.

Parmi toutes ces méthodes, on arrive à choisir les fonctions splines type plaque mince, en faisant aussi quelques commentaires.

Il est évident que les performances de la méthode sont bien dépendantes du nombre de points initiaux. Mais ce n'est pas seulement le nombre de points qui compte, mais aussi leur pouvoir informatif. Il est très important que les pics soient connus, ainsi que les points qui donnent le changement de gradient.

**B.** Pour tester le comportement de la méthode des fonctions spline dans le domaine de la *pollution atmosphérique*, nous avons simulé un champ de concentrations de polluant à partir des données réelles des émissions de la centrale thermique de Porcheville et des conditions météorologiques correspondantes (Météo France). Le modèle de calcul est un modèle gaussien, celui de Pasquill-Gifford, utilisant 6 classes d'atmosphères.

Nous avons calculé la concentration de SO<sub>2</sub> en 625 points représentant les nœuds d'un maillage régulier couvrant une surface d'environ 18 km<sup>2</sup>. Bien qu'il y a encore d'autres sources de pollution dans la proximité de Porcheville (usines Renault, La Sommec), sur cette zone, assez étendue, l'influence prédominante est celle de Porcheville. Si-non, on aurait superposé les champs de concentration de chaque source. A partir de ces 625 points, on a choisi 60 points de manière aléatoire et on a estimé à l'aide des fonctions splines les valeurs aux autres 565 points qui n'ont pas été pris en compte pour l'estimation. Une comparaison entre les 2 séries de valeurs (voir figure 4) met en évidence un coefficient de corrélation de 0,992.

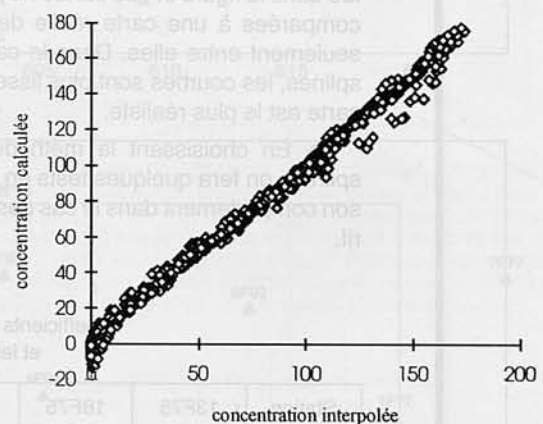


Figure 4.  
Corrélation entre les valeurs de concentration de polluant [µg/m<sup>3</sup>] calculées avec le modèle de Pasquill et les valeurs estimées à l'aide des fonctions spline. Coefficient de corrélation 0.992.  
*Correlation between the values of pollutant concentration[µg/m<sup>3</sup>] calculated with the Pasquill model and the values estimated by means of the spline functions. Correlation coefficient : 0.992.*

Pour valider la méthode des fonctions splines dans le cas particulier des données Airparif, il faudrait disposer d'une ou plusieurs distributions réelles de la concentration de polluant sur la région francilienne et comparer le champ estimé par les fonctions splines au champ réel. C'est le seul test qui peut montrer dans quelle mesure on peut reconstituer la distribution spatiale de la pollution dans la région des capteurs Airparif. Une simula-

Tableau II.  
Coefficients de corrélation et traitement des résidus aux points supplémentaires  
choisis sur la carte IGN.

| Méthode     | Coefficient corrélation | Moyenne résidus | Résidu minimum | Résidu maximum | Ecart type des résidus | Moyenne écarts abs |
|-------------|-------------------------|-----------------|----------------|----------------|------------------------|--------------------|
| Pondération | 0,922                   | - 0,018         | - 8,25         | 13,45          | 6,643                  | 5,72               |
| Shepard     | 0,888                   | 2,182           | - 10,75        | 19,21          | 7,861                  | 6,55               |
| Delaunay    | 0,920                   | 2,595           | - 7,17         | 18,98          | 6,767                  | 5,22               |
| Spline      | 0,965                   | 0,861           | - 6            | 9,5            | 4,466                  | 3,81               |

tion de la distribution de pollution pour la région parisienne et la petite couronne est difficile à mettre en œuvre, lorsque les sources majoritaires sont diffuses (par exemple le chauffage urbain), assez nombreuses et leurs émissions ne sont enregistrées que pour des périodes longues (quelques mois).

Malheureusement, on n'a pas les moyens de faire ce test, donc on ne pourra pas répondre sur la fiabilité de la méthode dans ce cas particulier. Donc, notre but est de choisir la méthode qui donne les meilleurs résultats parmi celles qu'on connaît.

a. Une comparaison visuelle des résultats obtenus à partir des données Airparif, en utilisant différentes méthodes d'approximation est présentée dans la figure 5. Les cartes ne peuvent pas être comparées à une carte réelle de pollution, mais seulement entre elles. Dans le cas des fonctions splines, les courbes sont plus lisses, l'aspect de la carte est le plus réaliste.

b. En choisissant la méthode des fonctions splines, on fera quelques tests en ce qui concerne son comportement dans le cas des données Airparif.

b1. On considère les mesures de concentration de  $\text{SO}_2$  enregistrées pour un jour de pointe (on a choisi le 17 Février 1994) par  $n$  stations qui se trouvent dans la zone délimitée par le grand rectangle, figure 6a. En utilisant les valeurs correspondantes aux  $(n-1)$  stations, on calcule une valeur pour la  $n$ -ième station et on fait une comparaison avec la valeur réellement mesurée. Ce test a été appliqué à plusieurs capteurs (dont le code apparaît dans le tableau 3) pour les 24 mesures horaires de ce jour de pointe. Les coefficients de corrélation entre les valeurs mesurées et les valeurs calculées sont présentés dans le tableau III.

Pour l'estimation, on a utilisé entre 18 et 21 points (à cause de l'absence des mesures).

Les résultats sont moins bons que ceux obtenus dans le cas du relief, surtout parce que le nombre de points de départ est moins important dans le cas de la pollution (de 18 à 21 points, par rapport au 60 points). Les résidus sont importants surtout dans les cas des pics, quand on a une sous-évaluation ou dans les cas où la station est entourée des capteurs mesurant des niveaux plus importants, quand on a une surévaluation.

Tableau III.  
Coefficients de corrélation ( $r$ ) entre les valeurs mesurées par Airparif et les valeurs estimées à l'aide des fonctions splines.

| Station | 13F75 | 18F75 | 4F94  | 1F92  | 1F77    | 14F75 | 8F75  | 4F93  | 71F75 |
|---------|-------|-------|-------|-------|---------|-------|-------|-------|-------|
| $r$     | 0,961 | 0,844 | 0,796 | 0,714 | - 0,122 | 0,851 | 0,919 | 0,477 | 0,728 |

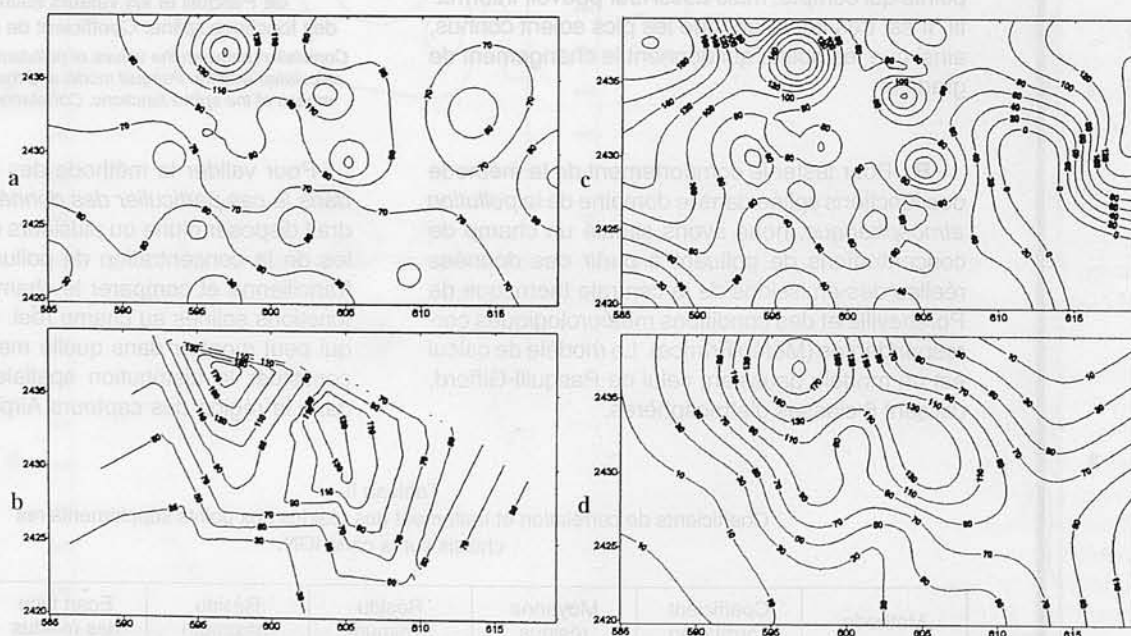
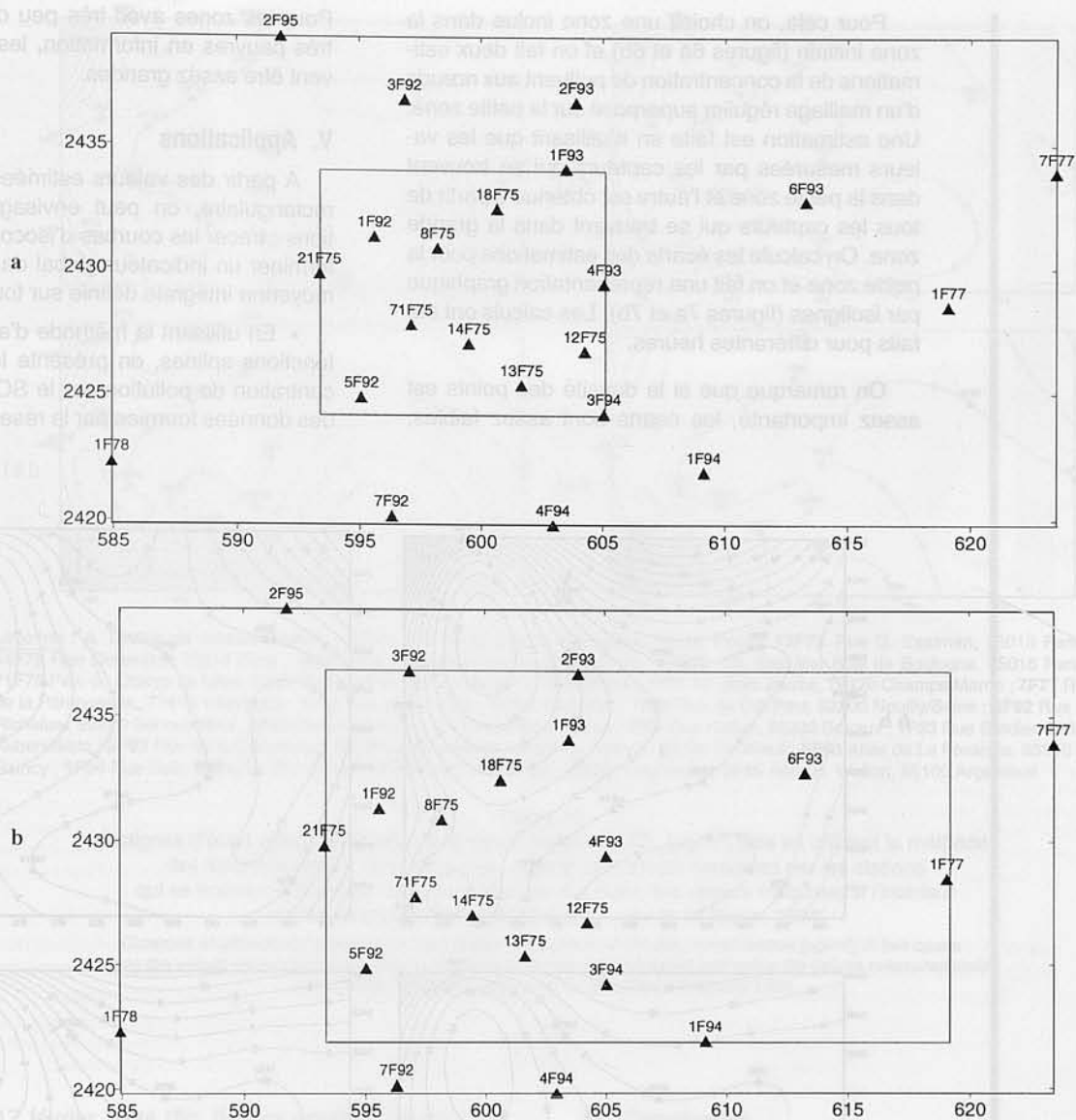


Figure 5.

Champs de concentration de  $\text{SO}_2$  [ $\mu\text{g}/\text{m}^3$ ], le 17 Février 1994, à 15 heures, région parisienne, obtenus en utilisant différentes méthodes d'approximation : a. pondération par rapport aux inverses des distances ; b. Shepard ; c. triangulation de Delaunay avec interpolation linéaire ; d. fonctions splines type plaque mince.

$\text{SO}_2$  concentration fields [ $\mu\text{g}/\text{m}^3$ ], the 17th of February 1994, 3 p.m., Paris region. Methods used : a. inverse distance pondereation method ; b. Shepard's method ; c. Delaunay triangulation method with linear interpolation ; d. spline method.



Légende : ▲ Station de mesure Airparif ; 12F75 Rue de la Brèche aux Loups, 75012 Paris ; 13F75 Rue G. Eastman, 75013 Paris ; 14F75 Rue Delambre, 75014 Paris ; 18F75 Rue Ferdinand Flocon, 75018 Paris ; 21F75 Ch. Gravillier Bois de Boulogne, 75016 Paris ; 71F75 Parc du Champ de Mars, 75007 Paris ; 8F75 Rue de Messine, 75008 Paris ; 1F77 Av. Jean Jaurès, 77420 Champs/Marne ; 7F77 Rue de la Pomponette, 77410 Villevaudé ; 1F78 Rue de Limoges, 78000 Versailles ; 1F92 Rue du Cdt Pilot, 92200 Neuilly/Seine ; 3F92 Rue de Richelieu, 92220 Gennevilliers ; 5F92 Rue Jules Ferry, 92130 Issy/Moulineaux ; 7F92 Rue Hodan, 92330 Sceaux ; 1F93 Rue Bordier, 93300 Aubervilliers ; 2F93 Rue de la Convention, 93120 La Courneuve ; 4F93 Av. Valvein, 93100 Montreuil ; 6F93 Allée de La Fontaine, 93340 Le Raincy ; 1F94 Rue Felix Maire, 94 000 Créteil ; 4F94 Rue P. Armangot, 94400, Vitry/Seine ; 2F95 Allée H. Wallon, 95100 Argenteuil.

Figure 6.

Stations de mesure Airparif et limites des zones utilisées pour tester le comportement de la méthode des fonctions splines en deux cas : quand on utilise seulement les capteurs emplacements dans la zone d'étude et quand on utilise aussi des capteurs situés à l'extérieur de cette zone.

a. zone 1; b. zone 2.

Airparif measurement stations and the limits of the regions used to test the spline method's behaviour in two cases : when we use only the stations which are in the studied region and second, when we use also stations outside the region. a. region 1 ; b. region 2.

En ce qui concerne le capteur 1F77, les estimations ont été très mauvaises, ce qui veut dire que sa présence est essentielle pour caractériser la pollution dans la zone où il est placé ; ses valeurs ne peuvent pas être reconstituées à partir d'autres points de mesure, soit parce que l'information dans

la zone est trop pauvre, soit parce qu'il a un comportement localisé.

b2. On se propose d'étudier la sensibilité de la fonction des limites de la zone d'étude.

Pour cela, on choisit une zone incluse dans la zone initiale (figures 6a et 6b) et on fait deux estimations de la concentration de polluant aux nœuds d'un maillage régulier superposé sur la petite zone. Une estimation est faite en n'utilisant que les valeurs mesurées par les capteurs qui se trouvent dans la petite zone et l'autre est obtenue à partir de tous les capteurs qui se trouvent dans la grande zone. On calcule les écarts des estimations pour la petite zone et on fait une représentation graphique par isolignes (figures 7a et 7b). Les calculs ont été faits pour différentes heures.

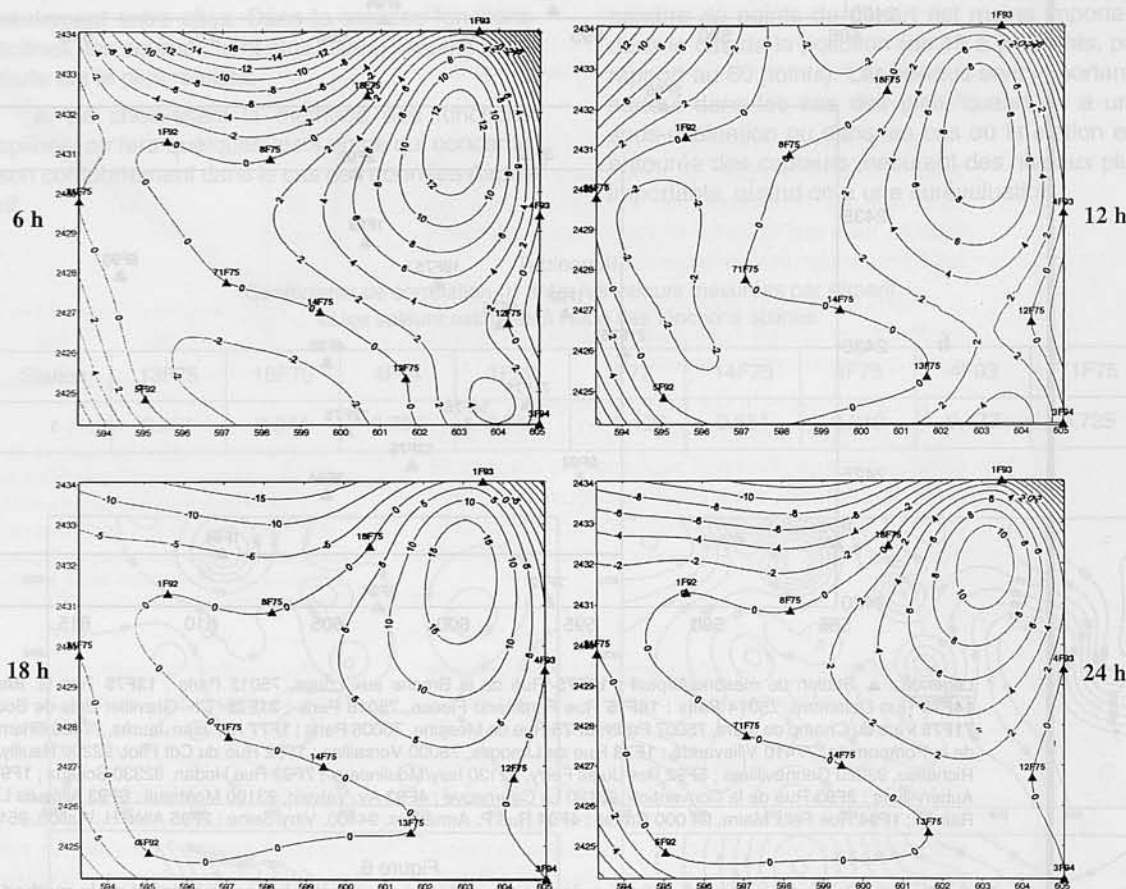
On remarque que si la densité des points est assez importante, les écarts sont assez faibles.

Pour les zones avec très peu de capteurs, donc très pauvres en information, les différences peuvent être assez grandes.

## V. Applications

A partir des valeurs estimées sur un maillage rectangulaire, on peut envisager deux applications : tracer les courbes d'isoconcentration et déterminer un indicateur global de pollution par une moyenne intégrale définie sur toute la région.

- En utilisant la méthode d'approximation des fonctions splines, on présente le champ de concentration de pollution par le  $\text{SO}_2$ , obtenu à partir des données fournies par le réseau Airparif pour le

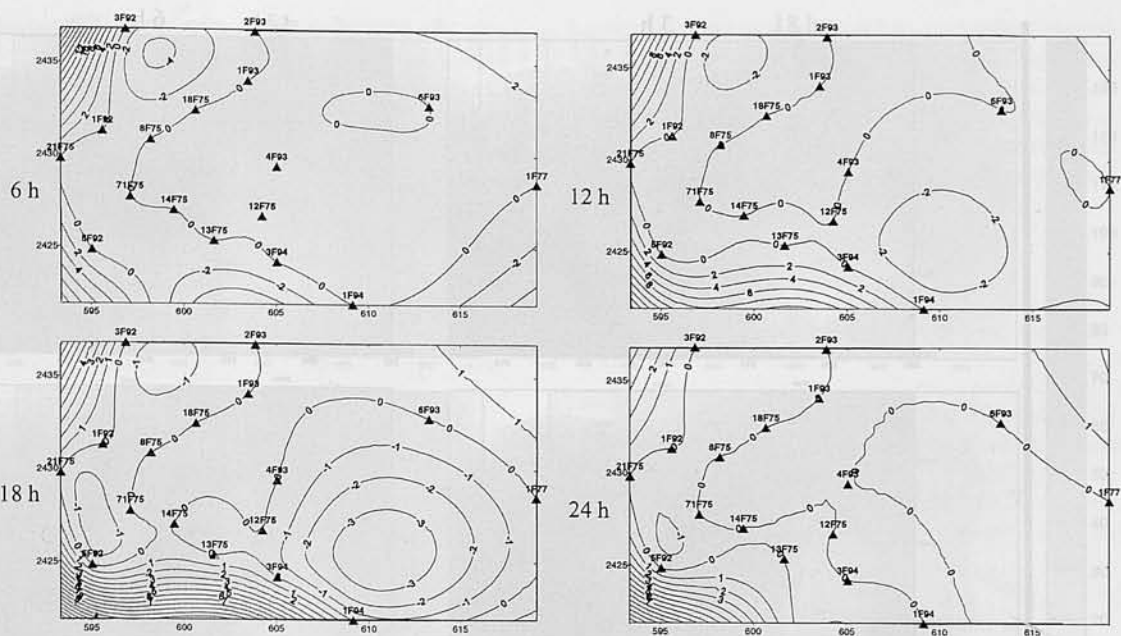


Légende : ▲ Station de mesure Airparif ; 12F75 Rue de la Brèche aux Loups, 75012 Paris ; 13F75 Rue G. Eastman, 75013 Paris ; 14F75 Rue Delambre, 75014 Paris ; 18F75 Rue Ferdinand Flocon, 75018 Paris ; 21F75 Ch. Gravierier Bois de Boulogne, 75016 Paris ; 71F75 Parc du Champ de Mars, 75007 Paris ; 8F75 Rue de Messine, 75008 Paris ; 1F77 Av. Jean Jaurès, 77420 Champs/Seine ; 7F77 Rue de la Pomponette, 77410 Villevaudé ; 1F78 Rue de Limoges, 78000 Versailles ; 1F92 Rue du Cdt Pilot, 92200 Neuilly/Seine ; 3F92 Rue de Richelieu, 92220 Gennevilliers ; 5F92 Rue Jules Ferry, 92130 Issy/Moulineaux ; 7F92 Rue Hodan, 92330 Sceaux ; 1F93 Rue Bordier, 93300 Aubervilliers ; 2F93 Rue de la Convention, 93120 La Courneuve ; 4F93 Av. Valvein, 93100 Montreuil ; 6F93 Allée de La Fontaine, 93340 Le Raincy ; 1F94 Rue Felix Maire, 94 000 Créteil ; 4F94 Rue P. Armangot, 94400, Vitry/Seine ; 2F95 Allée H. Wallon, 95100 Argenteuil.

Figure 7a.

Isolignes d'écart entre l'estimation de la concentration de  $\text{SO}_2$  [ $\mu\text{g}/\text{m}^3$ ] faite en utilisant la méthode des fonctions splines dans deux cas : à partir des valeurs mesurées par les stations qui se trouvent à l'intérieur de la zone d'étude et à partir des valeurs mesurées à l'intérieur et à l'extérieur de cette zone (zone 1), pour le 17 Février 1994.

Contours of differences between the estimations with splines of the  $\text{SO}_2$  concentration [ $\mu\text{g}/\text{m}^3$ ] in two cases : using the values measured by the stations situated inside the studied region and using the values measured inside and outside this region (region 1) for the 17th of February 1994.



Légende : ▲ Station de mesure Airparif ; 12F75 Rue de la Brèche aux Loups, 75012 Paris ; 13F75 Rue G. Eastman, 75013 Paris ; 14F75 Rue Delambre, 75014 Paris ; 18F75 Rue Ferdinand Flocon, 75018 Paris ; 21F75 Ch. Gravière Bois de Boulogne, 75016 Paris ; 71F75 Parc du Champ de Mars, 75007 Paris ; 8F75 Rue de Messine, 75008 Paris ; 1F77 Av. Jean Jaurès, 77420 Champs/Seine ; 7F77 Rue de la Pomponette, 77410 Villevaudé ; 1F78 Rue de Limoges, 78000 Versailles ; 1F92 Rue du Cdt Pilot, 92200 Neuilly/Seine ; 3F92 Rue de Richelieu, 92220 Gennevilliers ; 5F92 Rue Jules Ferry, 92130 Issy/Moulineaux ; 7F92 Rue Hodan, 92330 Sceaux ; 1F93 Rue Bordier, 93300 Aubervilliers ; 2F93 Rue de la Convention, 93120 La Courneuve ; 4F93 Av. Valvein, 93100 Montreuil ; 6F93 Allée de La Fontaine, 93340 Le Raincy ; 1F94 Rue Felix Maire, 94 000 Créteil ; 4F94 Rue P. Armangot, 94400, Vitry/Seine ; 2F95 Allée H. Wallon, 95100 Argenteuil.

Figure 7b.

Isolignes d'écart entre l'estimation de la concentration de  $\text{SO}_2$  [ $\mu\text{g}/\text{m}^3$ ] faite en utilisant la méthode des fonctions splines dans deux cas : à partir des valeurs mesurées par les stations qui se trouvent à l'intérieur de la zone d'étude et à partir des valeurs mesurées à l'intérieur et à l'extérieur de cette zone (zone 2), pour le 17 février 1994.

Contours of differences between the estimations with splines of the  $\text{SO}_2$  concentration [ $\mu\text{g}/\text{m}^3$ ] in two cases : using the values measured by the stations situated inside the studied region and using the values measured inside and outside this region (region 2), for the 17th of February 1994.

17 février 1994 (fig. 8). Les courbes ne sont pas tracées pour toute la région Ile-de-France, couverte par le réseau Airparif, mais seulement sur une zone où la densité des points est satisfaisante. Les cartes obtenues peuvent être utilisées par un logiciel d'animation, en réalisant une cartographie vidéo, qui permet de mettre en évidence la dynamique de la pollution.

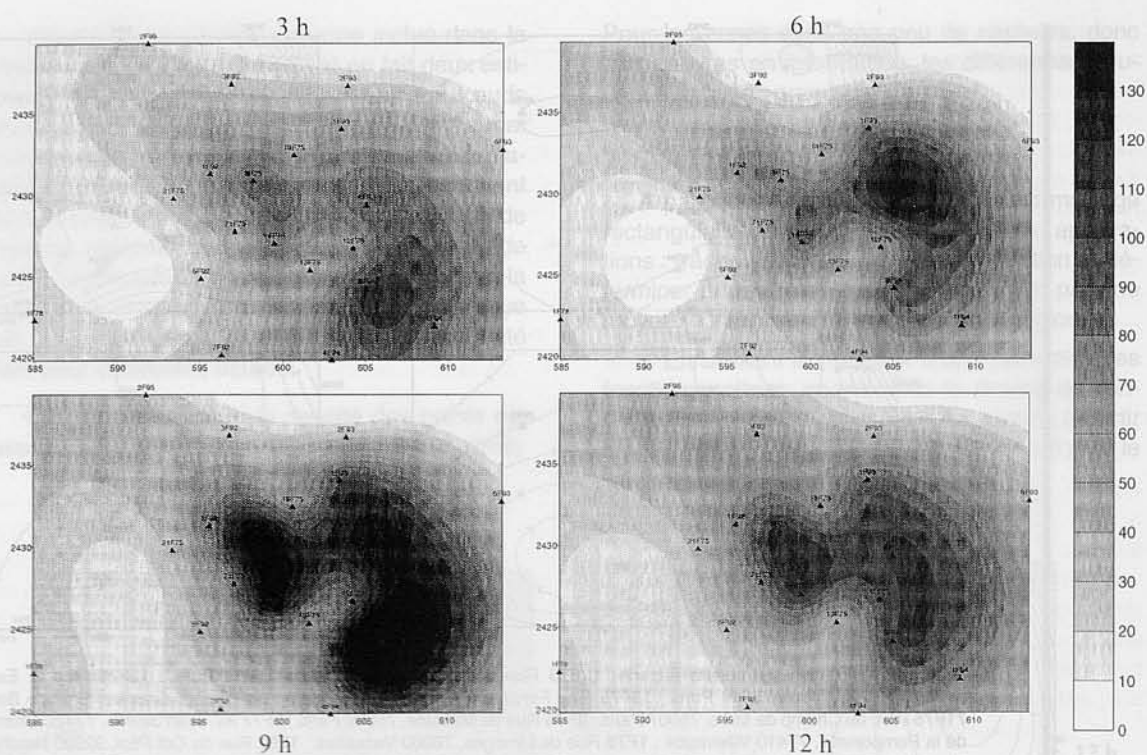
- Pour déterminer la concentration moyenne intégrale, on peut utiliser une méthode classique, comme celle de Simpson ou des trapèzes, qui permet le calcul du volume délimité par la surface d'estimation des concentrations et une base, qui est la surface plane contenant les coordonnées rectangulaires des points. En divisant ce volume par l'aire de la surface plane, on obtient une moyenne intégrale, qui peut être considérée comme indicateur global de la pollution pour toute la région étudiée. On va considérer cet indicateur comme une concentration moyenne. Il pourrait être utilisé pour montrer une variation temporelle globale, donc caractéristique pour toute la zone étudiée.

## VI. Conclusion

- La méthode d'approximation la mieux adaptée à l'estimation du champ de concentration d'un polluant à partir des valeurs discrètes mesurées par le réseau Airparif, est la méthode des fonctions spline type plaque mince. Elle permet d'obtenir une fonction qui correspond à la variation spatiale de la concentration et, en particulier, les valeurs estimées aux nœuds d'un maillage rectangulaire.

- Les performances de la méthode dépendent du nombre de points disponibles et de leur répartition spatiale, donc de leur pouvoir informatif. On ne peut pas estimer dans quelle mesure les distributions de polluants obtenues en utilisant les données Airparif correspondent à la réalité. Les tests faits nous ont permis de faire quelques remarques, mais pour arriver à des conclusions plus solides, un nombre plus important de tests s'impose et c'est l'objet d'une étude en soi.

- A partir des distributions spatiales de concentration de polluant, il est possible d'établir des corrélations liant la pollution et la structure urbaine.



Légende : ▲ Station de mesure Airparif ; 12F75 Rue de la Brèche aux Loups, 75012 Paris ; 13F75 Rue G. Eastman, 75013 Paris ; 14F75 Rue Delambre, 75014 Paris ; 18F75 Rue Ferdinand Flocon, 75018 Paris ; 21F75 Ch. Gravierier Bois de Boulogne, 75016 Paris ; 71F75 Parc du Champ de Mars, 75007 Paris ; 8F75 Rue de Messine, 75008 Paris ; 1F77 Av. Jean Jaurès, 77420 Champs/Seine ; 7F77 Rue de la Pomponette, 77410 Villevaudé ; 1F78 Rue de Limoges, 78000 Versailles ; 1F92 Rue du Cdt Pilot, 92200 Neuilly/Seine ; 3F92 Rue de Richelieu, 92220 Gennevilliers ; 5F92 Rue Jules Ferry, 92130 Issy/Moulineaux ; 7F92 Rue Hodan, 92330 Sceaux ; 1F93 Rue Bordier, 93300 Aubervilliers ; 2F93 Rue de la Convention, 93120 La Courneuve ; 4F93 Av. Valvein, 93100 Montreuil ; 6F93 Allée de La Fontaine, 93340 Le Raincy ; 1F94 Rue Felix Maire, 94 000 Créteil ; 4F94 Rue P. Armangot, 94400, Vitry/Seine ; 2F95 Allée H. Wallon, 95100 Argenteuil.

Figure 8a.

Champs de concentration de  $\text{SO}_2$  [ $\mu\text{g}/\text{m}^3$ ], région parisienne, le 17 Février 1994 et les stations de mesure Airparif. Les distributions ont été obtenues à partir des valeurs mesurées par ces stations, utilisant la méthode des fonctions splines. Pour les valeurs égales ou supérieures à  $130 \mu\text{g}/\text{m}^3$  est utilisée la même nuance (la plus foncée).

$\text{SO}_2$  concentration fields [ $\mu\text{g}/\text{m}^3$ ], Paris region, on the 17th of February 1994 and Airparif measurement stations. The distributions are obtained by means of the spline functions' method, using the values measured by these stations. For the values greater or equal to  $130 \mu\text{g}/\text{m}^3$ , the same tone is used (the darkest).

Les champs de concentration de polluant obtenus à partir des observations de terrain peuvent permettre la correction et la validation des données du satellite.

### Remerciements

Les auteurs tiennent à remercier M. Ph. Lameloise, M. Renaudot, M. Ch. Bournel Bosson du Réseau Airparif, pour les remarques et la mise à disposition des données de concentration de  $\text{SO}_2$  et M. Frédéric Delay pour son aide concernant la méthode de Krigage.

### Bibliographie

AVILA A., Dispersion de  $\text{SO}_2$  dans l'atmosphère autour d'une centrale thermoélectrique à Porcheville, DEA SPI, Université Paris XII, LETIEF, 1993.

BAKHVALOV N., Méthodes numériques, Edition Mir, Moscou, 1976.

BERSTEL J., PIN J. E., POCCHIOLA M., Triangulation de Delaunay, Mathématique et Informatique. Problèmes résolus, v. 2, Combinatoire et arithmétique, Mc Graw-Hill, 1991.

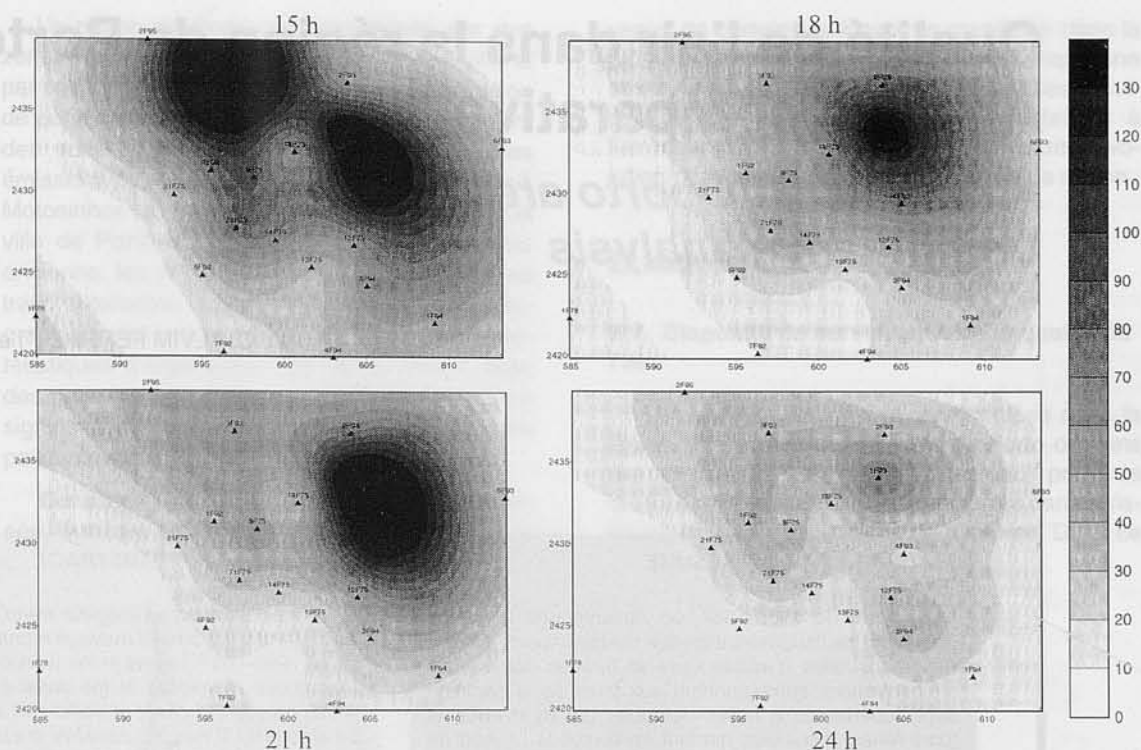
BESSEMULIN P., Urbanisation et météorologie, la météorologie, VI<sup>e</sup> série, n° 23, 1980.

CARASSO C., Lissage de données à l'aide de fonctions splines, Analyse numérique, Hermann, Paris, 1991, 357-415.

CHRISTAKOS G., THESING A., The Intrinsic Random Field Model in the Study of Sulfate Deposition Processes. Atmospheric Environment Vol. 27A, No. 10, 1993, 1521-1540.

CRESSIE N., Statistics for Spatial Data. J. Wiley, New York, 1991.

DEUTSCH C. V., JOURNEL A.G., Geostatistical Software Library and User's Guide, Oxford University Press, 1992.



Légende : ▲ Station de mesure Airparif ; 12F75 Rue de la Brèche aux Loups, 75012 Paris ; 13F75 Rue G. Eastman, 75013 Paris ; 14F75 Rue Delambre, 75014 Paris ; 18F75 Rue Ferdinand Flocon, 75018 Paris ; 21F75 Ch. Graviillier Bois de Boulogne, 75016 Paris ; 71F75 Parc du Champ de Mars, 75007 Paris ; 8F75 Rue de Messine, 75008 Paris ; 1F77 Av. Jean Jaurès, 77420 Champs/Seine ; 7F77 Rue de la Pomponette, 77410 Villevaudé ; 1F78 Rue de Limoges, 78000 Versailles ; 1F92 Rue du Cdt Pilot, 92200 Neuilly/Seine ; 3F92 Rue de Richelieu, 92220 Gennevilliers ; 5F92 Rue Jules Ferry, 92130 Issy/Moulineaux ; 7F92 Rue Hodan, 92330 Sceaux ; 1F93 Rue Bordier, 93300 Aubervilliers ; 2F93 Rue de la Convention, 93120 La Courneuve ; 4F93 Av. Valvein, 93100 Montreuil ; 6F93 Allée de La Fontaine, 93340 Le Raincy ; 1F94 Rue Felix Maire, 94 000 Créteil ; 4F94 Rue P. Armandot, 94400, Vitry/Seine ; 2F95 Allée H. Wallon, 95100 Argenteuil.

Figure 8b.

Champs de concentration de  $\text{SO}_2$  [ $\mu\text{g}/\text{m}^3$ ], région parisienne, le 17 Février 1994 et les stations de mesure Airparif. Les distributions ont été obtenues à partir des valeurs mesurées par ces stations, utilisant la méthode des fonctions splines. Pour les valeurs égales ou supérieures à  $130 \mu\text{g}/\text{m}^3$  est utilisée la même nuance (la plus foncée).

$\text{SO}_2$  concentration fields [ $\mu\text{g}/\text{m}^3$ ], Paris region, on the 17th of February 1994 and Airparif measurement stations. The distributions are obtained by means of the spline functions' method, using the values measured by these stations. For the values greater or equal to  $130 \mu\text{g}/\text{m}^3$ , the same tone is used (the darkest).

FRANKE R., NIELSEN G., Smooth Interpolation of Large Sets of Scattered Data, International Journal for Numerical Methods in Engineering, v.15, 1980, 1691-1704.

IONESCU A., Analyse de la dispersion des polluants en milieu urbain à partir des données Airparif, DEA CMPI, Université Paris XII, LETIEF, 1994.

JOURNAL A. G., HUIJBREGTS Ch., Mining Geostatistics. Academic Press, New York, 1978.

LAMELLOISE Ph., THIBAUT G., PETIT-COVIAUX F., La modernisation du dispositif de surveillance de la qualité

de l'air en Ile de France 1989-1993. Pollution atmosphérique, juillet-septembre 1991, 418-429.

RISLER J. J., Méthodes mathématiques pour la CAO, Masson Paris, 1991.

### Mots clés

Pollution atmosphérique ; Méthodes d'interpolation et de lissage ; Champs de concentration de polluants.