

ARTICLES

Représentativité spatiale d'une station de mesure de la pollution atmosphérique

Spatial representativeness of an air pollution measurement

Michel BOBBIA*, Anne CORI**, Chantal de FOUQUET**

Résumé

La notion de représentativité est mentionnée dans de nombreux textes législatifs sur la qualité de l'air. Une définition de la représentativité spatiale d'une station de mesure de la pollution de l'air par le dioxyde d'azote est proposée et testée sur l'agglomération rouennaise pour l'année 2005. Partant d'une notion empirique – une station de mesure est représentative d'un point géographique si les concentrations en ces deux sites sont « suffisamment proches » – nous proposons une formalisation mathématique de la représentativité. La technique du krigeage permet de ramener le problème de la délimitation de la zone de représentativité d'une station à la majoration d'une somme de deux termes : le premier est l'écart entre la moyenne annuelle de la concentration mesurée à la station et son estimation par krigeage au point courant, le second fait intervenir l'écart-type de l'erreur d'estimation. La mise en œuvre pratique fournit la zone de représentativité de chaque station. Certaines de ces zones peuvent se chevaucher ; en tout point des zones de recouvrement, on détermine alors la station la plus représentative. Enfin, la notion de représentativité est examinée en relation avec la probabilité de dépassement d'un seuil fixé.

Mots clés

Représentativité. Cartographie. Géostatistique. Variabilité spatiale. Pollution de l'air. Échantillonnage.

Abstract

The notion of representativeness is mentioned in many legislative texts on air quality monitoring. The work presented here aims at defining the spatial representativeness of a measurement station with regard to the annual average of NO₂, and testing this definition at Rouen urban area for the year 2005.

The work is carried out in the spatial probabilistic framework of geostatistics, which is well suited for such tasks. An intuitive definition of representativeness is written as follows: a measurement station is representative of a given geographical location if the concentrations at these sites are sufficiently similar. But pollution is only measured at the station, and remains unknown outside. The kriging technique gives an estimation of the pollution in each point and the standard deviation of the estimation error. Using these values, and a hypothesis of a normal-distributed-error, we can, for each station, select the points that satisfy the notion of representativeness. There will then be a good chance (at a fixed risk) that the concentrations within the "zone of representativeness" do not differ with more than a fixed deviation from the concentration measured at the station.

The practical implementation of these theoretical definitions is to produce maps of areas representative of the stations. It is also possible to determine, for any point of a region, the most representative station and its associated deviation (by assigning an acceptable risk to it). Lastly, for each area, it is also possible to indicate the probabilities of overshooting a threshold.

Keywords

Representativeness. Mapping. Geostatistics. Spatial variability. Air pollution. Sampling.

* Air Normand, 3 place de la pomme d'or, 76000 Rouen – E-mail : michel.bobbia@airnormand.fr

** École des Mines de Paris, Centre de géostatistique, 35 rue Saint-Honoré, 77305 Fontainebleau.

1. Introduction

En langage commun, la représentativité qualifie la qualité d'un effectif réduit à reproduire, d'une certaine façon, les caractéristiques d'une population plus vaste. En statistique inférentielle, le problème central consiste à déterminer certaines propriétés d'une population complète à partir d'un nombre de données réduit, ce qui n'est possible que si l'échantillon disponible est **représentatif**.

Cette notion de représentativité intervient en particulier dans les sondages ou pour la maîtrise des processus : à quelles conditions un échantillon peut-il être supposé représentatif d'une population ou d'un lot ? En chimie, les plans d'expérience visent à optimiser la représentativité de l'échantillon.

La notion de représentativité apparaît dans de nombreux textes légiférant la surveillance de la qualité de l'air. L'arrêté de surveillance de la qualité de l'air [1] impose dans son article 7 que les AASQA s'assurent que les mesures respectent les objectifs de qualité avec la représentativité demandée dans les directives filles [2]. Par exemple, l'annexe VI de la première directive fille précise que les sites de mesure doivent être localisés « de manière à fournir des renseignements sur les concentrations dans d'autres endroits d'une zone qui sont représentatifs du niveau d'exposition de la population générale ».

Conformément aux directives européennes, le guide ADEME de classification et d'implantation des stations [3], mentionne que « d'une manière générale, les points de prélèvements devraient être représentatifs, dans la mesure du possible, de sites similaires ne se trouvant pas à proximité immédiate ». Il reprend ensuite les aires de représentativité définies par l'agence européenne de l'environnement dans le cadre du réseau EUROAIRNET [4].

L'arrêté ATMO [5] stipule dans son article 5b que pour calculer les indices ATMO de la qualité de l'air sur une zone géographique, l'AASQA concernée sélectionne des stations fixes de telle sorte que la moyenne des mesures réalisées par ces stations soit représentative des concentrations et de leur évolution sur l'ensemble de la zone.

Cependant, aucune définition précise de la représentativité n'est donnée, et aucun document technique auquel les textes réglementaires pourraient renvoyer n'est actuellement disponible. Enfin, il faut compter avec les critiques du public : si l'on peut faire admettre de ne pas tout mesurer partout, il est clair que l'on ne mesure pas là où il faudrait, c'est-à-dire en particulier à proximité immédiate du domicile de chaque demandeur...

Nous traitons ici de la représentativité spatiale : existe-t-il une zone dont un site de mesure serait représentatif ? Autrement dit, peut-on délimiter une zone dans laquelle les concentrations auraient les mêmes propriétés, ou des propriétés analogues, à celles mesurées en une station, qui pourrait ainsi « représenter » cette zone ?

Il nous semble que la représentativité spatiale devrait dépendre de l'objectif visé ; pour un polluant donné, la représentativité spatiale d'une station de mesure n'est pas nécessairement identique pour la moyenne annuelle et pour une moyenne horaire. On peut en effet imaginer que la direction du vent détermine une zone de représentativité qui s'allonge « sous le vent », et donc variable avec la période de référence (annuelle, journalière ou horaire).

Formalisant une idée empirique, nous cherchons à définir la représentativité spatiale d'une station de mesure vis-à-vis de la moyenne annuelle en dioxyde d'azote NO₂ sur une agglomération. À cet effet, le cadre probabiliste spatial proposé par la géostatistique se montre particulièrement bien adapté. Comme illustration de l'approche théorique développée, nous examinons le cas du dioxyde d'azote sur l'agglomération rouennaise durant l'année 2005.

La première approche de la représentativité développée ici concerne la moyenne annuelle, qui est comparée à la valeur limite européenne pour le NO₂ : la directive 1999/30/CE demande notamment à ce que cette moyenne ne dépasse pas 40 µg/m³ à partir du 1^{er} janvier 2010, une marge de dépassement étant tolérée avant cette date. Les autres seuils réglementaires sont plus compliqués à traiter ; ils font l'objet d'un travail en cours dans la continuité de celui-ci.

Dans la suite, sauf mention contraire, le terme « concentration » désigne la moyenne annuelle de la concentration en dioxyde d'azote.

2. Objectifs et méthode

2.1 Définition de la zone de représentativité d'une station

La notion empirique de départ est la suivante : **une station de mesure est représentative d'un point géographique donné si les concentrations en ces deux sites sont suffisamment proches**. La zone de représentativité de la station est alors constituée par l'ensemble des points dont la concentration diffère peu de celle à la station.

Cette définition de la représentation procède bien d'un choix ; d'autres critères sont possibles, prenant par exemple en compte des dépassements de seuil ou la « dynamique » du phénomène. Une station pourrait être dite représentative des endroits où les chroniques présentent les mêmes variations temporelles, même si les niveaux moyens (la moyenne annuelle de la concentration) diffèrent. On pourrait aussi retenir les points où la concentration est toujours supérieure – ou à l'inverse inférieure – à celle mesurée à la station.

Nous retenons la définition suivante : l'écart entre la concentration en tout point de la zone de représentativité d'une station, et celle mesurée à la station, reste inférieur à une valeur « delta » fixée. Mathématiquement, notons x_0 les coordonnées de la station

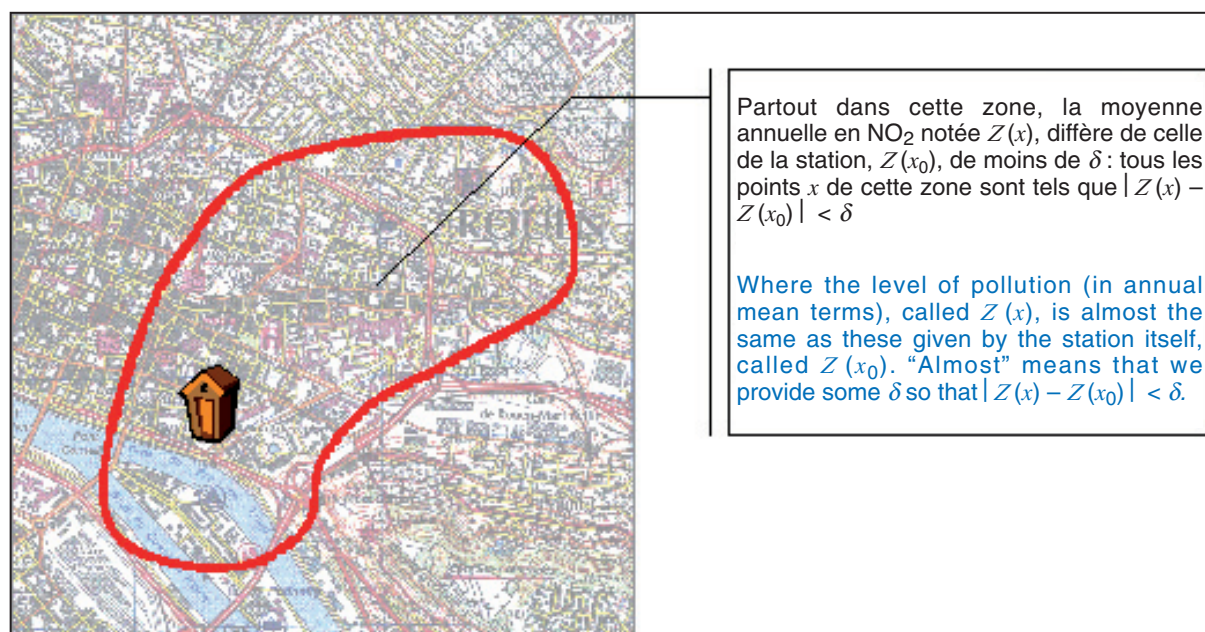


Figure 1.

Zone de représentativité de la station :

par définition, la pollution dans cette zone (en moyenne sur l'année) est égale, à δ près, à celle indiquée par la station.

Spatial area of representativeness of the air pollution measurement station is defined as all the points x around the station (in x_0) where the level of pollution (in annual mean terms), called $Z(x)$, is almost the same as these given by the station itself, called $Z(x_0)$. "Almost" means that we provide some δ so that $|Z(x) - Z(x_0)| < \delta$.

dans le plan, x le point géographique courant et $Z(x)$ la concentration en ce point, $Z(x_0)$ la concentration à la station ; la condition précédente s'écrit :

$$|Z(x) - Z(x_0)| < \delta \quad (1)$$

δ désignant une valeur que l'on se fixe (Figure 1). La zone de représentativité $A(x_0)$ de la station est alors définie par :

$$A(x_0) = \{x : |Z(x) - Z(x_0)| < \delta\}$$

2.1.1. Un peu de géostatistique

Dans le cadre géostatistique, la concentration est vue comme une réalisation d'une fonction aléatoire [6]. Au facteur $1/2$ près, le variogramme représente la moyenne de l'écart quadratique des concentrations mesurées en deux points, en fonction de leur distance (Figure 2). Le variogramme est un indicateur du degré de liaison spatiale. Dans certains cas, le variogramme atteint un « palier » : il reste constant aux distances supérieures à la « portée », qui correspond à la distance au-delà de laquelle la corrélation spatiale est nulle ou négligeable [6].

Or, nous cherchons la zone de représentativité $A(x_0)$ de la station implantée en x_0 . Bien sûr, le problème ne se résout pas directement, car la concentration, qui n'est mesurée qu'aux stations, reste inconnue ailleurs. À cause de l'erreur d'estimation, la concentration réelle $Z(x)$ diffère de son interpolation $Z^*(x)$, la différence provenant de l'erreur d'estimation $\varepsilon(x)$:

$$Z(x) = Z^*(x) + \varepsilon(x)$$

Pour que l'inégalité (1) soit satisfaite, il suffit que :

$$|Z^*(x) - Z(x_0)| + |\varepsilon(x)| < \delta$$

ou encore :

$$|\varepsilon(x)| < b(x) \text{ avec } b(x) = \delta - |Z^*(x) - Z(x_0)| \quad (2)$$

Le krigeage est un interpolateur dit « exact », c'est-à-dire qu'en tout point expérimental x_0 la valeur estimée est égale à la valeur mesurée : $Z^*(x_0) = Z(x_0)$; en ce point, l'erreur d'estimation est nulle : $\varepsilon(x_0) = 0$. Pour un tel interpolateur, l'inégalité (1) est donc vérifiée à la station ($x = x_0$) qui appartient de fait à sa propre zone de représentativité.

Si $b(x)$ est calculable, $\varepsilon(x)$ reste inconnu. Pour obtenir une formule applicable en pratique, nous allons raisonner par conditions suffisantes, **en majorant la valeur absolue de l'erreur d'estimation**. Nous cherchons à déterminer un ensemble $B(x_0)$ dans lequel l'inégalité (2) est vérifiée partout sauf en quelques points (c'est-à-dire à un certain risque statistique R près) et en-dehors duquel cette inégalité peut rester vérifiée localement.

Pour un écart δ admissible fixé, soit R le risque statistique, défini comme la probabilité que l'erreur ait une amplitude supérieure à la valeur $b(x)$:

$$R = P(|\varepsilon(x)| \geq b(x)) \quad (3)$$

Alors, pour tout point x de $B(x_0)$

$$P(|Z(x) - Z(x_0)| \geq \delta) \leq R,$$

Naturellement, la zone de représentativité empirique $B(x_0)$ ainsi obtenue sera d'autant meilleure qu'elle sera proche de la zone de représentativité réelle $A(x_0)$. C'est le cas si l'amplitude de l'erreur d'estimation $\varepsilon(x)$ est réduite et si l'on choisit un risque statistique R faible.

Parmi les estimateurs, on retient le krigeage*, qui est une combinaison linéaire sans biais et optimale des mesures disponibles, au sens où la variance de l'erreur d'estimation est minimale. En-dehors des stations, l'erreur $\varepsilon(x)$ reste inconnue, mais le krigeage en fournit l'espérance, nulle d'après la contrainte de non biais, ainsi que l'écart-type, qui est l'écart-type de krigeage, noté $\sigma(x)$.

Rappelons que le « quantile d'ordre α », d'une loi de probabilité est la valeur ayant la probabilité $1 - \alpha$ d'être dépassée. Notons q_α le quantile d'ordre α d'une variable gaussienne réduite (c'est-à-dire centrée et de variance unité). En particulier, le quantile d'ordre 0.975 est 1.96.

Pour une variable aléatoire gaussienne Y centrée et de variance σ^2 , et un ordre α supérieur à 0.5, on a :

$$P(|Y| < q_\alpha \cdot \sigma) = 2\alpha - 1 ;$$

en particulier, la variable gaussienne Y appartient à l'intervalle $[-1.96\sigma ; 1.96\sigma]$ avec la probabilité de 0.95 ; en effet

$$P(|Y| < 1.96\sigma) = 0.95$$

Supposons alors que l'erreur de krigeage $\varepsilon(x)$ est gaussienne, et considérons un risque R symétrique. Pour que la relation (3) soit vérifiée, il suffit de choisir :

$$b(x) = q_\alpha \cdot \sigma(x) \text{ avec } \alpha = 1 - R/2.$$

On obtient ainsi l'implication suivante :

$$\begin{aligned} q_\alpha \cdot \sigma(x) + |Z^*(x) - Z(x_0)| &\leq \delta \\ \Rightarrow \\ P(|Z(x) - Z(x_0)| < \delta) &\geq 2\alpha - 1 \end{aligned} \quad (4)$$

Concrètement, pour définir la zone de représentativité, on se fixe un risque d'erreur R , et on en déduit le quantile gaussien q_α avec $\alpha = 1 - R/2$; **à ce risque R près**, on sélectionne l'ensemble $B(x_0)$ des points x qui vérifient l'inégalité du premier membre de (4), dans laquelle tout est fixé, mesuré ou calculable par interpolation géostatistique : ils auront alors « de bonnes chances » (la probabilité $P = 1 - R$) que l'écart de leur concentration à celle de la station reste suffisamment petit, c'est-à-dire inférieur à δ .

2.1.2. Interprétation

D'après le second membre de l'implication (4), la probabilité que l'écart avec la concentration mesurée à la station reste inférieur à δ dépasse un seuil statis-

tique fixé. Par exemple, pour un risque de 5 % et $\delta = 5$, on pourra dire avec une probabilité supérieure à 95 %, que dans la zone de représentativité empirique $B(x_0)$ la concentration ne diffère pas de celle à la station de plus de 5 $\mu\text{g}/\text{m}^3$. On retient ainsi les points dont les concentrations sont assez proches de celle de la station, avec un certain risque « acceptable » de se tromper – risque lié au fait que, en dehors des stations, nous utilisons des estimations en lieu et place de mesures.

Classiquement, on retient un risque R statistique de 5 %. On a alors $\alpha = 0.975$, ce qui correspond au quantile gaussien $q_\alpha = 1.96$. L'erreur de krigeage reste inconnue, mais on sait, avec une probabilité supérieure à 95 %, que dans $B(x_0)$ elle reste inférieure à $1.96\sigma(x)$.

Pour constituer la zone de représentativité, on calcule en tout point x du domaine :

$$\zeta(x) = 1.96\sigma(x) + |Z^*(x) - Z(x_0)| ;$$

il suffit ensuite de conserver les points x tels que :

$$\zeta(x) \leq \delta \quad (5)$$

Ici $Z(x_0)$ est connue : c'est la moyenne annuelle mesurée à la station ; $Z^*(x)$ et $\sigma(x)$ sont calculés pour tout x par krigeage ponctuel à partir des mesures éparpillées dont on dispose, et δ est fixé. On peut donc déterminer l'ensemble des points x de l'espace qui vérifient l'inégalité du premier membre de (4). Ces points constituent la zone de représentativité de la station pour la moyenne annuelle en NO_2 , zone dans laquelle on pourra dire que la pollution est fournie par la station à plus ou moins delta près, ce qui est évidemment intéressant quand δ reste petit.

Remarque : L'hypothèse d'une loi gaussienne pour l'erreur d'estimation est classique, mais elle n'est pas nécessairement vérifiée expérimentalement. Cette hypothèse peut être considérée comme conventionnelle.

Le calcul développé précédemment peut être étendu sans poser l'hypothèse de loi gaussienne des erreurs d'estimation, en utilisant l'inégalité de Bienaymé-Tchebychev [7] pour majorer le quantile. Mais les zones de représentativité obtenues dans ce cas sont trop restreintes, et deviennent inutilisables.

2.2 Rattachement d'un point quelconque à une station

Les calculs étant effectués par station, les zones de représentativité obtenues peuvent se superposer partiellement. Un point du territoire peut donc être représenté par plusieurs stations, mais plus ou moins bien. Cela ouvre diverses perspectives, comme l'étude des redondances entre stations et l'optimisation de leur implantation. On peut alors inverser le point de vue : au lieu de chercher les points dont une station est représentative, nous pouvons réciproquement

* Plus précisément, il s'agit ici du krigeage « à moyenne inconnue ».

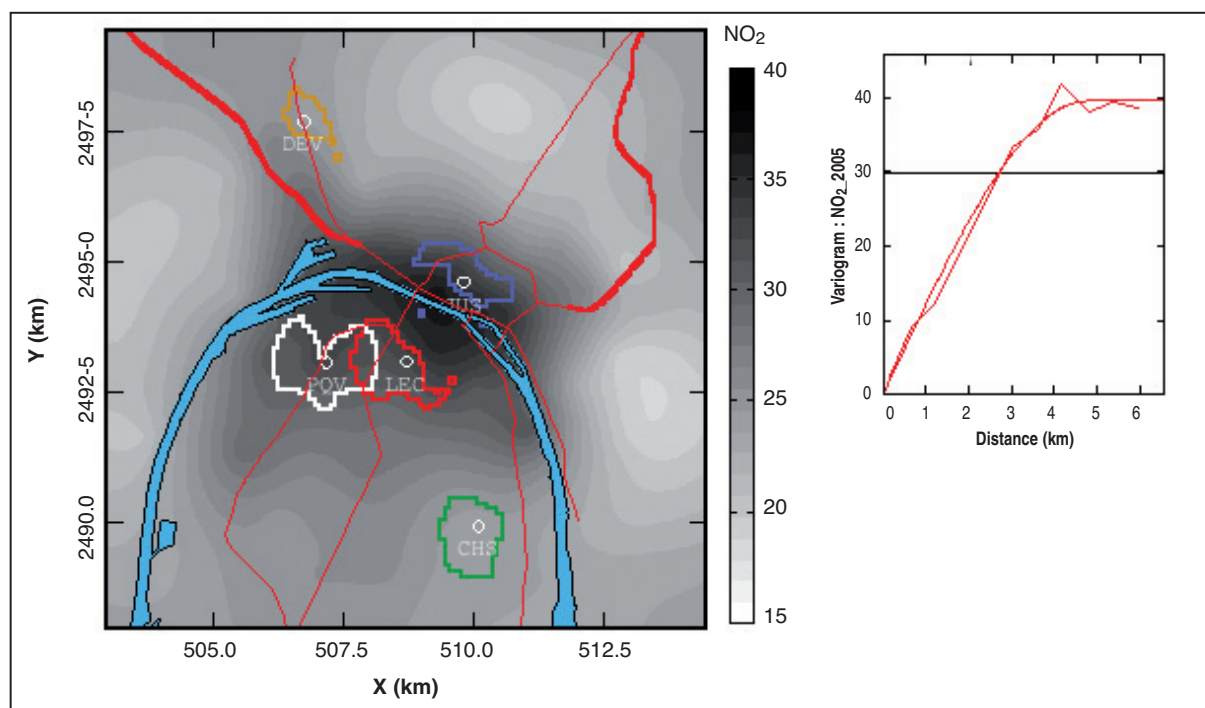


Figure 2.

Zones de représentativité pour l'année 2005 des stations rouennaises (figurées par les ronds), selon les choix suivants : $\delta = 6 \mu\text{g}/\text{m}^3$ et $P = 0.95$ (risque de 5 %). En fond est présenté la cartographie du NO_2 annuel moyen, en $\mu\text{g}/\text{m}^3$, établie avec le variogramme présenté en haut à droite.

La Seine et les principaux axes routiers sont reportés sur la carte pour une meilleure localisation.

Areas represented by the measurement stations (shown by circles) over Rouen city in 2005, with respect to annual NO_2 , on the basis of difference less than $6 \mu\text{g}/\text{m}^3$ and assessment risk of 5%. In the background of this map, we present the annual mean of NO_2 , expressed in $\mu\text{g}/\text{m}^3$. This map was established with the variogram shown in top-right corner. Seine River and main roads are placed on the map for easier localisation.

déterminer pour tout point du territoire la station la plus représentative, et l'écart η associé, à un risque statistique près.

En tout point x , on cherche la station notée x_{i_0} telle que $\zeta(x, x_i) = q_\alpha \cdot \sigma(x) + |Z^*(x) - Z(x_i)|$ est minimal (pour i parcourant les n stations), ce qui se note $x_{i_0} = \arg \min_{1 \leq i \leq n} \zeta(x, x_i)$.

$\eta = \zeta(x, x_{i_0})$ est un majorant de l'écart à la station de rattachement, au risque statistique choisi. η dépend de x ; notons-le alors η_x . Chaque point de l'espace est rattaché à une seule station (i_0), celle correspondant au plus petit majorant de l'écart.

En chaque point x , on peut ainsi dire quelle station est représentative à moins de $\eta_x \mu\text{g}/\text{m}^3$ près, avec un risque α de se tromper. Pour conserver des résultats raisonnables, on peut restreindre géographiquement les zones de représentativité à un rayon égal à la portée du variogramme, ou fixer une borne supérieure à $\zeta(x, x_{i_0})$. Dans ce cas, il peut parfois être impossible de trouver une station de rattachement pour certains points. Ces points non rattachables correspondent notamment aux zones sous-échantillonnées, ou à des milieux plus ou moins rares (dans le cas d'une variable auxiliaire, voir §2.4 ci-dessous) auxquels aucune station ne correspond.

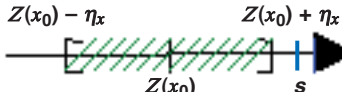
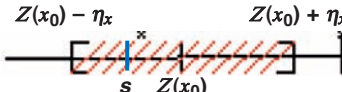
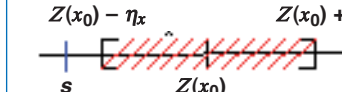
2.3 Risque de dépassement d'une valeur seuil

Au bout du compte, on sait que la concentration $Z(x)$ en un point x appartient à l'intervalle $[Z(x_0) - \eta_x; Z(x_0) + \eta_x]$ avec une probabilité P fixée. Par rapport à un seuil s donné (la valeur limite fixée par la réglementation par exemple), on peut dire que la concentration réelle en x est inférieure (respectivement supérieure) à s avec la probabilité P si $s > Z(x_0) + \eta_x$ (respectivement si $s < Z(x_0) - \eta_x$); si s se situe entre ces deux valeurs, alors on ne peut rien dire. Le tableau 1 résume les situations possibles.

La question du dépassement de seuil conduit à une autre définition de la zone de représentativité : une station est représentative d'une situation relativement à un seuil fixé (par exemple, une valeur limite). Une station est représentative de concentrations faibles, relativement à un seuil fixé, si dans sa zone de représentativité, les concentrations restent partout inférieures au seuil, à un risque statistique près ; inversement, la station peut être représentative de lieux où la pollution dépasse le seuil. Entre les deux, les incertitudes ne permettent pas de conclure : lorsque la concentration mesurée à la station est proche du seuil, on ne saura pas dire dans quelle zone ce seuil est ou non dépassé...

Tableau 1.

Situation de la pollution en un point x quelconque vis-à-vis d'un seuil.
 Situation of pollution in an unspecified point with respect to a threshold.

		
La moyenne annuelle en x est probablement dans l'intervalle, donc elle se situe en dessous du seuil : $Z(x) < s$, au risque 5 % de se tromper. <i>Pollution under threshold with 5 % risk</i>	La moyenne annuelle en x est probablement dans l'intervalle, donc on ne peut pas la situer par rapport au seuil. <i>Indeterminate.</i>	La moyenne annuelle en x est probablement dans l'intervalle, donc elle se situe au-dessus du seuil : $Z(x) > s$, au risque 5 % de se tromper. <i>Pollution over threshold with 5 % risk.</i>

Remarque : dans la modélisation précédente, l'intervalle de probabilité est symétrique autour de l'estimation par krigeage. Les bornes ainsi obtenues ne sont pas toujours réalistes, la borne inférieure pouvant par exemple devenir négative. Une modélisation plus fine est possible par la géostatistique non linéaire, les modèles mis en œuvre étant utilisables pour effectuer des simulations [9].

2.4 Faut-il limiter l'extension des zones de représentativité des stations ?

Au-delà de la portée du variogramme, les concentrations ne sont plus corrélées spatialement. Dans ces conditions, considérer une station comme représentative d'un point situé à une distance supérieure à la portée revient à poser des hypothèses fortes sur la loi de probabilité des concentrations en ces deux points.

Supposons la concentration stationnaire, au moins localement : le variogramme présente alors un palier, égal à la variance $C(0)$ des concentrations, qui est supposée constante dans le domaine étudié (Annexe 1). Dans ce cas, la variance de krigeage est majorée par $2C(0)$, et une condition suffisante pour que le premier membre de (4) soit vérifié est que :

$$2q_{\alpha} \cdot C(0) \leq \delta - |Z^*(x) - Z(x_0)|.$$

Logiquement, tout revient alors à comparer la variance de la loi de probabilité des concentrations, c'est-à-dire la dispersion autour de la moyenne, à l'écart admissible δ ainsi qu'à l'écart entre la concentration mesurée à la station et la concentration estimée. Il paraît pertinent que la zone de représentativité ne puisse dépasser la portée que lorsque la variance des concentrations reste « modérée ».

En pratique, restreindre le rayon de la zone de représentativité à la portée du phénomène évite d'étendre exagérément la zone de représentativité, jusqu'à inclure des sites situés dans des milieux très différents de celui dans lequel la station est implantée. Cela conduit à travailler à une échelle adaptée (dans notre cas, l'agglomération et l'année entière) en fonction des mesures disponibles (les stations fixes, toutes de même typologie de fond urbain). Quelques précisions sur la limitation de l'étendue de la zone de représentativité en fonction du variogramme sont fournies en annexe 1.

On remarque qu'en deux sites, suffisamment éloignés l'un de l'autre pour garantir l'absence de corrélation spatiale des concentrations, le phénomène peut tout de même se comporter de façon relativement identique. En effet, un milieu et un climat similaires devraient induire des niveaux de pollution assez semblables, par exemple dans les zones périurbaines de certaines agglomérations, qui sont en général des quartiers résidentiels. La géostatistique permet de tenir compte d'une analogie de milieux, décrits par une ou plusieurs variables auxiliaires qui sont introduites dans la modélisation [8]. Cette approche, présentée succinctement à l'annexe 2, a pour effet d'étendre les zones de représentativité, comme nous le verrons. Pour la mettre en œuvre, il faut connaître (partout si possible) des variables dites « auxiliaires », **suffisamment corrélées à la pollution**. La difficulté consiste à bien caractériser les environnements ainsi que leur homogénéité à travers ces variables... Attention donc aux situations où l'on se retrouve en extrapolation : la relation (idéalement linéaire) entre la pollution et la variable auxiliaire, calée sur les milieux où sont implantées les stations disponibles, ne reste pas nécessairement valide pour d'autres valeurs des variables auxiliaires.

3. Résultats et discussion

Les définitions théoriques que nous venons de poser sont mises en œuvre sur des résultats de mesures du NO_2 en situation de fond (c'est-à-dire à l'écart des voies de circulation importantes) effectuées pour l'année 2005 sur l'ensemble de l'agglomération rouennaise à l'aide de tubes à diffusion. La campagne regroupait 80 sites – $(x_i)_{1 \leq i \leq 80}$ – équipés pendant 6 périodes de 2 semaines, réparties uniformément dans l'année ; la moyenne annuelle en chaque site – $Z(x_i)$ – est considérée comme égale à la moyenne des 6 concentrations bihebdomadaires. La carte de la moyenne annuelle $Z^*(x)$ a été établie (Figure 2) par interpolation géostatistique [10], qui fournit également en tout point l'écart-type de l'erreur d'estimation $\sigma(x)$ (Figure 3).

3.1 Zones de représentativité

La méthodologie présentée permet de déterminer la zone de représentativité de chacune des 5 stations

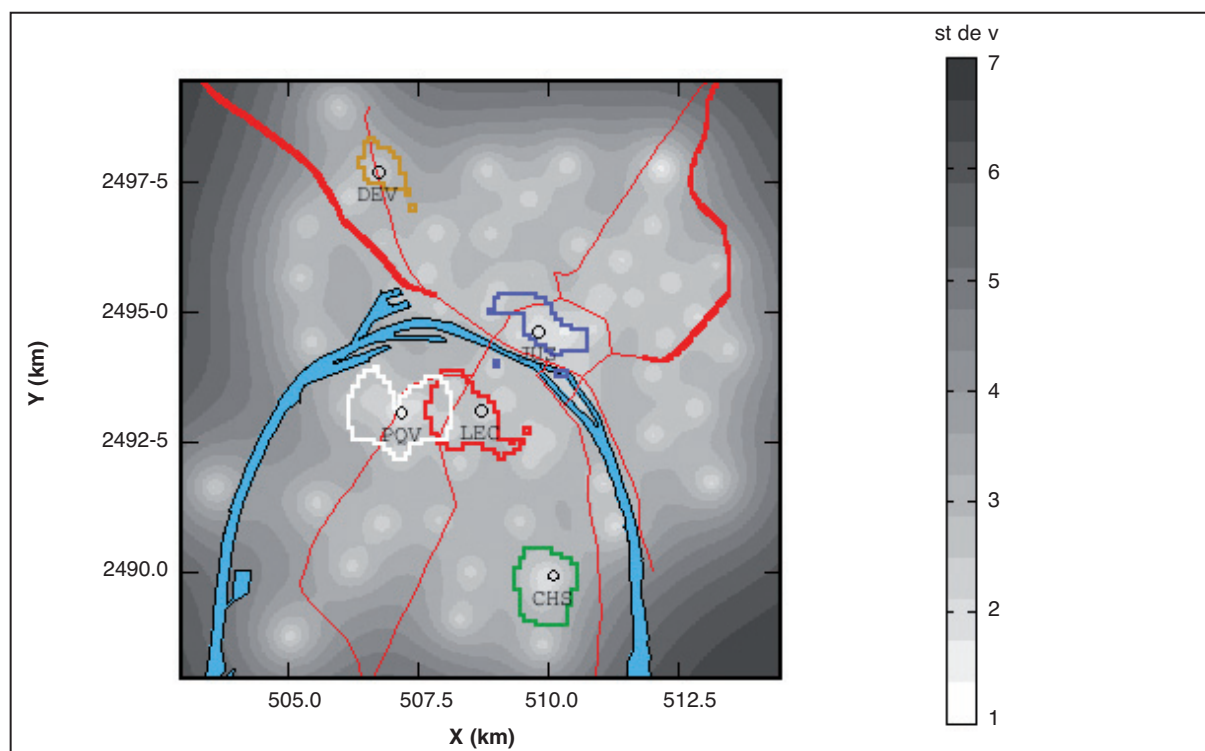


Figure 3.

Carte de l'écart-type d'estimation $\sigma(x)$, en $\mu\text{g}/\text{m}^3$.

On distingue nettement les points de mesure (où l'écart-type est le plus faible).

Map of the kriging standard deviation, $\sigma(x)$, in $\mu\text{g}/\text{m}^3$.

Measurement points (where standard deviation is smallest) are greatly visible.

urbaines de fond (DEV, JUS, LEC, PQV et CHS) de l'agglomération rouennaise, pour la moyenne annuelle en NO_2 .

La cartographie obtenue (Figure 2) doit être considérée avec précaution pour plusieurs raisons.

D'une part, on montre facilement que retenir un écart δ trop strict aurait pour effet de réduire la zone de représentativité à la seule station : en effet, pour un risque et un écart acceptable fixés, le premier membre $\zeta(x)$ de l'inégalité du premier membre de (4), varie entre une borne inférieure $\zeta_{\min}$ et une borne supérieure $\zeta_{\max}$. Or on cherche x tel que $\zeta(x) \leq \delta$; si δ est choisi inférieur à $\zeta_{\min}$, aucun point x ne serait sélectionné en-dehors de la station. Concrètement, une zone de représentativité à $\pm 2 \mu\text{g}/\text{m}^3$ (par exemple) peut ne pas aboutir si l'estimation spatiale n'est pas assez précise, en particulier si la variabilité spatiale à petite distance est importante, notamment en présence d'un important « effet de pépité » sur le variogramme expérimental. Dans cet exemple (où $\delta = 2$), il faudrait que $\sigma(x)$ soit largement inférieur à 1 pour que les zones ne soient pas réduites aux seules stations. Cela semble assez logique : dans le cas d'une estimation peu précise, il est illusoire d'espérer délimiter des zones à $\pm 2 \mu\text{g}/\text{m}^3$ si la technique d'estimation donne des résultats à $\pm 5 \mu\text{g}/\text{m}^3$ (par exemple).

D'autre part, la zone de représentativité dépend fortement de l'implantation des stations $\{x_i\}_{1 \leq i \leq n}$. La

figure 3 montre l'écart-type d'estimation pour la campagne de 2005. Autour des sites x_i instrumentés présentant une concentration voisine de celle à la station x_0 , on a $Z^*(x) \approx Z(x_0)$ et $\sigma(x)$ est très faible en raison de la proximité du point de donnée. Au contraire, loin des sites expérimentaux x_i , $\sigma(x)$ a tendance à être fort et donc les points vérifient difficilement l'inégalité du premier membre de (4) puisque $\zeta(x)$ augmente, ce qui ne signifie pas nécessairement que leur concentration réelle diffère beaucoup de celle de la station. Simplement nous n'en savons rien... La condition suffisante que nous avons retenue n'est pas une condition nécessaire : il y a probablement des points dont la concentration est proche de celle de la station (à moins de δ), mais nous ne pouvons pas les détecter, notamment si l'écart-type d'estimation est trop élevé, et en particulier s'ils sont trop éloignés des points de donnée.

Il s'agit là du comportement intrinsèque et incontournable de la méthode. Il faut garder présent à l'esprit le risque inhérent lié au cadre probabiliste : les calculs peuvent être corrects, mais il persiste une possibilité que la réalité soit tout autre (ce que Matheron appelle le risque d'erreur radicale [11]).

Il existe malgré tout diverses possibilités pour améliorer les résultats : le plan d'échantillonnage peut être densifié, les zones de représentativité peuvent être réduites en « coupant » à un certain pourcentage de la portée (Annexe 1) ou en ne conservant que des

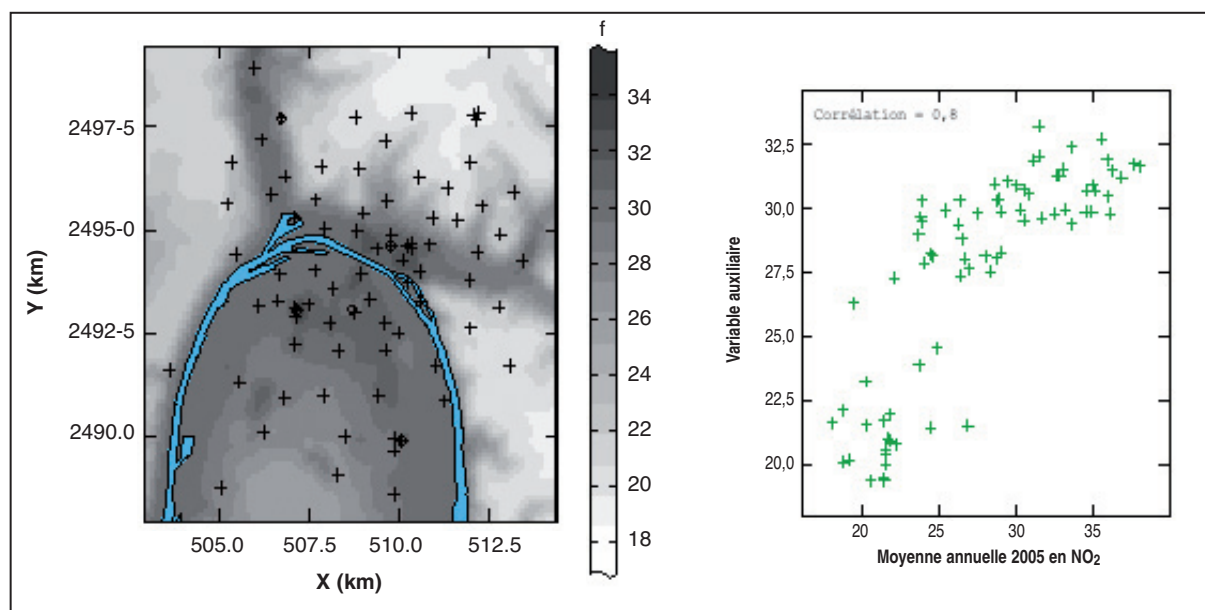


Figure 4.

Variable auxiliaire tenant compte de la population et du relief, sur l'ensemble de l'agglomération rouennaise (à gauche ; les croix figurent les points de mesure par tubes à diffusion et les ronds indiquent les stations fixes) et nuage de corrélation entre la pollution et cette variable auxiliaire (à droite).

Auxiliary variable taking into account population and elevation, all over Rouen city (left side ; cross point show passive sampling measurement sites and circles show fixed station) and correlation scatter diagram between pollution and the auxiliary variable (on the right).

zones connexes, le risque statistique peut être diminué, on peut intégrer des informations supplémentaires pour améliorer l'estimation (Annexe 2), etc.

Finalement, les zones de représentativité présentées à la figure 2, conduisent à trois constatations :

Constatation n° 1

Pour l'agglomération rouennaise, et pour l'année 2005, les zones de représentativité sont relativement **petites**. L'agglomération est loin d'être entièrement couverte par les cinq stations. Les contrastes importants sur le niveau de pollution (estimé) en sont à l'origine, d'autant plus que la portée du phénomène est réduite (Annexe 1). Si on considère la carte de pollution comme étant la plus juste possible (on ne peut pas avoir mieux), alors on est amené à conclure que les stations sont insuffisantes pour donner une vision de la pollution annuelle moyenne par le NO₂ sur l'ensemble de l'agglomération... L'intégration d'informations complémentaires à travers des variables auxiliaires permet d'améliorer la cartographie (Annexe 2). Pour le cas de la moyenne annuelle en NO₂ sur l'agglomération rouennaise en 2005, des informations complémentaires sont disponibles aux nœuds d'une grille régulière, notamment la densité de population et le relief. Une combinaison particulière de ces informations a pu être établie par régression multiple sur le NO₂ mesuré, puis par régularisation spatiale ; elle s'avère assez bien corrélée aux concentrations annuelles en NO₂. Le résultat, retenu comme variable auxiliaire, est montré à la figure 4.

Les zones de représentativité, délimitées selon le modèle utilisant la variable auxiliaire, sont montrées sur la figure 5, pour un risque statistique de 5 % et un écart δ ici réduit à 5,5 $\mu\text{g}/\text{m}^3$.

Ce qui pilote la définition de la zone de représentativité ici, ce sont à la fois des concentrations proches de celle de la station, mais aussi des valeurs de

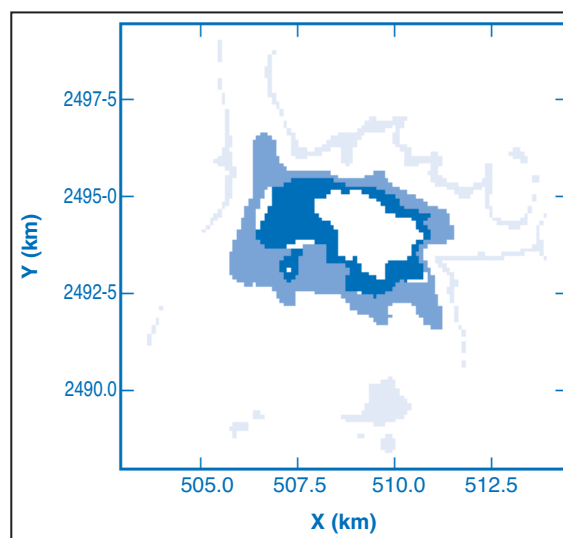


Figure 5.

Zones de représentativité pour les stations JUS (en bleu foncé), LEC et PQV (en bleu), CHS et DEV (en bleu très clair).

Representativeness areas for measurement stations JUS (in dark blue), LEC and PQV (in blue), and CHS and DEV (in light blue).

la variable auxiliaire relativement similaires. Si nous choisissons une variable auxiliaire qui soit caractéristique d'un milieu en terme de pollution (autrement dit, dont les valeurs sont bien corrélés aux concentrations), on pourra parler de **similitude** de milieu.

Sur la carte, on peut constater que :

- la zone centrale n'est pas représentée : les concentrations qui y sont mesurées ou estimées sont trop élevées pour pouvoir prétendre appartenir à une zone de représentativité ;
- en s'éloignant du centre vers la périphérie de l'agglomération, on passe dans des zones représentatives de niveaux **décroissants** de concentration en NO_2 . Les formes circulaires des zones de représentativité montrent bien les similitudes de milieu : plus on s'éloigne du centre de l'agglomération, moins le tissu urbain est dense et plus la pollution décroît (c'est *grosso modo* le cas de Rouen, avec une particularité au nord-ouest où une vallée urbanisée s'étend, bien visible sur les cartes) ;
- comme les zones autour des stations ne sont pas restreintes à un rayon égal à la portée du variogramme, les zones de représentativité associées à des stations fournissant des mesures voisines pour la moyenne annuelle sont presque identiques : c'est le cas pour les stations LEC et PQV (avec une moyenne annuelle mesurée d'environ $31 \mu\text{g}/\text{m}^3$), ainsi que pour les stations DEV et CHS (avec une moyenne annuelle mesurée d'environ $24 \mu\text{g}/\text{m}^3$).

Constatation n° 2

Il semble étonnant de constater qu'**aucune zone de représentativité ne recouvre les plus fortes valeurs** alors que la même étude conduite en 2000 montrait que la station du palais de justice (JUS) était alors bien placée (Figure 6) [9]. Les endroits de l'agglomération les plus touchés sont pourtant les zones à surveiller en priorité...

D'une année à l'autre, l'écart entre la concentration mesurée à la station et la concentration estimée à partir des tubes varie, de même que la variance de krigeage (qui dépend du variogramme) ; de nombreux paramètres peuvent modifier la répartition spatiale de la pollution, comme la configuration environnementale (PDU, ...), les conditions climatiques (vents dominants, ...), les changements socio-économiques (industries, démographie, ...). Cela signifie que, pour un même critère (la concentration moyenne annuelle en NO_2) les zones de représentativité peuvent différer d'une année à l'autre...

Ces zones de représentativité variables avec les années constituent une des limitations des développements proposés : la zone de représentativité est établie *a posteriori* et pour une année donnée, sa forme et sa taille dépendant des conditions environnementales (météorologie, émissions, ...). Il sera intéressant d'étudier la stabilité des zones de représentativité sur plusieurs années, encore faut-il pour cela disposer de mesures en nombre suffisant, ce qui

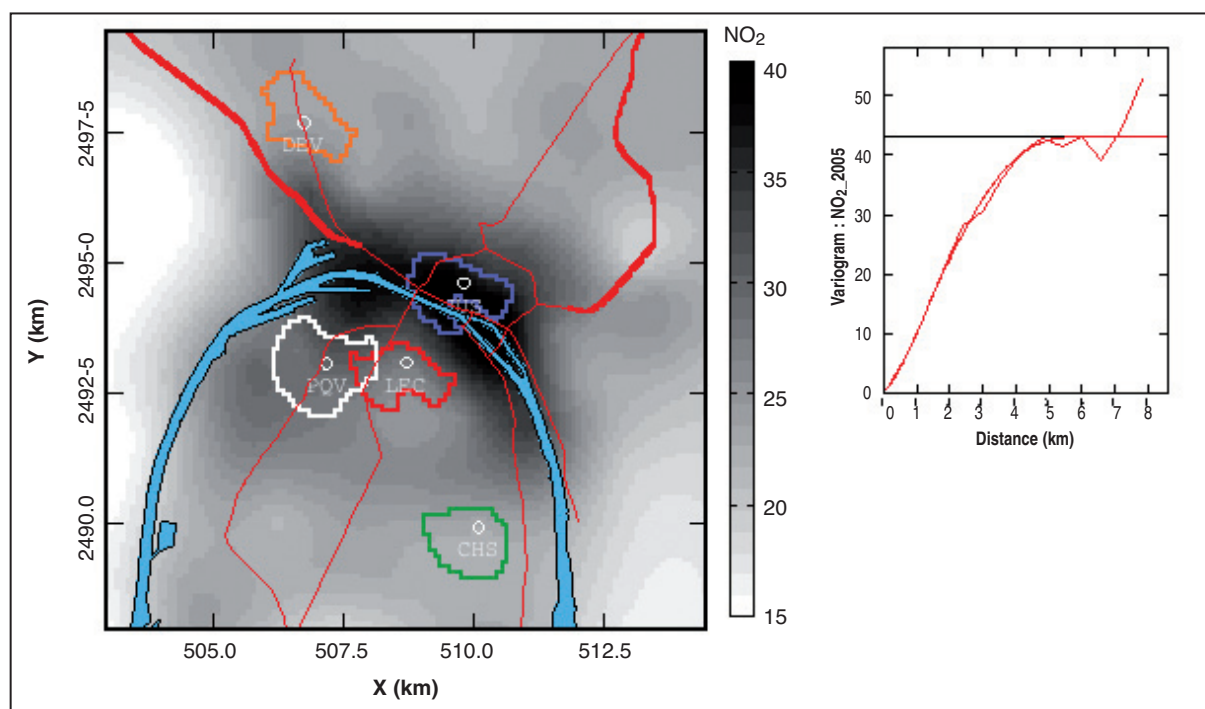


Figure 6.

Zones de représentativité pour l'année 2000, selon les mêmes critères $\delta = 6 \mu\text{g}/\text{m}^3$ et $P = 0.95$.

En fond est présentée la cartographie du NO_2 annuel moyen pour 2000, en $\mu\text{g}/\text{m}^3$.

L'échelle est identique à celle de la figure 2, et le variogramme est établi avec les données de l'année 2000.

Areas represented by the measurement stations, with respect to annual NO_2 for year 2000, on the basis of difference less than $6 \mu\text{g}/\text{m}^3$ and assessment risk of 5%.

In the background of this map, we present the 2000 annual mean of NO_2 , expressed in $\mu\text{g}/\text{m}^3$.

The scale is the same as the one used in figure 2 and the variogram is estimated with year 2000's data.

implique de poursuivre les campagnes de mesure par tubes passifs. Peut-être pourrait-on dégager des zones permanentes, ou encore des zones dépendantes de conditions environnementales (basées sur des roses de vent par exemple). Un travail dans ce sens a été initié en collaboration avec le Laboratoire Central de Surveillance de la qualité de l'air (LCSQA).

Constatation n° 3

Des zones de représentativité **se chevauchent** (pour les stations LEC et PQV). On peut donc parler de **redondance** pour ces stations, au moins dans l'intersection des zones de représentativité, endroit qui peut être rattaché à l'une ou l'autre station. Ceci devrait conduire à déplacer l'une des stations, de façon à optimiser la couverture spatiale du réseau de mesures (voir §4).

3.2 Rattachement d'un point à une station

Malgré les limites méthodologiques évoquées, des résultats intéressants sont obtenus, parmi lesquels l'obtention de zones de représentativité qui se recoupent partiellement. L'application de la méthodologie exposée en 2.2 conduit à la figure 7.

L'intérêt de cette façon de procéder est double. Tout d'abord, les zones de représentativité sont plus vastes. Puisque nous n'avons pas de seuil sur la différence entre la concentration estimée en un point et celle d'une station, les zones de représentativité s'étendent naturellement jusqu'à la portée du phénomène. Ensuite, et c'est là le plus important, il n'y a plus de recouvrement possible. Un point du territoire est rattaché à une seule station. Dans les zones où

on pouvait avoir une certaine redondance, on peut alors visualiser la station **prédominante**. Ainsi, les points appartenant simultanément à la zone de LEC et à celle de PQV, sont mieux représentés par la station LEC. Ce serait donc la station PQV qui serait à déplacer préférentiellement.

La critique qui peut être faite de cette approche est que, en un point x quelconque, l'écart maximal η_x peut être assez grand, et la station à laquelle ce point se rattache n'est peut-être pas si représentative que ça du point x en question. Intuitivement, *seul un écart faible dans un rayon raisonnable autour de la station* est compatible avec la représentativité et, plus l'écart est faible, mieux la pollution en x est représentée par la mesure en x_0 .

En pratique, on peut évidemment combiner les deux points de vue, en fixant une valeur maximale pour l'écart η_x : chaque point est alors rattaché à une station au plus (la plus représentative), mais les zones de représentativité des stations ne sont pas exagérément étendues.

3.3 Risque de dépassement

Dans l'exemple de l'agglomération de Rouen en 2005, les concentrations mesurées par les tubes à diffusion ne dépassent pas la valeur limite de $40 \mu\text{g}/\text{m}^3$; la carte qui en résulte ne montre donc pas de zones avec des concentrations supérieures au seuil. Dans le but d'illustrer la question, nous abaissons la valeur du seuil s à $30 \mu\text{g}/\text{m}^3$, de façon à ce que ne soient pas vides les zones dans lesquelles le dépassement de seuil est sûr à 95 %.

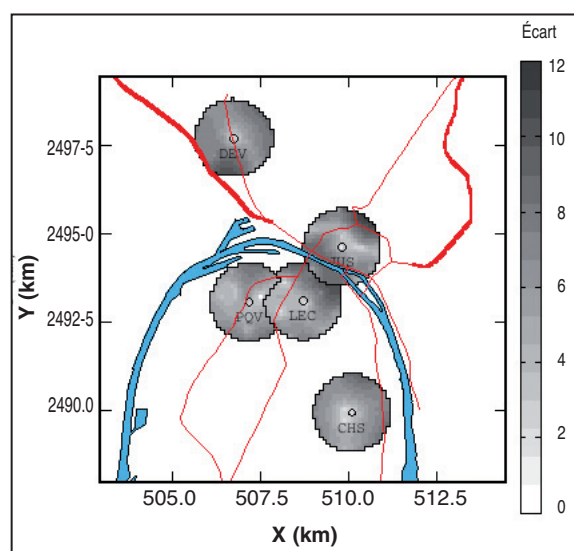


Figure 7.

Ensemble des zones rattachables aux stations, dans un rayon « acceptable » (voir §2.4 et annexe 1) de 1 km ; les écarts maximaux par rapport à la station de rattachement sont indiqués selon l'échelle (en $\mu\text{g}/\text{m}^3$).

Connecting each geographical point to a measurement station within a radius of 1 km leads to these areas, each related to a station, with maximum differences given against the connecting station by the scale (in $\mu\text{g}/\text{m}^3$).

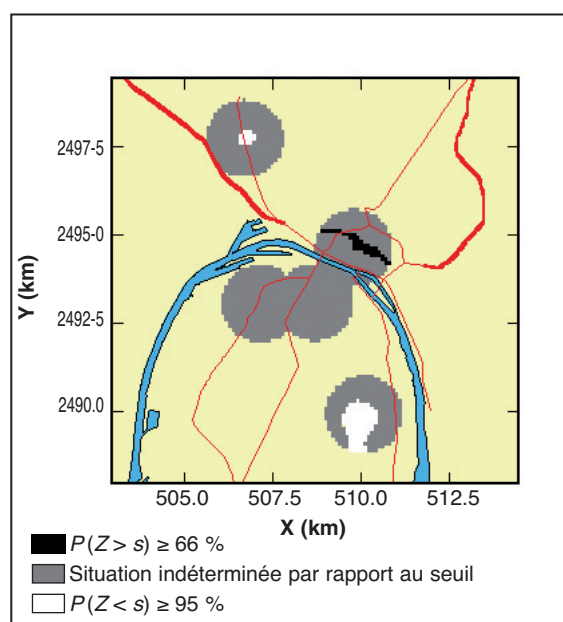


Figure 8.

Probabilités de dépassement ou non du seuil $s = 30 \mu\text{g}/\text{m}^3$ dans un rayon d'1 km.

Areas showing pollution overshooting a threshold $s = 30 \mu\text{g}/\text{m}^3$ with a probability of 66%, and showing pollution below the threshold with a probability of 95%, within 1 km radius.

Nous différencions les risques, en considérant qu'il est important de définir des zones où l'on est presque sûr que la concentration ne dépasse pas le seuil et des zones dans lesquelles le risque de dépassement n'est pas négligeable sans être forcément très élevé. Autrement dit on veut connaître les zones dans lesquelles, avec une forte probabilité, égale à 95 % par exemple, la concentration reste inférieure au seuil. Par ailleurs, nous avons choisi de définir les zones dans lesquelles la probabilité de dépassement du seuil est supérieure à 66 %. Les probabilités retenues sont finalement : $P(Z < s) \geq 0.95$ et $P(Z > s) \geq 0.66$. Les résultats sont montrés sur la figure 8.

Rappelons que les zones de représentativité ont été restreintes à la portée du phénomène. Il se trouve donc des parties de l'agglomération non couvertes par ces zones. En fait, les zones dans lesquelles le risque de dépassement est très faible ou plutôt fort sont très restreintes : l'indétermination est ici la règle (au seuil de $30 \mu\text{g}/\text{m}^3$).

4. Conséquences

À l'examen des zones de représentativité obtenues, il apparaît que les stations LEC et PQV pourraient être espacées. Nous avons vu que c'est sur la station PQV qu'il valait mieux porter l'effort de déplacement. Or, pendant l'exploitation des données de la campagne de mesure du NO_2 sur l'agglomération rouennaise en 2005, il s'est trouvé que la station LEC, située dans la cour d'une école, a été provisoirement arrêtée pour cause de travaux dans l'école. La question s'est alors posée de sa remise en service, ou de son déplacement (plutôt que PQV). En effet, lorsqu'on examine la répartition des concentrations aux alentours des deux

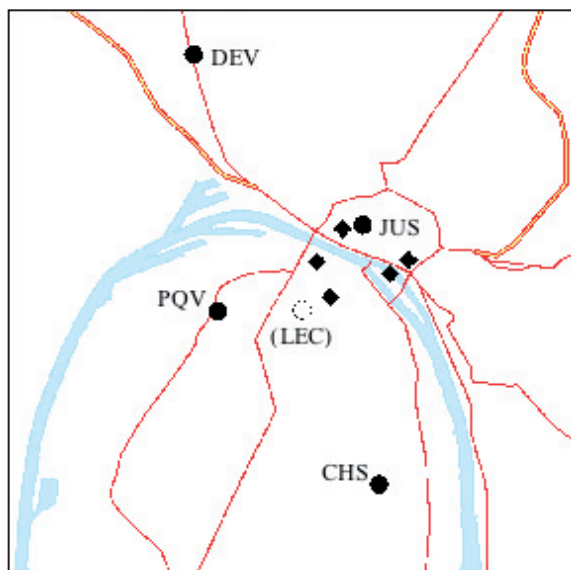


Figure 9.

Extension de la zone de surveillance au moyen de tubes à diffusion (signalés par un losange), exposés par périodes de 2 semaines en continu sur l'année.

Expansion of the survey area with diffusive samplers (shown by a losenge), which will be exposed by 2 weeks period continually over the year.

stations, il apparaît qu'il serait intéressant d'agir sur la station LEC en la déplaçant légèrement vers le nord-est, là où les concentrations sont plus importantes (Figure 2). En couvrant mieux le centre de l'agglomération, nous pourrions ainsi espérer avoir un site représentatif des fortes valeurs, site qui ne serait pas redondant avec une autre station. Mais comme l'implantation d'une station de fond en milieu urbain est assez longue et onéreuse, Air Normand a choisi dans un premier temps d'effectuer ces mesures à l'aide de tubes à diffusion, exposés par période de deux semaines, en continu. Plusieurs points de mesure supplémentaires ont ainsi été choisis au centre de l'agglomération afin de mieux surveiller la partie la plus touchée par la pollution de fond au NO_2 (Figure 9) et de vérifier le respect de la valeur limite à horizon 2010 en situation de fond.

Une étude de la même ampleur que celles de 2000 et 2005 devrait être menée en 2010 pour compléter les informations spatiales. L'exploitation de toutes ces mesures permettra le cas échéant de rechercher un site adéquat pour implanter une station fixe automatique dans cette partie de l'agglomération.

5. Conclusion

Une définition possible et intuitive de la représentativité des stations se transcrit en termes géostatistiques : dans la zone de représentativité de la station, la concentration ne diffère pas de plus d'un écart δ fixé de la valeur mesurée à la station (du moins dans le modèle), ce qui s'écrit : $|Z(x) - Z(x_0)| \leq \delta$.

La définition de zones de représentativité autour des stations doit s'accompagner de δ , l'écart maximal toléré entre concentrations (ici les moyennes annuelles) dans la zone et à la station. Cet écart ne peut pas être librement fixé. En pratique, on peut être amené à admettre une certaine tolérance, liée à la précision de l'estimation, afin d'éviter des zones de représentativité trop restreintes. En outre, le cadre probabiliste dans lequel les calculs sont développés exige la détermination du risque statistique accepté (5 % en général).

Une connaissance spatiale du phénomène, la plus complète possible, est primordiale puisque les zones vont finalement dépendre de l'incertitude d'estimation spatiale à travers l'écart-type de krigeage $\sigma(x)$; cet écart-type dépend à la fois de l'amplitude des fluctuations (palier du variogramme) et de la distance de corrélation (portée du variogramme).

Le formalisme mathématique aboutit à plusieurs zones possibles, selon le paramétrage mais aussi selon la façon dont on regarde les choses. Plutôt que contradictoires, les résultats sont souvent complémentaires. Ainsi, nous n'utilisons plus la portée du phénomène pour limiter les zones dans l'espace mais nous introduisons une variable auxiliaire qui caractérise le milieu. La notion de représentativité inclut donc la similarité des concentrations mais aussi des milieux.

Ce travail a conduit Air Normand à déployer des tubes à diffusion en plusieurs sites de façon à vérifier l'absence de dépassement de la valeur limite annuelle à horizon 2010 ; une étude complémentaire sera finalement menée pour, entre autre, déterminer un emplacement plus approprié pour la station LEC le cas échéant.

Un travail important reste à faire pour compléter les définitions de la représentativité, comparer les

zones suivant les années, et surtout, trouver un formalisme pour les autres valeurs limites comme les percentiles, voire pour les concentrations horaires dans le cas des seuils d'alerte. Il sera aussi très instructif d'appliquer ce genre d'approche au cas de l'indice ATMO. Le LCSQA en collaboration avec quelques AASQA s'est attelé à ces tâches, avec l'appui du centre de géostatistique de l'école des Mines de Paris.

Annexe 1

Choix d'un rayon maximal d'influence pour les stations

Nous cherchons à déterminer un rayon d'influence « acceptable » pour chaque station.

Notons m l'espérance (la moyenne de la loi de probabilité) de la concentration Z supposée stationnaire ; m est alors constante dans l'espace. La covariance entre $Z(x)$ et $Z(x+h)$ est supposée ne dépendre que de h (et pas de x) ; elle est notée $C(h)$. $C(0)$ est alors la variance de $Z(x)$, supposée égale à la variance de $Z(x+h)$.

Le coefficient de corrélation entre $Z(x)$ et $Z(x+h)$ est défini comme :

$$\rho(h) = \frac{C(h)}{\sqrt{\text{Var } Z(x)} \sqrt{\text{Var } Z(x+h)}} = \frac{C(h)}{C(0)}$$

On a donc $C(h) = \rho(h) \cdot C(0)$.

Or nous cherchons une zone dans laquelle $Z(x)$ et $Z(x+h)$ sont bien corrélés, pour que la variance du résidu soit faible. Nous pouvons par exemple considérer que c'est le cas si $\rho(h)^2 \geq 50\%$,

soit $\rho(h) \geq \sqrt{0.5} \approx 0.71$, ou encore $C(h) \geq 0.71 \cdot C(0)$.

Comme le variogramme vaut $\gamma(h) = C(0) - C(h)$, l'inégalité précédente s'écrit :

$\gamma(h) \leq C(0) - 0.71 \cdot C(0)$, c'est-à-dire $\gamma(h) \leq 0.29 \cdot C(0)$.

Avec le modèle de variogramme ajusté pour les concentrations moyennes annuelles en dioxyde d'azote sur l'agglomération rouennaise en 2005, on trouve $C(0) = 39.5$; l'inégalité précédente devient $\gamma(h) \leq 0.29 \times 39.5 = 11.455$, ce qui donne $h \leq 1$ km environ.

Annexe 2

Utilisation d'une variable auxiliaire

Lorsqu'une variable auxiliaire, notée f , connue partout dans l'espace (i.e. $\forall x \in X \subset \mathbb{R}^2$, $f(x)$ est donné) est bien corrélée à Z , on peut l'utiliser pour améliorer l'estimation de Z . Parmi les techniques existantes (cokrigage, krigeage des résidus, krigeage avec dérive externe, ... voir [8], [12] et [13]), nous choisissons pour simplifier d'effectuer la décomposition $Z(x) = f(x) + R(x)$, et c'est alors le résidu R que l'on estime :

$$R(x) = R^*(x) + \varepsilon_R(x), \forall x \in X \subset \mathbb{R}^2.$$

Pour déterminer les zones de représentativité, on doit donc disposer d'un échantillon de mesures $(Z(x_i))_{1 \leq i \leq n}$ aux n points de mesure $(x_i)_{1 \leq i \leq n}$, ainsi que des valeurs de la variable auxiliaire $(f(x_i))_{1 \leq i \leq n}$ aux points de mesure ; on peut donc y calculer les résidus $\{R(x_i) = Z(x_i) - f(x_i)\}_{1 \leq i \leq n}$. Pour que l'inégalité (1) soit satisfaite, il suffit alors que

$$|\varepsilon_R(x)| \leq \delta - |R^*(x) - R(x_0) + f(x) - f(x_0)|.$$

où $\varepsilon_R(x)$ est l'erreur d'estimation commise au point x . Cette erreur est évidemment inconnue, mais le krigeage permet d'en évaluer la variance $\sigma_{\varepsilon_R}^2(x)$.

Pour constituer la zone de représentativité avec la variable auxiliaire, sous l'hypothèse d'une erreur d'estimation de loi gaussienne, on sélectionne les points x tels que :

$$1.96\sigma_{\varepsilon_R}(x) + |R^*(x) - R(x_0) + f(x) - f(x_0)| \leq \delta \quad (4')$$

Dans cette expression, tout est connu : on peut donc facilement délimiter dans l'espace les zones de représentativité ainsi définies. Le formalisme du krigeage des résidus s'avère donc particulièrement commode pour obtenir une majoration de l'erreur.

Remarque : ce formalisme n'intègre pas l'incertitude liée à la régression entre concentration et variable auxiliaire, qui est établie à partir des données disponibles.

Références

1. Arrêté du 17 mars 2003 relatif aux modalités de surveillance de la qualité de l'air et à l'information du public (JO du 22 juillet 2003).
2. Directive 1999/30/CE du Conseil du 22 avril 1999 relative à la fixation de valeurs limites pour l'anhydride sulfureux, le dioxyde d'azote et les oxydes d'azote, les particules et le plomb dans l'air ambiant, Directive 2000/69/CE du Parlement européen et du Conseil du 16 novembre 2000 concernant les valeurs limites pour le benzène et le monoxyde de carbone dans l'air ambiant, Directive 2002/3/CE du Parlement européen et du Conseil du 12 février 2002 relative à l'ozone dans l'air ambiant.
3. Classification et critères d'implantation des stations de surveillance de la qualité de l'air. ADEME 2002.
4. EUROAIRNET site selection 1998. Technical report n° 16. European Environment Agency 1998 : 31 p.
5. Arrêté du 22 juillet 2004 relatif aux indices de la qualité de l'air (JO du 25 novembre 2004).
6. Arnaud M, Emery X. Estimation et interpolation spatiale. Méthodes déterministes et méthodes géostatistiques. *Hermès Science*, Paris, 2000 : 216 p.
7. Cori A. Rapport de stage : Représentativité spatiale des stations de mesure de la concentration moyenne annuelle en NO₂. Travail d'option, deuxième partie. École des Mines de Paris, 2005 : 55 p.
8. Wackernagel H. Multivariate Geostatistics (3rd edition). Springer-Verlag, Berlin, 2003 : 387 p.
9. De Fouquet C, Deraisme J, Bobbia M. Comment évaluer les risques de dépassements de seuil : les modèles géostatistiques et leurs applications en pollution atmosphérique. *Environnement, risque et santé* 2007 ; 6 (3) : 207-18.
10. Air Normand. Rapport d'étude n° E 05_02_06 : Mesures de NO₂ sur l'agglomération rouennaise – Sites urbains de fond. Air Normand, Rouen, 2006 : 17 p.
11. Matheron G. Estimer et choisir. Cahiers du Centre de morphologie mathématique de Fontainebleau, Fasc. 7, École des Mines de Paris, 1978.
12. Bobbia M, Pernellet V, Roth C. L'intégration des informations indirectes à la cartographie géostatistique des polluants. *Pollution Atmosphérique* 2001 ; 170 : 251-62.
13. Cori A. Rapport de stage : Utilisation de données environnementales pour la cartographie du NO₂ sur la région rouennaise. Travail d'option, première partie. École des Mines de Paris, 2005 : 60 p.

