



Avec quelle précision connaît-on la concentration d'un polluant atmosphérique dans une ville ? Les calculs statistiques simples fournissent des résultats rarement satisfaisants. Pour utiliser au mieux les mesures disponibles, et réduire ainsi les risques d'une erreur d'évaluation, les mathématiciens ont imaginé une nouvelle approche, la géostatistique.

# Des statistiques contre la pollution

Quelles relations entre l'évaluation d'un niveau de pollution et les mathématiques ? Que l'on cherche à caractériser la concentration en métaux lourds dans le sol d'une friche industrielle, la concentration en pesticides dans une nappe phréatique, ou encore la pollution atmosphérique, l'idéal serait de disposer de mesures couvrant la totalité de la région (ou de la période) considérée ; mais ce n'est jamais le cas. Comment estimer alors la valeur de la concentration là où elle n'a pas été mesurée ? Prenons l'exemple de la pollution atmo-

**Chantal de Fouquet**  
est maître de recherche  
au Centre de géostatistique  
de l'École des mines  
de Paris.  
chantal.de\_fouquet  
@ensmp.fr

sphérique urbaine en dioxyde d'azote ( $\text{NO}_2$ ), polluant principalement émis par la combustion des hydrocarbures (circulation automobile, chauffage). La réglementation européenne fixe comme objectif, pour 2010, une concentration moyenne annuelle inférieure à 40 microgrammes par mètre cube d'air ( $\mu\text{g}/\text{m}^3$ ). Mais sait-on obtenir des statistiques fiables ?

## Effectuer des mesures

Avant de tracer des cartes de concentration, il faut effectuer des mesures. Pour cela, on dispose, sur l'ensemble de l'agglomération, des « tubes à diffusion » : le  $\text{NO}_2$  qui les traverse déclenche une réaction chimique dont les produits sont analysés quantitativement. À Mulhouse et dans ses environs, une cinquantaine de tubes ont ainsi été implantés durant l'année 2001. Résultats : la concentration annuelle la plus basse enregistrée était de  $12 \mu\text{g}/\text{m}^3$  et la plus haute de  $40 \mu\text{g}/\text{m}^3$ . Dans l'ensemble, les mesures étaient proches de leur moyenne :  $22 \mu\text{g}/\text{m}^3$  [fig. 1].

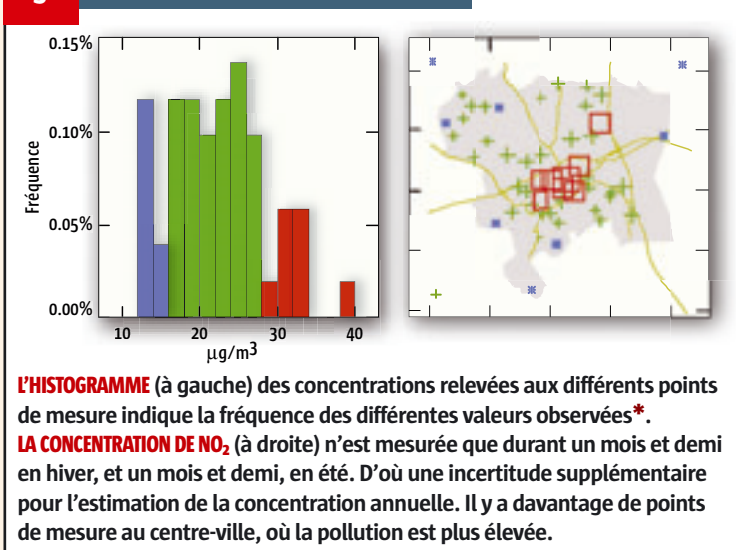
En reportant ces mesures sur une carte, on voit que la concentration est plus faible en périphérie et plus forte au centre-ville et vers le nord, où la circulation automobile, les installations de chauffage et les industries sont plus denses. La localisation des minimums et des maximums de concentration est liée à l'ori-

gine de la pollution, aux propriétés physicochimiques du polluant, ou à la topographie du milieu urbain.

## Irrégulier mais structuré

Pour évaluer de façon plus précise les valeurs fortes de la concentration en  $\text{NO}_2$ , davantage de tubes ont été implantés dans le centre de l'agglomération. Mais accroître le nombre de tubes là où les concentrations sont fortes augmente automatiquement la moyenne des mesures ! En effet, deux tubes rapprochés mesurent des concentrations plutôt voisines : on dit que les concentrations sont corrélées spatialement. Or, les calculs statistiques usuels supposent l'absence d'une telle corrélation. Conclusion : la simple moyenne des mesures n'est pas un indicateur fiable du niveau moyen de pollution. C'est l'observation de la corrélation spatiale des teneurs dans les gisements miniers et le constat de résultats d'exploitation systématiquement inférieurs aux prévisions qui ont amené le Français Georges Matheron à développer, au début des années soixante, une discipline nouvelle : la géostatistique. Son objet est l'estimation de phénomènes qui présentent un aspect très irrégulier, et cependant structuré. Le principe : même si une mesure en un point ne permet pas d'en déduire avec certitude la valeur en un

**Fig.1** La mesure des concentrations



point voisin, l'écart entre ces valeurs est d'autant plus réduit que les points sont proches. Pour étudier ces phénomènes, on fait appel à des modèles probabilistes : on ne travaille pas en décrivant la réalité avec exactitude, mais on considère un ensemble de réalités possibles.

En quoi consistent ces modèles qui décrivent et quantifient les corrélations spatiales ?  $x$  désignant un point du domaine d'étude, la concentration annuelle est représentée par une fonction numérique  $z(x)$ . Si la concentration était mesurée partout,  $z$  serait parfaitement connue en tout point.

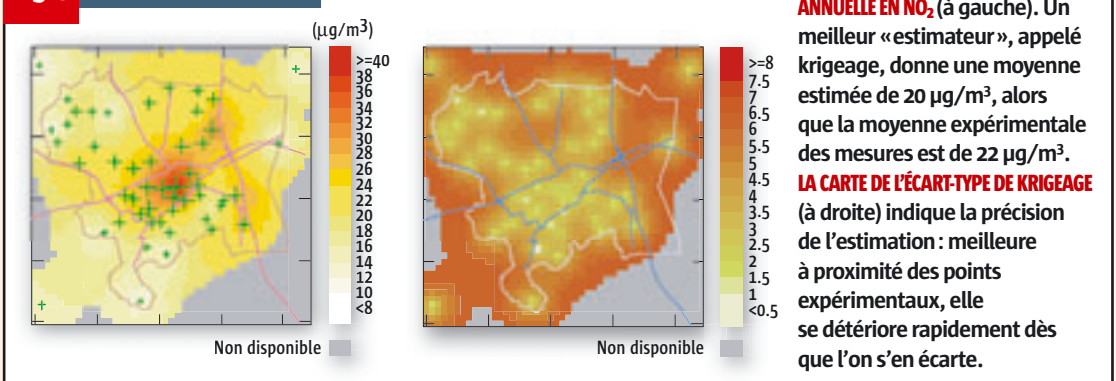
Pour plusieurs années successives, les concentrations se ressemblent, car leur variabilité spatiale est analogue. Nous les interprétons comme des tirages particuliers (ou « réalisations ») parmi toute une classe de fonctions numériques possibles, qui constituent une fonction aléatoire.

### L'exemple du dé

Qu'est-ce qu'une fonction aléatoire ? Faisons l'analogie avec le tirage d'un dé. La valeur pouvant apparaître sur la face supérieure est une variable aléatoire caractérisée par les fréquences d'apparition (ou probabilités) associées aux six valeurs possibles 1, 2, ..., 6. Chaque lancer du dé fournit une valeur numérique particulière, ou réalisation, par exemple 4. Une variable aléatoire est donc un « objet mathématique », accessible en pratique uniquement à travers ses réalisations.

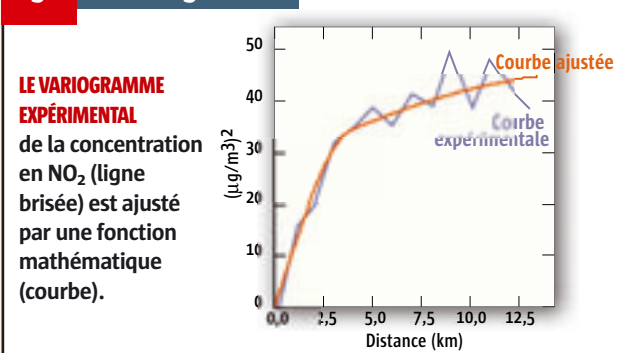
Revenons à la pollution. La fonction  $z$  qui représente la concentration pour l'année 2001, est interprétée comme une réalisation d'une « fonction aléatoire »  $Z$ , représentant « les variantes possibles de la

**Fig.3 Tracer des cartes**



\* Données ASPA, Association pour la surveillance et l'étude de la pollution atmosphérique en Alsace.

**Fig.2 Le variogramme**



concentration annuelle sur l'agglomération ». En deux points  $x$  et  $x'$ , les concentrations  $z(x)$  et  $z(x')$  sont ainsi vues comme des réalisations de deux variables aléatoires  $Z(x)$  et  $Z(x')$ , d'autant plus fortement liées que  $x$  et  $x'$  sont proches. On mesure cette liaison par le « variogramme », qui est, en fonction de la distance  $h$ , la moyenne (« l'espérance ») du carré de l'écart  $Z(x+h) - Z(x)$ . Cet écart est plutôt réduit quand  $h$  est petit, et il augmente avec  $h$ . Pour calculer « la meilleure estimation » de la concentration en un point non mesuré, il suffit de connaître le variogramme de la fonction aléatoire  $Z$ .

Comment procède-t-on ? À partir des mesures, on calcule les moyennes expérimentales du carré des écarts  $z(x+h) - z(x)$ , pour les différentes classes de distances  $h$  accessibles sur les données. Ce variogramme expérimental fournit beaucoup d'informations, par exemple sur la régularité des concentrations (quelle est l'amplitude des écarts des concentrations entre deux points voisins ?) ou sur une éventuelle anisotropie (existe-t-il une orientation préférentielle, due à la topographie de la ville, ou à la direction des vents dominants ?). À ce variogramme expérimental on superpose une fonction mathématique adéquate, le variogramme « théorique » qui quantifie la variabilité spatiale des concentrations [fig. 2]. Ce faisant, on détermine la fonction aléatoire  $Z$ .

### Calcul de l'estimation

Pour tracer des cartes, ou établir des comparaisons sur plusieurs années, on s'intéresse aux valeurs des concentrations moyennées, par exemple, sur des carrés

de 500 mètres de côté. Comme on ne mesure pas la concentration partout (ni de façon continue) dans l'agglomération, on a seulement accès à une estimation de la concentration annuelle aux différents points ou sur ces carrés. On calcule cette estimation en combinant toutes les mesures, pondérées par des coefficients, les « poids optimaux », qui sont déterminés grâce au variogramme, de façon que l'erreur d'estimation (l'écart entre la valeur réelle de la concentration et son estimation) soit la plus réduite possible. Le « meilleur estimateur » ainsi obtenu, appelé krigeage (en l'honneur des travaux précurseurs de Daniel Krige dans les mines d'or d'Afrique du Sud dans les années cinquante) permet de tracer des cartes [fig. 3].

### Affiner le modèle

Reste que, sans l'accès aux valeurs réelles, l'erreur d'estimation demeure inaccessible. Le variogramme permet néanmoins de calculer la variance ou l'écart-type de l'erreur d'estimation, qui indiquent les endroits où cette erreur est plutôt faible et ceux où elle est plutôt élevée. En plaçant cette nouvelle information sur une carte, on visualise ainsi les régions où l'estimation est imprécise et où il conviendrait donc de placer de nouveaux tubes de mesure.

Pour affiner le modèle, on peut modifier les fonctions aléatoires afin qu'elles intègrent des variables « explicatives » comme la densité locale de population ou la densité du bâti urbain. Les estimateurs multivariés permettent ainsi de relier les cartes de concentrations aux caractéristiques du milieu environnant. ■■