

→ R. Vincent (1), M. Grzebyk (2),
H. Wäckernagel (3), C. Lajaunie (3)

(1) service Evaluation et prévention du risque chimique

(2) Direction des Etudes et Recherches, du Centre de recherche de l'INRS, Nancy

(3) Ecole des Mines de Paris, Centre de Géostatistique

Application de la géostatistique à l'hygiène industrielle

Evaluation d'un cas d'exposition professionnelle au trichloroéthylène

APPLICATION OF GEOSTATISTICS TO INDUSTRIAL HYGIENE

ASSESSMENT OF A CASE OF OCCUPATIONAL EXPOSURE TO TRICHLOROETHYLENE

In industrial hygiene assessment of occupational exposure to chemical substances generally fails to take the autocorrelation of measurements into account. This may lead to substantial biases where exposure estimates are obtained using conventional statistics based on the hypothesis of independence. Geostatistics makes it possible to model exposure variability using the variogram and, through kriging, to obtain better, unbiased estimates of the mean and corresponding confidence intervals. This article presents a case study based on a series of trichloroethylene exposure measurements. The results obtained by a conventional statistical analysis method are compared with those obtained by the geostatistical method.

● geostatistics ● autocorrelation
● variogram ● occupational exposure
● chemical substance ● trichloroethylene ● sampling strategy

En hygiène industrielle, l'évaluation des expositions professionnelles aux agents chimiques ne prend généralement pas en compte l'autocorrélation des mesures. Cette situation peut conduire à obtenir des estimateurs biaisés en utilisant la statistique classique basée sur l'hypothèse d'indépendance. La géostatistique permet de modéliser, par le variogramme, la variabilité des expositions et d'obtenir par « krigage » des estimateurs plus performants avec des intervalles de confiance non biaisés. Cet article présente une étude de cas à partir d'une série de mesures d'exposition au trichloroéthylène. Les résultats d'une analyse classique sont comparés à ceux donnés par la méthode géostatistique.

● géostatistique ● exposition professionnelle ● trichloroéthylène
● produit chimique

Évaluer l'exposition professionnelle aux substances chimiques est une tâche complexe qui nécessite la mise en place d'une stratégie de prélèvement élaborée. Généralement, il s'agit pour l'hygiéniste industriel d'apporter un jugement sur une situation d'exposition à partir d'une série limitée de prélèvements atmosphériques, dont les résultats sont comparés aux valeurs limites d'exposition professionnelle (VLEP) recommandées ou réglementaires. En France, les concepts généraux relatifs à la stratégie d'évaluation des expositions aux agents chimiques sont décrits dans la norme européenne NF EN 689, de juillet 1995 [1].

La comparaison des résultats des prélèvements passe par une procédure statistique; celle qui est la plus souvent utilisée, du fait de sa simplicité, consiste à calculer la moyenne arithmétique, l'écart-type et un intervalle de confiance à l'aide du test de Student. Cette procédure est fondée sur l'hypothèse d'absence de corrélation ou encore d'indépendance entre les mesures.

Cependant, il apparaît que cette hypothèse d'indépendance des mesures n'est pas vérifiée, comme le montrent plusieurs études publiées ces dernières années, dans lesquelles on observe une corrélation entre les mesures réalisées durant des périodes plus ou moins longues [2 à 4]. Dans ce contexte, l'utilisation des techniques statistiques fondées sur une hypothèse (ici l'indépendance des mesures), qui n'est pas vérifiée, conduit à une sous-estimation de l'amplitude de l'intervalle de confiance. Cela peut entraîner des conclusions erronées lors de la comparaison de l'exposition à une VLEP.

Dans un article publié il y a maintenant dix ans, Prétat [4] suggérait d'appliquer les techniques de la géostatistique pour traiter les problèmes d'interprétation des résultats, liés à la corrélation des mesures en hygiène industrielle, et élaborer des stratégies de prélèvement adaptées qui prennent en compte cette corrélation, la durée de prélèvement, le nombre de mesures et l'intervalle de temps entre les mesures.

L'INRS, dans le cadre du projet «Estimation des nuisances», a instruit, en collaboration avec le Centre de Géostatistique de l'École des Mines de Paris, une étude visant à appliquer ces méthodes au domaine de l'hygiène industrielle [5]. Le présent article décrit à l'aide d'une étude de cas, l'utilisation de la géostatistique lors de l'évaluation de l'exposition professionnelle aux agents chimiques.

1. La géostatistique

Développée en France dans les années 60, par le professeur Georges Matheron, la géostatistique a été définie initialement comme étant l'application de méthodes et de théories statistiques à la géologie minière [6]. Au départ, ces techniques statistiques ont été développées pour décider, à partir de l'étude d'une série d'échantillons de minerai et de leur variabilité spatiale, d'exploiter ou non un gisement.

Dans son concept, cette approche est très voisine de celle utilisée en hygiène industrielle : juger d'une situation d'exposition à partir de la connaissance issue d'une série de prélèvements atmosphériques.

QU'EST-CE-QUE LA GÉOSTATISTIQUE ?

La géostatistique est un ensemble de méthodes statistiques dédiées à l'analyse de phénomènes répartis dans le temps ou dans l'espace. Dans cet article, le phénomène est la concentration en trichloroéthylène au cours du temps.

Mathématiquement, ce phénomène est une fonction qui dépend du temps $z(t)$ que l'on considère comme une réalisation d'une fonction aléatoire $Z(t)$.

La fonction $z(t)$ n'est connue qu'en un nombre de points donnés (t_n) et les valeurs $z(t_n)$ sont considérées comme une réalisation des variables aléatoires $Z(t_n)$. On considère donc 2 niveaux :

- le niveau expérimental, dans lequel on effectue les calculs explicites à l'aide des données $z(t_n)$,
- le niveau théorique, dans lequel on travaille avec la fonction aléatoire $Z(t)$.

Par la suite, la géostatistique a été appliquée dans différents domaines tels que : l'hydrogéologie, l'agronomie, la météorologie, l'océanographie, la sylviculture... et plus récemment dans le domaine de l'environnement et de la décontamination des sols en particulier.

2. Etude de cas

Les objectifs de cette étude étaient de vérifier l'existence de corrélation entre les mesures d'exposition d'une série, d'évaluer l'influence de cette corrélation sur l'analyse statistique des résultats (moyenne, écart-type et intervalle de confiance) et d'exploiter les informations issues de l'analyse géostatistique pour déterminer une stratégie optimale de prélèvement.

Après avoir décrit la série de mesures, nous leur appliquerons successivement la procédure statistique classique puis les méthodes géostatistiques, afin de comparer leurs résultats et mettre en évidence l'influence de la corrélation entre les mesures et l'évaluation de l'exposition. Pour cette étude de cas, les mesures ont été réalisées pendant une journée et à poste fixe, à proximité d'une machine à dégraisser fonctionnant au trichloroéthylène.

L'objectif de ces mesures était de disposer d'une série de résultats se prêtant à une analyse géostatistique. En raison de cet objectif, la stratégie retenue ne correspond pas aux pratiques générales de l'évaluation de l'exposition aux agents chimiques : mesures de longue durée (4-8 h) réalisées pendant plusieurs jours et/ou sur plusieurs salariés appartenant à un groupe d'exposition homogène.

3. Description des mesures d'exposition

Deux types de mesures ont été effectués :

- des mesures séquentielles réalisées chaque minute, avec une durée de prélèvement de 22 secondes, à l'aide d'un spectromètre photoacoustique Bruel et Kjaer BK 1302 (BK), préalablement étalonné pour une monopollution au trichloroéthylène. Durant la période de mesurage,

appelée aussi fenêtre d'observation en géostatistique, 477 mesures ont été réalisées de 8 h 39 à 17 h 06 ;

- des mesures séquentielles sur tube de charbon actif (TCA), effectuées toutes les 7 minutes et 1/2 en prélevant l'air à proximité immédiate de la sonde du spectromètre photoacoustique BK 1302. Les mesures ont été réalisées de 8 h 40 à 11 h 55 et de 13 h 30 à 17 h 00. Au total, 56 prélèvements ont été collectés. La quantification de l'exposition au trichloroéthylène a été effectuée en utilisant la technique classique de désorption au sulfure de carbone, puis d'une analyse par chromatographie en phase gazeuse avec détection par ionisation de flamme.

Pour mémoire, signalons ici que la valeur limite de moyenne d'exposition (VME) recommandée en France pour le trichloroéthylène a été fixée à 405 mg/m³.

L'ensemble des résultats est représenté sur la *figure 1*. De façon à comparer les résultats du spectromètre BK 1302 et les mesures sur tube de charbon actif, les mesures BK 1302 ont été moyennées sur 7 minutes. Les concentrations obtenues après modification du temps d'intégration (de 22 s à 7 min) en prenant la moyenne de 7 mesures consécutives, est une opération classique en géostatistique appelée changement de support. On observe alors une bonne concordance entre les 2 types de mesure. L'examen des résultats représentés sur la *figure 1* indique tout d'abord deux profils d'exposition totalement différents pour la matinée et l'après-midi. L'après-midi, la variabilité de l'exposition était importante en raison de l'ouverture de fenêtres, qui a induit une turbulence des flux atmosphériques dans l'atelier.

4. Analyse des résultats

4.1. Approche statistique classique sous hypothèse d'indépendance

Nous avons appliqué cette méthode à la série de mesures.

Le *tableau 1* indique les moyennes, écarts-types, étendues pour chaque type de mesure (BK et TCA), pour les périodes du matin et de l'après-midi et pour différentes durées d'intégration (support), ainsi que la borne supérieure de l'intervalle de confiance à 95 % de la moyenne.

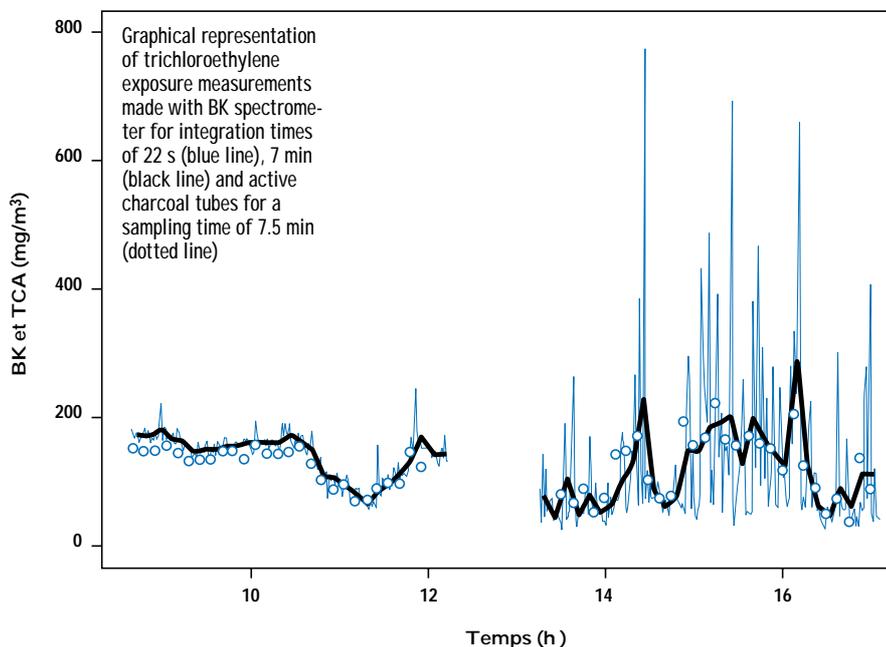
TABLEAU I

STATISTIQUES SUR LES RÉSULTATS DU MATIN ET DE L'APRÈS-MIDI PAR TYPE DE MESURE POUR DIFFÉRENTS SUPPORTS (mg/m³) -
STATISTICS ON MORNING AND AFTERNOON MEASUREMENT RESULTS FOR DIFFERENT INTEGRATION TIMES (mg/m³)

Mesure	Support	Nombre de mesures (n)	Moyenne arithmétique \bar{x}	Ecart-type (σ)	Mini	Maxi	Borne supérieure de l'intervalle de confiance à 95 %
MATIN							
BK	22 s	202	144	34	61	247	149
	7 min	28	147	32	69	184	159
	30 min	6	143	34	98	176	177
TCA	7,5 min	27	131	27	74	159	142
	30 min	6	132	28	86	155	160
APRÈS-MIDI							
BK	22 s	218	123	117	27	774	139
	7 min	31	123	60	49	294	145
	30 min	7	123	45	74	178	163
TCA	7,5 min	29	127	50	43	226	146
	30 min	6	129	40	77	183	165
SUR LA JOURNÉE							
TCA	7,5 min	56	129	41	43	217	140
	15 min	28	129	37	61	199	142
	30 min	14	130	30	76	183	146

La borne supérieure de l'intervalle de confiance est égale à : $\bar{x} + t \cdot \frac{\sigma}{\sqrt{n-1}}$ où \bar{x} est la moyenne arithmétique, σ correspond à l'écart-type, n au nombre de mesures et t est le fractile d'ordre $(1 - \frac{\alpha}{2})$ de la loi de Student à (n - 1) degrés de liberté (ici $\alpha = 0,05$).

Fig. 1. Représentation des mesures de trichloroéthylène par spectromètre BK sur support 22 s (ligne bleue), sur support 7 min (ligne noire) et par tubes de charbon actif (points, temps d'intégration 7,5 min)



L'approche statistique classique consiste à calculer les moyennes arithmétiques, les écart-types et les intervalles de confiance en utilisant les tables de test de Student. Cette procédure repose sur les deux hypothèses suivantes :

- les mesures suivent une loi normale de même moyenne et de même variance,
- il n'y a pas de corrélation entre les mesures.

Dans ce contexte, cette dernière hypothèse signifie que les mesures fluctuent de façon aléatoire autour de la moyenne constante.

Ces résultats indiquent que la moyenne arithmétique de l'exposition est plus élevée le matin que l'après-midi, ceci quelle que soit la valeur du support (durée d'intégration).

L'écart-type est cependant plus important l'après-midi que le matin.

Si l'hypothèse d'absence de corrélation était vérifiée, on conclurait pour des supports supérieurs à 4 min, à une différence non significative des moyennes d'exposition entre le matin et l'après-midi (fig. 2).

De même, on conclurait pour des supports supérieurs à 12 min, à une différence non significative des variances de l'exposition du matin et de l'après-midi (fig. 3).

Dans ce cas, une stratégie de prélèvement qui consisterait à effectuer des prélèvements consécutifs, d'une durée supérieure à 12 min et sur tube de charbon actif ne permettrait pas de différencier les situations d'exposition du matin et de l'après-midi, pourtant très différentes en variabilité si l'on considère les mesures séquentielles réalisées avec le spectromètre photoacoustique BK 1302.

On constate par ailleurs dans le tableau I que l'écart-type est pratiquement insensible au changement de durée d'intégration pour les mesures du matin, alors que sa valeur décroît quand la durée d'intégration augmente pour les mesures effectuées l'après-midi. Ce dernier résultat est en cohérence avec l'hypothèse d'indépendance. En effet, si cette hypothèse est vérifiée, la valeur de l'écart-type des mesures intégrées sur un temps modifié peut se déduire de celle des mesures élémentaires selon la relation :

$$\text{Ecart-type de } n \text{ mesures intégrées} = \frac{1}{\sqrt{n}}$$

écart-type des n mesures élémentaires

Dans le cas des séries du matin et de l'après-midi, on aurait dû trouver pour les écarts-types des mesures BK à 7 min

(7 mesures élémentaires de 22 s) et 30 minutes (30 mesures élémentaires à 22 s) des valeurs proches de :

$$\frac{34}{\sqrt{7}} = 12,8 ; \frac{34}{\sqrt{30}} = 6,2 \text{ le matin}$$

$$\frac{117}{\sqrt{7}} = 44,2 ; \frac{117}{\sqrt{30}} = 21,4 \text{ l'après-midi}$$

contre respectivement 32, 34 et 60, 45 (voir tableau I).

Ce constat est en contradiction avec les hypothèses sur lesquelles sont fondés les calculs d'intervalle de confiance du tableau I et mettent en évidence une présence de corrélation entre les mesures : en fait, la concentration mesurée à l'instant $T + \Delta t$ est dépendante de la concentration mesurée à l'instant T .

4.2. Application des méthodes géostatistiques

4.2.1. Etude de la corrélation des séries de mesures du matin et de l'après-midi

Pour étudier la corrélation des mesures d'une série chronologique en fonction du temps, la géostatistique utilise une fonction $\gamma(\delta t)$ appelée variogramme.

Le variogramme est une fonction qui quantifie la corrélation entre des mesures séparées par un intervalle de temps δt ($z(t)$ et $z(t + \delta t)$).

Pour chaque intervalle t considéré, il représente la moitié de la valeur moyenne du carré de la différence des expositions

Fig. 2. Intervalles de confiance (en bleu) pour le test sur la différence entre la moyenne BK du matin et celle de l'après-midi, en fonction du support. Sous hypothèse d'indépendance (non vérifiée), on conclut pour des temps d'intégration supérieurs à 4 min, à une différence non significative

Confidence intervals (blue line) obtained when testing difference between BK average in the morning and in the afternoon. For integration time over 4 min, there is no significant difference under (unconfirmed) independence hypothesis

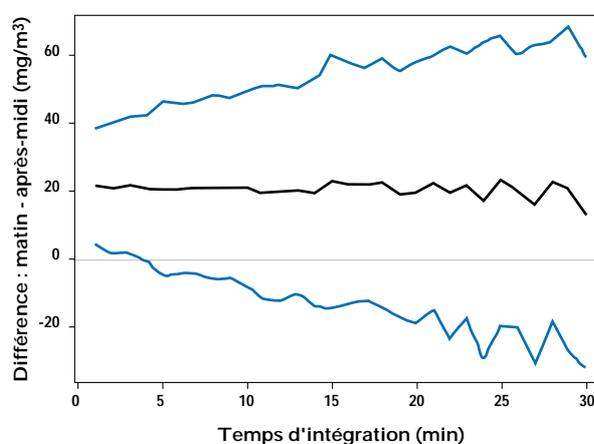


Fig. 3. Intervalles de confiance (en bleu) pour le test sur le rapport entre la variance du matin du BK et celle de l'après-midi (en trait noir), en fonction du temps d'intégration. Sous hypothèse d'indépendance (non vérifiée), on conclut pour des temps d'intégration supérieurs à 12 min à une différence non significative des variances

Confidence intervals (blue line) obtained when testing difference between BK variance in morning and in afternoon (black line) as a function of integration time. For integration time over 12 minutes, there is no significant difference between variances under (unconfirmed) independence hypothesis

prises à des instants quelconques mais espacés de t :

$$\gamma(\delta_t) = \frac{1}{2} E \left[(Z(t + \delta_t) - Z(t))^2 \right]$$

E correspond à l'espérance mathématique. Ainsi, $\gamma(0) = 0$.

Plus $\gamma(\delta_t)$ est petit, plus des mesures prises à des instants espacés de δt sont semblables en moyenne, donc fortement corrélés. A l'inverse, plus $\gamma(\delta_t)$ est grand, plus la corrélation entre des mesures espacées de t est faible.

L'hypothèse d'absence de corrélation se traduit par un variogramme tel que : $\gamma(0) = 0$; $\gamma(\delta_t) =$ constante différente de 0 si $\delta_t \neq 0$. Un tel variogramme est appelé effet de pépite (fig. 4a). Pour simplifier, on écrira par la suite que les données sont autocorrélées si le variogramme n'est pas un effet de pépite (fig. 4b).

La fonction d'autocorrélation est une autre fonction couramment utilisée pour quantifier la corrélation temporelle entre les mesures d'une série. On montre que le variogramme a des propriétés mathématiques plus intéressantes que la fonction d'autocorrélation [4]. Pour une série de mesures, un variogramme peut se déduire de la fonction d'autocorrélation. A l'inverse, dans certains cas, un variogramme peut caractériser la corrélation d'une série sans qu'une fonction d'autocorrélation existe [7].

Dans ce cadre géostatistique, le processus d'analyse de la corrélation d'une série consiste à :

- calculer le variogramme expérimental à l'aide des données recueillies ;
- décrire de façon qualitative le variogramme expérimental (c'est l'analyse variographique) ;
- ajuster le variogramme expérimental par un modèle de variogramme.

Les sections suivantes illustrent cette analyse à l'aide des mesures effectuées avec le spectromètre BK 1302 et des tubes de charbon actif.

Mesures BK

Pour les mesures de spectrométrie, les séries du matin et de l'après-midi sont analysées séparément.

LE VARIOGRAMME

La fonction aléatoire $Z(t)$ est caractérisée par son variogramme. Il s'agit de la fonction :

$$\gamma(\delta_t) = \frac{1}{2} E \left[(Z(t + \delta_t) - Z(t))^2 \right] \quad \forall t$$

où le symbole E représente l'espérance mathématique. Cette fonction mesure la dépendance entre des mesures espacées de d_t ; si les mesures sont indépendantes, alors $g(d_t)$ est une constante égale à la variance de $Z(t)$. On remarque au passage que cette fonction ne dépend que de l'incrément d_t mais pas de l'instant t ; on dit qu'on se place dans une hypothèse de stationnarité intrinsèque d'ordre 2.

La fonction g est appelée variogramme théorique de la fonction aléatoire $Z(t)$.

A l'aide de l'ensemble des données, on calcule le variogramme expérimental g^* de la façon suivante :

- on définit un pas de temps d_0 ,
- pour chaque multiple du pas de temps kd_0 , on détermine l'ensemble des paires de points de mesures :

$$T_k = \left\{ (t_1, t_2), kd_0 - \frac{1}{2} < |t_1 - t_2| < kd_0 + \frac{1}{2} \right\}$$

- on calcule la moyenne arithmétique de l'écart quadratique entre toutes les paires de valeurs $(z(t_1), z(t_2))$, telles que :

$$(t_1, t_2) \in T_k$$

c'est-à-dire :

$$\gamma^*(k\delta_0) = \frac{1}{2n_k} \sum_{(t_{11}, t_{12}) \in T_k} (z(t_{11}) - z(t_{12}))^2$$

où n_k est le nombre d'éléments de T_k .

Le variogramme expérimental n'est calculé que pour un ensemble limité d'incrémentes $(kd_0, k = 1, \dots, N)$ où Nd_0 est égal à la moitié de la durée échantillonnée.

Il n'est donc pas utilisable en tant que tel et doit être exprimé par une équation pour permettre en particulier de lui attribuer une valeur pour tous les incréments d_t . Cette fonction est obtenue par l'ajustement sur le variogramme expérimental d'une courbe d'expression théorique connue $\tilde{\gamma}$. Cette courbe est appelée variogramme modélisé.

La courbe pour modéliser le variogramme doit avoir des propriétés mathématiques particulières. On se limite généralement à un jeu restreint de modèles de base tels que :

J l'effet de pépite (voir fig. 4a)

$$\tilde{\gamma}(\delta_t) = a \quad \forall \delta_t \neq 0$$

J le modèle exponentiel (voir fig. 4b)

$$\tilde{\gamma}(\delta_t) = a \cdot e^{-b|\delta_t|}$$

J le modèle triangulaire (voir fig. 4b)

$$\tilde{\gamma}(\delta_t) = b|\delta_t| \quad \text{si } \delta_t < \frac{a}{b} \quad \text{et}$$

$$\tilde{\gamma}(\delta_t) = a \quad \text{si } \delta_t > \frac{a}{b}$$

Fig. 4a. Variogramme à effet de pépite : absence d'autocorrélation. Les points correspondent au variogramme expérimental et la ligne au variogramme modélisé

Variogram with nugget effect: absence of autocorrelation (points represent experimental variogram and line represents model variogram)

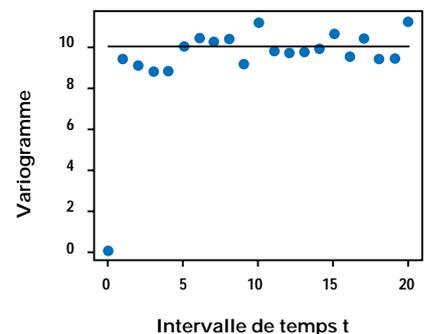
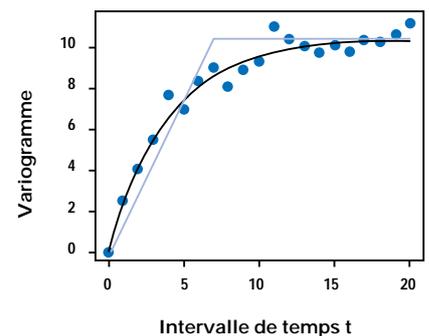


Fig. 4b. Variogramme avec présence d'autocorrélation (ligne noire : modèle exponentiel, ligne bleue : modèle triangulaire)

Variogram corresponding to autocorrelated data (black line: exponential model, blue line: triangular model)



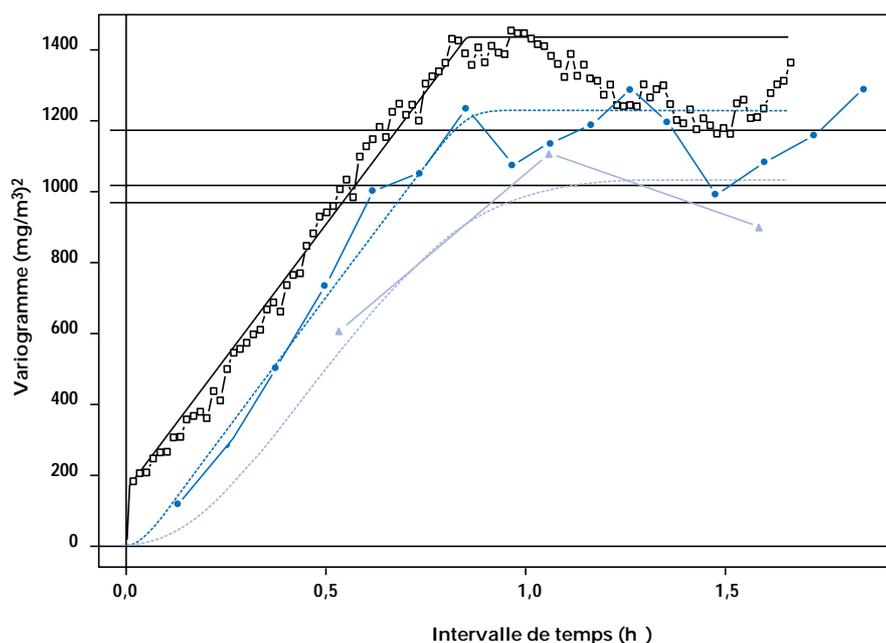


Fig. 5. Variogrammes du matin : ajustement, avec un modèle triangulaire de portée 51 min plus un effet de pépite, de trois variogrammes régularisés (sur supports 22 s, 7 min, 30 min). L'effet de pépite vaut 160 (mg/m³)² au support 22 s
Morning variograms: fitting, with a triangular model (51 min range and nugget effect), for three regularised variograms (integration times: 22 s, 7 min, 30 min). The nugget effect corresponds to 160 (mg/m³)² for an integration time of 22 s

Analyse variographique des mesures du matin

Sur ce variogramme (fig. 5), on peut noter un certain nombre de traits caractéristiques de la structure temporelle :

1. Le palier : c'est la valeur de l'asymptote horizontale dans la fenêtre retenue, ici la matinée. On peut considérer que les valeurs du variogramme expérimental se stabilisent autour d'une valeur voisine de 1300, 1200, 1000 (mg/m³)² pour les 3 supports. L'existence de ce palier permet de conclure que durant la période d'observation, l'intégralité de la variabilité temporelle a été observée. Le palier mesure la variabilité temporelle maximale de la série.

2. La portée : c'est l'intervalle de temps à partir duquel le variogramme atteint son palier. Sur la figure 5, les portées aux 3 supports sont respectivement aux alentours de 0,8 h, 0,9 h et 1,1 h. La portée correspond à l'intervalle de temps au delà duquel des observations consécutives peuvent être considérées comme non-corrélées : dans le cas d'un effet de pépite, la portée est nulle. Si cet intervalle de temps minimal entre deux mesures n'est pas respecté, les mesures sont alors corrélées entre elles. Dans ce cas, l'utilisation de statistiques classiques, ne prenant pas cet effet en considération, donnera des intervalles de confiance erronés.

3. L'effet de pépite : c'est le décalage à l'origine du variogramme. Sur la figure 5, sa valeur est d'environ 160 (mg/m³)² pour le support de 22 s et pratiquement nulle pour les 2 autres supports.

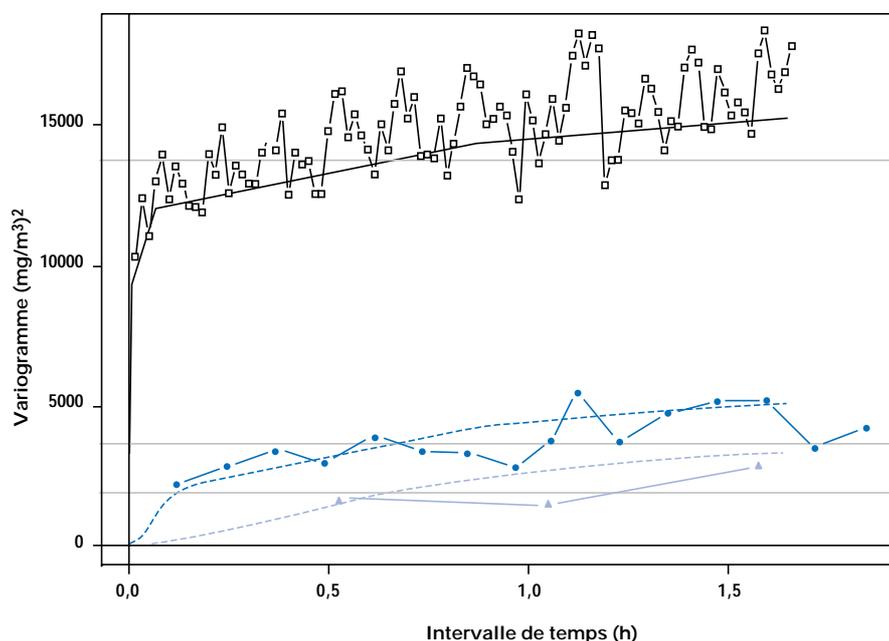


Fig. 6. Variogrammes de l'après-midi : ajustement, avec un modèle triangulaire gigogne (de portées 4 min, 53 min, 1 h 43 min) et un effet de pépite, de trois variogrammes régularisés (sur supports 22 s, 7 min, 30 min). L'effet de pépite vaut 9000 (mg/m³)² au support 22 s
Afternoon variograms: fitting, with a triangular nested model (ranges of 4 min, 53 min, 1 h 43 min, and nugget effect), for three regularised variograms (integration times: 22 s, 7 min, 30 min). The nugget effect corresponds to 9000 (mg/m³)² for an integration time of 22 s

La structure temporelle de la série se divise en 2 composantes indépendantes :

- la première relative à l'effet de pépîte : elle correspond à la composante non corrélée (à l'échelle des 22 s) de la série;
- la seconde, relative au palier et à la portée ; elle correspond à la composante autocorrélée de la série.

Cette dernière implique une corrélation entre les mesures. On peut comparer la part relative de ces 2 composantes sur la variabilité de la série pour les 3 supports. Pour le support de 22 s, l'effet de pépîte représente environ 11 % (160/1400) de la variabilité contre 89 % pour la partie structurée. Pour le support de 7 min et 30 min, l'effet de pépîte à pratiquement disparu. Dans ce cas précis, il n'est pas correct d'utiliser l'approche statistique classique pour évaluer la situation d'exposition.

Ajustement d'un modèle de variogramme

L'ajustement par un modèle de variogramme qui tient compte de ces observations conduit à un modèle à 2 structures : un effet de pépîte d'amplitude $160 \text{ (mg/m}^3\text{)}^2$ et un modèle triangulaire (fig. 4b) croissant linéairement entre 0 et la portée égale à 51 min et d'amplitude $1240 \text{ (mg/m}^3\text{)}^2$, de sorte que le palier global du variogramme modélisé est égal à $1400 \text{ (mg/m}^3\text{)}^2$.

Analyse variographique des données de l'après-midi

Sur la *figure 6* sont représentés les variogrammes des mesures réalisées l'après-midi. Par rapport aux données du matin (fig. 4), on constate que les profils de variogramme sont très différents :

- l'effet de pépîte est important : $9000 \text{ (mg/m}^3\text{)}^2$ pour le support de 22 s et est pratiquement absent des variogrammes des supports 7 min et 30 min. Cela correspond au profil d'exposition très variable au cours de l'après-midi pour les mesures de 22 s (fig. 1) ;
- sur le variogramme, on constate une croissance rapide, jusqu'à 5 min, puis une croissance lente ; dans ce cas, le palier n'est pas atteint. Cela signifie que la période d'observation (4 h) n'est pas suffisamment longue pour décrire l'ensemble de la variabilité de l'exposition.

Ajustement d'un modèle de variogramme

L'ajustement conduit à un variogramme à 4 structures :

- un effet de pépîte d'amplitude $9000 \text{ (mg/m}^3\text{)}^2$;
- trois variogrammes triangulaires, de portée 4 min, 53 min et 1 h 43.

Dans ce cas, deux types d'action seront envisageables :

- accroître la période d'observation afin de quantifier de façon fiable un processus dont la variabilité va au delà de la fenêtre d'observation ;
- réduire la variance globale en stratifiant les données sur la base d'une analyse de l'activité, puis calculer le variogramme de chaque série.

Mesures TCA

Les variogrammes de la série de mesures TCA sont représentés *figure 7*. Les variogrammes expérimentaux ont été calculés pour les supports de 7,5 min, 15 et 30 min.

Le variogramme indique une structure temporelle avec une portée de 1 h 24 min dans le cas d'un support à 7 min 30. L'effet de pépîte pour ce même support vaut $470 \text{ (mg/m}^3\text{)}^2$. Cet effet devient négligeable pour des supports plus importants.

Les variogrammes sont ajustés par un modèle à 2 structures. Un effet de pépîte de $470 \text{ (mg/m}^3\text{)}^2$ et un modèle triangulaire de portée 1 h 24 et d'amplitude $2000 \text{ (mg/m}^3\text{)}^2$, ce qui porte le palier à $2470 \text{ (mg/m}^3\text{)}^2$ pour le support de 7 min 30.

4.2.2. Estimation de la moyenne et de l'intervalle de confiance

Ces calculs de variogrammes montrent que les données sont autocorrélées.

Dans ces conditions, l'écart type d'estimation de la moyenne arithmétique n'est plus égal à :

$$\frac{\sigma}{\sqrt{n-1}}$$

valable uniquement en l'absence de corrélation entre les mesures.

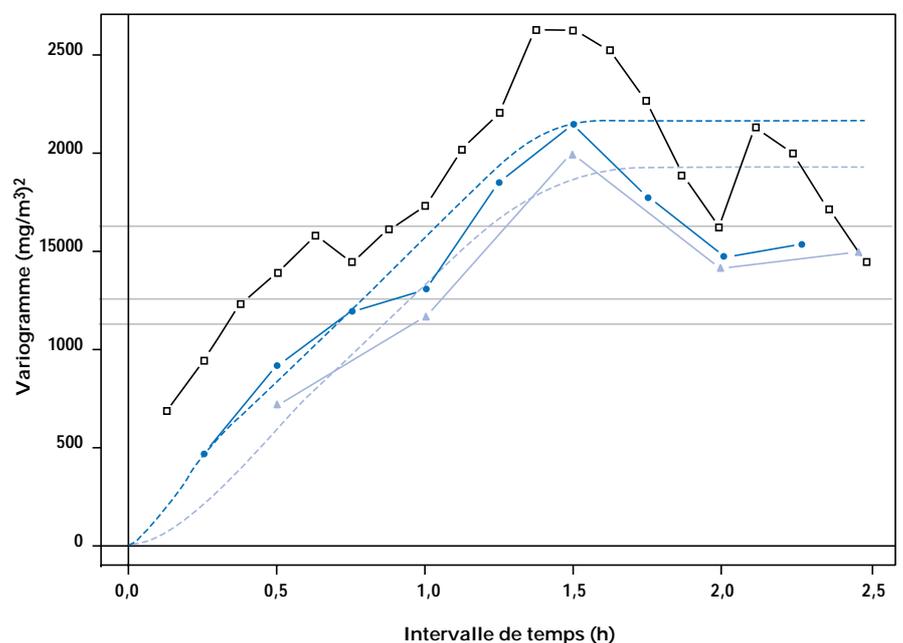


Fig. 7. Ajustement avec un modèle triangulaire de portée 1 h 24 plus un effet de pépîte, de trois variogrammes de TCA régularisés (sur supports 7 min 30, 15 min, 30 min). L'effet de pépîte vaut $470 \text{ (mg/m}^3\text{)}^2$ au support 7 min 30

Triangular model fitting with a range of 1 h 24 min and a nugget effect for three regularised variograms corresponding to active charcoal measurements (integration times: 7.5 min, 15 min, 30 min). The nugget effect corresponds to $470 \text{ (mg/m}^3\text{)}^2$ for an integration time of 7.5 min

De façon à tenir compte de cette corrélation, les calculs de moyenne, d'écart-type d'estimation et d'intervalle de confiance sont effectués à l'aide du « krigeage » [7, 8]. C'est la technique d'estimation de la géostatistique fondée sur la connaissance du variogramme.

Les nouveaux estimateurs obtenus avec la méthode de krigeage figurent dans le *tableau II*.

Les valeurs moyennes d'exposition obtenues à l'aide du calcul classique et les valeurs moyennes résultant du calcul par krigeage en tenant compte de la corrélation sont équivalentes (voir tableaux I et II). Par contre, les bornes supérieures de l'intervalle de confiance à 95 % sont très différentes (*tableau III*).

Pour tous les types de mesure (BK ou TCA), et quelle que soit la période considérée, les résultats montrent qu'en supposant indépendantes des mesures qui ne le sont pas, des bornes supérieures de l'intervalle de confiance de l'exposition sont systématiquement sous-estimées. Cette sous-estimation est d'environ 35 % dans le cas des mesures de l'après-midi. Cette situation, lorsque l'on est amené à comparer des résultats d'exposition à une VLEP, peut conduire à établir un diagnostic erroné.

5. Discussion et conclusion

Dans tous les cas analysés ici, l'analyse de données d'exposition au trichloroéthylène, à l'échelle d'une journée de travail, a mis en évidence une structure temporelle caractérisée par des distances d'autocorrélation à l'échelle de la minute et de l'heure. Ce constat a déjà été signalé dans d'autres travaux menés sur ce sujet [3]. De plus, les résultats montrent que la corrélation des données dépend de la durée de prélèvement (ou temps d'intégration) et qu'il est possible, par la méthode géostatistique, d'analyser simultanément ces deux aspects de la structure temporelle.

Dans ces conditions, une estimation d'exposition réalisée sans tenir compte de la corrélation des mesures conduit à sous-estimer l'intervalle de confiance de l'exposition moyenne et risque d'entraîner la prise de décisions erronées vis-à-vis des exigences réglementaires.

Du fait de l'autocorrélation entre les mesures, cette étude de cas montre les

TABLEAU II
MOYENNES ET VALEURS DE LA BORNE SUPÉRIEURE DE L'INTERVALLE DE CONFIANCE À 95 % DÉTERMINÉES PAR KRIGEAGE - MEANS AND UPPER BOUNDS OF THE 95% CONFIDENCE INTERVAL DETERMINED BY KRIGING

Type	Support	n	Moyenne (mg/m ³)	Ecart-type d'estimation (mg/m ³) σ_E	Borne supérieure (*) de l'intervalle de confiance à 95 % (mg/m ³)
MATIN					
BK	22 sec	202	145	16	176
	7 min	28	146	15	176
	15 min	14	150	17	183
	30 min	6	148	15	179
APRÈS-MIDI					
BK	22 sec	218	117	36	188
	7 min	31	117	52	219
	15 min	15	108	57	222
	30 min	7	123	62	244
SUR LA JOURNÉE					
TCA	7 min 1/2	56	125	21	166
	15 min	28	130	28	183
	30 min	14	125	44	213

(*) Borne supérieure de l'intervalle de confiance = $m + 1,96 \sigma_E$ (écart-type d'estimation).

TABLEAU III
ÉTENDUES DES BORNES SUPÉRIEURES DE L'INTERVALLE DE CONFIANCE À 95 % OBTENU SOUS HYPOTHÈSE D'INDÉPENDANCE ET PAR KRIGEAGE EN TENANT COMPTE DE L'AUTOCORRÉLATION - RANGES OF UPPER BOUNDS OF THE 95% CONFIDENCE INTERVAL OBTAINED UNDER THE INDEPENDENCE HYPOTHESIS AND BY KRIGING, TAKING AUTOCORRELATION INTO ACCOUNT

Type de mesure	Période	Hypothèse d'indépendance	Krigeage
BK	matin	149-177	176-183
BK	après-midi	139-163	188-244
TCA	journée	140-146	166-213

difficultés pratiques pour caractériser l'exposition et pour la comparer à une VLEP.

On peut remarquer qu'il existe 2 stratégies d'apparence simple pour s'affranchir des effets de l'autocorrélation :

- soit effectuer des prélèvements à des instants aléatoires répartis sur la période de référence ;
- soit effectuer des prélèvements à des instants distants d'un intervalle de temps supérieur à la portée de l'autocorrélation.

De telles stratégies d'échantillonnage autorisent l'emploi de méthodes statistiques classiques, puisque le plan d'échan-

tillonnage conduit à des mesures non corrélées. Mais ce ne sont pas forcément des plans optimaux vis-à-vis de la précision de la comparaison à une VLEP et la réalisation pratique de tels plans n'est pas aussi évidente :

- peut-on réellement faire des prélèvements aléatoires ?
- quel intervalle de temps choisir pour être sûr d'être au delà de la portée, et cet intervalle est-il compatible avec les contraintes techniques du prélèvement ?

Les périodes d'observation étudiées ici ont été limitées à une journée de travail pour des raisons pratiques et de faisabilité.

LE KRIGEAGE DE LA MOYENNE

Le krigeage permet d'estimer la moyenne m de la fonction aléatoire $Z(t)$. Dans le cadre de l'hypothèse d'indépendance entre les valeurs $z(t_i)$, qui correspond au cas où le variogramme théorique est un effet de pépite, le meilleur estimateur linéaire au sens des moindres carrés est la moyenne arithmétique. En l'absence de cette hypothèse, il faut alors utiliser le krigeage de la moyenne, noté m_K .

Cet estimateur est, comme la moyenne arithmétique, une combinaison linéaire des valeurs :

$$m_K = \sum_i \lambda_i z(t_i)$$

Le problème consiste à choisir les poids λ_i de telle sorte que :

- l'estimateur soit sans biais ($E(m_K) = m$), cela conduit à la contrainte :

$$\sum_i \lambda_i = 1$$

- la variance d'estimation de m_K soit la plus faible ; la variance de m_K s'exprime en fonction du variogramme modélisé :

$$\sigma_K^2 = \sum_i \sum_j \lambda_i \lambda_j \tilde{\gamma}(t_i - t_j)$$

Ces deux conditions conduisent au système d'équations du krigeage de la moyenne :

$$\begin{cases} \sum_i \lambda_i \tilde{\gamma}(t_i - t_j) + \mu = 0 & \forall j \\ \sum_i \lambda_i = 1 \end{cases}$$

dans lequel μ est le multiplicateur de Lagrange.

Pour estimer l'exposition moyenne à long terme, il faudrait tenir compte de la variabilité inter-journée, celle-ci pouvant être aussi gouvernée par une autocorrélation. Dans ce contexte, les méthodes d'analyse présentées ici peuvent être aussi appliquées. Seule la collecte des données est le facteur limitant pour étudier l'autocorrélation à long terme, mais ce travail permettrait de spécifier les critères pertinents à prendre en compte pour définir une stratégie de mesurage de l'exposition à long terme : par exemple, exposition moyenne sur une année de travail.

En effet, l'estimation de l'exposition à long terme, à partir de mesures collectées sur des périodes d'observation limitées, est possible en tenant compte de la structure temporelle de l'exposition.

Pour exploiter pratiquement cette possibilité et en déduire une stratégie de prélèvement optimale, il est nécessaire d'acquérir des connaissances sur les relations entre les conditions d'exposition (analyse du travail) et les typologies de variogrammes. Ces connaissances impliquent d'étudier les caractéristiques des situations d'exposition en relation à des classes de variogrammes, afin de permettre « d'inverser le processus » et d'être en mesure, à terme, de définir *a priori* une stratégie de prélèvement adaptée à un type de situation d'exposition ou de travail.

Cette démarche a déjà été proposée en employant les fonctions d'autocorrélation [9]. Mais en employant l'approche géostatistique, la même démarche est possible avec un outil plus général, le variogramme, dont l'estimation ne demande pas plus de mesures que celle d'une fonction d'autocorrélation.

Une autre voie d'approche pratique de cette question serait d'étudier la relation entre des mesures d'ambiance et des mesures d'exposition individuelle. En effet, des mesures d'ambiance, réalisées en point fixe et en continu puis analysées avec les méthodes géostatistiques appliquées aux corrélations spatiales [10], associées à une analyse d'activité adaptée, devraient renseigner le préleveur sur le type de variation prévisible de l'exposition individuelle et donc guider le choix d'une stratégie de mesurage. Cette étude ferait intervenir les corrélations temporelles mais aussi spatiales.

Ce travail ouvre de nombreuses pistes et peut susciter des prolongements dans plusieurs directions d'études. Il permet de mieux situer les intérêts et les limites de la géostatistique dans le domaine de l'hygiène industrielle [5]. La géostatistique offre un cadre plus cohérent que l'analyse statistique classique basée sur l'hypothèse d'indépendance, mais implique une connaissance approfondie des situations d'exposition pour caractériser l'autocorrélation.

Remerciements

Les auteurs remercient très sincèrement Mme A. GOOD et Mr C. VALLET du Laboratoire interrégional de chimie du Val Rosay (Lyon) pour la réalisation de la campagne de mesures d'exposition au trichloroéthylène.

Reçu en août 1998, accepté en février 1999.

BIBLIOGRAPHIE

1. NF EN 689 - Atmosphères des lieux de travail. Conseils pour l'évaluation de l'exposition aux agents chimiques aux fins de comparaison avec des valeurs limites et stratégie de mesurage. Paris-La Défense, AFNOR, juillet 1995, 34 p.
2. FRANCIS M., SELVIN S., SPEAR R., RAPPA-PORT S. - The effect of autocorrelation on the estimation of worker's daily exposures. *American Industrial Hygiene Association Journal*, 1989, 50, pp. 341-350.
3. KUMAGAI S., MATSUNAGA I., KUSAKA Y. Autocorrelation of short term and daily exposure levels in workplaces. *American Industrial Hygiene Association Journal*, 1993, 54, pp. 341-350.
4. PREAT B. - Application of geostatistical methods for estimation of the dispersion variance of occupational exposures. *American Industrial Hygiene Association Journal*, 1987, 48, pp. 877-884.
5. WACKERNAGEL H., LAJAUNIE C., THIERY L., VINCENT R. et GRZEBYK M. - Applying geostatistics to exposure monitoring data in industrial hygiene. In : SOARES A., GOMEZ-HERNANDEZ J., FROIDEVAUX R. (eds.) *geoENV I : Geostatistics for Environmental Applications*, 1997, pp. 463-476, Amsterdam, Kluwer.
6. WACKERNAGEL H. - Fondements de la géostatistique. Séminaire estimation des nuisances, approche statistique, Nancy 17 mars 1993. Paris, INRS, coll. Notes scientifiques et techniques, NS 104, pp. 1-14.
7. WACKERNAGEL H. - Multivariate geostatistics : an introduction with applications. Berlin, Springer-Verlag, 1995, 256 p.
8. MATHERON G. - Géostatistique. *Encyclopedia Universalis*, 1995, Corpus 10, pp. 387-390, Paris.
9. KUMAGAI S. et MAISUNAGA I. - Approaches for estimating the distribution of short term exposure concentrations for different averaging times. *Annals of Occupational Hygiene*, 1994, 38, pp. 815-825.
10. SCHNEIDER T., PETERSON O., NIELSEN A.A., WINDFELD K. - A geostatistical approach to indoor surface sampling strategies. *Journal of Aerosol Science*, 1990, 4, pp. 555-567.

