

Échantillonnage et reconstitution de données dans l'espace et le temps

Laure MALHERBE
Giovanni CARDENAS
Chantal de FOUQUET
Claire FAUCHEUX

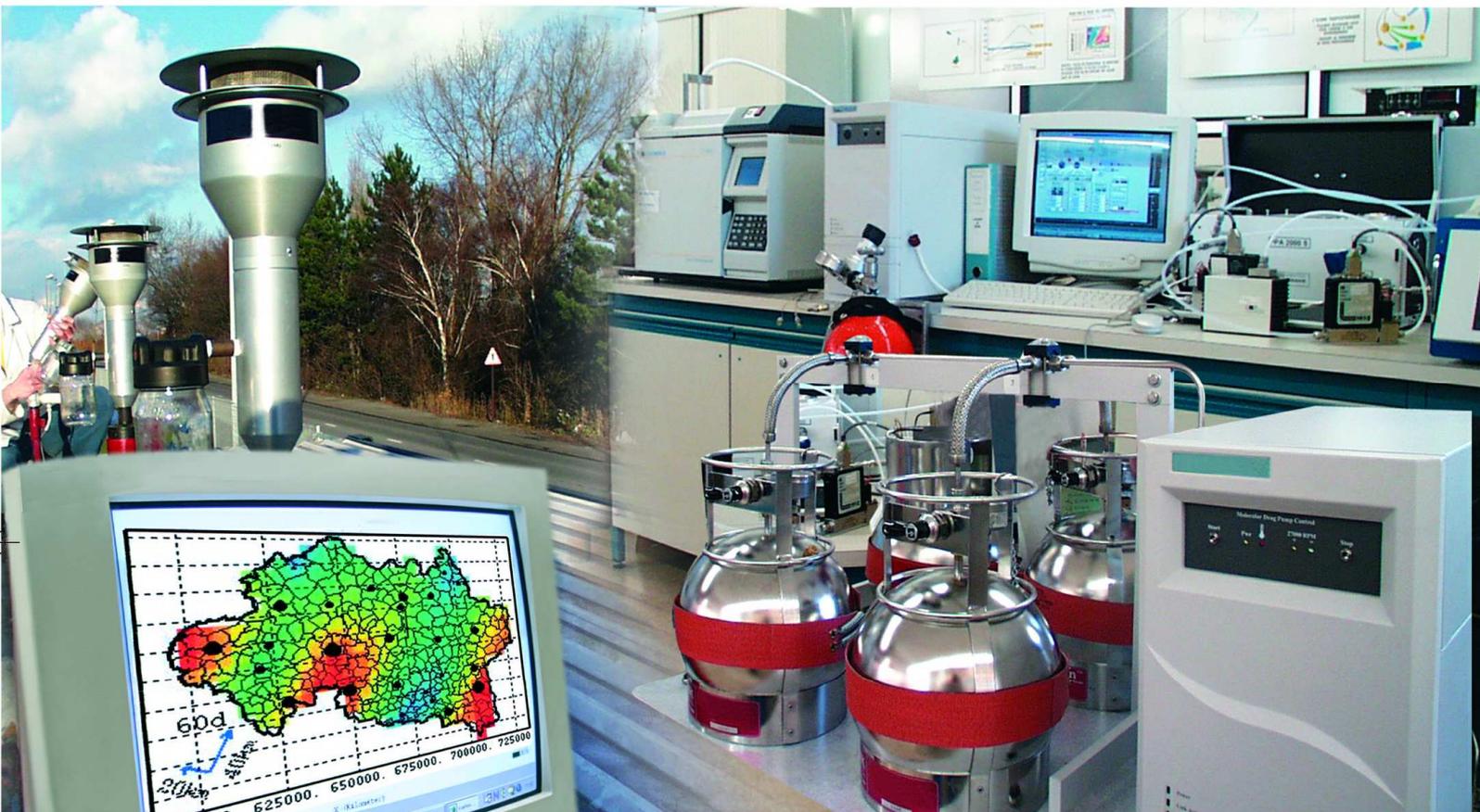
Mars 2009

R090331LMAL

Centre de Géosciences
Géostatistique
MINES ParisTech, Fontainebleau, France



Laboratoire Central de Surveillance de la Qualité de l'Air



Modélisation - Traitements numériques

Echantillonnage et reconstitution de données dans l'espace
et le temps

Décembre 2008

Programme 2008

L. MALHERBE, G. CARDENAS





PREAMBULE

Le Laboratoire Central de Surveillance de la Qualité de l'Air

Le Laboratoire Central de Surveillance de la Qualité de l'Air est constitué de laboratoires de l'Ecole des Mines de Douai, de l'INERIS et du LNE. Il mène depuis 1991 des études et des recherches finalisées à la demande du Ministère chargé de l'environnement, sous la coordination technique de l'ADEME et en concertation avec les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA). Ces travaux en matière de pollution atmosphérique supportés financièrement par la Direction Générale de la Prévention et des Risques du Ministère de l'Ecologie, de l'Energie, du Développement durable et de l'Aménagement du Territoire sont réalisés avec le souci constant d'améliorer le dispositif de surveillance de la qualité de l'air en France en apportant un appui scientifique et technique aux AASQA.

L'objectif principal du LCSQA est de participer à l'amélioration de la qualité des mesures effectuées dans l'air ambiant, depuis le prélèvement des échantillons jusqu'au traitement des données issues des mesures. Cette action est menée dans le cadre des réglementations nationales et européennes mais aussi dans un cadre plus prospectif destiné à fournir aux AASQA de nouveaux outils permettant d'anticiper les évolutions futures.



Echantillonnage et reconstitution de données dans l'espace et le temps

Laboratoire Central de Surveillance
de la Qualité de l'Air

Thème : Modélisation - Traitements numériques

Programme financé par la
Direction Générale de la Prévention et des Risques (DGPR)

2008

INERIS : L. MALHERBE, G. CARDENAS

MINES ParisTech / ARMINES : Ch. de FOUQUET, C. FAUCHEUX

Ce document comporte 19 pages (hors couverture et annexes)

	Rédaction	Vérification	Approbation
NOM	L. MALHERBE	L. ROUÏL	M. RAMEL
Qualité	Ingénieur Direction des Risques Chroniques	Responsable du pôle DECI Direction des Risques Chroniques	Responsable LCSQA/INERIS Direction des Risques Chroniques
Visa			

TABLE DES MATIÈRES

RÉSUMÉ	6
1. INTRODUCTION	7
2. ECHANTILLONNAGE ET RECONSTITUTION DE DONNÉES DANS LE TEMPS	7
3. ECHANTILLONNAGE ET RECONSTITUTION DE DONNÉES DANS L'ESPACE	9
3.1 Introduction	9
3.2 Plan d'échantillonnage initial	9
3.2.1 Généralités	9
3.2.2 Méthode.....	11
3.3 Ajustement du plan d'échantillonnage a l'issue de la première campagne .	16
3.4 Campagnes de surveillance ultérieures	16
3.4.1 Principes généraux.....	17
3.4.2 Méthode.....	18
4. CONCLUSION	18
5. REFERENCES	19
6. LISTE DES ANNEXES	19

RÉSUMÉ

Depuis plusieurs années, le LCSQA s'intéresse aux questions posées par l'élaboration d'une stratégie d'échantillonnage pour l'estimation de la répartition des concentrations atmosphériques de polluants dans le temps et l'espace. Ce rapport synthétise les résultats des travaux conduits en 2008.

En ce qui concerne l'aspect temporel, le LCSQA a participé activement avec des AASQA au groupe de travail *Plans d'échantillonnage et reconstitution de données* animé par l'ADEME et a assuré le secrétariat de ce groupe. Les travaux réalisés ont conduit à la rédaction d'un guide méthodologique et au développement d'un logiciel permettant d'appliquer les méthodes mises au point. Pour aider les AASQA à prendre en main ces outils et à assimiler les concepts statistiques sous-jacents, l'INERIS a organisé, avec l'appui du groupe de travail, des sessions de formation pour l'ensemble des AASQA (fin 2008-début 2009).

Pour ce qui est de l'aspect spatial, deux questions ont été approfondies en collaboration avec ARMINES : celle de l'élaboration d'un plan d'échantillonnage initial en tenant compte de variables auxiliaires et celle de la surveillance sur plusieurs années. Dans la première situation, la méthode proposée consiste schématiquement :

- à sélectionner des variables auxiliaires corrélées à l'aide des stations fixes ou de toute autre information disponible ;
- à définir un maillage régulier (maille carrée ou motif centré) d'après la connaissance qu'on a du polluant et les préconisations existantes : l'espace géographique est ainsi parcouru;
- à compléter ce maillage de façon que toute la gamme des valeurs prises par les variables auxiliaires soit représentée dans l'échantillonnage : ainsi l'espace des variables auxiliaires est-il lui aussi entièrement parcouru ; dans le modèle géostatistique, les relations avec ces variables auxiliaires pourront être correctement calées ;
- à contrôler la variance d'estimation en effectuant des essais de krigeage à partir de valeurs fictives et de différents modèles de variogramme.

Cette méthode est illustrée sur un jeu de données de benzène (concentrations urbaines de fond).

Dans la seconde situation, on suppose que la campagne de mesure initiale et la vérification a posteriori de l'échantillonnage ont permis de définir un plan d'échantillonnage de référence. Etant donné une campagne conduite selon ce plan, la méthode consiste à tirer parti des corrélations entre saisons ou entre années pour diminuer le nombre de points dans les campagnes de mesure suivantes. Les recommandations relatives au plan d'échantillonnage initial s'appliquent encore à l'échantillonnage réduit qui doit couvrir à la fois l'espace géographique et l'espace des variables auxiliaires.

1. INTRODUCTION

Les campagnes de mesure font partie intégrante des moyens de surveillance mis en œuvre par les AASQA.

L'objectif peut être de recueillir une information sur la qualité de l'air en un point précis du territoire, par exemple au centre d'une commune dépourvue de station fixe. La question est alors de déterminer le nombre, la durée et la répartition temporelle des mesures à effectuer en ce point afin d'y estimer de manière fiable la variable d'intérêt (ex : concentration moyenne annuelle).

Dans d'autres cas, la campagne a pour objet d'évaluer la distribution spatiale des concentrations sur une zone et de produire une cartographie. Selon le type de variable à représenter (moyenne saisonnière, moyenne annuelle...), la question de l'échantillonnage temporel se pose encore. Mais le problème est également de définir le nombre minimum de points de mesure que doit comprendre la campagne et la façon la plus adéquate de les répartir dans l'espace.

S'agissant de l'échantillonnage dans le temps et de la reconstitution de données annuelles, le LCSQA a participé aux travaux du GT *Plans d'échantillonnage et reconstitution de données*. Dans ce cadre, il a organisé des sessions de formation sur ce sujet. Ces actions sont brièvement décrites au chapitre 2. Pour de plus amples informations, on consultera les documents publiés par le GT (cf. références citées dans ce chapitre).

En ce qui concerne l'échantillonnage spatial, le LCSQA a produit en 2007 un premier guide de recommandations (Wroblewski et al., 2007) en fonction du polluant et du type de zone considérés. En 2008 une étude a été engagée en collaboration avec ARMINES (Centre de Géosciences/ équipe de géostatistique de l'Ecole des Mines de Paris) afin de compléter ces préconisations. Deux points ont été abordés : la prise en compte de variables auxiliaires dans la définition d'un plan d'échantillonnage ; le réajustement de plans d'échantillonnage pour une surveillance sur plusieurs années. Le chapitre 3 offre une rapide synthèse de ces études. Le rapport d'ARMINES est joint en annexe.

2. ECHANTILLONNAGE ET RECONSTITUTION DE DONNÉES DANS LE TEMPS

De janvier 2006 à janvier 2009, l'EMD et l'INERIS ont contribué aux activités du GT *Plans d'échantillonnage et reconstitution de données*, groupe de travail animé par l'ADEME et composé de représentants des AASQA et du LCSQA. Ils ont participé aux développements méthodologiques et l'INERIS a assuré le secrétariat du groupe. L'année 2008 a permis de finaliser le guide technique sur la planification de l'échantillonnage et la reconstitution de données ainsi que les programmes informatiques associés.

Enfin, avec la participation des membres du GT, le LCSQA a organisé des sessions de formation aux méthodes et aux outils élaborés par le GT. Trois sessions destinées à l'ensemble des AASQA ont eu lieu les 4 et 5 décembre 2008, les 15 et 16 décembre 2008 et les 13 et 14 janvier 2009. Elles alternaient exposés théoriques et travaux pratiques sur logiciel. Les présentations préparées pour cette occasion sont consultables sur le site Internet du LCSQA (<http://www.lcsqa.org/thematique/traitements-numeriques/modelisation/documentation>).

2.1 PRÉSENTATION DU GUIDE

Ce document, qui sera prochainement édité, peut être consulté en ligne sur le site Internet du LCSQA (<http://www.lcsqa.org/thematique/traitements-numeriques/modelisation/documentation>). Il comprend deux parties.

La première montre comment, à partir d'informations existantes, il est possible de construire un plan d'échantillonnage capable de saisir correctement la variabilité temporelle des concentrations. La méthodologie développée repose sur l'exploitation de données caractéristiques du polluant et du type de site étudiés et fait appel aux principes statistiques de la théorie des sondages. Elle comprend plusieurs étapes :

- analyse des contraintes de qualité et de ressources ;
- analyse de la variabilité temporelle des concentrations ;
- définition et dimensionnement du plan d'échantillonnage ;
- évaluation des ressources nécessaires à la mise en œuvre du plan choisi et contrôle de la faisabilité de ce plan.

La seconde partie décrit trois méthodes de reconstitution de données : la méthode dite des « Plans de Sondage », la méthode « ISO » issue de la norme ISO 9359 et la méthode de régression linéaire. Celles-ci permettent, selon des approches différentes, de tirer parti des informations disponibles (données d'échantillonnage, variables auxiliaires) et d'obtenir des indicateurs fiables de la qualité de l'air (ici des estimations de concentrations moyennes annuelles et de leur intervalle de confiance). L'intérêt et les contraintes de chaque méthode sont présentés.

Une dizaine d'annexes (état des pratiques en France et en Europe, définitions statistiques, fiches techniques sur les méthodes, cas d'étude) complète ces deux chapitres.

2.2 OUTILS INFORMATIQUES

Les méthodes d'échantillonnage et de reconstitution peuvent être appliquées grâce à des programmes R développés conjointement par ATMO Poitou-Charentes, AIR Pays de la Loire et l'INERIS. Afin d'en permettre l'utilisation par le plus grand nombre, l'INERIS a développé une interface logicielle accessible par le site Internet du LCSQA. Celle-ci comprend :

- un onglet destiné à la planification de l'échantillonnage (<http://www.lcsqa.org/thematique/traitements-numeriques/modelisation/planification-temporelle-de-echantillonnage>);

- un onglet destiné à la reconstitution (estimation d'une concentration moyenne annuelle et de son incertitude : <http://www.lcsqa.org/thematique/traitements-numeriques/modelisation/reconstitution-de-donnees>).

Différents fichiers sont également proposés en téléchargement : fichier Excel pour l'analyse temporelle préalable à la planification, exemples de fichiers d'entrée, mode d'emploi illustré de l'interface.

3. ECHANTILLONNAGE ET RECONSTITUTION DE DONNÉES DANS L'ESPACE

3.1 INTRODUCTION

En 2006 et 2007, le LCSQA a émis un guide de recommandations sur le choix d'une maille d'échantillonnage en fonction du polluant (NO₂, benzène, ozone) et du domaine d'étude (zones régionales/rurales, grandes et moyennes agglomérations, zones industrielles, aéroportuaires, de proximité automobile) (Wroblewski et al., 2007).

Notre objectif est d'affiner ces préconisations en distinguant plusieurs stades dans l'élaboration des campagnes :

- Etat initial : conception d'un plan d'échantillonnage dans une zone qui n'a jamais fait l'objet de mesures ; conduite de la première campagne
- Contrôle à l'issue de la première campagne : réajustement du plan d'échantillonnage (définition d'un schéma de référence) et s'il est besoin, réalisation de nouvelles mesures.
- Saisons et/ou années suivantes : réduction du schéma de référence et mise en place d'une surveillance selon un plan d'échantillonnage allégé.

Le rapport d'ARMINES, joint en annexe (annexe 1, chapitre II), présente les principales considérations théoriques qui doivent guider la mise en œuvre de ces étapes. Les paragraphes 3.2 à 3.4 ci-après en reprennent les principaux éléments. Deux exemples traités respectivement par ARMINES et l'INERIS - échantillonnage du benzène à Bordeaux (annexe 1, chapitre III) et du NO₂ à Reims (cf. figures ci-après) - illustrent la conception initiale d'un plan d'échantillonnage. Notons que la pollution de fond est plus particulièrement examinée. L'étude LCSQA de 2009 sur la cartographie urbaine de haute résolution fournira des compléments sur l'échantillonnage à proximité des routes.

3.2 PLAN D'ÉCHANTILLONNAGE INITIAL

3.2.1 GÉNÉRALITÉS

La définition d'un premier plan d'échantillonnage repose sur les données suivantes :

- Mesures en continu des stations fixes disponibles ;
- Variables auxiliaires connues dans tout le domaine ;
- Informations sur la variabilité spatiale du polluant.

Elle doit satisfaire à plusieurs exigences :

- Les **objectifs de la cartographie** et les **critères de précision** doivent être clairement définis. En particulier, les choix d'échantillonnage peuvent différer suivant que l'on adopte un critère de précision absolue ou relative (cf. 3.2.2, étape 3b).
- Malgré les coûts et les contraintes techniques, il convient de rechercher un **certain surdimensionnement de l'échantillonnage**. Cette préconisation est d'autant plus pertinente qu'il s'agit d'alléger ultérieurement l'échantillonnage et d'établir une surveillance régulière de la zone par des campagnes de moindre envergure. On ne pourra réduire l'échantillonnage de manière satisfaisante que si l'on acquiert au départ une connaissance suffisamment détaillée des concentrations.
- Les points d'échantillonnage **doivent se répartir non seulement dans l'espace géographique mais également dans l'espace des variables auxiliaires** potentiellement corrélées aux concentrations (cela signifie que toute la gamme des valeurs prises par ces variables sur le domaine d'étude doit être représentée dans l'échantillonnage).

3.2.2 MÉTHODE

Les étapes de la construction d'un plan d'échantillonnage, telles qu'elles ont été définies en collaboration avec ARMINES, sont synthétisées ci-après. La description plus détaillée de la méthode est fournie dans le rapport d'ARMINES (annexe 1).

Etape 1	Données exploitables, méthode
Recensement des informations disponibles et sélection de variables auxiliaires corrélées aux concentrations	Bibliographie, données d'autres campagnes réalisées dans des conditions comparables Etude des corrélations entre les mesures des stations fixes et les variables auxiliaires.

Etape 2	Données exploitables, méthode
Définition d'un maillage initial régulier : <ul style="list-style-type: none">• Délimitation du domaine d'étude• Choix d'un type et d'une taille de maille	<p>La taille de maille pourra être contrainte par des impératifs budgétaires qui limitent le nombre de points.</p> <p>Mais dans tous les cas, elle sera choisie en cohérence avec les caractéristiques du polluant :</p> <ul style="list-style-type: none">• Guide LCSQA 2007 : indications sur le choix d'une taille de maille en fonction du polluant et du type de zone (rurale, urbaine, ...) (Figure 1)• Données sur la portée spatiale du phénomène (études antérieures, portée des variables auxiliaire). <p>Plusieurs types de maillage sont possibles (cf. Annexe 1). Si en pratique, la maille régulière est la plus fréquemment adoptée (Figure 1), un schéma à maille centrée peut se révéler économiquement intéressant : il réduit de moitié le nombre de points tout en maintenant le même espacement dans les directions diagonales (Figure 2).</p>

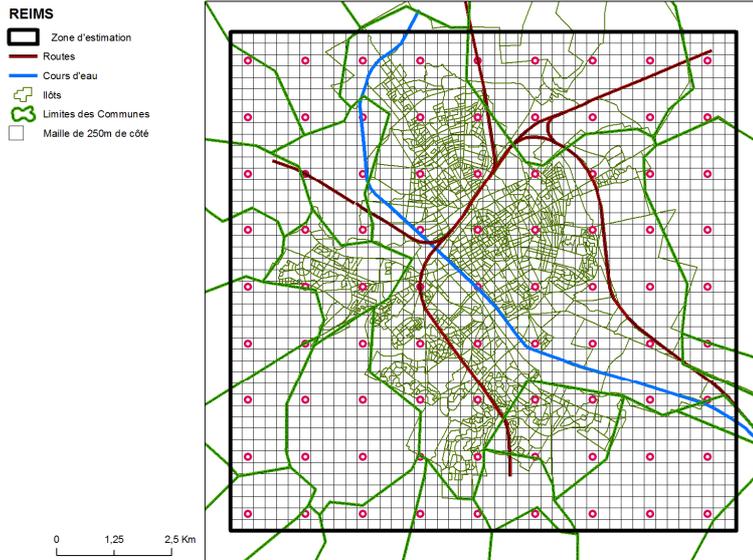


Figure 1 – Echantillonnage du NO₂ à Reims. Maillage initial choisi : la taille de la maille (~1,25 km) s'accorde avec les préconisations relatives à l'échantillonnage du NO₂ en zone urbaine (guide LCSQA, 2007)

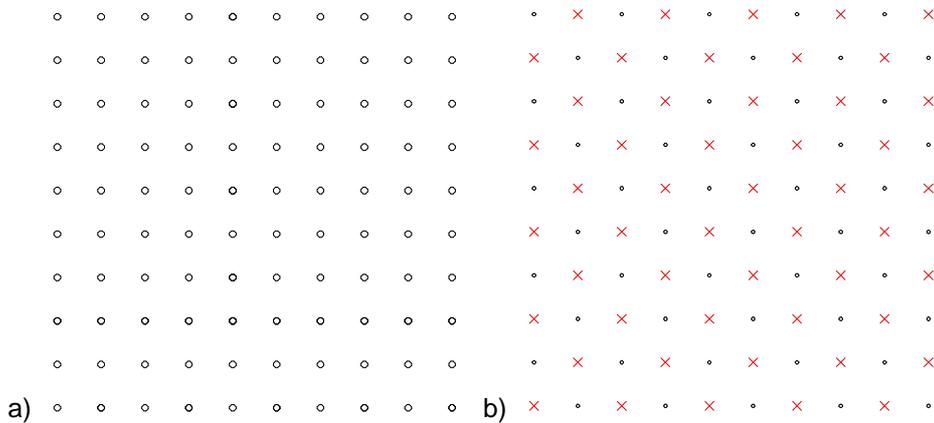


Figure 2 – A gauche : Maille carrée (N points : o) ; à droite : Maille centrée (N/2 points : o ou x)

Etape 3	Méthode
a) Ajustement de l'échantillonnage en fonction des variables auxiliaires	<p>Pour chaque variable auxiliaire sélectionnée, on s'assure graphiquement (histogrammes, nuages de corrélation) que l'échantillonnage permet de parcourir toute la gamme des valeurs prises sur le domaine d'étude.</p> <p>Si ce n'est pas le cas, on ajoute des points là où des manques sont mis en évidence ; on en retire éventuellement (en vérifiant que cette suppression n'a pas d'effets contraires) dans d'autres zones. Une procédure d'ajustement est proposée dans l'exemple de l'annexe 1 (cf. partie III, 5^e étape).</p> <p>Rem. : Cette étape conduit généralement à resserrer l'échantillonnage dans les zones où se situent les valeurs les plus importantes des variables auxiliaires (queues des histogrammes) (Figure 4, Figure 4).</p>
b) Ajustement éventuel de l'échantillonnage en fonction du critère de précision visé	<p>Critère de précision absolue (la précision visée est exprimée en unité de concentration) : on cherche à réduire l'écart-type d'estimation. Si la variabilité croît avec la concentration (effet proportionnel, cf. annexe 1, § II.2) ce critère s'accorde avec les ajustements sur les variables auxiliaires : il conduit à densifier l'échantillonnage dans les zones supposées les plus polluées.</p> <p>Critère de précision relative (la précision visée est exprimée en pourcentage de la concentration) : on cherche à réduire l'écart-type d'estimation rapporté à la concentration estimée. Ce critère peut conduire à des ajustements artificiels qui consistent à densifier l'échantillonnage là où de plus faibles concentrations sont attendues ou à espacer l'échantillonnage là où de plus fortes concentrations sont attendues. Il paraît préférable de privilégier un ajustement de l'échantillonnage en fonction des variables auxiliaire.</p>
c) Ajustements complémentaires	<p>Si le budget le permet, des resserrements locaux de l'échantillonnage permettront d'apprécier la variabilité des concentrations à petite distance. Pour une évaluation correcte de cette variabilité sur l'ensemble du domaine, on densifiera l'échantillonnage (si ce n'est déjà fait) dans des zones de valeurs élevées et de valeurs plus faibles (au moins une zone de valeurs intermédiaires) des variables auxiliaires.</p>

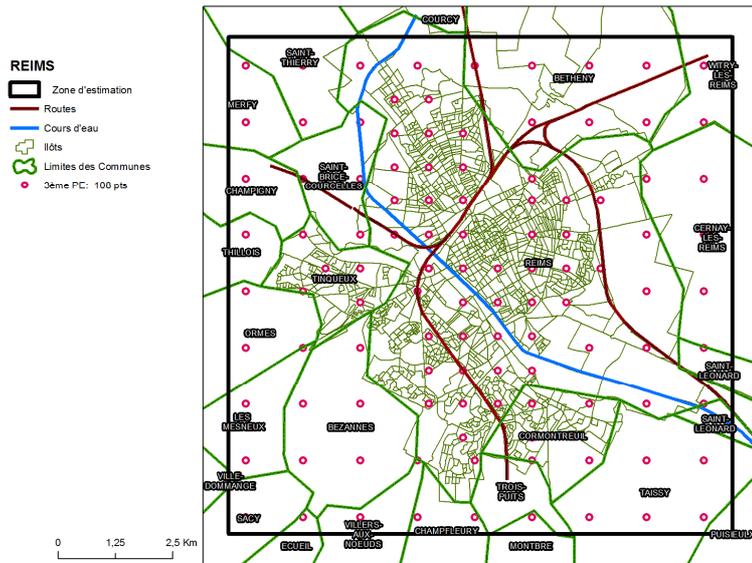


Figure 3 – Echantillonnage du NO₂ à Reims. Maillage ajusté en fonction des variables auxiliaires sélectionnées : logarithme translaté de la densité de population, densité de territoires artificialisés et densité d'émissions de NO_x dans un rayon de 1 km. Cet ajustement conduit à un resserrement de la maille dans les parties du domaine où ces trois variables sont plus élevées.

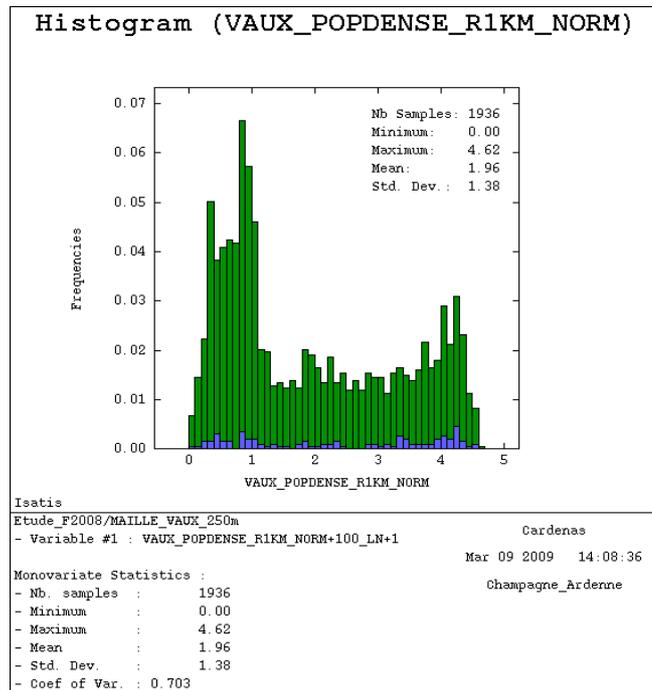
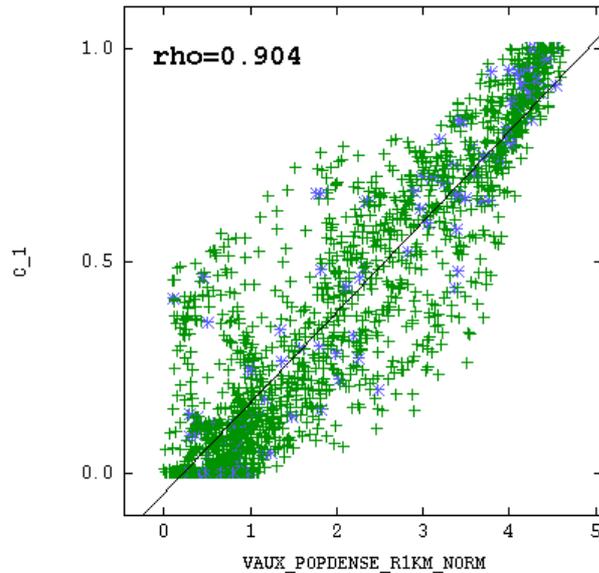


Figure 4 – Echantillonnage du NO₂ à Reims. On vérifie que l'échantillonnage (partie bleue de l'histogramme) permet de parcourir l'étendue des valeurs de chaque variable auxiliaire. Exemple de la densité de population.



Isatis

Figure 5 – Echantillonnage du NO₂ à Reims. Les nuages de corrélation sont un autre moyen de contrôler la représentativité de l'échantillonnage (points en bleu) vis-à-vis des différentes variables auxiliaires. Exemple de la densité de population (en abscisse) et de la densité de territoires artificialisés (en ordonnée).

Etape 4	Méthode
Vérification du plan d'échantillonnage	<p>A partir de la bibliographie, choix de différents variogrammes.</p> <p>Attribution d'une valeur quelconque aux points d'échantillonnage (Rappel : la variance de krigeage ne dépend pas des valeurs mesurées).</p> <p>Pour chaque variogramme prédéfini, réalisation d'un krigeage en dérive externe avec, en dérive, la ou les variables auxiliaires sélectionnées.</p> <p>Examen des cartes de variance de krigeage associées.</p> <p>La variance de krigeage ne permet pas de contrôler de façon absolue la précision atteinte puisque les paliers des variogrammes sont fixés plus ou moins arbitrairement ; en revanche, elle permet de l'évaluer de manière relative à l'intérieur du domaine. En fonction des cartes de variance obtenues, des points d'échantillonnage pourront être ajoutés dans les zones de variance élevée ou au contraire retirés des zones de faible variance.</p> <p>Retour à l'étape 3a) pour vérifier le nouveau schéma d'échantillonnage.</p>

3.3 AJUSTEMENT DU PLAN D'ÉCHANTILLONNAGE A L'ISSUE DE LA PREMIÈRE CAMPAGNE

Le plan d'échantillonnage initial est fondé sur plusieurs hypothèses relatives à la variable de concentration ; ces hypothèses concernent en particulier la portée spatiale du phénomène et les corrélations avec les variables auxiliaires.

Même si le plan d'échantillonnage a été construit avec soin, rien ne garantit qu'elles soient toutes vérifiées. Il est nécessaire de s'en assurer à la suite de la première campagne. Ce contrôle peut porter sur la campagne entière (à l'issue de toutes les périodes de mesure) ou intervenir au cours de celle-ci (par exemple après la première saison de mesure).

Dans tous les cas, des données de mesure sont disponibles pour cette vérification. Elles permettent en particulier de calculer un variogramme expérimental. Notons que ce dernier correspondra éventuellement à une période différente de celle de référence, par exemple à l'une seulement des deux saisons.

En étudiant les résultats de la validation croisée et la carte d'écart-type de krigeage, on cherchera également à voir si la précision d'estimation recherchée est ou non atteinte.

Si les résultats de cette évaluation sont favorables, le schéma d'échantillonnage initial sera jugé pertinent et défini comme « schéma de référence ». Dans le cas encore plus favorable où la précision souhaitée est très largement atteinte, le schéma de référence pourra même être allégé par rapport au plan initial (ex : passage d'une maille carrée à une maille centrée). On évaluera l'effet de cet allègement en effectuant un nouveau krigeage avec le variogramme modélisé, quitte à compléter ce krigeage par une étude de sensibilité (par exemple, pour le NO₂, le variogramme des concentrations estivales peut différer sensiblement de celui des concentrations hivernales, et donc aussi du variogramme des concentrations annuelles).

En revanche, si l'analyse des données recueillies révèle des insuffisances, le plan d'échantillonnage sera réajusté. Avant d'envisager une diminution du nombre de points comme il est décrit au paragraphe 3.4, on devra conduire préalablement une campagne de mesure – ou des mesures complémentaires – selon ce nouveau plan.

3.4 CAMPAGNES DE SURVEILLANCE ULTÉRIEURES

Cette troisième partie porte sur la réduction de l'échantillonnage pour une surveillance sur le long terme.

La technique du « cokrigeage ordinaire » ou du « cokrigeage ordinaire avec dérive externe » permet d'estimer les concentrations en utilisant :

- les données de la période étudiée ;

- les données d'une ou de plusieurs périodes antérieures ;
- éventuellement une ou plusieurs variables auxiliaires.

Par cette méthode, on peut ainsi bénéficier, pour l'estimation des concentrations d'une période, de l'exploitation des informations - potentiellement plus nombreuses - d'une campagne antérieure.

Supposons qu'une campagne de référence ait été conduite au cours d'une saison ou d'une année. L'idée est de tirer parti du cokrigeage pour diminuer l'échantillonnage au cours de la saison ou des années suivantes.

Cette méthode peut s'appliquer efficacement entre deux saisons d'une même année (Fouquet et al., 2008 ; Gallois et al., 2005, exemple de Mulhouse ; Perdrix et Cárdenas, 2006, annexe 4, exemple de Lille). Elle peut s'appliquer aussi entre deux années différentes (Cárdenas et Malherbe, 2007, §3.3.1 et annexe 7, exemple de Rouen). Plusieurs configurations sont alors possibles, selon les données disponibles par période et les corrélations mises en évidence :

- cokrigeage entre les moyennes annuelles des années N et N+k ;
- cokrigeage entre l'hiver ou l'été de l'année N+k et la saison de l'année N qui lui est la mieux corrélée ;
- cokrigeage à trois variables entre l'hiver et l'été de l'année N+k et la saison de l'année N qui leur est la mieux corrélée ;
- cokrigeage à quatre variables entre les deux saisons de l'année N+k et les deux saisons de l'année N.

Pour des indications sur l'application du cokrigeage, on se reportera aux références précitées. Nous nous intéressons ici à la question de l'échantillonnage spatial. Deux questions se posent :

- combien de points d'échantillonnage faut-il conserver au minimum ?
- où peut-on judicieusement supprimer des points d'échantillonnage ?

3.4.1 PRINCIPES GÉNÉRAUX

La réponse aux deux questions précédentes dépend grandement des caractéristiques du plan d'échantillonnage initial, des niveaux de concentration mesurés et de l'évolution des concentrations dans le temps (variations saisonnières, variations interannuelles). On peut toutefois énoncer quelques grands principes :

- Les points conservés doivent être en nombre suffisant (environ une quarantaine par agglomération) pour permettre le calcul et la modélisation d'un variogramme multivariable.
- Ils doivent encore se répartir dans tout l'espace des variables auxiliaires : du schéma de référence, nous proposons de conserver, tout en l'élargissant, le maillage de base, et de préserver une partie des sites supplémentaires, ceux qu'on juge indispensables au calage des relations entre concentrations et variables auxiliaires.

- Avant d'établir l'échantillonnage sur l'année N+k, il faut vérifier que le domaine d'étude n'a pas subi de modification majeure : changement dans les émissions, l'occupation du sol, ... Si tel est le cas, il sera probablement nécessaire d'ajouter des sites de mesure en conséquence (retour à l'étape 3a) du paragraphe 3.2.2).
- Si un critère de précision absolue est visé, sous réserve que les exigences précédentes soient respectées, la dégradation de l'échantillonnage pourra porter plus particulièrement sur les zones où la variance de krigeage est la plus faible (zones mises en évidence par l'exploitation de la campagne de référence) et sur les périodes où la variabilité spatiale des concentrations est moindre (par exemple, comme c'est souvent le cas, l'été pour le NO₂).

3.4.2 MÉTHODE

Pour tester différentes dégradations du plan d'échantillonnage et en évaluer l'impact sur la précision d'estimation, le rapport d'ARMINES propose de s'appuyer sur une approximation majorante de la variance de krigeage. La méthode est détaillée d'un point de vue théorique dans la partie II.5 de l'annexe 1. On pourra en proposer une application concrète dans les travaux futurs du LCSQA.

4. CONCLUSION

Echantillonnage temporel

Les travaux réalisés par le GT *Plans d'échantillonnage et reconstitution de données* se sont achevés à la fin de l'année 2008. Les méthodes développées ont été diffusées auprès des AASQA à l'occasion de sessions de formation ; la documentation et les outils informatiques correspondants ont été mis à disposition sur le site du LCSQA (www.lcsqa.org). En 2009, le LCSQA continuera à apporter aux AASQA une assistance technique sur ces sujets et une aide à l'utilisation du logiciel.

Echantillonnage spatial

Deux points ont été approfondis en complément du guide élaboré en 2007 par le LCSQA : la prise en compte de variables auxiliaires dans l'élaboration d'un premier plan d'échantillonnage ; la mise en place d'une surveillance sur le long terme. Une méthode générale assortie de principes et de recommandations a été mise au point par le Centre de Géosciences de l'École des Mines de Paris, dans le cadre d'une collaboration entre l'INERIS et ARMINES. Le premier point a été mis en pratique sur un jeu de données de benzène (Bordeaux, AIRAQ). Le second pourra être appliqué à l'occasion de prochaines études afin d'en faciliter la compréhension. D'autre part, comme en fait état l'analyse bibliographique d'ARMINES, on constate un engouement pour des techniques d'optimisation de

l'échantillonnage telles que le « recuit simulé ». ARMINES propose de comparer cette dernière méthode à l'approche plus pragmatique qu'il a développée. La question de l'échantillonnage en situation de proximité sera abordée plus en détail en 2009, dans l'étude sur la cartographie urbaine de haute résolution.

5. REFERENCES

Cárdenas G., Malherbe L., 2007. Représentativité des stations de mesure du réseau national de surveillance de la qualité de l'air : application des méthodes géostatistiques à l'évaluation de la représentativité spatiale des stations de mesure de NO₂ et O₃. Rapport LCSQA, www.lcsqa.org.

Fouquet C. (de), Gallois D., Perron G., 2007. Geostatistical characterization of the nitrogen dioxide concentration in an urban area. Part I: Spatial variability and cartography of the annual concentration. *Atmospheric Environment*, 41, n°32, pp.6701-6714.

Gallois D., de Fouquet C., Le Loc'h G., Malherbe L., Cárdenas G., 2005. Mapping annual nitrogen dioxide concentrations in urban areas. O. Leuangthong and C.V. Deutsch (eds), *Geostatistics Banff 2004*, pp. 1087-1096

Perdrix E., Cárdenas G., 2006. Méthode de surveillance des concentrations de NO₂ : cartographie automatique à partir de stations fixes et prise en compte de la proximité. Rapport LCSQA, www.lcsqa.org.

Wroblewski A., Riffault V., Malherbe L., Perdrix E., 2007. Adaptation des plans d'échantillonnage aux objectifs des campagnes. Rapport LCSQA, www.lcsqa.org.

6. LISTE DES ANNEXES

Repère	Désignation	Nombre de pages
Annexe 1	Recommandations pour les schémas d'échantillonnage des campagnes de mesure de la qualité de l'air Rapport d'ARMINES (MINES ParisTech - Centre de Géosciences/Géostatistique)	29

ANNEXE 1