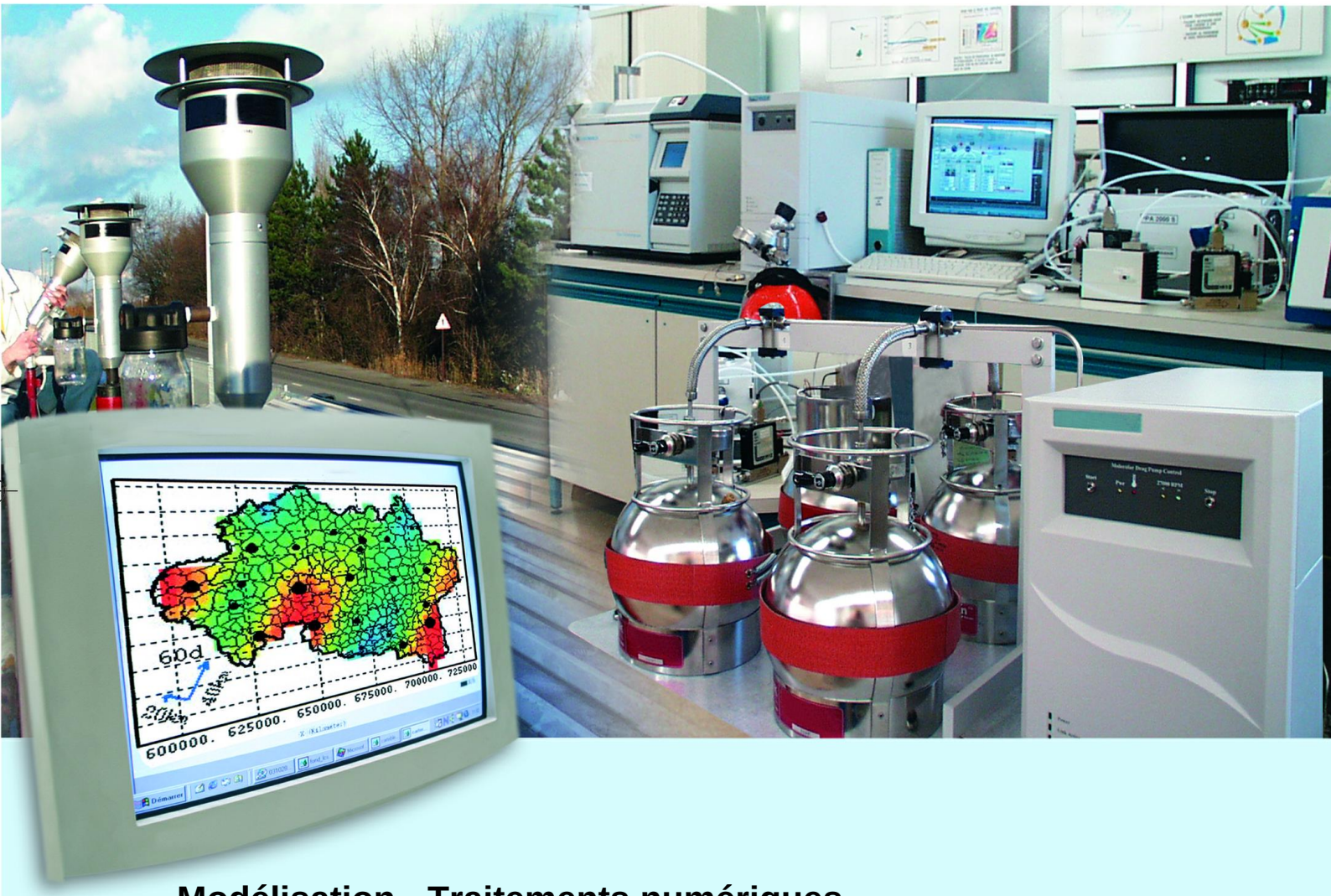
LCSQA

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

INERIS

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 le Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer 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 le
Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer (MEEDDM)

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

RESUME

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 DONNEES 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 PRESENTATION 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 DONNEES 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 (NO_2, 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 NO_2 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'ECHANTILLONNAGE INITIAL

3.2.1 GENERALITES

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 METHODE

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 : • Délimitation du domaine d'étude • Choix d'un type et d'une taille de maille	La taille de maille pourra être contrainte par des impératifs budgétaires qui limitent le nombre de points. Mais dans tous les cas, elle sera choisie en cohérence avec les caractéristiques du polluant : • 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). 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).

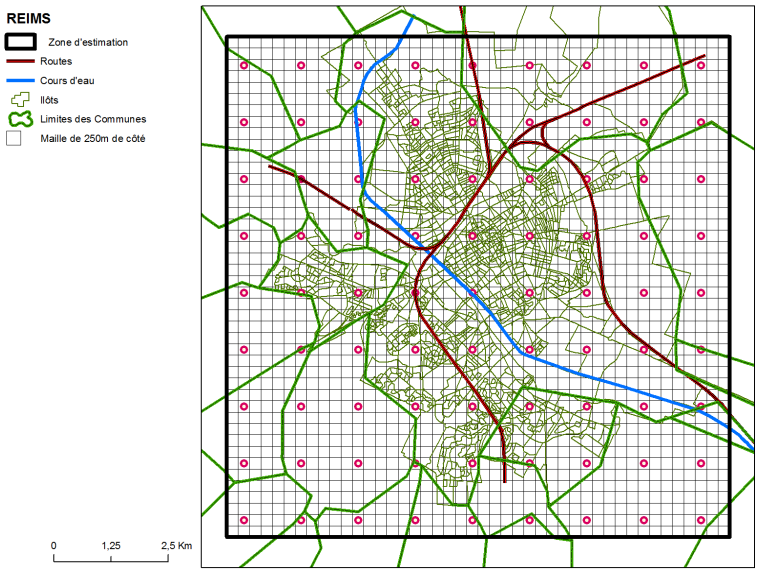


Figure 1 – Echantillonnage du NO_2 à Reims. Maillage initial choisi : la taille de la maille (~1,25 km) s'accorde avec les préconisations relatives à l'échantillonnage du NO_2 en zone urbaine (guide LCSQA, 2007)

a) b)

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	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. 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). 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).
b) Ajustement éventuel de l'échantillonnage en fonction du critère de précision visé	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. 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.
c) Ajustements complémentaires	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.

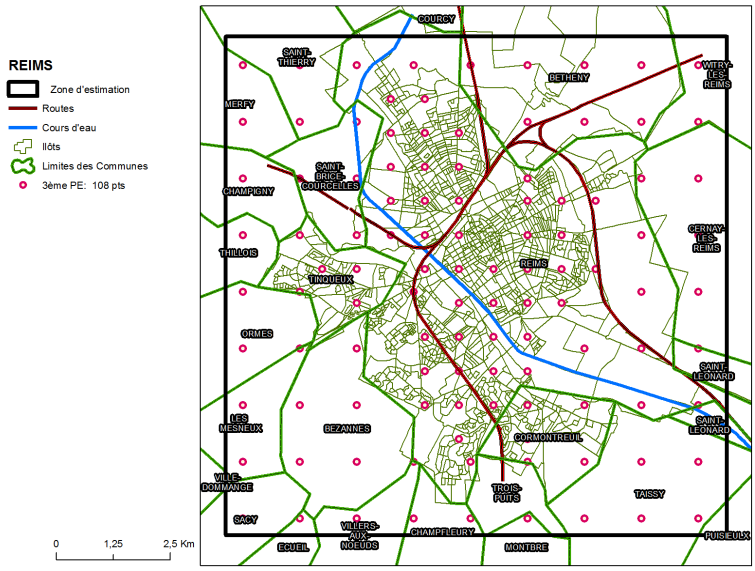


Figure 3 – Echantillonnage du NO_2 à Reims. Maillage ajusté en fonction des variables auxiliaires sélectionnées : logarithme translaté de la densité de population, densité de territoires artificialisés et densité d'émissions de NOx dans un rayon de 1 km. Cet ajustement conduit à un resserrement de la maille dans les parties du domaine où ces trois variables sont plus élevées.

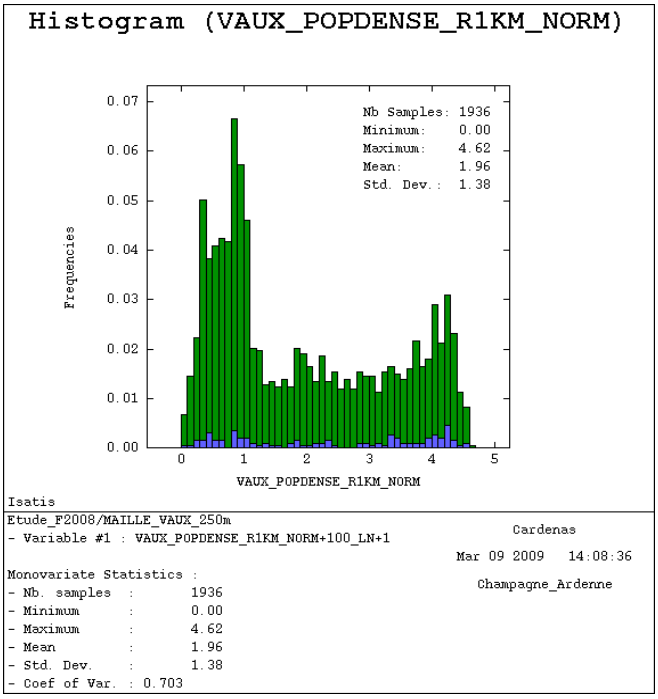


Figure 4 – Echantillonnage du NO_2 à Reims. On vérifie que l'échantillonnage (partie bleue de l'histogramme) permet de parcourir l'étendue des valeurs de chaque variable auxiliaire. Exemple de la densité de population.

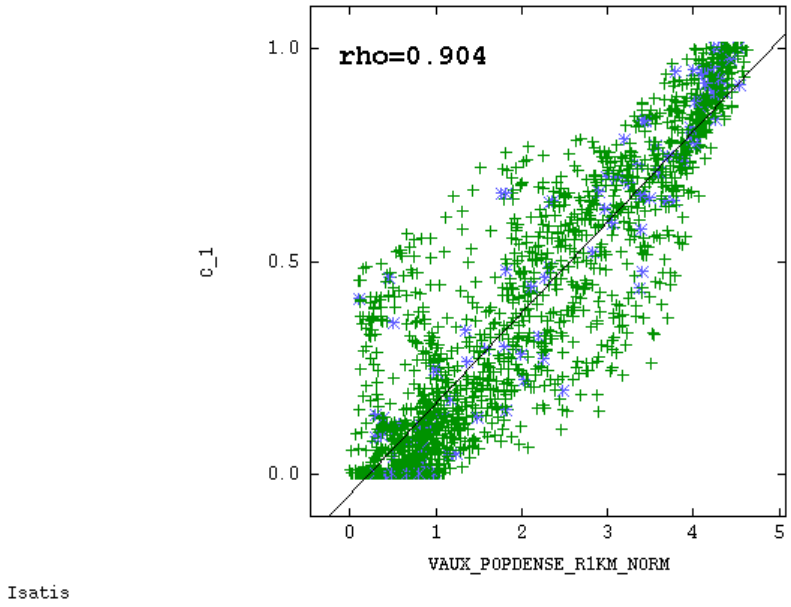


Figure 5 – Echantillonnage du NO_2 à Reims. Les nuages de corrélation sont un autre moyen de contrôler la représentativité de l'échantillonnage (points en bleu) vis-à-vis des différentes variables auxiliaires. Exemple de la densité de population (en abscisse) et de la densité de territoires artificialisés (en ordonnée).

Etape 4	**Méthode**
Vérification du plan d'échantillonnage	A partir de la bibliographie, choix de différents variogrammes. Attribution d'une valeur quelconque aux points d'échantillonnage (Rappel : la variance de krigeage ne dépend pas des valeurs mesurées). Pour chaque variogramme prédéfini, réalisation d'un krigeage en dérive externe avec, en dérive, la ou les variables auxiliaires sélectionnées. Examen des cartes de variance de krigeage associées. 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. Retour à l'étape 3a) pour vérifier le nouveau schéma d'échantillonnage.

3.3 AJUSTEMENT DU PLAN D'ECHANTILLONNAGE A L'ISSUE DE LA PREMIERE 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 NO_2, 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 ULTERIEURES

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 GENERAUX

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 NO_2).

3.4.2 METHODE

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'Ecole 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

Recommandations pour les schémas d'échantillonnage des campagnes de mesure de la qualité de l'air

Chantal de FOUQUET

Claire FAUCHEUX

Mars 2009

Centre de Géosciences
Géostatistique
MINES ParisTech, Fontainebleau, France

TABLE DES MATIERES

Introduction

Les campagnes de mesure de la qualité de l'air, à l'échelle de l'agglomération, du département ou de la région, nécessitent des moyens conséquents. En complément de la connaissance de terrain, qui reste indispensable, quelles recommandations méthodologiques formuler pour que le choix des sites de mesure réponde aux objectifs de cartographie d'une moyenne temporelle avec une précision fixée ? Deux situations sont examinées.

1. Étant donnée une zone qui n'a jamais fait l'objet de campagne. Peut-on proposer aux AASQA des lignes directrices pour établir « au mieux » un plan d'échantillonnage spatial qui tienne compte des variables auxiliaires et de toute autre information disponible (connaissance que l'on a du polluant, de la portée du phénomène, des niveaux de concentrations, etc.) ?

Un rapport LCSQA de 2007 recommande des dimensions de maillage pour différentes typologies : rurale, urbaine, zones industrielles, etc. D'autres préconisations restent à examiner, notamment l'utilisation des informations auxiliaires et l'alternance des mesures aux points expérimentaux.

2. Étant donnée une zone dans laquelle un échantillonnage a été effectué durant l'année N. Comment réduire judicieusement cet échantillonnage les années suivantes pour pouvoir, avec un nombre réduit de points, produire des cartographies annuelles suffisamment précises?

En particulier, peut-on proposer un échantillonnage adéquat pour l'estimation de la moyenne annuelle de la concentration de l'année N+1 par cokrigeage avec dérive externe de l'hiver N, l'été N, l'hiver N+1 et l'été N+1 ?

Une revue bibliographique est d'abord présentée. Les critères sont ensuite précisés et des recommandations sont données, en insistant sur les principes. Nous n'avons pas cherché à « optimiser » le nombre et l'implantation des mesures. Les recommandations ont pour objet d'adapter la reconnaissance à l'objectif visé. Pour plusieurs raisons (pertes inévitables, précision), une certaine redondance est souhaitable, au moins pour le schéma initial. Enfin, un exemple pratique est présenté.

I. Synthèse bibliographique

Les articles suivants (ici classés thématiquement) ne constituent pas une bibliographie exhaustive, mais présentent une réflexion intéressante sur la définition ou la modification d'un schéma d'échantillonnage. Ont également été résumés les quelques articles examinés qui traitaient de l'estimation simultanée de la pollution de fond et de proximité.

- Brus D., Heuvelink G., 2007. Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138 pp. 86-95

- Emery X., Hernández J., Corvalán P., Montaner D., 2008. Developing a costeffective sampling design for forest inventory. VIII International Geostatistics Congress, GEOSTATS 2008. Ortiz & Emery, eds. Volume 2, pp. 1001-1010

- Kanaroglou P., Jerrett M., Morrison J., Beckerman B., Arain A., Gilbert N., Brook J., 2005. Establishing an air pollution monitoring network for intra-urban population exposure assessment: a location allocation approach. Atmospheric environment 39 pp. 2399-2409

- Fuentes M., Chaudhuri A. et Holland D.M., 2007. Bayesian entropy for spatial sampling design of environmental data. Environmental and Ecological Statistics 14 (3), pp. 323-340

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

- Wroblewski A., Riffault V., Perdrix E., Malherbe L., 2007. Adaptation des plans d'échantillonnage aux objectifs des campagnes. Echantillonnage spatial – guide de recommandations. Rapport LCSQA, Ecole des mines de Douai. INERIS – DRC – 07 – 84894 – 17725A

- Jeannée N., Fayet S., Mary L., Fromage-Mariette A, Cabero C., Perron G. et Armangaud A., 2006. Cartographie de la qualité de l'air en agglomération : comment intégrer pollution de fond et

pollution de proximité. 2ème conf. Environnement & Transports, incl. Le 15ème coll. Transports et pollution de l'air, Reims, France, 12-14 juin 2006 – Actes n°107, Vol. 1, Inrets ed., Arcueil, France, pp. 303-310

- Malherbe L., Cárdenas G., Colin P. et Yahyaoui A., 2008. Using different spatial scale measurements in a geostatistically based approach for mapping atmospheric nitrogen dioxide concentrations. Application to the French Centre region. Environmetrics 19 (7), pp. 751-764

a) Brus & Heuvelink. 2007

Le nombre de sites de mesures étant fixé, les auteurs utilisent la procédure itérative de recuit simulé pour trouver une implantation « optimale ». Le critère d'optimalité retenu est la moyenne spatiale de la variance de krigeage en dérive externe sur le domaine à estimer.

Au lieu de traiter séparément le calage de la régression linéaire et l'estimation des résidus, le krigeage universel, en fait le krigeage avec dérive externe (cf. équation. 1 de l'article) est préconisé. D'après le théorème d'additivité des variances de krigeage, cet estimateur minimise la somme de deux variances d'estimation, celle des résidus (à moyenne connue) et celle de la dérive (G. Matheron, La théorie des variables régionalisées et ses applications. Cahiers du Centre de Morphologie Mathématique, Fascicule 5. Ecole des mines de Paris, 1970).

La démarche suppose connues les variables auxiliaires, mais aussi le variogramme des résidus. Le schéma initial est obtenu par « surface de réponse ». Les auteurs comparent, pour le critère d'optimalité retenu (variance d'estimation prévue par le modèle) le krigeage avec dérive externe au krigeage sans dérive et à moyenne inconnue (krigeage ordinaire) et à la régression linéaire multiple. Les deux termes de la décomposition de la variance de krigeage sont aussi discutés.

L'exemple porte sur l'estimation du niveau piézométrique moyen d'une nappe, avec trois variables auxiliaires (altitude relative, profondeur et densité de drainage). Pour le critère d'optimalité choisi, le krigeage avec dérive externe est toujours le meilleur, le classement des deux autres dépendant de la taille n de l'échantillon :

- n petit, l'échantillonnage suivant un critère de régression linéaire est classé 2ème,
- n grand, l'échantillonnage suivant un critère de krigeage des résidus est classé 2ème.

Le gain de précision du krigeage avec dérive externe par rapport à l'estimateur classé 2ème reste modeste, l'écart au 3ème étant plus important pour un effectif réduit. La comparaison de plusieurs optimisations indépendantes montre la similarité des schémas par rapport aux extrema des variables auxiliaires.

L'influence du modèle de variogramme reste en suspens. Le temps de calcul n'est pas discuté. A noter aussi que la comparaison porte sur la variance d'estimation annoncée par le modèle, et non sur un écart quadratique expérimental.

b) Emery et al, 2008

L'objectif est l'ajout du nombre minimum d'échantillons à un schéma régulier initial, afin d'améliorer localement la précision de l'estimation. S'agissant d'une modification d'un schéma existant, le variogramme est connu. L'incertitude locale n'est pas quantifiée par la variance de krigeage, mais par la largeur de l'intervalle contenant 95% des valeurs, cet intervalle étant évalué par simulations conditionnelles. L'algorithme itératif par recuit simulé minimise une fonction objectif dépendant du nombre de points à rajouter et de la largeur de l'intervalle d'incertitude.

L'exemple présenté porte sur des plantations d'arbustes au nord du Chili. Importants, les temps de calcul peuvent être réduits par quelques astuces. La comparaison montre qu'un critère fondé sur l'écart-type de krigeage conduirait à ajouter beaucoup plus d'échantillons. Mais un éventuel effet proportionnel n'est pas pris en compte pour le krigeage.

c) Kanaroglou et al., 2005

Cet article a donné lieu à un commentaire par N. Kumar (2009) suivi d'une réponse des auteurs.

Les auteurs cherchent à localiser de façon optimale un réseau dense de stations de mesures de la qualité de l'air à l'aide d'informations auxiliaires comme l'occupation du sol et la densité de population, pour en déduire un modèle d'exposition d'une population à risque.

Les informations auxiliaires servent à construire une « surface de demande », fondée sur une technique de stratification, qu'ils pondèrent en fonction de la variabilité locale des auxiliaires et de la densité de population à risque. Ils appliquent ensuite un algorithme d'allocation optimale, dont la pertinence est mise en doute par N. Kumar.

Les auteurs présentent une succession d'algorithmes, mais sans bien préciser les concepts sous-jacents. A noter que le critère de variabilité locale est finalement assez analogue à celui d'une variance de dispersion en présence d'un effet proportionnel.

d) Fuentes et al., 2007

Les auteurs cherchent à définir des réseaux de surveillance de la pollution de l'air possédant de bonnes capacités prédictives, tout en minimisant les coûts. Pour cela, une méthode pour classer différents réseaux en fonction de leur coût et de l'information statistique qu'ils contiennent est présentée. A partir d'un réseau initial, un algorithme bayésien fondé sur un critère d'entropie détermine des sous réseaux optimaux.

Un deuxième aspect porte sur la non stationnarité des données environnementales avec la création d'une covariance non stationnaire à partir d'une famille de processus stationnaires ; la technique repose entre autres sur la méthode de Reversible Jump Markov Chain Monte Carlo (RJMCMC) pour estimer les paramètres.

L'objectif proposé est de diminuer le nombre de points d'un réseau existant en sélectionnant le réseau le plus intéressant parmi tous les sous réseaux possibles (cette méthode pourrait d'ailleurs être utilisée pour compléter un réseau existant en sélectionnant les points les plus informatifs à ajouter). Le critère pour sélectionner les sous réseaux les plus intéressants du point de vue statistique (à nombre d'échantillons et donc à coût fixé) est fondé sur l'entropie et donne la priorité aux sites associés à des fortes concentrations.

L'exemple présenté part d'un réseau de mesure de l'ozone sur 513 sites, aux États-Unis.

L'entropie est d'abord définie puis un cadre bayésien est utilisé pour sélectionner un sous réseau du réseau de mesures existant. Le choix se fait en calculant la densité a posteriori (« posterior predictive density ») et en retenant le schéma d'entropie maximale. De cette manière, les sites associés à de grandes incertitudes, plus difficiles à « prédire », sont préférentiellement retenus.

Pour un schéma donné, les auteurs définissent ensuite une fonction dépendant de l'entropie et d'un critère d'utilité donnant la priorité aux mesures ayant des concentrations proches du seuil de pollution. L'utilité est calculée pour chacun des sites du sous réseau proposé et ces utilités sont ensuite sommées pour avoir « l'utilité » du schéma entier. On retient alors le sous réseau maximisant la fonction regroupant entropie et utilité.

Le problème soulevé est celui des temps de calcul : il est en effet impossible d'étudier tous les sous réseaux possibles pour un réseau initial comportant 513 points. L'algorithme de recuit simulé évite de devoir considérer l'ensemble des sous réseaux.

La dernière partie porte sur la modélisation de la non stationnarité : le domaine est divisé en petites régions définies par des centres de gravité (grâce à la méthode de RJMCMC) qui déterminent des régions pour lesquelles le phénomène est stationnaire.

e) Gallois et al., 2005

Trois estimations de la moyenne annuelle, utilisant des variables auxiliaires, sont comparées : krigeage en dérive externe à partir des « mesures annuelles » calculées comme la moyenne des mesures saisonnières, moyenne de deux krigeages saisonniers avec dérive externe et enfin cokrigeage en dérive externe de la moyenne annuelle à partir des mesures saisonnières, qui fournit également la variance d'estimation de la moyenne annuelle. Le principe est l'amélioration du krigeage en utilisant les mesures disponibles pour une seule saison seulement. Pourvu que le modèle multivariable soit satisfaisant, les estimations à partir des mesures saisonnières sont préférables, car elles tiennent compte de l'ensemble des mesures et pas seulement des mesures communes aux deux saisons.

Deux exemples sont présentés :

- Montpellier :
  variables auxiliaires : densité de population dans des rayons de 200, 1000 ou 1500 m et émissions d'oxydes d'azote sur une grille kilométrique,
  143 sites de mesure, trois saisons,
  l'ACP est utilisée pour déterminer les auxiliaires les plus corrélées ;
- Mulhouse :
  variables auxiliaires : densité de population et émissions d'oxyde d'azote à maille kilométrique et occupation du sol avec une résolution de 200 m (utilisation du tissu urbain continu)
  79 sites de mesure, deux saisons.

L'accent est mis sur l'importance du contexte local pour déterminer les relations entre auxiliaires et concentrations et sur le danger lié à l'extrapolation du modèle. Dans les deux cas les corrélations entre les saisons sont élevées et il y a toujours au moins une variable auxiliaire bien corrélée aux concentrations.

La comparaison des trois méthodes par validation croisée permet de quantifier le gain en précision apporté par le cokrigeage et l'utilisation des variables auxiliaires. Les résultats montrent peu de différences selon les dérives retenues. Par contre, parmi les trois estimations, le krigeage à partir des mesures annuelles est beaucoup moins précis, car les mesures communes aux deux saisons ne sont pas très nombreuses (seulement 50 à Mulhouse). Le cokrigeage des moyennes annuelles par les mesures saisonnières présente donc un grand intérêt.

Pour optimiser les campagnes en exploitant la forte corrélation des mesures saisonnières (« été » et « hiver ») les auteurs suggèrent de conserver 40% de données communes, 30% des sites étant échantillonné seulement en hiver et les 30% restants seulement en été. Les données retirées le sont aléatoirement. Les cartes obtenues sont très similaires et la précision de l'estimation n'est que peu diminuée. Il faut garder suffisamment de points communs aux deux saisons pour vérifier la validité du modèle multivariable et distribuer les points aussi bien dans l'espace géographique que dans l'espace des variables auxiliaires.

f) Wroblewski et al., 2007 (rapport LCSQA)

A partir d'une étude bibliographique et des résultats de plusieurs campagnes pour le dioxyde d'azote NO_2, l'Ozone O_3 et le benzène C_6H_6, ce rapport donne des recommandations en terme de :

- nombre minimum de points expérimentaux pour le calcul d'un variogramme : une trentaine ;
- type de maillage, régulier ou stratifié, et d'espacement des points de mesures, en fonction du polluant considéré, du champ à cartographier, du milieu (typologie, taille de l'agglomération, etc.) et de la portée attendue.

Les dimensions de maille indiquées constituent une référence utile pour la détermination d'un maillage initial.

g) Jeannée et al., 2006

L'objectif est la cartographie intégrant à la fois pollution de fond et de proximité. Une première approche consiste à superposer, à la cartographie de la pollution de fond, les concentrations estimées le long des routes par un modèle de rue (STREET, résolution de 25 m). Cette approche pose deux problèmes : il n'y a pas de spatialisation du modèle de rue alors que l'effet de la pollution automobile peut s'étendre jusqu'à plusieurs centaines de mètres, et il n'y a pas de garantie de cohérence autre que qualitative entre les résultats du modèle de rue et les mesures de proximité. La décroissance des niveaux de pollution autour des axes est en réalité exponentielle, atteignant 100 m au maximum en milieu urbain fermé et jusqu'à 200 m en milieu ouvert.

Les auteurs proposent de réaliser la spatialisation des « sur-concentrations » liées au réseau routier par krigeage simple avec un variogramme exponentiel. La cartographie finale est obtenue en superposant la carte de pollution de fond et la carte en proximité routière. Cette méthode répond aux deux problèmes précédents.

Les données utilisées proviennent de la campagne AIRMARAIX sur l'agglomération de Toulon et concernent le dioxyde d'azote, avec 73 sites de fond et 30 sites de proximité automobile. Les variables auxiliaires utilisées sont les émissions en oxydes d'azote et les données d'occupation du sol.

La cartographie de la pollution de fond est tout d'abord réalisée, par krigeage en dérive externe. En parallèle, les concentrations modélisées par STREET sur le réseau routier sont discrétisées tous les 20m puis corrigées par krigeage en dérive externe en raison de leur corrélation moyenne avec les mesures réelles. Elles sont ensuite spatialisées par krigeage simple avec un modèle exponentiel et deux portées différentes selon la proportion de bâti. Les cartes de fond et de proximité sont enfin combinées pour obtenir la cartographie finale.

Les difficultés ou inconvénients mis en évidence sont les suivants :

- la quantification de l'incertitude associée à la cartographie est difficile car l'incertitude associée aux concentrations modélisées par STREET est inconnue ;
- seul le réseau routier principal est modélisé par STREET alors que le reste du réseau peut également participer à la pollution ;
- il est délicat de différencier les milieux fermés et ouverts pour définir la portée du modèle utilisé pour spatialiser les concentrations en proximité routière (il pourrait par ailleurs être intéressant d'introduire le type de rue « canyon » en milieu urbain très dense).

Souple, la méthode fournit des résultats réalistes et se transpose sans difficulté. Elle est aussi applicable en l'absence de modèle de rue, en exploitant des variables auxiliaires telles que les émissions d'oxydes d'azote sur le réseau ou simplement la position du réseau routier.

h) Malherbe et al., 2008

L'objectif est de représenter simultanément la pollution de fond et la pollution de proximité automobile en utilisant des variables auxiliaires, sachant que la résolution des cartes de pollution de fond est trop faible pour prendre en compte les gradients de concentrations près des sources. De ce fait, les cartes sous-estiment fortement les niveaux de pollution le long des réseaux routiers. La question principale est de savoir comment traiter en même temps deux composantes de la pollution, de portées différentes.

Les données étudiées (région Centre) comportent 49 sites de fond et 19 sites de proximité automobile. Le polluant considéré est le dioxyde d'azote ; les variables auxiliaires disponibles sont les émissions d'oxydes d'azote, la densité de population et l'occupation du sol.

La méthodologie proposée est décomposée en quatre étapes :

- étude préalable permettant de trouver les variables explicatives à utiliser pour la cartographie de la pollution de fond ;
- estimation des concentrations de fond par krigeage des résidus, krigeage en dérive externe ou cokrigeage colocalisé ;

- calcul des écarts entre les concentrations des sites de proximité automobile et les concentrations estimées en ces sites sur la cartographie de fond. Modélisation de ces écarts par une régression linéaire utilisant les émissions d'oxydes d'azote ;
- raffinement de la grille d'estimation le long des axes routiers et correction des concentrations de fond le long de ces axes grâce à la régression établie à l'étape précédente.

Les variables auxiliaires utilisées dans l'exemple sont le taux d'urbanisation et les émissions d'oxydes d'azote dans un rayon de 2 km, ainsi que la densité de population dans un rayon de 5 km.

La comparaison des méthodes pour la cartographie de la pollution de fond révèle que le krigeage des résidus et le krigeage en dérive externe donnent de meilleurs résultats et des cartes plus réalistes que le krigeage ordinaire, qui n'utilise pas d'information auxiliaire. Les critères de comparaison par validation croisée sont les statistiques des erreurs relatives et la corrélation entre valeur estimée et valeur mesurée.

Les exemples montrent que la méthode fournit une cartographie satisfaisante, intégrant à la fois pollution de fond et de proximité automobile. Il faut toutefois veiller à ne pas extrapoler le modèle de proximité routière, la distance à la route restant en pratique inférieure à quelques mètres.

Autres références bibliographiques

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

Castelier E., 1993. Dérive externe et régression linéaire. Compte rendu des journées de géostatistique – Cahiers de géostatistique – Fascicule 3 – Ecole des Mines de Paris – 1993 – pp. 448-59.

Kumar N. An optimal sampling design for intra-urban population exposure assessment. Atmospheric Environment 43 (2009) pp. 1153-1155.

Lavancier F., Caïni F., Gazeau A. 2003. Plan de sondage pour mesures mobiles de la pollution atmosphérique. Pollution atmosphérique, 180 pp. 551-565.

Wu H.W.Y. et Chan L.Y., 1997. Comparative study of air quality surveillance networks in Hong-Kong. Atmospheric Environment 31 (7), pp. 935-945.

II. Quelques principes pour la définition des campagnes

Quelques points-clés sont d'abord rappelés : objectifs de la campagne, contexte.

Dans la suite, un site désigne un « point » de mesure, comportant généralement plusieurs tubes ; le champ désigne le domaine à cartographier.

II.1 Choix de critères

Objectifs de la campagne

- *Estimation d'une moyenne temporelle* (annuelle, pluri-horaire, etc.) avec une précision fixée, ou *surveillance des zones* supposées *à forte concentration*, en particulier *en référence à une valeur réglementaire* à ne pas dépasser ?

- Substances visées : une ou plusieurs. Dans ce cas : lesquelles ? L'objectif est-il identique pour toutes les substances ?

- Pollution de fond ou simultanément de fond et de proximité ?

- Le champ d'étude est-il fixé (limites administratives) ?

- Nous examinons l'estimation d'une moyenne temporelle de la pollution de fond, pour une substance.

Critères de précision

- *Précision absolue* définie à partir de l'écart-type de krigeage σ_K (par exemple : variance de l'erreur d'estimation, intervalle de probabilité conventionnel à $\pm 2\sigma_K$) ou *précision relative*, rapportée à la concentration estimée ? L'écart-type d'estimation en dépendant, préciser le support de la grandeur à estimer.

- Un *maximum admissible* de la précision a-t-il été fixé ? Si oui, porte-t-il sur l'écart-type de krigeage σ_K (ou s'y ramène-t-il conventionnellement, dans le cas d'un intervalle) ou sur l'écart-type relatif ?

II.2 Campagne initiale : contexte

- Des *stations fixes* sont-elles disponibles dans le champ ?

Si oui, l'implantation a-t-elle été préférentielle (à proximité d'une source) ou non (réseau de surveillance de la pollution de fond, par exemple), et suivant quelle typologie (industrielle, proximité routière par exemple) ?

En présence de plusieurs stations fixes, la relation entre moyenne et variance temporelles par station (par exemple : les 52 moyennes hebdomadaires durant une année) peut montrer une régionalisation spatiale. Au lieu de la variance temporelle par station, qui inclut les composantes périodiques, on peut aussi examiner le palier des composantes temporelles non périodiques des variogrammes temporels par station.

Les stations fixes fournissent un ordre de grandeur local de la moyenne temporelle, utile lorsqu'un effet proportionnel est attendu. Enfin, ces stations donnent aussi une indication grossière (a priori plutôt préférentielle, donc à ne pas généraliser sans précaution) de la liaison entre concentration et information auxiliaire. Les stations fixes présentent l'avantage d'examiner les relations entre concentration et informations auxiliaires pour différents supports temporels, comme la période visée (par exemple, l'année) et la période de mesure (par exemple, une ou plusieurs quinzaines).

- D'après la bibliographie, s'attend-on à la présence d'un *effet proportionnel ?*

L'effet proportionnel est un conditionnement de la variance locale à la moyenne locale. Le variogramme s'écrit $\gamma(x, x+h) = \omega^2(x)\gamma_0(h)$, où le facteur $\omega(x)$ dépend de la moyenne locale de la concentration.

Notons Z la concentration, Z*(x) l'estimation par krigeage, $\sigma_k(x)$ l'écart-type de (l'erreur de) krigeage.

En présence d'un effet proportionnel, l'écart-type de krigeage prend la forme $\sigma_k^0(x).\omega(x)$, et l'écart-type relatif $\sigma_k^0(x).\omega(x)/Z^*(x)$. Si $\omega(x)$ est proportionnel à la moyenne locale, un critère de précision relative revient, à un facteur près, à travailler avec $\sigma_k^0(x)$. Dans le cas général où $\omega(x)$ est approché par $[Z^*(x)]^\beta$, l'écart-type de krigeage rapporté à l'estimation s'écrit $\frac{\sigma_k^0(x).\omega(x)}{Z^*(x)} = \sigma_k^0(x).\left[Z^*(x)\right]^{\beta-1}$. A même configuration de données, de fortes concentrations détériorent la précision relative si β > 1 (car $\beta - 1 > 0$) et l'améliorent si β < 1.

Remarque : Le resserrement des mesures dans les zones de fortes concentrations fréquemment observé peut viser

- un objectif de comparaison à une valeur réglementaire ;
- un critère de précision absolue en présence d'un effet proportionnel direct où, à configuration de données fixée, la variance de krigeage croît avec la concentration locale.

- Les variables auxiliaires disponibles, retenues d'après l'expérience ou la bibliographie, présentent-elles une *anisotropie* notable ? Si oui, l'anisotropie possible de la concentration pourrait-elle lui être liée ?

- Le budget alloué (ou l'expérience en d'autres lieux dans un contexte analogue) majore-t-il le nombre de sites de mesures, ou ce nombre résulte-t-il du canevas proposé ?

- La dimension du maillage et le support (ponctuel, bloc….) sont-ils fixés ?

II.3. Campagne initiale : canevas

a) Principe

- Un certain *suréchantillonnage initial*, sur un ensemble de sites implantés spatialement aux nœuds d'une grille régulière, est nécessaire pour
  - caler les relations entre concentrations et informations auxiliaires ;
  - calculer les variogrammes spatiaux dans de « bonnes conditions » ;
  - obtenir une première estimation si possible « suffisamment précise » pour pouvoir ensuite réduire l'information sans manquer d'éventuelles particularités (implicitement, il est donc fait référence à la notion de surveillance des valeurs élevées).

- Compléter le maillage spatial systématique par une couverture dans l'espace des variables auxiliaires sur le domaine à cartographier, de façon à éviter d'extrapoler la relation qui sera établie avec la concentration.

b) Étapes

i) Examen des informations disponibles : variables auxiliaires, stations fixes

- Choix des variables auxiliaires et représentation restreinte au champ : cartographies des auxiliaires, nuages de corrélation entre auxiliaires, histogrammes et quantiles.

Les variables auxiliaires sont par exemple les émissions, la densité de population, etc. ou leurs logarithmes translatés ; se référer à la bibliographie. Le nuage de corrélation et le coefficient r^2 associé, ou dans le cas de plus de trois variables, l'analyse en composantes principales, sont utiles pour préciser les différences ou les éventuelles redondances entre ces variables.

- Si des stations fixes sont disponibles, examen de la relation entre concentration et informations auxiliaires, et adaptation éventuelle de ces dernières (choix entre une variable ou sa transformée logarithmique, par exemple). Calcul des moyennes temporelles pour la période visée, et ajustement éventuel des dates et de la durée des mesures. Examen de la variabilité temporelle : apparaît-elle liée à la moyenne temporelle, est-elle régionalisée ?

- Calcul des variogrammes simples et croisés entre auxiliaires : présence de composantes spatiales analogues, amplitude relative du palier des structures de différentes échelles ? présence d'anisotropie, analogue ou non pour ces différentes variables ?

ii) Maillage spatial complété d'après les variables auxiliaires

Suivant les contraintes : budget, maille élémentaire pour la cartographie, … choix d'un maillage régulier, éventuellement anisotrope (calé pour la direction et le rapport, suivant l'anisotropie des informations auxiliaires). Les recommandations d'espacement indiquées par le « guide de recommandations » (2007) établi par le LCSQA pour l'échantillonnage spatial fournissent des références.

Le maillage carré ou rectangulaire (en présence d'anisotropie) n'est pas nécessairement le plus favorable. Un dispositif « *à maille centrée* » peut se montrer économique (cf. ***Figure 1***). Plus rarement utilisé, *le maillage hexagonal*, fondé sur une triangulation régulière est également intéressant ; sa mise en œuvre est-elle problématique ?

Visualisation des valeurs en ces sites pour les variables auxiliaires, et détection sur les histogrammes (par exemple, via les quantiles à 5% ou les déciles) ou les nuages de corrélation, des plages de valeurs du champ total qui ne sont pas échantillonnées. Implantation de sites supplémentaires par sélection d'après les valeurs des auxiliaires, par exemple en recherchant un resserrement du maillage initial.

Remarques :

Les plages de valeurs des variables auxiliaires non échantillonnées dans la première étape sont par exemple les « queues de distribution » associées aux valeurs fortes des émissions ou de la densité de population. Leur échantillonnage amène alors à resserrer les mesures au centre de l'agglomération.

En admettant que les structures de grande portée observées sur les variables auxiliaires persistent pour les concentrations, vérifier que l'espacement des sites ainsi obtenu leur est compatible.

Si le budget le permet, resserrements locaux du maillage pour préciser le comportement du variogramme aux petites distances. Localiser les resserrements dans une (ou des) zone(s) de valeurs fortes des auxiliaires (cf. les queues d'histogrammes), ainsi qu'une zone intermédiaire.

Un critère de précision relative pourrait amener à espacer les sites là où de fortes concentrations sont attendues. Il semble préférable de privilégier l'échantillonnage dans l'espace des auxiliaires.

iii) Vérification, à l'aide du krigeage en dérive externe

Choisir une palette de quelques variogrammes « conventionnels », suivant la bibliographie ou systématiquement (modèle linaire, de pente 1, schéma sphérique ou exponentiel de portée 1/2, 1/4 et 1/8 du champ). Si nécessaire, calcul des variables auxiliaires aux points expérimentaux situés en-dehors des nœuds de grille des auxiliaires (remarque : le support des informations auxiliaires est supposé identique pour les données et pour la grille d'estimation). Ceci est pertinent pour un krigeage ponctuel.

Pour ces modèles conventionnels, calculer la variance de krigeage en dérive externe, également conventionnelle, avec un voisinage assez grand ; il suffit pour cela d'effectuer un krigeage en affectant une valeur quelconque (0 par exemple) à la concentration aux futurs points expérimentaux. La variance de krigeage ainsi calculée montre l'influence des auxiliaires sur la précision. L'étude de sensibilité au variogramme (modèle, palier, portée, …) permet de repérer les éventuelles zones critiques, dans lesquelles il faut rajouter des sites. En effet, en dérive externe, la variance d'estimation dépend des valeurs des auxiliaires aux points expérimentaux et sur la grille d'estimation. Elle tend à augmenter en cas « de divergence » (à un facteur et à une constante près) entre valeurs aux points expérimentaux et valeurs sur la grille d'estimation. Réciproquement, il peut être éventuellement possible de réduire l'échantillonnage là où la variance de krigeage conventionnelle est faible, tout en respectant les contraintes (échantillonnage des queues de distribution des auxiliaires, resserrements locaux pour l'inférence du variogramme).

On peut éventuellement effectuer une étude de sensibilité au choix des variables auxiliaires.

Remarque importante :

A cette étape, la variance de krigeage « réelle » reste évidemment inconnue.

II.4 Ajustement à l'issue de la première campagne

Les mesures issues de la première campagne permettent :

- d'établir et de quantifier les relations entre concentrations et variables auxiliaires, en sélectionnant les variables les plus informatives. Écarter une variable auxiliaire jugée pas ou peu utile peut avoir des conséquences sur le plan d'échantillonnage, si celle-ci était prise en compte dans l'étape de balayage des informations auxiliaires.
- de vérifier la présence d'un éventuel effet proportionnel, et dans l'affirmative, de le quantifier.
- d'inférer le variogramme des concentrations pour l'intervalle de temps considéré, directement (pour les résidus d'une régression) ou indirectement (dans un modèle avec dérive externe).

A l'aide du variogramme ainsi ajusté, le calcul de la variance de krigeage permet de vérifier si la précision recherchée est ou non atteinte.

- Si la précision n'est pas atteinte localement, tester si le resserrement local du maillage par ajout de quelques stations peut suffire ;
- Si le problème concerne une grande partie du champ, un resserrement du maillage initial (en ajoutant par exemple un échantillon central à une maille régulière carrée ou rectangulaire) risque d'augmenter de façon importante le nombre de mesures. Pour une campagne ultérieure, il peut être plus économique de définir un nouveau maillage, tout en conservant le maximum de points communs.

Réciproquement, si la précision est largement atteinte, réduire l'information (en passant d'une maille carrée de côté d à un dispositif carré centré de côté 2d, ou d'un maillage avec motif centré à un maillage sans motif centré) tout en préservant le balayage des informations auxiliaires. Un krigeage à partir de l'information réduite permet alors :

- de vérifier si les cartes de concentrations estimées ne sont pas modifiées de façon importante ;
- de vérifier la qualité de l'estimation aux points expérimentaux susceptibles d'être supprimés. En particulier, on vérifiera le non biais et si la moyenne de l'écart standardisé (l'écart entre mesure et estimation, rapporté à l'écart-type de krigeage) est proche de l'unité.

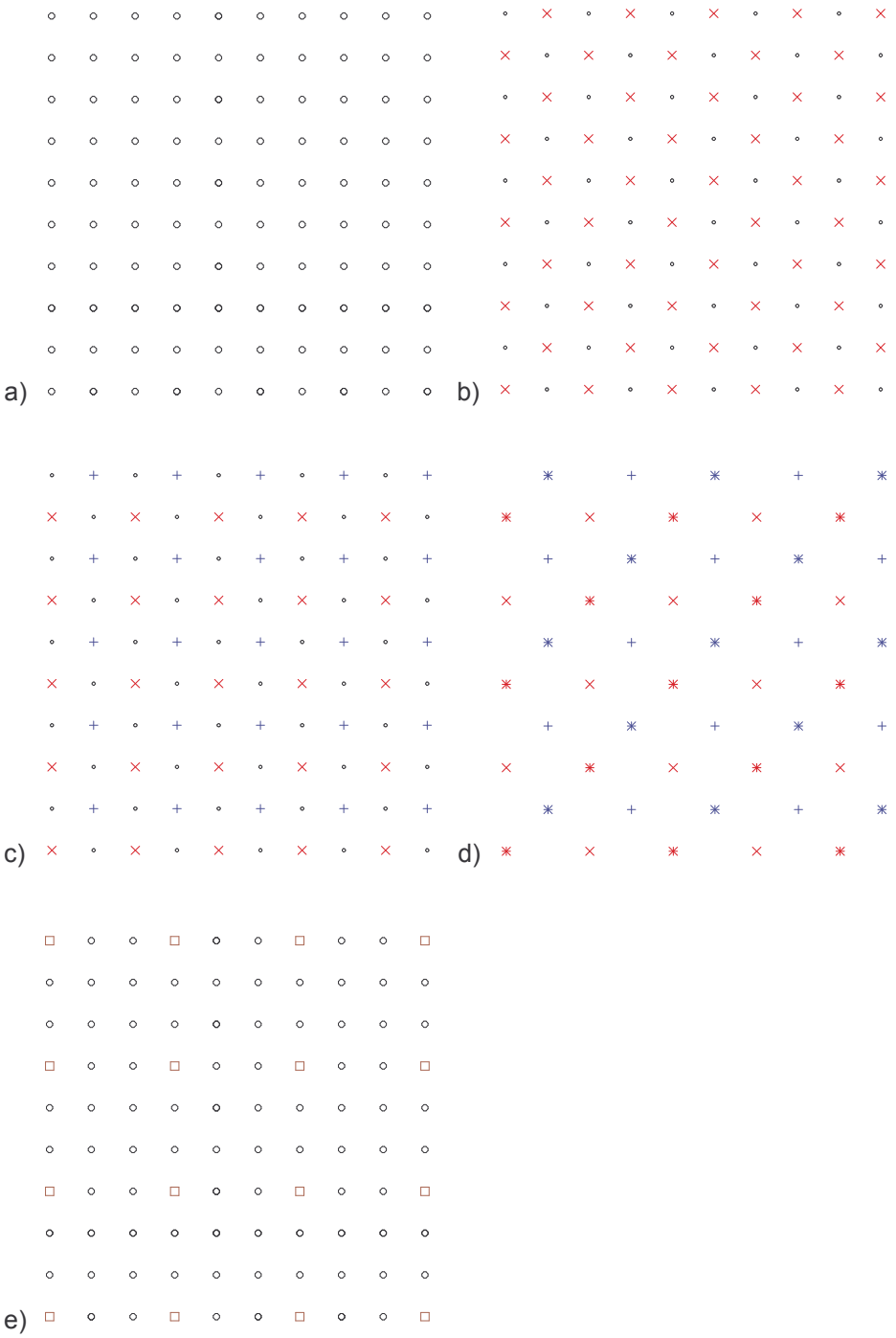


*Figure 1 : Maille initiale carrée. a) maillage initial ; b) maillage x ou o : réduction de 1/2, avec apparition d'un motif centré ; c) maillage x ou + : réduction de ¼ du maillage initial ; d) maillage * ou * : réduction de 1/8 ; e) maillage □ : réduction de 1/9.*

II.5. Campagnes de surveillance ultérieures

Les campagnes de mesures sont supposées « saisonnières » pour l'estimation et la cartographie de la moyenne annuelle de la concentration.

a) Décomposition de la variance d'estimation de la moyenne annuelle

La moyenne annuelle Z étant estimée par $Z=\ell Y+\ell' Y'$, Y et Y' désignant les moyennes saisonnières, la variance d'estimation s'écrit :

$$\begin{aligned} Var\left[Z-Z^*\right] &= \ell^2\, Var\left[Y-Y^*\right]+\left(\ell'\right)^2 Var\left[Y'-Y'^*\right]+2\,\ell\,\ell'\,Cov\left(\left[Y-Y^*\right],\left[Y'-Y'^*\right]\right) \\ &= \ell^2\,\sigma_Y^2+\left(\ell'\right)^2\sigma_{Y'}^2+2\,\ell\,\ell'\,r\,\sigma_Y\,\sigma_{Y'} \end{aligned} \qquad \text{(Eq. 1)}$$

r désignant le coefficient de corrélation des *erreurs* d'estimation saisonnières au point courant ; r est en fait de la forme r(x). σ désigne la variance d'estimation, par cokrigeage, d'une moyenne saisonnière par les mesures des deux saisons. Généralement on pose ℓ=ℓ'=1/2, mais la pondération peut être ajustée par krigeage temporel.

En cas de fort contraste des paliers des variogrammes saisonniers, il convient de tenir compte du rapport entre σ_Y et $\sigma_{Y'}$. Deux cas se présentent.

i) Pour réduire au mieux cette variance en vue d'atteindre une précision fixée, il convient de réduire les termes les plus importants, après vérification du signe et de l'ordre de grandeur de r. Par exemple, en modèle de corrélation intrinsèque et en l'absence de dérive (et pour ℓ=ℓ'=½), on augmentera le nombre de mesures pour la saison admettant le palier le plus élevé.

ii) Au contraire, si l'échantillonnage disponible permet d'atteindre la précision voulue pour la moyenne annuelle, pour réduire le nombre de mesures sans trop détériorer la variance d'estimation, il convient plutôt d'augmenter la variance d'estimation saisonnière la plus faible, c'est-à-dire celle dont le palier du variogramme est le plus bas. On réduira le nombre de mesures pour cette saison.

Si les paliers saisonniers sont peu contrastés, les deux saisons jouent un rôle à peu près équivalent. Pour réduire le nombre de mesures sans trop détériorer la précision, on pourra alterner les mesures saisonnières, avec des points en hiver et d'autres en été (cf. point b) ci-après). Ce sera d'autant plus efficace que la corrélation entre les concentrations saisonnières est plus élevée.

b) Sites de mesures annuelles ou saisonnières

Un ensemble de sites annuels est défini pour caler les relations avec les auxiliaires et comparer les mesures saisonnières. Les sites restants sont mesurés durant l'une des saisons seulement, en alternance ou non. Les différents scénarios de mesures sont schématisés de la façon suivante :

hiver + été		avec en %	x	
hiver	été		$y=p(100-x)$	$z=(1-p)(100-x)$

avec x+y+z=100. x détermine le pourcentage de sites « annuels », mesurés en hiver et en été ; p détermine la proportion des autres stations mesurées pour l'une des saisons, par exemple en hiver. Si N_{points} désigne le nombre de points expérimentaux, le nombre total de mesures est $N= N_{points}$ (2x+y+z)% ; la réduction d'échantillonnage obtenue en mesurant certains points durant l'une des saisons seulement est $\Delta N= N_{points}$ (100-x)%.

Posant y=p.(100-x), alors z=(1-p)*(100-x) ; p est la proportion des mesures « saisonnières » hivernales, et 1-p la proportion des mesures « saisonnières » estivales.

Exemple : En agglomération, les concentrations de NO_2 sont généralement supérieures en hiver, et de variabilité également supérieure en hiver. Le variogramme des concentrations hivernales est supérieur et de portées différentes de celui des concentrations estivales.

La solution proposée par Gallois et al. (2005) consiste à conserver 40% de sites annuels, et 30% respectivement en hiver ou en été seulement.

Dans cet exemple: x=40, y=z=30. La réduction d'échantillonnage est de $\frac{\Delta N}{2N_{points}} = \frac{100-X}{2}$ %=30%.

A même échantillonnage, la variance de cokrigeage est supérieure pour la moyenne hivernale par rapport à l'estivale (mais c'est l'inverse en relatif). Avec un critère d'écart-type d'estimation de la moyenne annuelle, il est en fait intéressant de réduire plus fortement l'écart-type d'estimation de la concentration hivernale. De plus, pour une surveillance simultanée des valeurs hivernales plus fortes, les points mesurés l'été seulement sont-ils réellement utiles ? Sachant que les valeurs estivales sont inférieures, et de variogramme également inférieur, au lieu de réduire de façon analogue le nombre de points expérimentaux en été et en hiver en les alternant, *on peut aussi bien tolérer une plus forte dégradation de la précision de l'estimation estivale*. Un deuxième scenario consisterait par exemple à conserver 100% des sites en hiver, et à n'en mesurer que 40% l'été. La réduction de 30% de l'information totale reste inchangée, mais avec une meilleure surveillance des fortes concentrations hivernales. Le non biais reste assuré par les conditions de non biais du cokrigeage à « moyenneS inconnueS », du moins tant que le nombre de stations estivales reste suffisant.

Ce deuxième scénario correspond à x=60, y=40, z=0, avec le même ΔN que précédemment.

Il paraît difficile de fixer des règles précises, à cause de la diversité des situations possibles : critères, niveaux saisonniers et amplitude des variations saisonnières notamment. Les contraintes de terrain peuvent aussi conduire à privilégier la pérennité de certains sites selon des critères propres, par exemple une moindre exposition aux éventuelles dégradations. Seules quelques recommandations sont donc données.

c) Hypothèses

La précision (absolue ou relative) pour la période visée (l'année) est désormais supposée partout atteinte par la première campagne (saisonnière), à un éventuel ajustement du schéma de reconnaissance près.

Deux cas sont examinés :

- l'ajustement pour une saison différente suivante (§ d, par exemple, campagne estivale à partir d'une première campagne hivernale). Le variogramme spatial de la saison suivante reste inconnu (de même que le variogramme croisé entre les saisons), ainsi que la liaison entre concentration saisonnière et auxiliaires ; ils peuvent différer sensiblement de ceux de la saison précédente ;
- après deux campagnes saisonnières de l'année N, le dimensionnement des deux campagnes saisonnières de l'année N' ; les variogrammes spatiaux simples et croisés entre saisons, ainsi que les liaisons entre concentrations saisonnières et auxiliaires sont alors supposés connus, aux fluctuations temporelles près.

L'estimation est effectuée par cokrigeage des mesures « saisonnières », supposées assez fortement corrélées entre elles. L'ordre de grandeur du coefficient de corrélation des concentrations saisonnières est accessible, de façon indicative, à l'aide des stations fixes dont l'implantation peut être préférentielle.

La moyenne temporelle des différentes périodes (années, saisons) peut présenter des variations sensibles. L'ordre de grandeur des variations saisonnières (été, hiver) peut être là encore fourni par les mesures aux stations fixes disponibles durant plusieurs années. Ce rapport peut varier avec la typologie (par exemple, les variations saisonnières du NO_2 sont généralement atténuées sous influence routière). Il est donc important de recaler les relations entre variables auxiliaires et concentrations en tenant compte des différentes typologies ; l'ensemble des typologies doit donc être mesuré à chaque campagne. Il est aussi préférable d'estimer la concentration « saisonnière » en un point d'une typologie à l'aide de mesures provenant de sites de même typologie, car les concentrations et les variables auxiliaires peuvent être supposées « cohérentes » en ces points expérimentaux et au point courant.

Sous réserve de validation expérimentale, nous proposons de privilégier un espacement des points de la grille régulière (compensé par le cokrigeage), tout en conservant des points pour le

balayage des variables auxiliaires. Par exemple, fixer des points « obligatoires » du maillage initial, sur la maille régulière ou rajoutés en complément :

- pour le calage ou le contrôle de la relation entre concentrations et variables auxiliaires (cf. le balayage du nuage pour deux auxiliaires ou l'échantillonnage stratifié de l'histogramme dans le cas d'une seule) ;
- pour le calcul des variogrammes expérimentaux simples et croisés.

Une quarantaine de sites « obligatoires » semble raisonnable ; anticipant l'inévitable dégradation de l'information, quarante-cinq à cinquante sites sont alors à prévoir. Ce chiffre est à ajuster suivant le niveau des concentrations, leur variabilité spatiale (amplitude des fluctuations, cf. le palier éventuel du variogramme ; portées, cf. les préconisations données par LCSQA dans le rapport 2007), le degré de liaison entre concentrations et auxiliaires.

A l'issue des tests nécessaires, la démarche suivante pourra être modifiée ou précisée.

d) Configuration de la campagne saisonnière suivante

L'ordre de grandeur des variations saisonnières, ainsi que du coefficient de corrélation des moyennes saisonnières est accessible à l'aide des stations fixes. Le variogramme de la moyenne saisonnière visée et la relation entre concentrations et auxiliaires restent inconnus.

Commencer par ajuster le plan initial, cf. § II.4.

Se référer à la bibliographie et aux stations fixes : les concentrations et la variabilité spatiale (absolue, ou relative, selon le critère retenu) sont-elles plutôt croissantes ou décroissantes entre les deux saisons ? Faudrait-il plutôt densifier ou réduire l'information ?

- Si une dégradation de la précision est attendue (exemple : campagne estivale puis hivernale pour le NO_2, avec un critère de valeur maximale de l'écart-type de krigeage), poser des hypothèses simplificatrices « pessimistes » de façon à majorer la variance de cokrigeage de la moyenne annuelle, et vérifier si le critère reste atteint.

Exemple de calcul approché : dans l'équation (1), la variance de *cokrigeage* de chacune des moyennes saisonnières (exemple : hiver estimé par été et hiver) est majorée par la variance de *krigeage* respective (exemple : hiver estimé par hiver seulement), dont l'une est connue. Appliquer un facteur A « pessimiste » pour majorer l'autre (cf. bibliographie, ou en s'aidant des moyennes saisonnières ou des variances aux stations fixes). On a alors

$$Var\left[Z - Z^*\right] < \left(\ell^2 + A^2\left(\ell'\right)^2 + 2\,\ell\,\ell'\,A\right).\sigma_Y^2$$

et pour ℓ=ℓ'=1/2

$$Var\left[Z - Z^*\right] < \tfrac{1}{4}\left(1 + A\right)^2.\sigma_Y^2$$

On vérifie alors que le terme de droite de l'inégalité reste inférieur à la limite fixée.

- Si une amélioration est attendue, modifier le schéma d'échantillonnage (comme ci-dessus) seulement si le critère visé est (a priori) largement atteint (sur l'ensemble du champ, ou plus localement).

e) Configuration des campagnes saisonnières de l'année N' à partir des campagnes saisonnières de l'année N

Une première étape de contrôle et d'ajustement du plan d'échantillonnage est nécessaire si le milieu a été modifié de façon importante entre les deux années, par exemple en cas de changement majeur des infrastructures ou du plan de circulation. Dans ce cas, vérifier (à l'aide des auxiliaires disponibles concernées, avant et après changement du milieu) la présence de points expérimentaux dans la zone modifiée et à proximité ; si nécessaire, rajouter des points en vue d'une meilleure description du milieu, et selon les critères retenus.

Les mesures hivernales et estivales de l'année N, fournissent :

- les relations saisonnières entre auxiliaires et concentrations ;
- les variogrammes saisonniers simples et croisés pour l'année N.

Utiliser quelques stations fixes en contexte analogue, pour vérifier sur les années antérieures la corrélation établie entre les mesures saisonnières par tubes pour l'année N. Si les ordres de grandeur sont comparables, extrapoler à l'aide des stations fixes les corrélations saisonnières entre les années N et N'. Si les résultats divergent, il est préférable de rechercher l'origine de cette divergence. On peut aussi se caler aux corrélations les moins bonnes. Examiner aussi l'amplitude des fluctuations des concentrations saisonnières sur plusieurs années.

La démarche suivante fournit un échantillonnage non optimal en vue d'une surveillance pluri-annuelle. Des approximations en simplifient la mise en œuvre.

i) A partir des mesures saisonnières de l'année N, par exemple (hiver, été)$_N$, définir un échantillonnage saisonnier quasi « optimal » pour le cokrigeage par les seules mesures de l'année N+1 (par exemple). En effet, au cours du temps, la corrélation avec l'année N tend à décroître, et ce schéma de référence (hiver, été)$_{REF}$ pourra s'avérer utile.

Cet échantillonnage intègre les points jugés obligatoires, d'après les auxiliaires. Il peut être construit d'après l'équation (1), en alternant des mesures si les paliers saisonniers sont équilibrés (y et z ≠ 0), ou en utilisant les contrastes de palier pour réduire les mesures sans trop dégrader la variance de cokrigeage annuelle, tout en maintenant la surveillance locale des zones de niveau le plus élevé (y ou z peut alors être nul).

ii) Dégrader le schéma (hiver, été)$_{REF}$ pour tenir compte de la corrélation temporelle avec les mesures saisonnières (hiver, été)$_N$ effectuées durant l'année N. Cette étape pourra se transposer à la définition du schéma pour une année ultérieure (hiver, été)$_{N+k}$ à partir des mesures des années précédentes, avec un ajout éventuel de sites du schéma de référence (hiver, été)$_{REF}$.

En théorie, il faudrait effectuer un cokrigeage à quatre variables ((hiver, été)$_N$, (hiver, été)$_{N+1}$), recommandé pour l'estimation effective de la concentration moyenne annuelle de l'année N+1. Pour simplifier le calcul, un majorant approché de chacun des termes de l'équation (1) est recherché :

- poser r=1 pour le terme croisé ;
- pour chaque saison, rechercher (bibliographie, stations fixes) la saison a priori la mieux corrélée (exemple : pour l'hiver N+1, l'hiver N ou l'été N ?).

A l'aide du modèle variographique été-hiver pour l'année N, ou d'un modèle simplifié (corrélation intrinsèque) pour les deux hivers (ou les deux étés), calculer pour chaque saison la variance de cokrigeage saisonnier bivariable, supposée majorer la variance de cokrigeage à quatre variables. Eventuellement, comparer au cokrigeage à partir d'une saison mieux informée (exemple : pour l'été N+1, comparer le cokrigeage par l'été N supposé le mieux corrélé, au cokrigeage par l'hiver N comportant des mesures plus nombreuses).

On obtient alors un majorant pour chacun des termes $\sigma_{\text{Hiver},N+1}$, $\sigma_{\text{Ete},N+1}$, et par suite de la variance de la moyenne annuelle :
$Var\left[Z - Z^*\right] \le \ell^2 \sigma^2_{Hiver,N+1} + \left(\ell'\right)^2 \sigma^2_{Ete,N+1} + 2\,\ell\,\ell'\,\sigma_{Hiver,N+1}\,\sigma_{Ete,N+1}$. On réduit alors les mesures des saisons N+1, d'après la variance de cokrigeage saisonnier, en équilibrant la réduction de $\sigma_{\text{Hiver},N+1}$ et $\sigma_{\text{Ete},N+1}$.

Remarque : La covariance des erreurs de cokrigeage est calculable à l'aide du modèle de variogrammes simples et croisés. En pratique, ce calcul n'est peut être pas très commode. Une validation croisée ad hoc (construite à cet effet) fournit une évaluation empirique de cette corrélation. Si la corrélation r est faible, en tenir compte pour un majorant plus fin de la variance d'estimation de la moyenne de l'année N+1.

Les tests devront vérifier si les majorations proposées sont valides et ne sont pas trop grossières.

III. Exemple pour une campagne initiale

La construction d'un schéma d'échantillonnage pour une première campagne de mesure est généralement fondée sur des connaissances préalables de la zone à cartographier. On commence ici par construire une maille régulière, afin de faciliter l'étude variographique. Elle est ensuite complétée par des points découlant de l'étude des variables auxiliaires disponibles. Pour évaluer cette méthodologie, le résultat est comparé à une campagne d'échantillonnage déjà disponible.

L'étude est réalisée sur des données de l'agglomération de Bordeaux (AIRAQ). On dispose de mesures de dioxyde d'azote et de benzène pour l'été 2004 et l'hiver 2005 et de mesures de benzène pour l'été 2001 et l'hiver 2002 (cf. Annexes). Dans cette partie, seules les données de benzène de l'été 2004 sont utilisées. Disposant de deux quinzaines de mesures (22 juillet au 4 août 2004 et 4 août au 18 août 2004), les valeurs attribuées à la saison sont des moyennes des valeurs des quinzaines.

Seules les données de fond et de proximité industrielle sont ici prises en compte. Les données de proximité automobile, beaucoup plus élevées et devant être traitées spécifiquement, feront l'objet d'une prochaine étude. Une variable auxiliaire expliquant cette proximité automobile, telle que les émissions de NO_X ou le réseau routier, est nécessaire.

La démarche repose sur différentes étapes constituées d'allers-retours entre l'espace géographique et l'espace défini par les variables auxiliaires.

1ère étape : Délimitation de la zone à cartographier

La délimitation de la zone à cartographier est l'étape préliminaire à toute construction d'un schéma d'échantillonnage. En l'absence de connaissances a priori sur l'agglomération de Bordeaux, on considère que la zone à cartographier est la zone reconnue lors de la campagne de l'été 2004 (Figure 2).

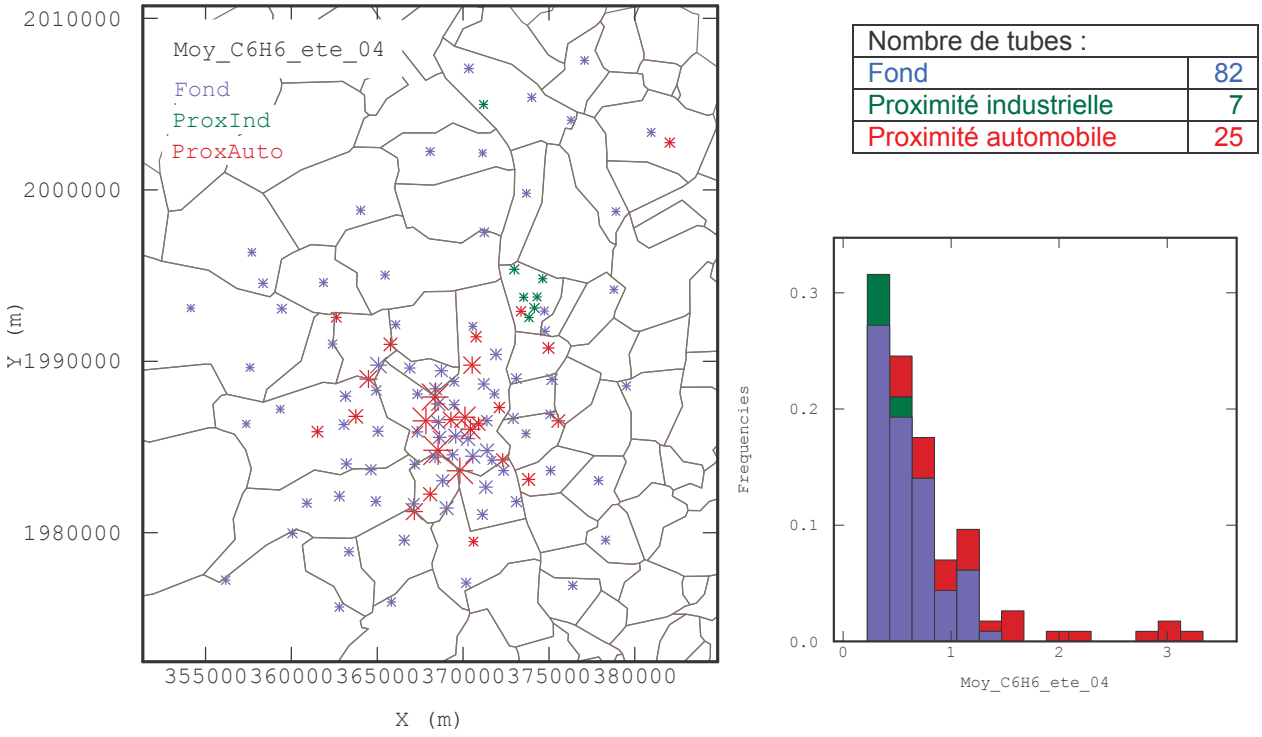


Figure 2 : Carte d'implantation de la campagne d'échantillonnage réalisée pour le benzène durant l'été 2004. En bleu : tubes mesurant la pollution de fond ; en vert : proximité industrielle et en rouge : proximité automobile. Le tableau présente le nombre de points de chaque catégorie. L'histogramme des concentrations reprend le même code couleur.

On considère la zone suivante : X compris entre 354 000 et 385 000 et Y compris entre 1 975 000 et 2 008 000. De plus, on souhaite réaliser une cartographie sur une maille d'estimation de 250 m.

2ème étape : Création d'un premier échantillonnage régulier

Ne disposant d'aucune information, le plus simple consiste à proposer un échantillonnage régulier, qui sera ensuite complété et modifié.

Dans un premier temps il faut donc choisir le nombre de tubes à implanter et, compte tenu de la zone à cartographier, déterminer la dimension de la maille régulière initiale. Pour l'agglomération de Bordeaux, le nombre d'échantillons total de la campagne réelle est de 114 dont 89 données de fond et de proximité industrielle ; la première grille d'échantillonnage comportera ici 90 échantillons. La zone à cartographier mesurant 31 par 33 km, on implante un échantillon tous les 4 km (Figure 3).

Remarques :

- si la portée du phénomène est déjà connue, on peut suivre la recommandation de Wroblewski A. et al. (2007) et créer un échantillonnage avec une maille égale au tiers de cette portée,
- si l'on sait que le phénomène est anisotrope (par l'étude des variables auxiliaires par exemple), il est intéressant de créer une maille non plus carrée mais rectangulaire, qui tienne compte des directions principales et du rapport d'anisotropie.

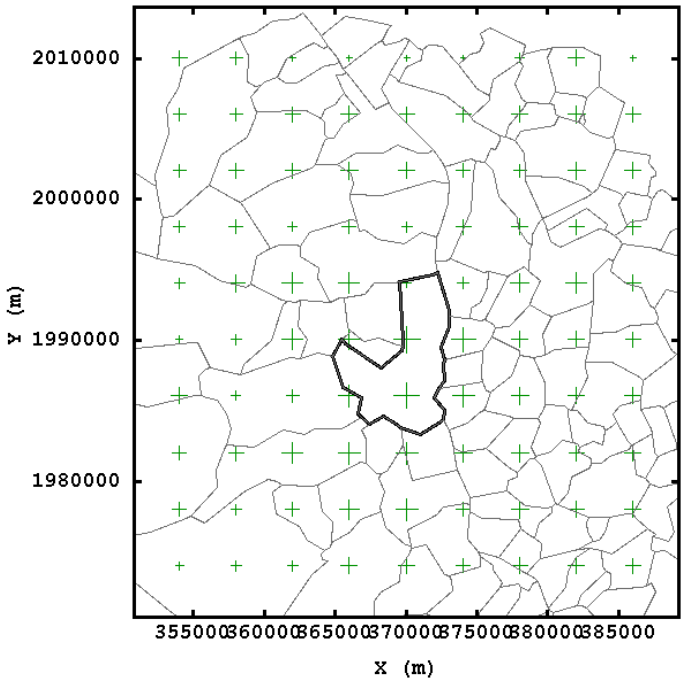


Figure 3 : Première proposition d'échantillonnage : maille régulière, à intervalles de 4km (commune de Bordeaux en gras, la taille des croix est proportionnelle à la densité de population dans un rayon de 100 m).

3ème étape : Étude des variables auxiliaires

On dispose de plusieurs variables calculées à chaque point expérimental de la campagne réelle et à chaque point de la maille d'estimation à 250 m. Les variables utilisées sont les suivantes :

- densité de population, calculée dans un rayon de 100, 250, 500 ou 1000 m,
- occupation du sol d'après la base Corine Land Cover de 1990, disponible selon trois niveaux de détail. On associe à chaque point la proportion de chaque catégorie d'occupation du sol dans un rayon de 100, 250, 500 ou 1000 m.

On commence par choisir les variables ou combinaisons de variables utilisées pour établir le schéma d'échantillonnage. On sélectionne a priori la densité de population et la proportion de tissu urbain

continu, ce qui revient à travailler dans un espace des variables auxiliaires à deux dimensions. Disposer des résultats d'une campagne réelle permet de choisir de manière optimale un rayon de 100 m pour la densité de population et un rayon d'un kilomètre pour la proportion de tissu urbain continu ; assez classiquement, on utilise plutôt le logarithme translaté des variables (ln(1+z)). La Figure 4 présente le nuage de corrélation et les histogrammes des logarithmes translatés de la densité de population et de la proportion de tissu urbain continu pour les 28 336 points de la grille d'estimation (mailles de 250 m) ; ce nuage va servir à améliorer le schéma d'échantillonnage.

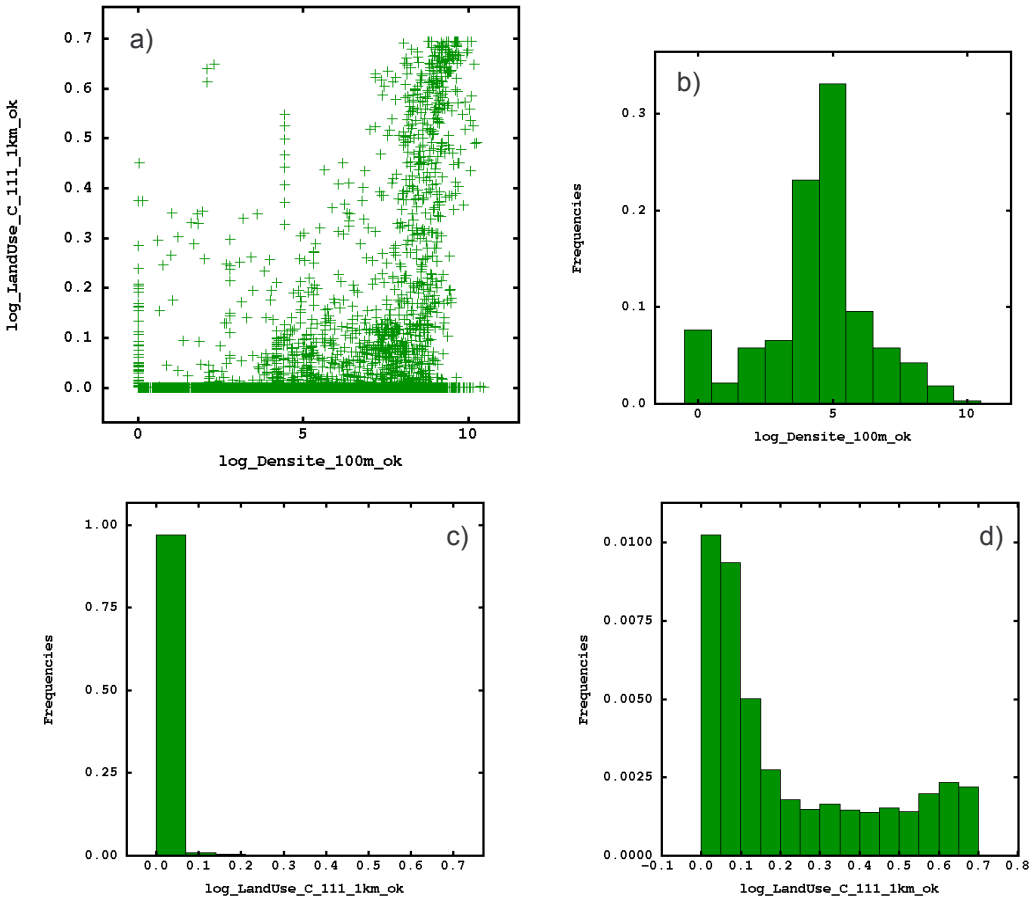


Figure 4 : a) Nuage de corrélation entre les logarithmes translatés de la densité de population dans un rayon de 100 m et de la proportion de tissu urbain continu dans un rayon de1 km ; b) Histogramme du logarithme translaté de la densité de population ; c) Histogramme du logarithme translaté de la proportion de tissu urbain continu ; d) Histogramme du logarithme translaté de la proportion de tissu urbain continu sans les valeurs nulles.

Supposant ces variables corrélées aux concentrations en benzène, on cherche un schéma d'échantillonnage qui parcourt l'espace géographique mais également le nuage des auxiliaires. En effet, un balayage suffisant du nuage des variables auxiliaires, corrélées aux concentrations en benzène, devrait fournir des échantillons représentant la variabilité des concentrations ainsi que les concentrations extrêmales. De cette manière, les dangers des estimations réalisées en extrapolation du modèle bivariable sont évités.

Remarques sur la Figure 4 :

La corrélation entre les deux variables étant faible, elles apportent une information différente et se complètent pour participer à l'estimation de la concentration en benzène. Il n'y a pas de redondance d'information.

La proportion de tissu urbain continu présente de nombreuses valeurs nulles (partout où le tissu urbain continu est distant de plus de 1 km du point considéré). Ceci aura des conséquences lors du krigeage en dérive externe.

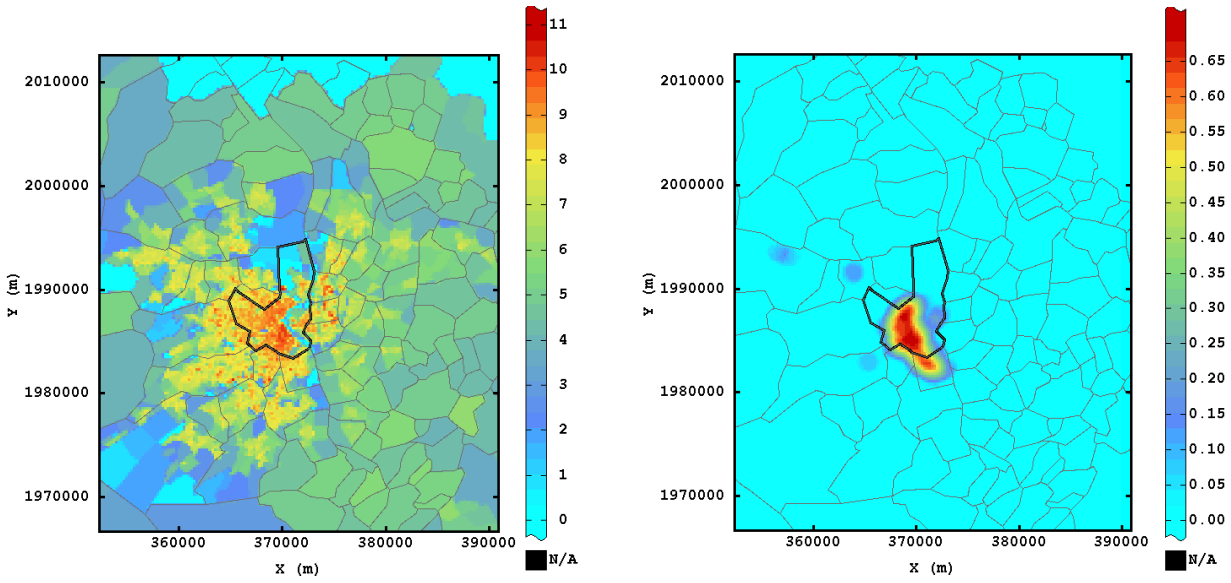


Figure 5 : Carte des logarithmes translatés de la densité de population (à gauche) et de la proportion de tissu urbain continu (à droite).

Le nuage de corrélation entre les deux variables auxiliaires comporte manifestement plusieurs populations. Comme les corrélations entre densité de population, proportion de tissu urbain continu et concentration en benzène sont positives, les points « en haut à droite » du nuage des variables auxiliaires devraient correspondre aux concentrations en benzène les plus élevées. Cette partie du nuage de corrélation correspond effectivement au centre de l'agglomération de Bordeaux (Figure 6).

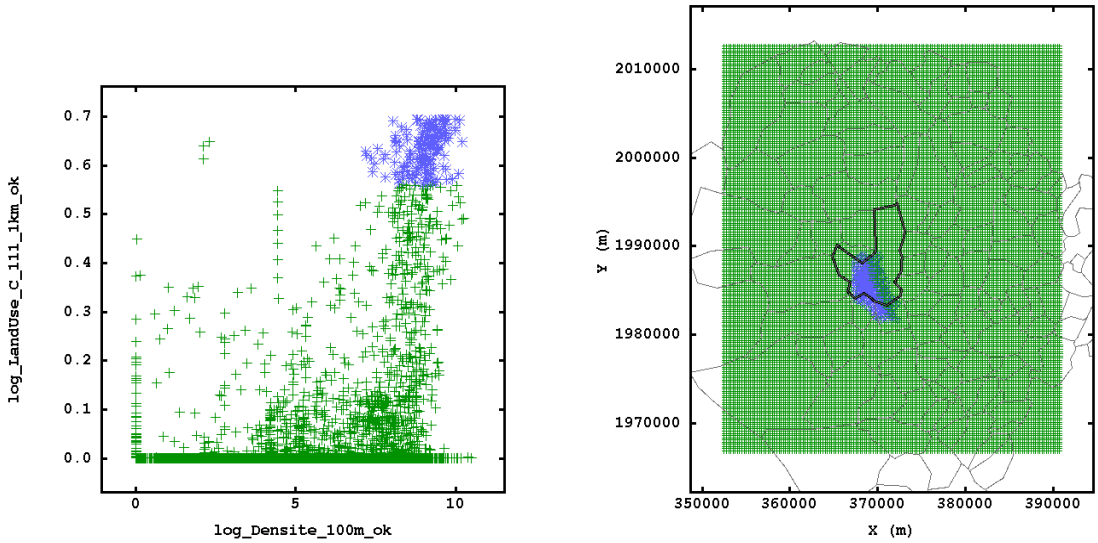


Figure 6 : Détection du centre de l'agglomération à l'aide du nuage des variables auxiliaires.

4ème étape : Projection sur le nuage des variables auxiliaires et deuxième schéma

Afin d'améliorer le schéma d'échantillonnage, on le projette sur le nuage des auxiliaires pour l'ensemble de la zone à cartographier (Figure 7). Le nuage des auxiliaires n'est pas entièrement parcouru, de nombreux points sont situés en périphérie de l'agglomération (à l'extérieur de la zone cartographiée lors de la campagne réelle) ou correspondent à des valeurs nulles de la proportion de tissu urbain continu. A l'opposé, la partie en haut à droite du nuage est peu échantillonnée.

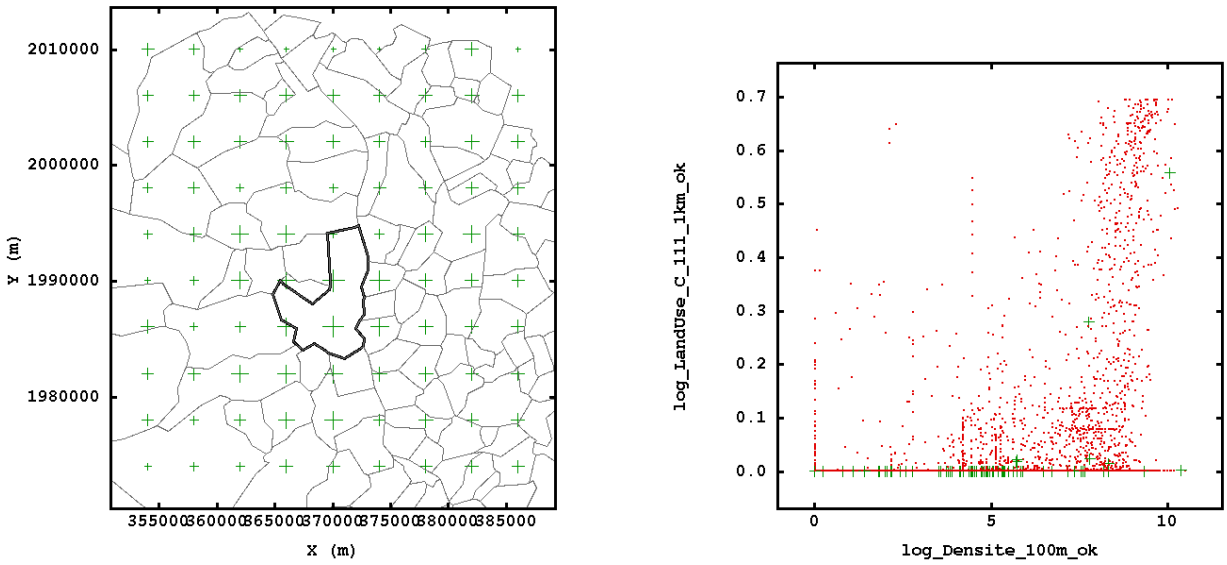


Figure 7 : Schéma d'échantillonnage à maille régulière d'intervalles de 4 km (à gauche) et projection des 90 points sur le nuage des auxiliaires à droite (points rouges : points non échantillonnées – croix vertes : 90 points du schéma d'échantillonnage).

L'étape suivante consiste à supprimer des points de la périphérie pour se rapprocher de la zone à cartographier et à densifier l'échantillonnage dans le centre de l'agglomération. Pour cela 25 points sont retirés en périphérie et l'intervalle de la maille d'échantillonnage est divisé par deux dans le centre de l'agglomération, ce qui conduit à ajouter 23 points. L'échantillonnage ainsi obtenu est présenté Figure 8.

La suppression de points du premier schéma a conduit à la suppression de points dans la partie du nuage où les croix vertes (points retenus) étaient les plus nombreuses. L'ajout des points du centre ville permet de mieux parcourir le nuage des auxiliaires. Cette modification jugée bénéfique est retenue pour la suite. Cependant, le nuage des auxiliaires n'est pas encore entièrement parcouru, la dernière étape consiste à ajouter quelques points supplémentaires afin d'y remédier.

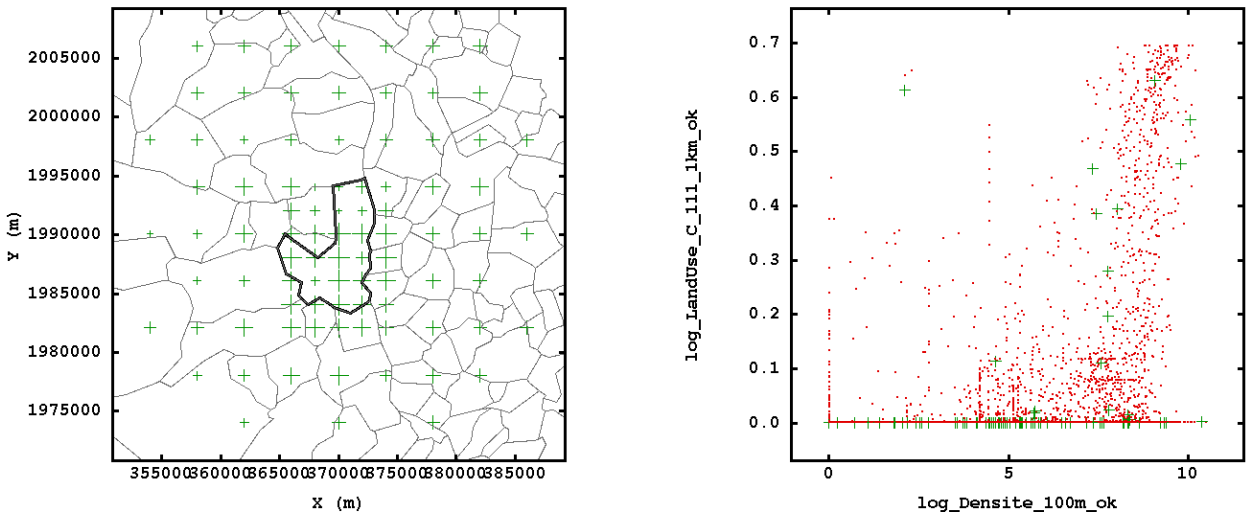


Figure 8 : Deuxième schéma d'échantillonnage, comprenant moins de points en périphérie et une maille régulière resserrée à 2 km dans le centre de l'agglomération (à gauche) ; projection des points de ce schéma d'échantillonnage sur le nuage des auxiliaires (à droite).

5ème étape : Parcours du nuage des auxiliaires, troisième schéma

Dans cette étape on sélectionne des points du nuage des auxiliaires à ajouter à l'échantillonnage de manière à le parcourir entièrement. Ici, ces points sont ajoutés, portant le total à 108 tubes, ce qui correspond à plus de tubes que la campagne réelle ; il aurait été possible de supprimer encore des points en périphérie de la zone à cartographier afin de se ramener à un nombre proche du nombre d'échantillons de la campagne réelle car la zone est ici un peu plus étendue.

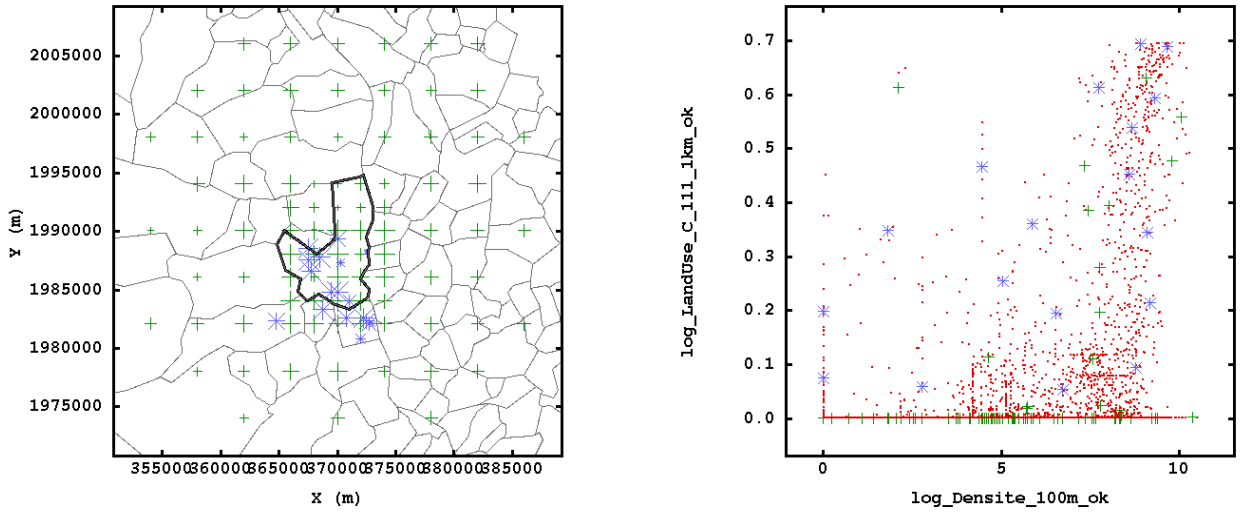


Figure 9 : Troisième schéma d'échantillonnage. En rouge : points de la grille à estimer non échantillonnés, en vert : échantillons disposés sur deux mailles régulières imbriquées, en bleu : ajout des échantillons permettant le parcours du nuage des auxiliaires en entier.

Le schéma d'échantillonnage ainsi obtenu permet de remplir les critères fixés : une base régulière simplifiant l'étude variographique et des échantillons supplémentaires disposés de manière à avoir des points dans toutes les parties du nuage des auxiliaires.

La dernière étape d'ajout des points à partir du nuage des auxiliaires est ici réalisée empiriquement. Il est possible d'automatiser cette étape, par exemple par un échantillonnage stratifié du nuage des auxiliaires : il suffit pour cela de quadriller ce nuage (quelle qu'en soit la dimension) et de tirer au hasard dans chaque case un ou plusieurs points. Le nombre de points à tirer dépend du rapport du nombre de points présents dans cette case au nombre total de points du nuage et du nombre de points appartenant déjà au schéma d'échantillonnage. Ainsi les parties denses du nuage des auxiliaires sont préférentiellement échantillonnées et tout le nuage est parcouru, en tenant compte des points déjà présents dans les mailles régulières. Il est nécessaire alors de bien vérifier la présence de points dans les cases en bordure du nuage. Cette méthode permet de bien échantillonner les différents modes que les histogrammes des variables auxiliaires pourraient présenter.

Améliorations possibles :

- Il est recommandé d'implanter un tube passif à proximité immédiate d'une ou, mieux, de chaque station fixe, afin de caler les tubes par rapport aux stations. Cela détermine alors un ou plusieurs emplacements pour des tubes ; projeter les stations fixes sur le nuage des auxiliaires aide à sélectionner éventuellement celles où il est utile d'installer un tube passif.
- Si le nombre d'échantillons possible est suffisamment important, il est intéressant de localiser une ou deux croix de sondage (avec des échantillons espacés de 125 ou 250 m par exemple) afin de préciser le comportement du variogramme à l'origine. La disposition des croix de sondage peut se faire à partir du nuage des auxiliaires afin de les placer dans des parties intéressantes et distinctes.

L'échantillonnage obtenu n'est pas « optimal ». Cependant même pour une méthode optimale, il faudrait tenir compte de la perte d'information pouvant survenir (dégradations, avaries). Tous les points étant importants il faut pouvoir les compenser par d'autres sites.

Enfin, les points d'échantillonnage sont supposés implantés sur une grille théorique. Il faut ensuite placer les tubes au plus près de ces coordonnées en tenant compte des réalités du terrain (bâtiments, rivières,…), ce qui modifie légèrement l'emplacement des points et donc les valeurs des variables auxiliaires associées. Il convient de vérifier que l'échantillonnage du nuage des auxiliaires ne s'en trouve pas trop modifié.

6ème étape : Validation du schéma d'échantillonnage proposé

Grâce à la campagne de mesure réalisée sur l'agglomération de Bordeaux il est maintenant possible, mais aussi nécessaire, de valider le schéma d'échantillonnage proposé.

La validation est réalisée en comparant le schéma proposé au schéma réel (Figure 10). Ne disposant pas des valeurs de concentration en benzène pour le nouveau schéma, le krigeage est effectué en affectant des valeurs nulles à tous les points, puisque la variance de krigeage ne dépend pas des valeurs aux points expérimentaux. Le critère de comparaison est l'écart-type d'estimation dans le modèle, obtenu après krigeage en dérive externe.

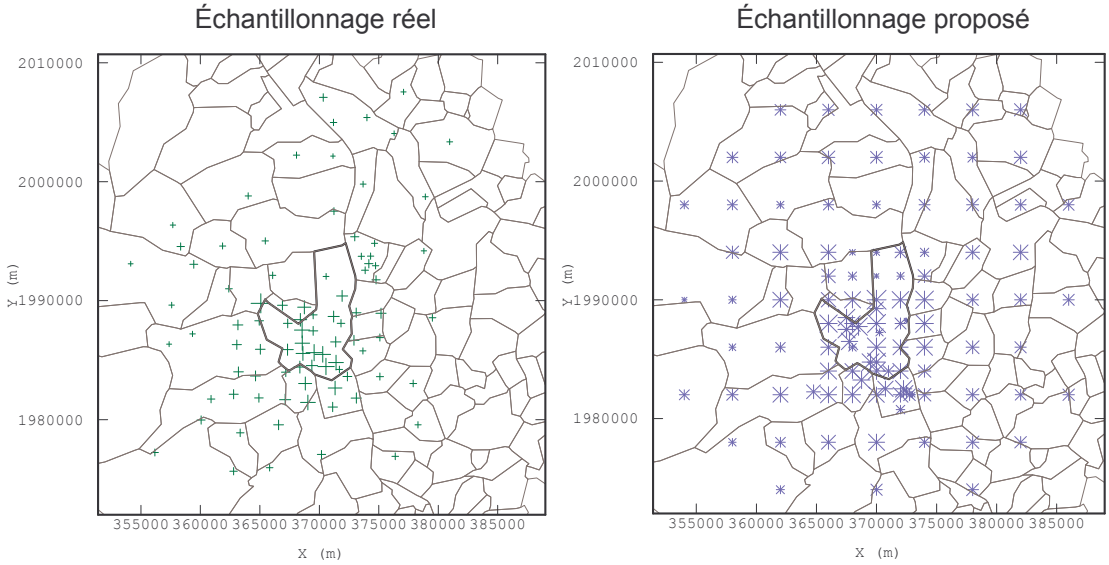


Figure 10 : Cartes des schémas d'échantillonnage réel (à gauche) et proposé (à droite).

La première étape du krigeage en dérive externe est réalisée à l'aide du module Non-stationary Modeling du logiciel Isatis®. Il faut en effet connaître le variogramme des résidus dans le voisinage de chaque point à estimer. Les seuls résidus accessibles étant ceux calculés pour la régression globale, différents modèles de variogrammes pour différentes dérives (plusieurs variables auxiliaires ou une combinaison des variables auxiliaires) vont être testés dans le module Non-stationary modeling.

Les dérives testées ici sont la densité de population avec la proportion de tissu urbain continu, une combinaison linéaire de ces deux variables et une combinaison linéaire de la densité de population et des variables d'occupation du sol suivantes (code Corine Land Cover entre parenthèses) :

- tissu urbain continu (111)
- tissu urbain discontinu + espaces verts artificialisés, non agricoles (112+14)
- zones industrielles ou commerciales et réseaux de communication + mines, décharges et chantiers (12+13)
- territoires agricoles + forêts et milieux semi-naturels (2+3)
- zones humides +surfaces en eau (4+5)

Les régressions linéaires multiples sont établies à partir de toutes les données de la campagne réelle disponibles (de fond et de proximité industrielle). La première conduit à un coefficient de corrélation de 0,77 avec les concentrations en benzène, la seconde à un coefficient de corrélation de 0,82. Pour cette seconde régression, le découpage et les regroupements des classes d'occupation

du sol ont été déterminés a priori, sans connaissance particulière de l'agglomération ; ils fournissent un résultat presque aussi bon qu'une régression linéaire multiple avec toutes les variables d'occupation du sol du niveau le plus détaillé. Les nuages sont présentés Figure 11.

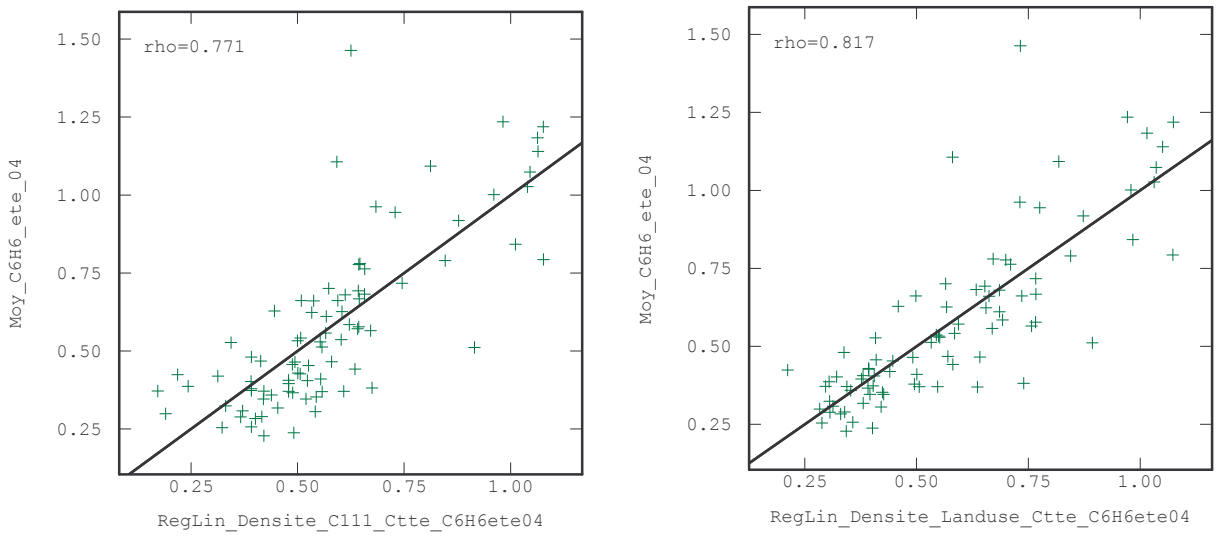


Figure 11 : Nuages de corrélation entre les concentrations en benzène et les régressions multiples réalisées (avec seulement la proportion de tissu urbain continu à gauche, avec 5 classes de proportions d'occupation du sol à droite).

Les modèles de variogrammes élémentaires testés sont un effet de pépite, un modèle linéaire et trois modèles sphériques de portées 8, 12 et 15 km (le champ ayant une dimension d'environ 30 km).

A l'issue de cette étape, trois krigeages en dérive externe sont réalisés. Les modèles de variogrammes des résidus sont les suivants (Tableau 1).

Variables en dérive	Modèle de variogramme des résidus
Densité de population Proportion de tissu urbain continu	Effet de pépite : 0,03142 Modèle sphérique de portée 15 km : 0,03894
Régression 1	Effet de pépite : 0,03756
Régression 2	Effet de pépite : 0,03695

Régression 1 – Variables explicatives : densité de population et proportion de tissu urbain continu.
Régression 2 – Variables explicatives : densité de population ; proportion de tissu urbain continu ; tissu urbain discontinu + espaces verts artificialisés, non agricoles ; zones industrielles ou commerciales et réseaux de communication + mines, décharges et chantiers ; territoires agricoles + forêts et milieux semi-naturels ; zones humides + surfaces en eau.

Tableau 1 : Modèles de variogrammes des résidus retenus pour les krigeages en dérive externe.

Ces modèles sont ensuite vérifiés par validation croisée. La Figure 12 présente la validation croisée réalisée pour la deuxième régression linéaire utilisée en dérive externe.

Remarque :

Le variogramme des résidus utilisé pour le krigeage en dérive externe a un palier beaucoup plus faible et un rapport effet de pépite sur structure sphérique plus élevé que le variogramme des résidus calculé sur le jeu de données entier (Effet de pépite de 1,3 et modèle sphérique de portée 15 km et de palier 3,3).

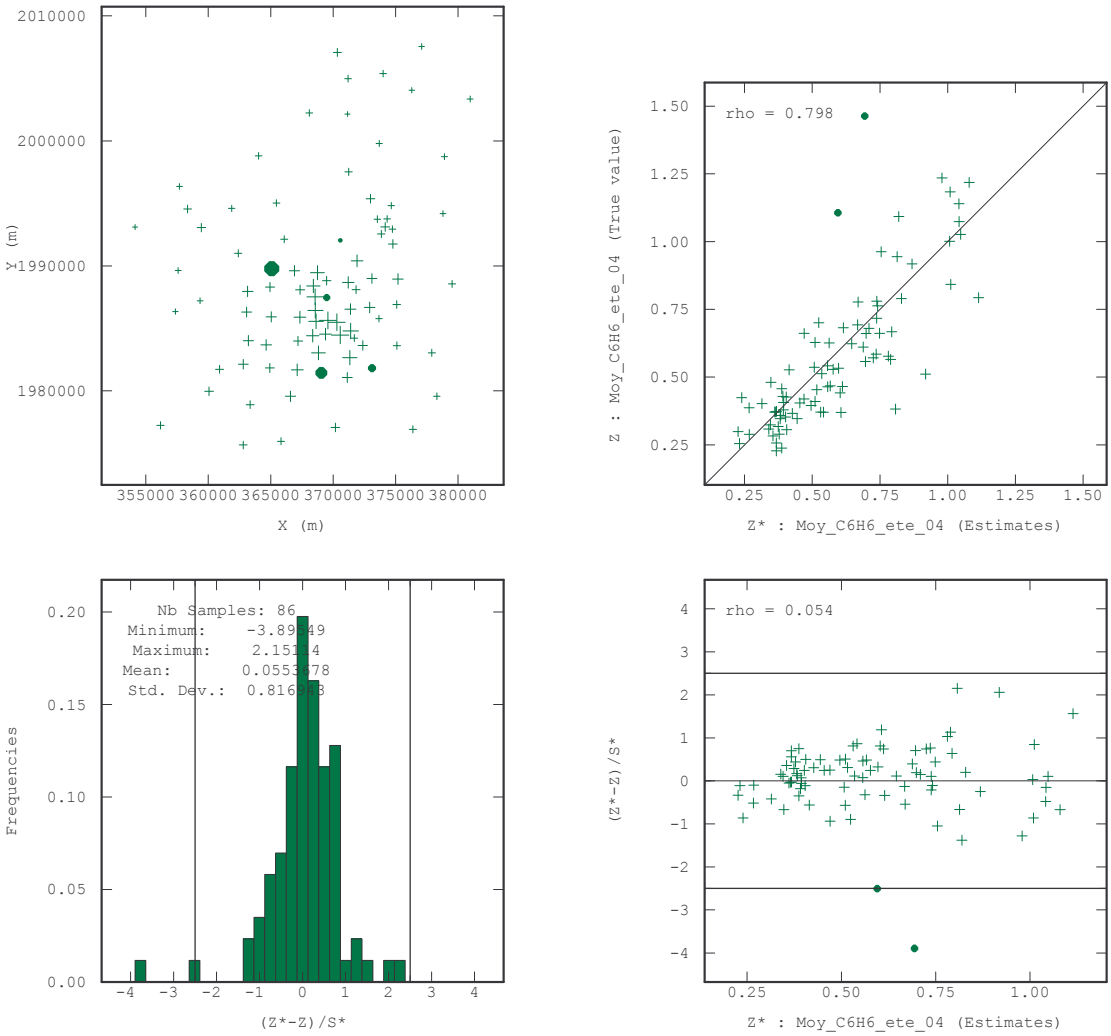


Figure 12 : Validation croisée du modèle de variogramme des résidus retenu pour le krigeage en dérive externe en utilisant la seconde régression linéaire.

En réalité, seules les deux régressions linéaires permettent sans problème la comparaison des deux schémas d'échantillonnage. La forte proportion de valeurs nulles de la variable proportion de tissu urbain continu dans l'agglomération empêche la réalisation du krigeage sur toute la zone à cartographier en utilisant cette variable en tant que dérive. Considérons un point de la maille d'estimation : on souhaite utiliser deux dérives, l'une fonction de la densité, l'autre fonction de la proportion de tissu urbain continu. L'utilisation de la densité de population ne pose aucun problème. Par contre, lorsque dans le voisinage du point à estimer tous les points de données ont des proportions de tissu urbain continu nulles, le krigeage en utilisant cette dérive ne peut pas se faire (problème d'inversion de matrice). L'estimation en utilisant les deux dérives ne peut se faire que sur une partie de la zone à cartographier (Figure 13).

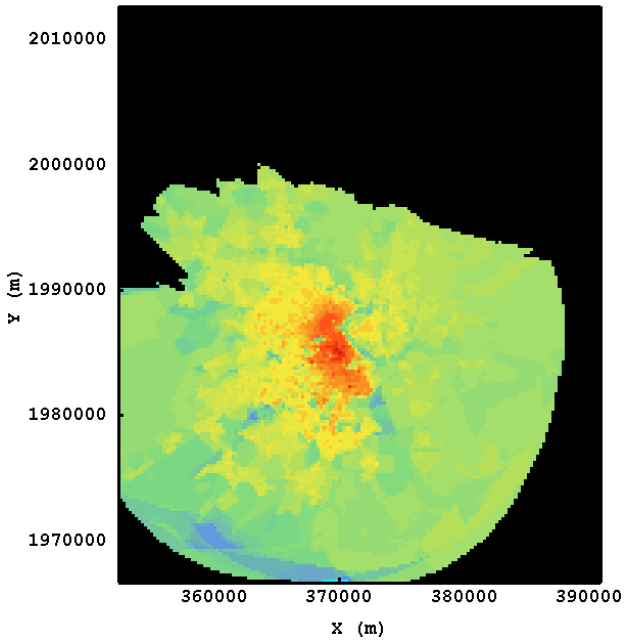


Figure 13 : Carte d'estimation des concentrations en benzène en utilisant deux dérives sur toute la zone à cartographier (la partie noire ne peut être estimée).

Un moyen de remédier à ce problème est de considérer deux zones :

- dans la première, les deux dérives sont utilisées ; au voisinage d'un point à estimer, au moins une valeur (ou en pratique, quelques valeurs) de proportion de tissu urbain continu est non nulle ;
- la seconde pour laquelle, considérant un point à estimer, tous les points contenus dans le voisinage ont des proportions de tissu urbain continu nulles. Dans cette zone la proportion de tissu urbain continu n'apporte aucune information, on n'utilise alors que la densité de population comme dérive.

Il reste ensuite à recaler ces deux zones l'une part rapport à l'autre. Cette étape n'est pas aisée à réaliser et peut engendrer des problèmes de continuité aux abords de la frontière entre les deux zones. Dans le cas d'une variable auxiliaire présentant de nombreuses valeurs nulles, il est plus simple de créer une régression linéaire multiple, même si le schéma d'échantillonnage initial a été établi avec deux variables distinctes.

La comparaison des deux schémas est effectuée en considérant les deux régressions linéaires multiples. Les statistiques des écart-types d'estimation sont données dans le Tableau 2. A titre d'information, et même si l'estimation ne peut être réalisée que pour un nombre très restreint de points de la grille, les statistiques des écart-types sont également données pour l'utilisation des deux dérives. Le voisinage de krigeage choisi présente un rayon de 15 km avec un minimum exigé de 5 points pour réaliser le krigeage en dérive externe.

Dérives utilisées	Nombre[1]		Minimum		Maximum		Moyenne		Variance	
	SR	SP	SR	SP	SR	SP	SR	SP	SR	SP
Densité – Proportion de tissu urbain continu	15083	17631	0,05	0,05	0,79	1,47	0,12	0,09	0,01	0,00
Régression 1	25835	27395	0,04	0,05	0,95	0,56	0,11	0,08	0,02	0,00
Régression 2	25835	27395	0,04	0,04	0,57	0,68	0,09	0,07	0,00	0,00

[1] : nombre de points de la grille où l'estimation est réalisée
SR : schéma réel – *SP* : schéma proposé
Régression 1 – Variables explicatives : densité de population et proportion de tissu urbain continu.
Régression 2 – Variables explicatives : densité de population ; proportion de tissu urbain continu ; tissu urbain discontinu + espaces verts artificialisés, non agricoles ; zones industrielles ou commerciales et réseaux de communication + mines, décharges et chantiers ; territoires agricoles + forêts et milieux semi-naturels ; zones humides +surfaces en eau.

Tableau 2 : Statistiques élémentaires des écart-types d'estimation pour les deux schémas d'échantillonnage et les trois krigeages en dérive externe réalisés.

Les statistiques montrent que les deux schémas donnent des résultats relativement semblables (dans le modèle). On note tout de même une diminution de la moyenne et de la variance des écart-types avec le nouveau schéma d'échantillonnage proposé. Cela pourrait venir du fait que l'on ait plus de points, surtout en périphérie mais montre que le nouveau schéma est au moins aussi bon que le schéma d'échantillonnage réalisé à Bordeaux. Pour compléter la comparaison, la Figure 14 montre les cartes d'écart-type d'estimation associées aux deux schémas d'échantillonnage, avec pour dérive l'une ou l'autre des deux régressions linéaires.

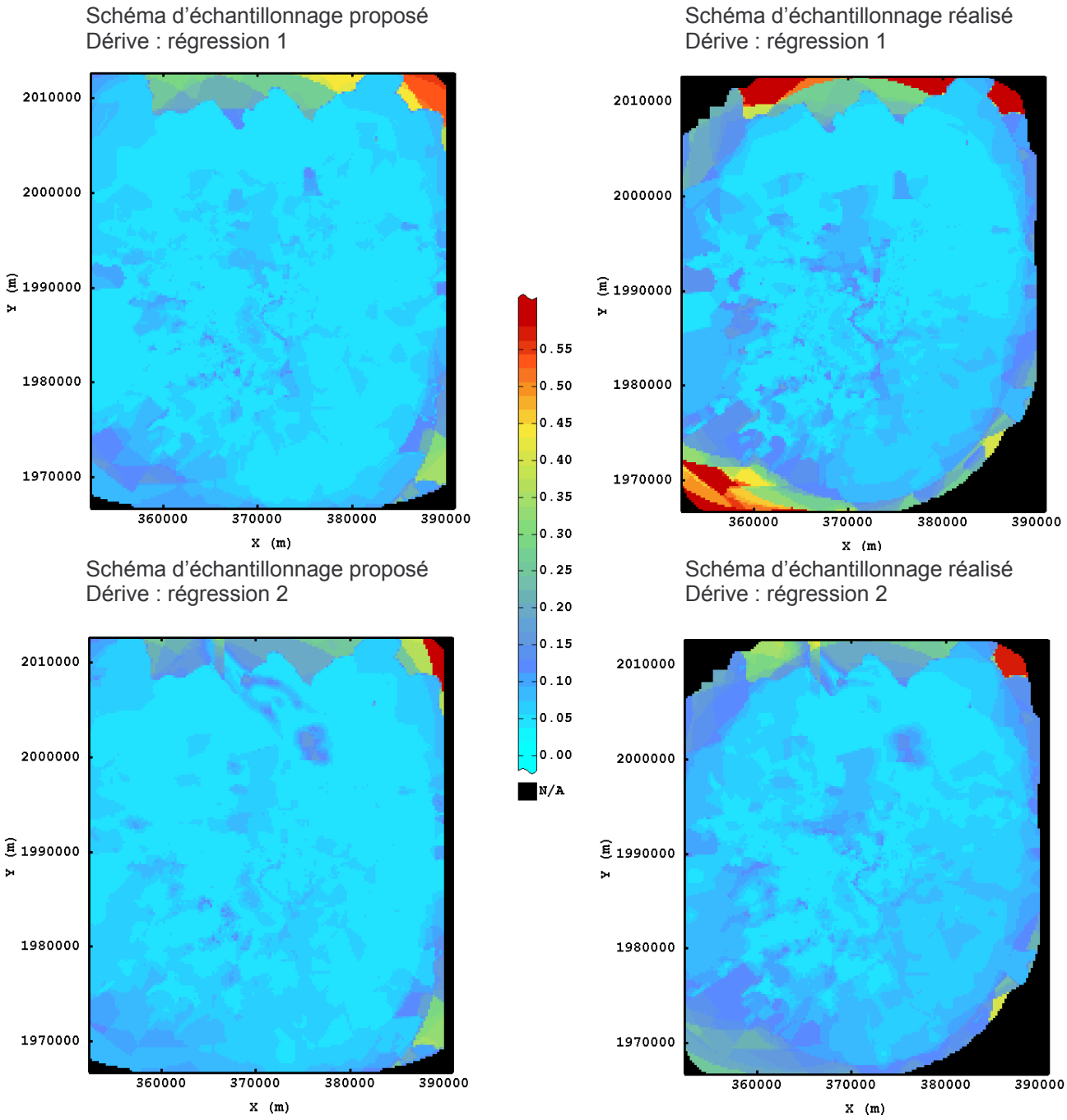


Figure 14 : Cartes d'écart-type d'estimation des deux schémas d'échantillonnage.

Ces cartes d'écart-type d'estimation montrent un léger avantage en faveur du schéma d'échantillonnage proposé. En effet les valeurs sont légèrement plus faibles, mais surtout elles sont plus homogènes sur la zone à cartographier. Cela montre l'intérêt de l'échantillonnage régulier non seulement dans l'espace géographique mais aussi dans l'espace des variables auxiliaires. On

constate aussi que l'écart-type est un peu plus faible avec l'utilisation de la deuxième régression linéaire fondée sur toutes les classes d'occupation du sol.

L'utilisation des régressions linéaires paraît donc intéressante. Elle nécessite cependant :

- de connaître les variables qui vont être incluses dans le modèle. A ce titre, il serait intéressant de vérifier que ce découpage en classes d'occupation du sol est pertinent sur d'autres villes ou agglomérations,
- de connaître les coefficients affectés à chacune des variables. La régression est effectuée une fois que l'on connaît les valeurs de concentrations or il pourrait être utile de la connaître avant (notamment si l'on veut définir un nuage des variables auxiliaires comprenant une régression). La régression est ici créée à partir des données en benzène de l'été 2004 ; étant donnée la forte corrélation saisonnière avec l'hiver 2005, on obtient une corrélation entre les données de l'hiver et la régression linéaire établie à partir des données de l'été de 0,77 ce qui est bon (même si ce n'est pas la régression optimale) et autorise l'utilisation de la même régression linéaire. Il faudrait vérifier sur d'autres jeux de données ce comportement. A l'inverse, on ne peut pas utiliser cette même régression avec les données de dioxyde d'azote du même été : le coefficient de corrélation linéaire n'est que de 0,38.

Apport supplémentaire des variables auxiliaires :

Les variables auxiliaires peuvent également être utilisées pour avoir une première idée de la portée des concentrations. Ainsi, on observe bien sur les deux variogrammes simples (Figure 15) des concentrations et de la deuxième régression linéaire une portée d'environ 8 km. Par contre, le comportement à l'origine et jusqu'à 5 km environ est totalement différent, la variable auxiliaire étant beaucoup plus régulière (d'autant plus qu'elle provient d'une régularisation dans un rayon de 100 ou 1000 m). Le variogramme croisé a quant à lui un comportement très proche de celui de la régression linéaire.

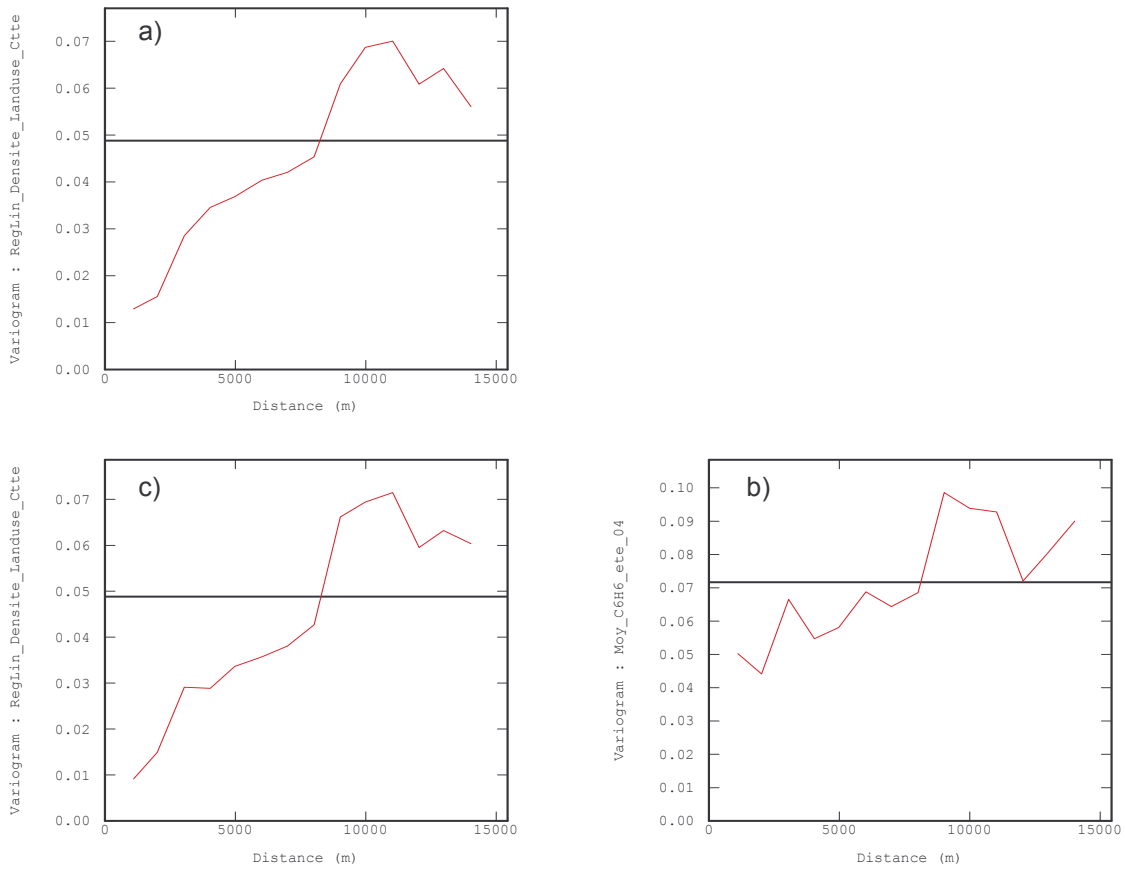
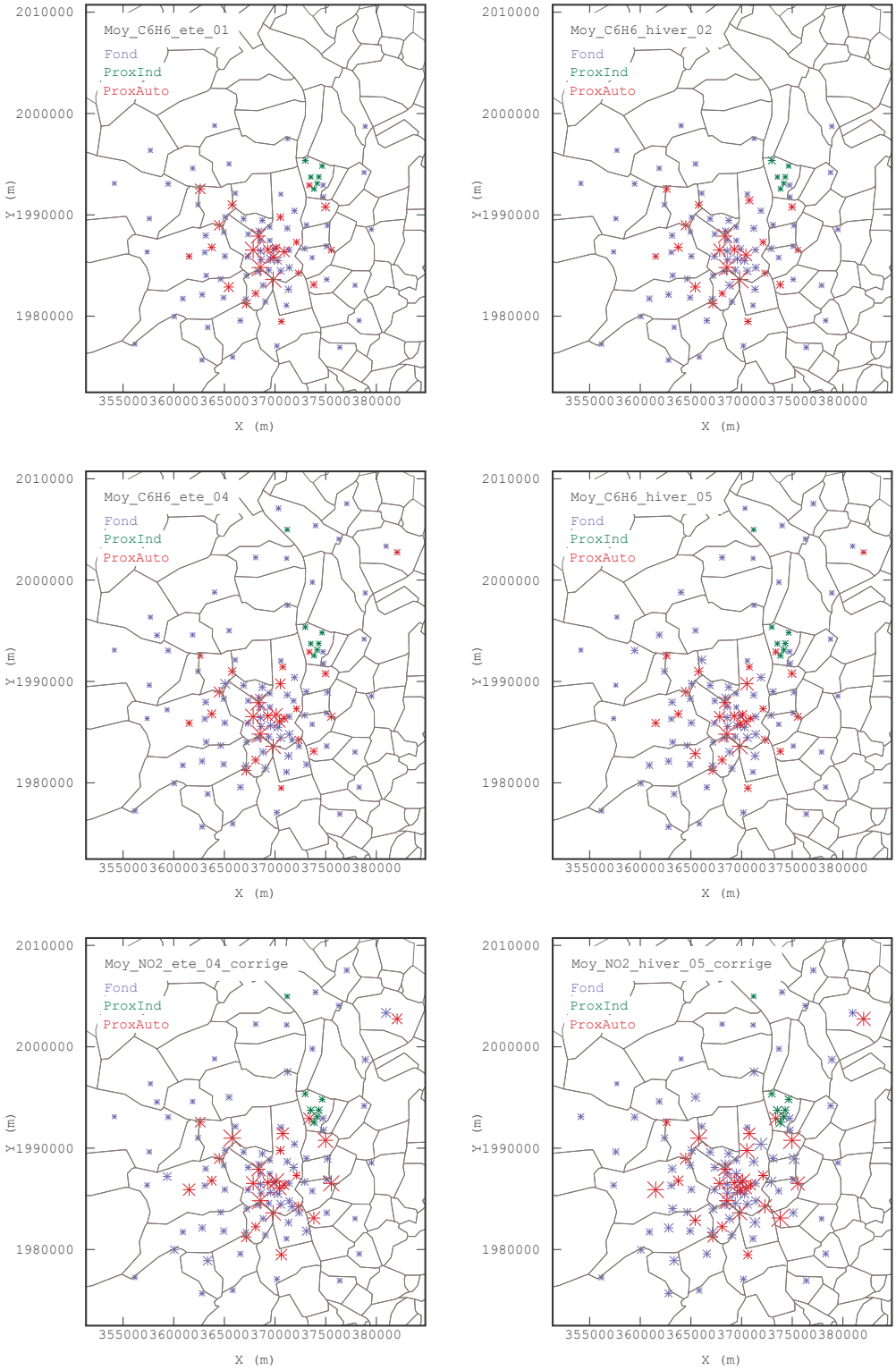


Figure 15 : Variogrammes simples a) de la deuxième régression linéaire utilisée en dérive externe, b) des concentrations en benzène de l'été 2004 et c) variogramme croisé.

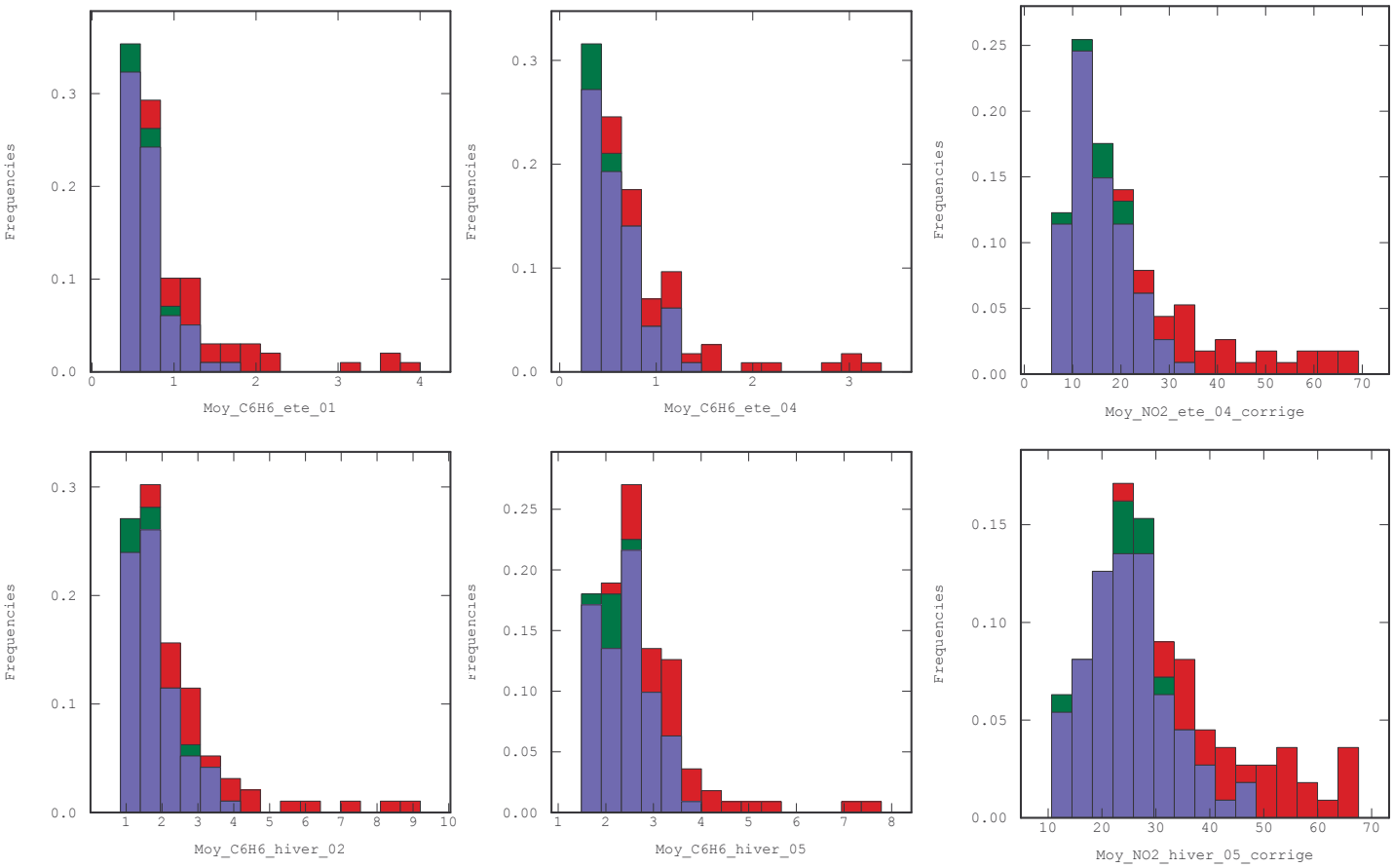
Annexes



Cartes d'implantation des mesures réalisées sur l'agglomération de Bordeaux pour le benzène (étés 2001 et 2004 et hivers 2002 et 2005) et pour le dioxyde d'azote (été 2004 et hiver 2005).

Statistiques élémentaires et histogrammes :

VARIABLE	Nombre	Minimum	Maximum	Moyenne	Écart-type	Variance	Coefficient de variation
C_6H_6 Été 2001	99	0,35	4,00	0,92	0,70	0,49	0,76
C_6H_6Hiver 2002	96	0,85	9,20	2,25	1,47	2,17	0,65
C_6H_6Été 2004	114	0,23	3,33	0,76	0,57	0,32	0,75
C_6H_6Hiver 2005	111	1,49	7,78	2,68	1,02	1,03	0,38
NO_2 Été 2004	114	5,70	69,15	21,53	14,11	198,96	0,66
NO_2 Hiver 2005	111	10,71	67,50	30,49	13,19	173,93	0,43



Histogrammes des concentrations en benzène (étés 2001 et 2004 et hivers 2002 et 2005) et en dioxyde d'azote (été 2004 et hiver 2005). En bleu : données de fond ; en vert : données de proximité industrielle, en rouge : données de proximité automobile.