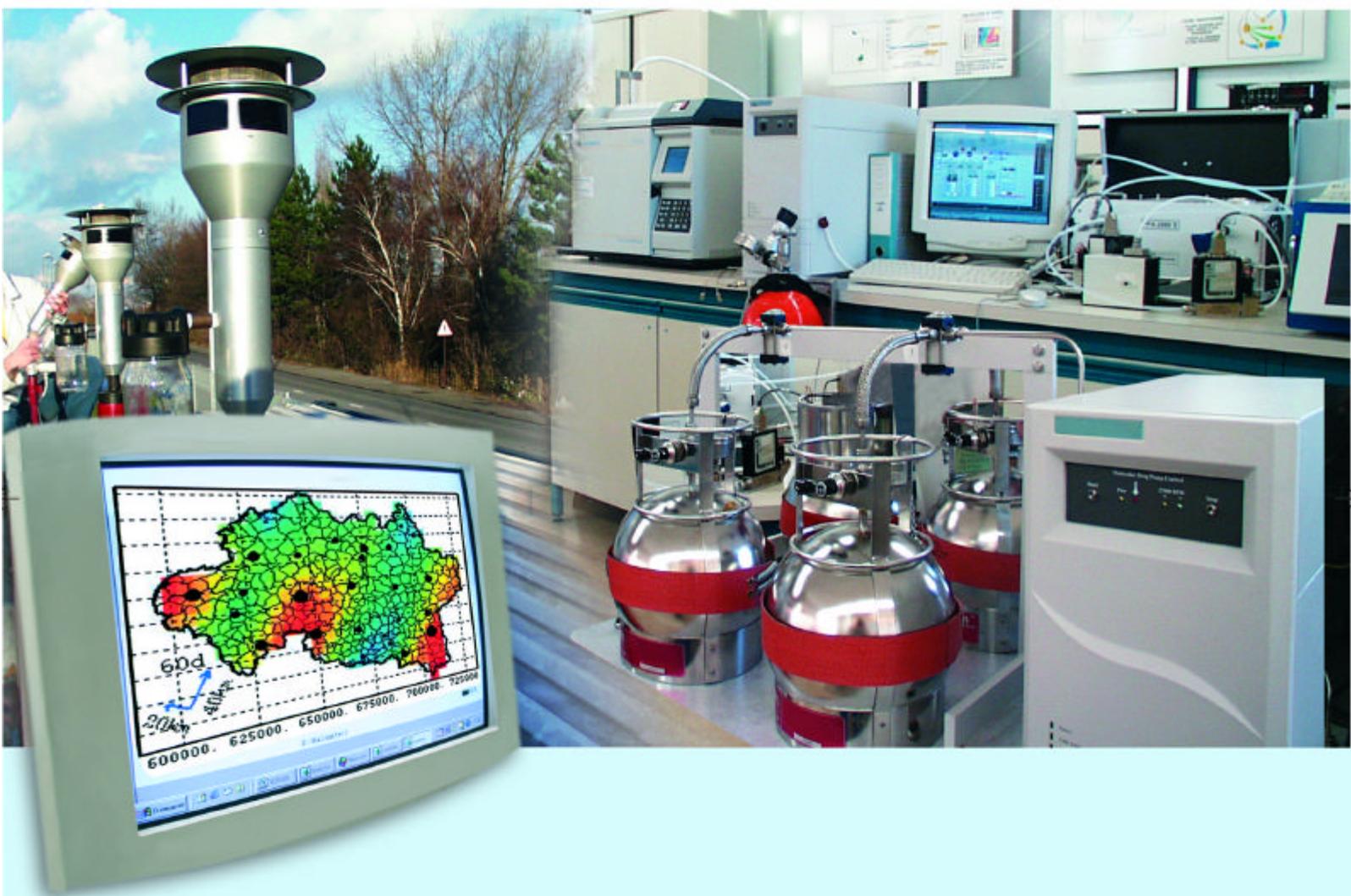




Laboratoire Central de Surveillance de la Qualité de l'Air



Etude n°12 - Géostatistiques et aspects intemporels

Géostatistiques et prise en compte de l'aspect temporel

Novembre 2004

Convention : 04000087

Laure MALHERBE - Giovanni CARDENAS





GEOSTATISTIQUE ET PRISE EN COMPTE DE L'ASPECT TEMPOREL

Laboratoire Central de Surveillance
de la Qualité de l'Air

Etude n°12

Convention n° 04000087

Financée par la Direction de la Prévention des
Pollutions et des Risques (DPPR)

NOVEMBRE 2004

Giovanni CARDENAS
Laure MALHERBE

Ce document comporte 63 pages (hors couverture et annexes).

	Rédaction	Vérification	Approbation
NOM	Giovanni CARDENAS Laure MALHERBE	Olivier SAINT-JEAN Laurence ROUIL	Martine RAMEL
Qualité	Ingénieur Etudes et Recherches Direction des Risques Chroniques	Responsable unité 2IEN Ingénieur Etudes et Recherches Direction des Risques Chroniques	Coordination LCSQA Direction des Risques Chroniques
Visa			

TABLE DES MATIERES

1. CONTENU DE LA FICHE.....	5
2. RÉSUMÉ.....	8
3. INTRODUCTION.....	9
3.1 CONTEXTE.....	9
3.2 PLAN DE L'ÉTUDE.....	9
4. ETAT DES LIEUX	10
4.1 EXPÉRIENCE DES AASQA	10
4.1.1 <i>Stratégies d'échantillonnage</i>	10
4.1.2 <i>Cartographie saisonnière ou annuelle</i>	11
4.2 ELÉMENTS BIBLIOGRAPHIQUES	19
4.3 AVANCÉES MÉTHODOLOGIQUES : COKRIGEAGE DE LA MOYENNE ANNUELLE, STRATÉGIE D'ÉCHANTILLONNAGE – CAS DU DIOXYDE D'AZOTE.....	19
4.3.1 <i>Cokrigage des moyennes saisonnières et de la moyenne annuelle</i>	20
4.3.2 <i>Variabilité temporelle</i>	20
4.3.3 <i>Optimisation de l'échantillonnage</i>	21
5. ETUDE DE CAS : CARTOGRAPHIE DU DIOXYDE D'AZOTE À BOURG-EN-BRESSE.....	22
5.1 DÉMARCHE THÉORIQUE	22
5.2 PRÉSENTATION DES DONNÉES.....	23
5.3 L'ALTITUDE.....	26
5.4 LA DISTANCE À LA ROUTE.....	27
5.5 LA DENSITÉ DE POPULATION	28
5.6 RELATION ENTRE MOYENNES SAISONNIÈRES DU DIOXYDE D'AZOTE.....	34
5.7 ANALYSE VARIOGRAPHIQUE DE L'HIVER.....	35
5.8 ESTIMATIONS DE L'HIVER.....	38
5.9 ANALYSE VARIOGRAPHIQUE DE L'ÉTÉ.....	41
5.10 ESTIMATIONS DE L'ÉTÉ.....	44
5.11 LE COKRIGEAGE ENTRE MOYENNES SAISONNIÈRES	46
5.11.1 <i>Cokrigage Ordinaire</i>	46
5.11.2 <i>Ajustement des variogrammes simples et croisés</i>	48
5.11.3 <i>Validation croisée des modèles ajustés</i>	50
5.11.4 <i>Résultats des cokrigeages saisonniers</i>	51
5.12 KRIGEAGE DE LA MOYENNE ANNUELLE.....	54
5.12.1 <i>Krigeage de la moyenne annuelle expérimentale</i>	54
5.12.2 <i>Moyenne des cokrigeages saisonniers</i>	56
6. CONCLUSION	59
7. RÉFÉRENCES	61
8. LISTE DES ANNEXES	63

LISTE DE FIGURES

Figure 1 : Localisation géographique des échantillonneurs passifs dans l'agglomération de Bourg-en-Bresse	23
Figure 2 : Carte de l'Altitude de Bourg-en-Bresse et nuages de corrélation avec les concentrations saisonnières et annuelle du dioxyde d'azote	26
Figure 3 : Nuages de corrélation entre les concentrations saisonnières et annuelle du dioxyde d'azote et la distance à la route la plus proche des tubes passifs	27
Figure 4 : Vue en perspective de la densité de population de Bourg-en-Bresse	28
Figure 5 : Histogrammes des moyennes saisonnières et annuelle du dioxyde d'azote, et de la densité de population migrée aux tubes	30
Figure 6 : Carte de densité de population	30
Figure 7 : Nuages de corrélation entre les moyennes saisonnières et annuelle du dioxyde d'azote et la densité de population	32
Figure 8 : Nuages de corrélation entre les concentrations saisonnières et annuelle du dioxyde d'azote et la densité de population le long des axes coordonnées	33
Figure 9 : Comparaison entre les moyennes saisonnières	34
Figure 10 : Variogrammes directionnels et carte variographique des concentrations hivernales	35
Figure 11 : Variogramme omnidirectionnel de l'hiver	36
Figure 12 : Nuages des corrélation valeurs vraies - valeurs estimées, modèles hivernales	38
Figure 13 : Cartes des estimations de la moyenne hivernale	39
Figure 14 : Carte des écarts- types d'estimation d'une moyenne hivernale	40
Figure 15 : variogrammes omnidirectionnels pour le NO2 estivale (avec et sans 2 points extrêmes)	41
Figure 16 : variogrammes directionnels pour le NO2 estivale	42
Figure 17 : Deux modèles de variogrammes testés	42
Figure 18 : Validation croisée des trois modèles de variogrammes choisis	43
Figure 19 : Cartes d'estimation de la moyenne estivale	44
Figure 20 : Cartes des écarts-types des estimations de la moyenne estivale	45
Figure 21 : Covariogrammes des variables brutes	49
Figure 22 : Nuage de corrélation entre valeurs vraies et valeurs estimées par cokrigeage en dérive externe	50
Figure 23 : Cokrigeage des moyennes saisonnières	52
Figure 24 : Ecart-type d'estimation par cokrigeage des moyennes saisonnières	52
Figure 25 : Principales statistiques de la moyenne annuelle du dioxyde d'azote	54
Figure 26 : Cartes du krigeage ordinaire et de l'écart-type de krigeage de la moyenne annuelle	55
Figure 27 : Comparaison des cartes de la moyenne annuelle obtenue par deux méthodes	56
Figure 28 : Nuages de corrélation entre valeurs vraies et valeurs estimées par deux méthodes de la moyenne annuelle	57
Figure 29 : Variogramme de la moyenne estivale, covariance croisée et variogramme de la moyenne hivernale	2
Figure 30 : Nuage de corrélation du krigeage et du cokrigeage sur la moyenne hivernale	4
Figure 31 : Nuage de corrélation du krigeage et du cokrigeage sur la moyenne estivale	4
Figure 32 : Krigeage Ordinaire de la moyenne estivale	5
Figure 33 : Cokrigeage Ordinaire de la moyenne estivale	5
Figure 34 : Krigeage Ordinaire de la moyenne hivernale	5
Figure 35 : Cokrigeage Ordinaire de la moyenne hivernale	6
Figure 36 : Présentation des nuages de corrélation pour les deux méthodes de prise en compte de la moyenne	7
Figure 37 : Carte de la moyenne des cokrigeages saisonniers	8
Figure 38 : Cartes du krigeage de la moyenne annuelle expérimentale	8

LISTE DE TABLEAUX

Tableau 1 – Campagnes de mesure dans les AASQA : différentes stratégies d'échantillonnage	11
Tableau 2 : Cartographie saisonnière ou annuelle.....	13
Tableau 3 : Méthodes d'estimation de la moyenne saisonnière ou annuelle.....	19
Tableau 4 : Statistiques des huit campagnes de mesure du dioxyde d'azote	24
Tableau 5 : Statistiques des huit campagnes de mesure du dioxyde d'azote	25
Tableau 6 : Statistiques des deux sélections des tubes passifs en fonction de la valeur de la densité de population.....	29
Tableau 7 : Moyennes des concentrations par classe de densité de population	32
Tableau 8 : Paramètres de la droite de régression entre les moyennes saisonnières du dioxyde d'azote et la densité de population.....	33
Tableau 9 : Validation croisée des modèles pour l'azote hivernale.....	37
Tableau 10 : Statistiques des estimations du NO2 hivernale.....	40
Tableau 11 : Statistiques des écarts- types d'estimation du NO2 hivernale.....	40
Tableau 12 : Validation croisée plusieurs modèles de dérive externe, moyenne estivale	43
Tableau 13 : Statistiques des estimations de la moyenne estivale	45
Tableau 14 : Statistiques des écarts-types des estimations de la moyenne estivale.....	45
Tableau 15 : Paliers du modèle de corégionalisation linéaire.....	49
Tableau 16 : Statistiques de la validation croisée du modèle multivariable.....	50
Tableau 17 : Ecart-type d'estimation par cokrigage	52
Tableau 18 : Statistiques des estimations par cokrigage des moyennes saisonnières (ensemble du domaine et sélections des écarts- types plus petites).....	53
Tableau 19 : Statistiques du krigeage ordinaire de la moyenne annuelle.....	55
Tableau 20 : Statistiques de la moyenne des deux cokrigeages saisonniers.....	56
Tableau 21 : Statistiques de la validation croisée du modèle multivariable.....	57
Tableau 22 : Paliers du modèle de corégionalisation linéaire sous « GA ».....	2
Tableau 23 : Récapitulatif des statistiques de krigeage et de cokrigage des données de la pollution	3
Tableau 24 : Récapitulatif des statistiques de la validation croisée des deux méthodes d'estimation de la pollution moyenne annuelle.....	7

1. CONTENU DE LA FICHE

Thème III : TRAVAUX numériques

13 : GEOSTATISTIQUE ET PRISE EN COMPTE DE L'ASPECT TEMPOREL

PROGRAMME NOUVEAU

CONTEXTE ET OBJECTIFS

Hormis dans les régions équipées d'un réseau dense d'analyseurs, les cartes de concentration issues de l'application des méthodes géostatistiques sont établies à l'aide de données d'échantillonnage par tubes à diffusion. De ce fait, elles représentent le plus souvent des situations moyennes sur la durée des campagnes, soit quelques semaines en été et/ou en hiver. La dimension temporelle n'est pas encore prise en compte au moment du krigeage, afin d'extrapoler ces cartes à des périodes plus longues, typiquement la saison ou l'année comme le demandent les directives européennes.

Des travaux destinés à satisfaire aux directives ont été toutefois réalisés ou engagés dans certaines AASQA. Ils sont généralement découplés de la géostatistique et consistent à estimer par la modélisation, en amont de l'analyse géostatistique, des concentrations moyennes annuelles (ex : AIRPARIF-cartographie du benzène-, ATMO Poitou-Charentes).

D'autre part, si l'élaboration d'une stratégie d'échantillonnage temporel destinée à estimer au mieux une moyenne saisonnière ou annuelle a fait déjà l'objet d'études, notamment au sein du groupe de travail Moyens Mobiles, cette question n'a encore jamais été abordée du point de vue de la géostatistique et des besoins du krigeage.

Dans l'étude confiée en 2002/2003 par l'INERIS au Centre de Géostatistique de l'Ecole des Mines de Paris, des réflexions ont été amorcées sur l'intégration du temps dans l'estimation géostatistique et sur le possible couplage des aspects spatiaux et temporels lors de la mise en œuvre du krigeage. Ces premiers éléments semblent ouvrir des perspectives intéressantes quant à une définition plus judicieuse de stratégies d'échantillonnage et à une meilleure utilisation des méthodes géostatistiques pour répondre aux exigences réglementaires.

RAPPEL DES TRAVAUX 2003

Etude de l'application des méthodes géostatistiques à la représentation de la qualité de l'air :

- ≪ finalisation d'un guide méthodologique sur l'usage des techniques de la géostatistique linéaire ;
- ≪ suivi d'une étude du Centre de Géostatistique de l'Ecole des Mines de Paris sur la cartographie du dioxyde d'azote à l'échelle d'une agglomération ;
- ≪ développement d'une méthodologie d'évaluation des incertitudes associées aux méthodes d'interpolation géostatistiques.
- ≪ Participation au groupe de travail Moyens Mobiles :
- ≪ évaluation de la méthode des plans de sondage pour la définition de stratégies d'échantillonnage temporel et l'estimation de moyennes annuelles

PROGRAMME PROPOSE POUR 2004

Il est proposé de recenser et de synthétiser les expériences des AASQA relatives à la prise en compte de la composante temporelle dans l'élaboration de campagnes de mesure et dans la réalisation de cartes moyennes de concentration.

Dans ce contexte, une analyse du problème spatio-temporel et des différentes méthodes employées pour extrapoler des cartes dans le temps pourra être menée en collaboration avec les AASQA impliquées sur ce sujet et avec le Centre de Géostatistique de l'Ecole des Mines de Paris. Elle pourra s'appuyer sur les études conduites au sein des groupes de travail Moyens mobiles et Tubes passifs et sur la littérature scientifique qui existe en la matière.

Les possibilités et les limites des logiciels de géostatistique (Isatis et Geostatistical Analyst) pour traiter le problème spatio-temporel seront également examinées.

Selon les conclusions de cette analyse, des calculs complémentaires sur une ou deux régions tests pourront être effectués en collaboration avec des AASQA volontaires. Ils auront pour objectif de compléter les travaux existants et d'évaluer la possibilité d'un couplage espace/temps dans l'application des méthodes géostatistiques et l'établissement de cartes moyennes.

Tous les résultats de ce travail seront consignés dans un document synthétique sous forme de recommandations et de conseils pratiques à l'intention des AASQA

COLLABORATIONS

AASQA , EMD, Centre de Géostatistique de l'Ecole des Mines de Paris

DUREE DES TRAVAUX

Il s'agit d'une nouvelle étude d'une durée prévisionnelle d'un an

2. RESUME

Compte tenu

- de l'usage désormais répandu des outils géostatistiques dans les AASQA,
- des contraintes réglementaires auxquelles doivent répondre les cartographies,

il est apparu utile d'établir un bilan des stratégies et techniques mises aujourd'hui en œuvre pour produire des cartes moyennes sur des périodes longues (années, saison) et d'explorer si nécessaire une ou plusieurs d'entre elles.

Une synthèse des expériences des AASQA met en avant le rôle prépondérant des campagnes de mesure par tubes passifs et montre des pratiques communes dans la stratégie d'échantillonnage. S'il s'appuie toujours sur le krigeage, le mode de traitement des données varie selon les associations :

- Les cartes produites représentent la situation moyenne pendant la campagne ; l'exploitation de données de stations fixes peut aider à en apprécier la représentativité temporelle.
- Les données d'entrée de la cartographie sont préalablement extrapolées sur l'année, grâce à des mesures de stations fixes.
- Les données d'entrée de la cartographie ne sont pas extrapolées mais une correction, fonction des mesures de stations fixes, est appliquée aux concentrations estimées dans tout le domaine.

Dans la première approche, le facteur temporel est ignoré ou considéré a posteriori, dans les deux suivantes, il est découplé du problème spatial. Une autre méthodologie, développée par le Centre de Géostatistique pour la cartographie du dioxyde d'azote et illustrée ici sur un cas d'étude, permet de rapprocher ces deux aspects au moyen d'un cokrigeage. Si, comme la première méthode, elle suppose que les mesures reflètent correctement les situations de pollution saisonnières, elle améliore sensiblement les estimations spatiales et ouvre la voie à des stratégies d'échantillonnage plus judicieuses dans l'espace et le temps.

3. INTRODUCTION

3.1 CONTEXTE

Les techniques géostatistiques ont aujourd'hui fait leur preuve dans l'exploitation des données de qualité de l'air, comme outils d'analyse et d'estimation. A partir de mesures ponctuelles de concentration (instantanées si elles proviennent de stations fixes, intégrées sur une à plusieurs semaines si elles sont issues de tubes à diffusion), les méthodes de la géostatistique permettent d'obtenir une représentation bidimensionnelle de la situation de pollution observée. La plupart du temps, seules les données de tubes sont en nombre suffisant pour établir une cartographie ; les cartes estimées reflètent alors une situation moyenne sur quelques semaines en été et/ou en hiver. Or, en vue de répondre aux besoins de la réglementation, il est nécessaire d'effectuer des bilans saisonniers ou annuels, ce qui suppose d'extrapoler ces cartes à de plus longues périodes.

Face à de telles contraintes, ce travail a pour objet d'étudier la façon dont la dimension temporelle est ou peut être prise en compte dans les cartographies. Il s'appuie sur les expériences menées dans les AASQA et sur les travaux réalisés en 2003 et 2004 par le Centre de Géostatistique de l'Ecole des Mines de Paris.

3.2 PLAN DE L'ETUDE

Ce rapport se compose de deux parties :

- la première dresse un état des lieux des moyens et techniques utilisés par les AASQA pour réaliser des cartographies moyennes et des travaux méthodologiques engagés sur le sujet ; elle est complétée par une brève revue bibliographique.
- la seconde est une étude de cas : le cokrigage temporel, dont l'intérêt a été montré dans des études conduites par le Centre de Géostatistique (Fouquet, 2003) est mis en œuvre sur une nouvelle zone géographique.

4. ETAT DES LIEUX

4.1 EXPERIENCE DES AASQA

Ce paragraphe aborde le traitement du problème temporel dans l'élaboration de campagnes d'échantillonnage et l'exploitation des données collectées. **Ce bilan porte spécifiquement sur les campagnes de mesure conduites à des fins de cartographie.** Les informations qu'il contient sont tirées d'échanges téléphoniques ou de discussions entre l'INERIS et les AASQA et de documents publiés sur ce sujet.

4.1.1 Stratégies d'échantillonnage

Différentes stratégies d'échantillonnage dans le temps ont été identifiées et sont présentées dans le Tableau 1.

Polluant	Stratégie de mesure	Fondement de cette stratégie	But
NO ₂	<p>Moyens : Tubes à échantillonnage passif</p> <p>Périodes de mesure : Trois à quatre quinzaines de mesure consécutives en hiver et en été</p>	<p>Stratégie mentionnée dans le guide <i>Echantillonneurs passifs NO₂</i>. La moins contraignante en terme de personnel.</p> <p>Utilisée dans de nombreuses AASQA.</p>	Cartographies saisonnières et annuelles
	<p>Moyens : Tubes à échantillonnage passif</p> <p>Périodes de mesure : Plusieurs quinzaines de mesure réparties sur toute l'année</p>	<p><i>Guide technique EEA, 1998*</i> : durée totale d'échantillonnage = 20% de la période de référence (ex: ASCOPARG) + tests préalables sur des données de stations fixes (ex: ASCOPARG, AIRMARAIX)</p> <p>Méthode des plans de sondage (ex : ATMO Poitou-Charentes)</p>	Cartographie annuelle
	<p>Moyens : Tubes à échantillonnage passif</p> <p>Périodes de mesure : Tubes renouvelés tout au long de l'année (ex : LIG' AIR)</p>		Cartographie annuelle
O ₃	<p>Moyens : Tubes à échantillonnage passif</p> <p>Périodes de mesure : Plusieurs semaines de mesure consécutives en été</p>		Cartographie estivale

	<p>Moyens : Tubes à échantillonnage passif</p> <p>Périodes de mesure : Plusieurs semaines de mesure réparties sur la saison estivale (ex: 6 fois 1 semaine)</p>	<p><i>Guide technique EEA, 1998:</i> durée totale d'échantillonnage = 20% de la période de référence + tests préalables sur des données de stations fixes (ex: ASCOPARG)</p>	<p>Cartographie estivale</p>
	<p>Moyens : Camion laboratoire</p> <p>Périodes de mesure : Passage du camion en divers sites de la zone étudiée. On obtient en ces points des séries partielles et non simultanées de données qui complètent les informations délivrées par les stations fixes (ex: ATMO Poitou-Charentes)</p>		<p>Cartographie automatique horaire et cartographie d'indicateurs annuels (AOT, dépassement du seuil de $120 \mu\text{g}/\text{m}^3$ sur 8h, dépassements horaires)</p>
Benzène	<p>Moyens : Tubes à échantillonnage passif</p> <p>Périodes de mesure : Trois à quatre quinzaines de mesure consécutives en hiver et en été</p>		<p>Cartographies saisonnières et annuelles</p>

Tableau 1 – Campagnes de mesure dans les AASQA : différentes stratégies d'échantillonnage

4.1.2 Cartographie saisonnière ou annuelle

En général, les concentrations moyennes saisonnières ou annuelles utilisées dans la cartographie sont estimées comme la moyenne des mesures effectuées sur la saison ou sur l'année. Dans le cas d'un échantillonnage passif, les cartes qui en découlent représentent une situation sur quelques semaines dont la représentativité par rapport à la période entière est le plus souvent admise ou contrôlée sur des mesures d'analyseurs.

Des méthodes de correction préalables ou consécutives au traitement géostatistique et fondées sur l'exploitation de mesures de stations fixes ont été cependant testées par quelques AASQA.

Le tableau suivant recense ces approches et indique pour chacune l'exploitation géostatistique qui est faite des données. Une description plus détaillée des méthodes est fournie dans les paragraphes 3.1.2.1 à 3.1.2.3.

Polluant	Variable à cartographier	Données exploitées	Exploitation géostatistique	Traitement des cartes d'estimation
NO ₂	Moyenne saisonnière	Valeurs moyennes sur les quinzaines de mesure de la saison considérée	Krigeage avec ou sans variables auxiliaires <i>Approche la plus courante</i>	
	Moyenne saisonnière	Valeurs moyennes sur les quinzaines de mesure de la saison considérée Valeurs moyennes sur les quinzaines de mesure de l'autre saison	Cokrigeage, avec ou sans variable auxiliaire : les concentrations moyennes sur une saison sont estimées dans le domaine d'étude en tenant compte des données de l'autre saison <i>Approche développée par le Centre de Géostatistique et testée à l'ASPA</i>	
	Moyenne annuelle	Moyenne des mesures disponibles sur l'année	Krigeage avec ou sans variables auxiliaires <i>Approche la plus courante</i>	
	Moyenne annuelle	Moyenne des mesures disponibles sur l'année	Krigeage avec ou sans variables auxiliaires	Application d'un facteur correctif, fonction du type d'environnement, en tout point de la grille d'estimation (ORAMIP)
	Moyenne annuelle	Moyenne des mesures disponibles sur l'année, corrigée à l'aide des données de stations fixes (application d'un facteur, AIRMARAIX, régression linéaire ou non linéaire, AIRPARIF, technique des plans de sondage, ATMO PC)	Krigeage avec ou sans variables auxiliaires	

	Moyenne annuelle	Moyenne des mesures disponibles sur l'été Moyenne des mesures disponibles sur l'hiver	Cokrigeage de la moyenne annuelle par les moyennes saisonnières <i>Approche développée par le Centre de Géostatistique et testée à l'ASPA</i>	
Ozone	Indicateurs annuels (AOT, dépassements de seuils)	Séries partielles de données horaires issues de moyens mobiles Séries complètes de données horaires issues de stations fixes Séries de données météorologiques	Krigeage avec ou sans variables auxiliaires des concentrations horaires	
Benzène	Moyenne annuelle	Moyenne des mesures disponibles sur l'année	Krigeage avec ou sans variables auxiliaires	
	Moyenne annuelle	Moyenne annuelle estimée par un modèle empirique calé en chaque point d'échantillonnage sur les données des tubes (AIRPARIF, 2000)	Krigeage avec ou sans variables auxiliaires	
	Moyenne annuelle	Moyenne des mesures disponibles sur l'année, corrigée par un modèle de régression établi sur les données de stations fixes (AIRPARIF)	Krigeage avec ou sans variables auxiliaires	

Tableau 2 : Cartographie saisonnière ou annuelle

4.1.2.1 Estimation de concentrations moyennes aux points d'échantillonnage

4.1.2.1.1 Approche la plus simple : moyenne des mesures ; pas de correction

Estimation des concentrations moyennes aux points de mesure

Les moyennes saisonnières sont estimées par les moyennes arithmétiques des données de tubes calculées sur chaque saison :

$$\hat{y}_{saison} = \frac{1}{N_s} \sum_{i=1}^{N_s} y_i$$

y_i : concentration intégrée sur la période d'exposition du tube

La moyenne annuelle est estimée par la moyenne des données de tubes sur l'année entière :

$$\hat{y}_{an} = \frac{1}{N_a} \sum_{i=1}^{N_a} y_i$$

Des critères sur le nombre minimal de données valides nécessaires par saison peuvent être imposés.

De façon évidente, la qualité de l'estimation dépend étroitement de la stratégie d'échantillonnage utilisée.

Erreur d'estimation

Aux points des stations fixes, l'erreur d'estimation peut être approchée *a posteriori* par l'écart relatif entre la moyenne des données sur la période de mesure et la moyenne réelle sur la saison ou l'année (ASCOPARG) :

$$e = 100 \frac{|y - \hat{y}|}{y}$$

La **théorie des sondages**, sous réserve que l'échantillonnage ait été réalisé de façon aléatoire (semaines ou quinzaines de mesure réparties aléatoirement dans chaque saison ou dans l'année) et que les données en chaque point soient suffisamment nombreuses (au moins une dizaine par an pour une estimation annuelle), permet de calculer un **intervalle de confiance** autour des valeurs estimées (Lavancier et al., 2003) :

- La variance de l'estimateur \hat{y} s'écrit : $\text{var}(\hat{y}) = \frac{N-n}{n.N} S^2$ avec $S^2 = \frac{1}{N-1} \text{var}(y)$.
- L'intervalle de confiance à 95% autour de la moyenne est $IC_{95\%} = \left[\hat{y} - 1,96 \sqrt{\frac{N-n}{n.N} S^2} ; \hat{y} + 1,96 \sqrt{\frac{N-n}{n.N} S^2} \right]$.
- La largeur $L_{IC_{95\%}}$ de cet intervalle vaut donc $L_{IC_{95\%}} = 2 \cdot 1,96 \sqrt{\frac{N-n}{n.N} S^2}$.

4.1.2.1.2 Correction de la moyenne grâce aux données d'une station fixe

Cette méthode nécessite la présence d'une ou plusieurs stations fixes jugées représentatives du comportement des polluants aux points d'échantillonnage.

a) Application d'un coefficient correctif

Une procédure par règle de trois a été développée par AIRMARAIX lors de campagnes ayant pour but de qualifier l'état zéro de la qualité de l'air autour de projets d'axes routiers (autoroutes, tunnels...) (AIRMARAIX, 2001).

En un point d'échantillonnage, la concentration moyenne annuelle de dioxyde d'azote ou de benzène est estimée par la formule :

$$\hat{C}_i^{an} = \left(\frac{C^{an}}{C^{période}} \right)_{SF} \cdot \left(\frac{C^{période}}{C_T^{période}} \right)_{SF'} \cdot C_i^{période}$$

$\left(\frac{C^{an}}{C^{période}} \right)_{SF}$: rapport moyen sur l'ensemble des stations fixes considérées entre concentration moyenne annuelle et concentration moyenne sur la période de mesure

$\left(\frac{C^{période}}{C_T^{période}} \right)_{SF'}$: rapport moyen sur l'ensemble des stations fixes équipées de tubes entre concentration mesurée par l'analyseur et concentration mesurée par échantillonnage passif.

$C_i^{période}$: concentration moyenne sur la période de mesure au point d'échantillonnage i

b) Application de la théorie des sondages

Les plans de sondage offrent là encore un cadre théorique pour corriger les estimations à l'aide des données d'une station fixe (indicées $_{aux}$ dans les formules).

Trois méthodes sont proposées :

- ***l'estimateur redressé par la différence*** : le \hat{y} -estimateur \hat{y} est corrigé par un facteur additif.

$$\hat{y}_D = \hat{y} + (\bar{y}_{aux} - \hat{y}_{aux})$$

- ***l'estimateur redressé par le quotient*** : le \hat{y} -estimateur est corrigé par un facteur multiplicatif.

$$\hat{y}_Q = \hat{y} \frac{\bar{y}_{aux}}{\hat{y}_{aux}}$$

- ***l'estimateur redressé par régression*** : le \hat{y} -estimateur est corrigé par régression linéaire.

$$\hat{y}_R = \hat{y} + \hat{b}(\bar{y}_{aux} - \hat{y}_{aux})$$

A chaque estimateur est associée une expression de l'intervalle de confiance à 95% de la moyenne (Lavancier et al., 2003). L'estimation de cet intervalle tient compte de la similarité – mesurée par la covariance- entre la série échantillonnée et la série auxiliaire.

c) Correction par un modèle de régression

Cette méthode, mise au point et utilisée par AIRPARIF, repose sur l'exploitation des données d'analyseurs en présence d'un réseau dense de stations (une douzaine de stations pour le benzène, environ vingt-cinq stations pour le NO₂).

Employée dans de multiples études pour le dioxyde d'azote et le benzène (cartographie de la qualité de l'air dans des agglomérations, autour d'aéroports, d'industries...), elle consiste à :

- établir un modèle de régression, linéaire ou non linéaire, entre les concentrations moyennes sur l'année mesurées par les stations fixes et les concentrations moyennes sur la campagne mesurées par ces stations :

$$\hat{y}_a^{SF} \approx f(\bar{y}_{campagne}^{SF})$$

\hat{y}_a^{SF} : concentration annuelle estimée par régression

$\bar{y}_{campagne}^{SF}$: concentration moyenne sur la campagne

- appliquer ce même modèle aux concentrations délivrées par les tubes :
 $\hat{y}_a^i \approx f(\bar{y}_{campagne}^i)$

La comparaison pour les stations fixes entre les concentrations annuelles modélisées \hat{y}_a^{SF} et observées y_a^{SF} fournit une indication en pourcentage de l'incertitude d'estimation.

Afin de pouvoir s'appliquer à des sites de fond urbain, de proximité automobile ou de caractéristiques intermédiaires entre ces deux catégories, le modèle de régression du NO₂ incorpore à la fois des données de stations urbaines de fond et des données de stations trafic.

Pour le benzène, à défaut de surveillance continue hors du cœur de Paris, le modèle construit sur les données urbaines est supposé valable aux points d'échantillonnage de type périurbain ou rural.

4.1.2.1.3 Approche par modélisation

a) Modélisation empirique

Un modèle d'accumulation-dispersion a été élaboré par AIRPARIF en 2000 afin de dresser une cartographie moyenne annuelle du benzène à partir de données de tubes et en l'absence d'un réseau dense de stations fixes¹. La démarche adoptée est la suivante :

- Modélisation des concentrations journalières de benzène mesurées par deux stations fixes urbaines. Construction d'un modèle général de la forme :

$$\hat{y}_j \approx F_{aj} \times F_{dj} \times F_s$$

¹ Par suite de l'extension de son réseau de surveillance du benzène, AIRPARIF utilise aujourd'hui la méthode de correction par régression présentée précédemment.

F_{aj} : facteur d'accumulation journalière exprimant les variations des concentrations en fonction du jour de la semaine. Le facteur retenu est la moyenne des facteurs de chaque station.

F_{dj} : facteur de dispersion journalière exprimant les variations des concentrations en fonction de la vitesse de vent. Le facteur retenu est la moyenne des facteurs de chaque station.

F_s : facteur saisonnier exprimant les variations des concentrations en fonction du mois. Le facteur retenu est la moyenne des facteurs de chaque station.

- Modélisation des concentrations moyennes sur quinze jours aux stations fixes. Construction d'un modèle propre à chaque station i :

$$\hat{y}_q^i = A^i + B^i \cdot \bar{y}_j$$

A^i et B^i sont estimés par régression linéaire sur trois années de mesures ; \bar{y}_j est la moyenne sur quinze jours des concentrations journalières estimées par le modèle général.

Evaluation de la précision des modèles sur une année test. Cette précision est de l'ordre de 5% pour l'estimation de la moyenne annuelle.

- Modélisation des concentrations moyennes sur quinze jours aux points d'échantillonnage. Construction d'un modèle spécifique à chaque point k :

$$\hat{y}_q^k = A^k + B^k \cdot \bar{y}_j$$

A^k et B^k sont estimés par régression linéaire sur les six mesures disponibles (trois quinzaines en hiver, trois quinzaines en été).

- Application des modèles : en chaque point k , le modèle propre à ce site permet de reconstituer la série annuelle des concentrations $\{\hat{y}_{qt}^k, t = 1 \dots 26\}$. La concentration moyenne annuelle est estimée par la moyenne de cette série.
- Utilisation des concentrations moyennes estimées comme données d'entrée du krigeage.

b) Modélisation statistique

Reconstitution de séries temporelles

La modélisation statistique est employée par ATMO Poitou-Charentes pour exploiter les données de **moyens mobiles** et compléter les mesures d'ozone délivrées par les stations fixes. Elle a pour objet de reconstituer aux points d'échantillonnage des séries annuelles de concentrations horaires d'ozone afin :

- de réaliser des cartes horaires de concentration ;
- d'utiliser ces cartes pour cartographier divers indicateurs annuels.

Différents modèles statistiques de reconstitution ont été testés et comparés : régression linéaire et non linéaire, analyse de covariance (ANACOVA), modèle neuro-flou. Avec un nombre relativement limité de variables (concentrations d’ozone et de NOx, vitesse et direction de vent, température, heure du jour), l’analyse de covariance fournit des résultats prometteurs (Cottier, 2003).

Estimation d’une concentration moyenne et de son incertitude

Si elle n’est pas encore régulièrement employée par les AASQA, une méthode de reconstitution temporelle a été élaborée à l’Ecole des Mines de Douai dans le cadre du GT Moyens Mobiles. S’appuyant sur la norme *ISO 9359 – Qualité de l’air ambiant – Echantillonnage aléatoire stratifié pour l’évaluation de la qualité de l’air ambiant*, elle consiste à prendre en compte l’influence des variations de paramètres météorologiques sur les mesures de concentration. Un outil statistique a été mis au point et évalué sur des données de stations fixes pour différents polluants (SO₂, NO, NO₂, O₃, CO, PM₁₀, PM_{2,5}) et différentes zones géographiques. Un guide d’utilisation à destination des AASQA sera rédigé en 2005 dans le cadre des travaux du LCSQA.

4.1.2.2 Traitement des cartes de concentration

Dans une étude de cartographie du NO₂ préalable à un projet urbain, l’ORAMIP a mis en œuvre une procédure de correction des cartes de krigeage. Cette procédure consiste :

- à calculer pour les stations fixes de la zone d’étude le rapport entre la concentration annuelle et la concentration moyenne de NO₂ sur la période d’échantillonnage; ce rapport diffère selon qu’il s’agit de stations de type urbain ou trafic ;
- à appliquer, suivant le type d’environnement, l’un ou l’autre de ces rapports aux concentrations estimées sur la grille.

4.1.2.3 Bilan

Le tableau suivant récapitule l’ensemble des méthodes décrites précédemment et le type de donnée requis par chacune.

Méthode d’estimation de la moyenne saisonnière ou annuelle aux points d’échantillonnage	Méthode de correction des cartographies	Données d’entrée pour la cartographie	Données supplémentaires nécessaires
Règle de 3		Mesures de tubes	Tubes installés sur des stations fixes
Théorie des sondages avec correction		Mesures de tubes	Une ou plusieurs stations fixes représentatives du comportement des polluants dans la zone

			d'échantillonnage
Modélisation empirique		Mesures de tubes	Une ou plusieurs stations fixes Données de vent
Modélisation statistique		Mesures de camion laboratoire	Une ou plusieurs stations fixes Données météorologiques
	Application d'un facteur correctif	Mesures de tubes	Stations fixes de types urbain et trafic

Tableau 3 : Méthodes d'estimation de la moyenne saisonnière ou annuelle

4.2 ELEMENTS BIBLIOGRAPHIQUES

L'estimation de concentrations moyennes sur le long terme à partir de mesures ponctuelles dans le temps n'est pas abordée explicitement dans la littérature scientifique. En pollution de l'air, le problème spatio-temporel est centré avant tout sur la cartographie horaire de l'ozone (Christakos et Vyas, 1998). Ces auteurs mettent en évidence le caractère irréaliste d'une séparation espace/temps et la nécessité de coupler ces deux aspects (Snepvangers et al., 2003). Dans la méthode qu'ils ont développée, l'estimation des concentrations est fondée non seulement sur les observations de l'instant considéré mais également sur les observations effectuées à d'autres périodes. Elle repose sur le calcul et la modélisation d'une covariance généralisée, qui représente simultanément les corrélations dans l'espace et le temps. D'autres polluants et d'autres approches ont fait l'objet d'études (ex : Kyriakidis et Journel, 2001 : cartographie des dépôts moyens mensuels de sulfate sur l'Europe); dans tous les cas cependant, les méthodologies présentées ne répondent pas exactement à notre problème :

- plus que de produire une estimation moyenne sur le long terme, leur but est de prendre en compte plus finement la dynamique du polluant afin de représenter l'évolution des concentrations dans l'espace et le temps.
- les données exploitées, issues de réseaux de stations fixes, ne correspondent pas aux jeux de données dont disposent généralement les AASQA pour la cartographie.

4.3 AVANCEES METHODOLOGIQUES : COKRIGEAGE DE LA MOYENNE ANNUELLE, STRATEGIE D'ECHANTILLONNAGE – CAS DU DIOXYDE D'AZOTE

Les résultats présentés dans ce paragraphe sont issus de travaux conduits par le Centre de Géostatistique de l'École des Mines de Paris en collaboration avec le LCSQA/INERIS (convention INERIS/ARMINES, 2003) et avec l'ASPA (2004). Ces travaux, réalisés à l'aide du logiciel Isatis (Géovariances), sont largement développés dans des rapports d'étude (Gallois, 2004, Fouquet, 2003, Le Loc'h, 2003). Consacrés à la **cartographie saisonnière et annuelle du dioxyde d'azote en milieu urbain**, ils ont abouti à des résultats concrets sur :

- les moyens d'améliorer l'estimation des concentrations moyennes annuelles ;
- l'évaluation de la variabilité temporelle ;
- la définition d'une stratégie d'échantillonnage spatial et temporel plus efficace.

4.3.1 Cokrigage des moyennes saisonnières et de la moyenne annuelle

La pratique générale pour cartographier les concentrations estivales (ou hivernales) est d'interpoler entre les mesures moyennes sur l'été (ou sur l'hiver), indépendamment de l'autre saison. Quant à la cartographie des concentrations annuelles, elle est couramment réalisée à partir des mesures moyennes sur l'année et éventuellement, d'une ou plusieurs variables auxiliaires.

Or plusieurs exemples (Mulhouse, Montpellier) mettent en évidence des **corrélations élevées entre les mesures estivales et hivernales de NO₂ qu'il se révèle judicieux d'exploiter**. De façon schématique, l'estimations des concentrations saisonnières et annuelles procède de la façon suivante :

- recherche de variables explicatives pour chaque saison.
- cokrigage des moyennes saisonnières avec dérive externe. Les variables auxiliaires mises en dérive peuvent être différentes pour chaque saison. Les variogrammes sont ajustés indirectement par validation croisée.
- cokrigage de la moyenne annuelle : cela revient à prendre la moyenne des deux estimations saisonnières précédemment obtenues.

Une comparaison par validation croisée avec les techniques de krigeage habituellement utilisées montrent la supériorité de cette méthodologie.

Celle-ci a l'avantage d'**intégrer de façon plus pertinente l'ensemble des informations disponibles** et de fournir des **cartographies saisonnières et annuelles cohérentes entre elles**. En revanche, elle suppose que l'échantillonnage est représentatif des situations de pollution observées dans les différentes saisons. Son usage est pour l'instant réservé au **dioxyde d'azote** ; son extension à d'autres polluants (le benzène par exemple) n'a pas encore été évaluée.

4.3.2 Variabilité temporelle

La variance de l'erreur d'estimation spatiale est donnée en tout point de la grille par la carte de variance de krigeage. En revanche, la variance de l'erreur d'estimation temporelle ne peut être évaluée qu'aux sites de mesure fixes, par comparaison sur plusieurs années entre la moyenne annuelle exacte et la moyenne annuelle estimée. Une méthode d'approximation de la variance temporelle hors des stations fixes, qui repose sur le calcul et la modélisation des variogrammes temporels de ces stations, a été développée. Les résultats obtenus demandent une certaine vigilance ; cette méthode n'est pas ici détaillée.

4.3.3 Optimisation de l'échantillonnage

Dans le cas du NO₂, les corrélations élevées entre les concentrations hivernales et estivales et leur prise en compte par le cokrigeage peuvent être mises à profit pour répartir de façon plus judicieuse les points d'échantillonnage. Ainsi la suppression sur des ensembles disjoints de sites de 30% des tubes en été et de 30% des tubes en hiver ne fait pas perdre significativement en précision (Fouquet, 2004, Gallois, 2004). Pourvu qu'il existe suffisamment de points communs aux deux saisons pour calculer le covariogramme, ce résultat montre que l'on peut :

- conserver une même qualité d'estimation avec un nombre réduit de tubes ;
- améliorer la précision de l'estimation à nombre de tubes égal, en répartissant différemment ces derniers selon la saison et en augmentant ainsi le nombre global de sites de mesure.

5. ETUDE DE CAS : CARTOGRAPHIE DU DIOXYDE D'AZOTE A BOURG-EN-BRESSE

5.1 DEMARCHE THEORIQUE

L'exploitation des corrélations entre les concentrations estivales et hivernales et la réalisation d'un cokrigeage entre moyennes saisonnières s'est révélée une démarche efficace pour améliorer la précision des cartographies de NO₂ (cf. §4.3).

Afin d'évaluer si cette approche est aisément transposable à d'autres contextes géographiques, il a été décidé de l'appliquer à une nouvelle zone d'étude. La cartographie du dioxyde d'azote dans l'agglomération de Bourg-en-Bresse (Ain), où des mesures par tubes passifs ont été conduites durant l'été 2001 et l'hiver 2001/2002, a été retenue comme cas d'application.

Par ailleurs, nous avons à disposition d'autres variables comme l'occupation du sol, l'altitude, le réseau routier et la densité de population.

Le déroulement de l'application d'une méthodologie de traitement géostatistique de ces données est proposé dans les paragraphes qui suivent.

L'objectif a été de comparer deux méthodes permettant de calculer la moyenne annuelle du dioxyde d'azote urbain à Bourg-en-Bresse. La première méthode, présentée et détaillée dans le paragraphe 5.12.1, consiste en une estimation directe de la moyenne annuelle expérimentale par krigeage en dérive externe. La seconde méthode, présentée et détaillée dans le paragraphe 5.12.2, consiste à prendre la moyenne des estimations saisonnières obtenues par cokrigeage en dérive externe.

A l'issue de plusieurs analyses, seule la densité de population a été retenue comme variable auxiliaire, car elle est connue partout sur la grille d'estimation et se révèle raisonnablement corrélée avec les concentrations de dioxyde d'azote.

La méthode de la validation croisée a été utilisée pour choisir les modèles variographiques lors des différentes estimations par krigeage ainsi que pour comparer les résultats des deux méthodes d'estimation de la moyenne annuelle du dioxyde d'azote.

Les graphiques ont été obtenus à l'aide du logiciel Isatis (Géovariances) qui a servi à mener ce travail. Des éléments sur le logiciel Geostatistical Analyst (ESRI), qui a été mis parallèlement en œuvre, sont fournis dans l'annexe 1.

5.2 PRESENTATION DES DONNEES

Afin de cartographier la pollution par l'azote dans l'agglomération de Bourg-en-Bresse (Figure 1), l'association Air de l'Ain et des Pays du Savoie a réalisé huit campagnes de mesures par tubes passifs, quatre pendant l'été 2001 et quatre pendant l'hiver 2001-2002. Les principales statistiques sont affichées dans le Tableau 4.

Les tubes ont été exposés pendant quatre quinzaines consécutives durant chaque saison (été 2001 et hiver 2001-2002), le but étant d'estimer au final une moyenne annuelle.

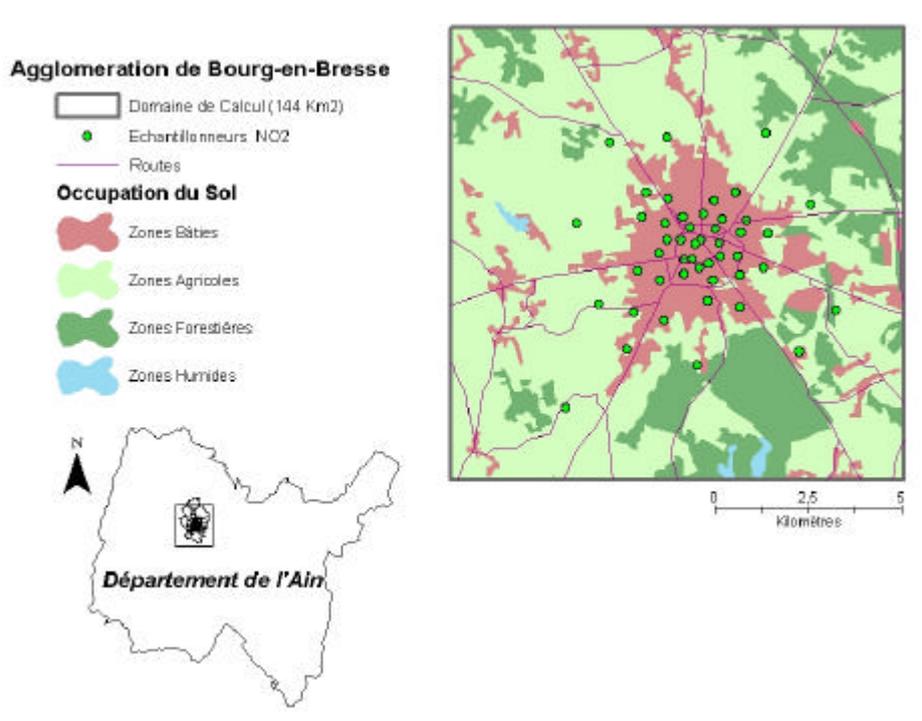


Figure 1 : Localisation géographique des échantillonneurs passifs dans l'agglomération de Bourg-en-Bresse

Dans la Figure 1, on constate que la plupart des tubes sont situés en zone urbaine (34 tubes) les autres tubes (13) sont localisés dans une auréole extérieure à plus de 3 Km du centre ville, ces sites correspondent à des zones boisées ou agricoles. Il y a donc un mélange des types de sites : 72% des tubes sont situés dans des zones urbaines et 28% des tubes dans des zones rurales ou périurbaines.

Pour les tubes exposés durant l'été 2001, les valeurs du dioxyde d'azote varient de 7 à 34 $\mu\text{g}/\text{m}^3$, la moyenne est plus forte pour les première et troisième quinzaines (17,6 et 17,8 $\mu\text{g}/\text{m}^3$). D'autre part, la variance est faible pour la quatrième quinzaine d'été, à cause sans doute d'une faible valeur maximale (22,8 $\mu\text{g}/\text{m}^3$). Les coefficients de variation sont très proches pour les quatre quinzaines (valeurs entre 27% et 30%).

Pour les tubes exposés durant l'hiver 2001-2002, les valeurs du dioxyde d'azote varient de 16,6 à 60,3 $\mu\text{g}/\text{m}^3$, concentrations presque deux fois supérieures à celles de l'été.

La seconde quinzaine hivernale présente une valeur maximale de 60,3 $\mu\text{g}/\text{m}^3$, sa moyenne ainsi que sa variance sont les plus fortes (respectivement 44,2 $\mu\text{g}/\text{m}^3$ et 47 ($\mu\text{g}/\text{m}^3$)²) mais son coefficient de variation (16%) est le plus petit parmi toutes les quinzaines mesurées. En hiver, hormis la seconde quinzaine, les coefficients de variation oscillent entre 18% et 21%, soit près de 10% de moins que pour les mesures estivales.

VARIABLE	Date où commence la période de mesures	Nb	V Minimale	V Maximale	Moyenne	Variance	Coefficient de variation (?/m)
ETE 1	19-06-01	38	8.4	34.4	17.6	24.8	0.28
ETE 2	05-07-01	42	7.7	30.0	13.6	16.2	0.30
ETE 3	19-07-01	46	8.7	32.7	17.8	23.8	0.27
ETE 4	02 au 20-08-01	44	6.8	22.8	11.9	12.0	0.29
HIVER 1	20-12-01	41	19.2	45.7	30.4	29.7	0.18
HIVER 2	08-01-02	45	30.4	60.3	44.2	47.0	0.16
HIVER 3	21-01-02	46	18.6	43.0	29.1	36.9	0.21
HIVER 4	04 au 18-02-02	44	16.6	42.7	27.0	30.4	0.20

Tableau 4 : Statistiques des huit campagnes de mesure du dioxyde d'azote

Avec des concentrations plus faibles, les mesures estivales de dioxyde d'azote présentent une plus grande variation qu'en hiver.

A partir de ces données, les deux moyennes saisonnières et la moyenne annuelle ont été calculées. Ce calcul a été effectué en ne prenant en compte que les sites où les mesures d'au moins deux quinzaines par saison sont disponibles. ***Il s'agit bien d'une moyenne arithmétique des données expérimentales et non pas d'une reconstitution des données.***

Les statistiques des moyennes saisonnières et de la moyenne annuelle sont affichées dans le Tableau 5. Elles confirment que le coefficient de variation est plus fort en été mais que la moyenne et la variance des concentrations sont plus fortes en hiver.

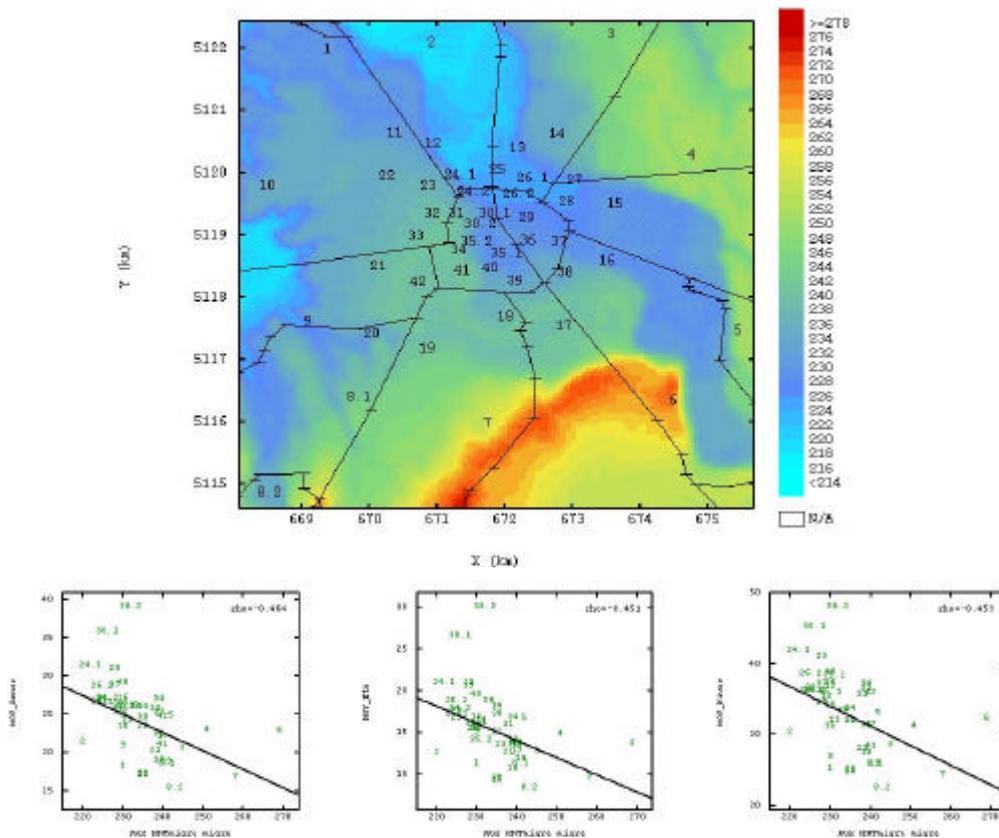
<i>VARIABLE</i>	<i>Nb</i>	<i>V Minimale</i>	<i>V Maximale</i>	<i>Moyenne</i>	<i>Variance</i>	<i>Coefficient de variation (?/m)</i>
Moyenne de l'été 2001	46	7.9	29.6	15.1	16.9	0.27
Moyenne de l'hiver 2001-2002	46	21.9	47.5	32.7	30.5	0.17
Moyenne annuelle	45	16.4	38.6	24.1	23.1	0.20

Tableau 5 : Statistiques des huit campagnes de mesure du dioxyde d'azote

5.3 L'ALTITUDE

La figure 2 présente la carte de l'altitude de l'agglomération de Bourg-en-Bresse et les nuages de corrélation entre l'altitude et les concentrations moyennes saisonnières et annuelle du dioxyde d'azote.

La plupart des tubes sont situés à une altitude comprise entre 220 et 240 mètres ; les nuages de corrélation montrent des coefficients de corrélation faibles (autour de 0.45) et une relation inverse entre l'altitude et la concentration des tubes. Mais ces statistiques sont influencées par les trois tubes localisés en zone boisée au sud du domaine et dont l'altitude est entre 250 et 270 mètres. La variable altitude ne sera pas prise en compte pour la cartographie du dioxyde d'azote car elle ne semble pas influencer de manière déterminante les concentrations.



Dans la carte de l'altitude les lignes représentent le réseau routier et les numéros (ainsi comme dans les nuages de corrélation) correspondent aux échantillonneurs passifs.

En ordonnée : à gauche : moyenne annuelle ; au milieu : moyenne estivale, à droite: moyenne hivernale

En abscisse : altitude migrée au tubes passifs

La ligne droite noire représente la régression linéaire de la concentration en fonction de l'altitude

Figure 2 : Carte de l'Altitude de Bourg-en-Bresse et nuages de corrélation avec les concentrations saisonnières et annuelle du dioxyde d'azote

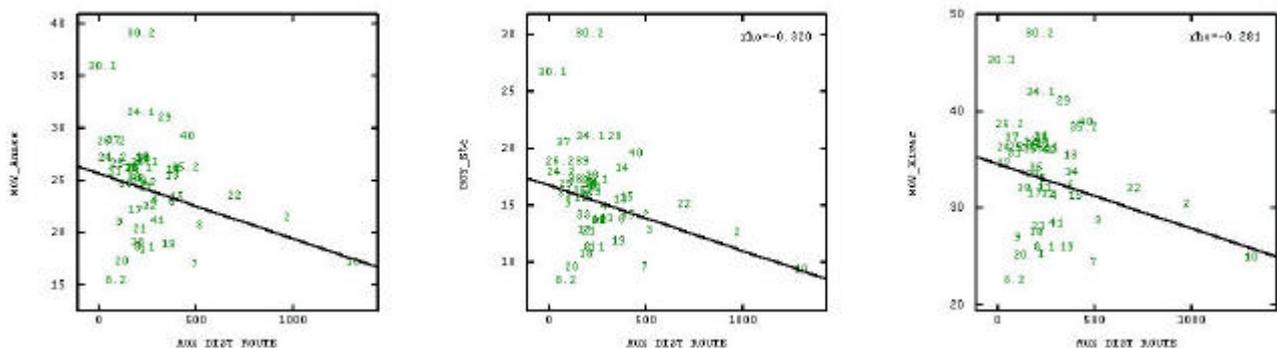
5.4 LA DISTANCE A LA ROUTE

Les gaz d'échappement de véhicules constituent une des principales sources de pollution du dioxyde d'azote. Il est donc intéressant de vérifier la relation qu'il existe entre les concentrations mesurées par les tubes et une variable présentant un lien avec l'impact de la circulation automobile.

Les figures 1 et 2, font apparaître le réseau des principales routes qui desservent l'agglomération, ainsi qu'un périphérique. A l'aide d'un SIG, la distance entre les différents tubes passifs et la route la plus proche de chaque tube a été calculée. La figure 3 montre les nuages de corrélation respectifs, pour les différentes situations temporelles étudiées.

On observe à nouveau des coefficients de corrélation faibles (autour de 0,3) et une relation inverse entre la distance à la route et les concentrations des échantillonneurs. Mais la plupart des tubes se trouvent à une distance de 0 à 500 mètres d'une route principale, et les statistiques sont influencées par 2 tubes se trouvant à plus de 900 mètres d'une route.

On remarque que cette analyse ne prend pas en compte le type de route et que l'on suppose le trafic constant. En effet, une analyse plus précise pourrait être faite en affectant à chaque route un indice comme le TMJA (Trafic Moyen Journalier Annuel en véhicules par jour). Le TMJA donne le nombre journalier de véhicules passant sur un tronçon de voie calculé sur une base annuelle. Dans notre cas, cette information n'était disponible que pour quelques routes et ce type d'analyse n'a pas été mené.



Les numéros correspondent aux échantillonneurs passifs (voir carte de la figure 2)

La ligne droite représente la régression linéaire de la concentration en fonction de la distance à la route

Figure 3 : Nuages de corrélation entre les concentrations saisonnières et annuelle du dioxyde d'azote et la distance à la route la plus proche des tubes passifs

5.5 LA DENSITE DE POPULATION

La densité de population a été calculée sur une grille de 50 mètres de résolution à l'aide d'un SIG, les estimations seront réalisées sur ce maillage.

Les valeurs de densité de population sur la maille ont été calculées à l'aide des données par îlots urbains. Un îlot est une sorte de division administrative des communes qui représente un ou plusieurs quartiers. Les informations disponibles étaient ainsi le nombre d'habitants et la surface en km² des îlots composant l'agglomération de Bourg-en-Bresse.

La Figure 4 montre une vue en perspective. La composante verticale est la densité de population calculée selon la maille de 50 mètres de côté. Le centre ville, à plus forte densité de population, est clairement représenté, ainsi que les zones périphériques où la densité de population est quasiment nulle.

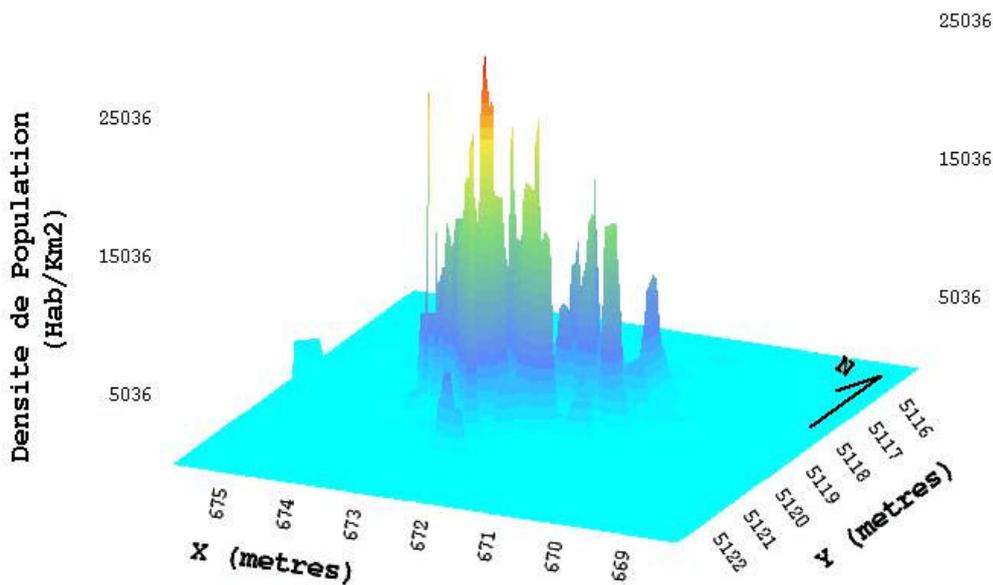


Figure 4 : Vue en perspective de la densité de population de Bourg-en-Bresse

Aux points où sont situés les tubes on a attribué la valeur de densité de population de la maille la plus proche (migration). La Figure 6 montre les histogrammes de la densité de population attribuée (ou migrée) aux tubes et ceux des concentrations saisonnières du dioxyde d'azote à Bourg-en-Bresse; en bleu sont identifiés les tubes où la densité de population est égale à zéro.

Le Tableau 6 indique que les moyennes des concentrations saisonnières et annuelle sont systématiquement plus faibles pour les tubes localisés en périphérie (dont la population est égale à zéro, tableau supérieur) que pour les tubes localisés à l'intérieur de la zone urbaine (tableau inférieur).

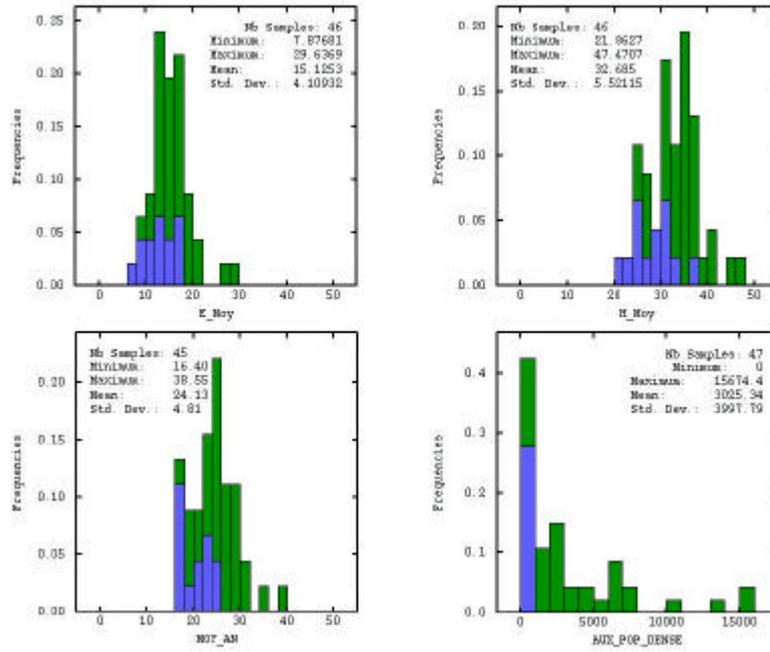
D'après les histogrammes en bleu de la Figure 5, on s'aperçoit que, pour les trois périodes étudiées, les concentrations sont comprises entre les valeurs minimales et les valeurs moyennes (tubes où la densité de population est égale à zéro).

L'histogramme de la densité de population migrée est fortement asymétrique avec une queue de valeurs extrêmes à plus de 10 000 hab/Km². Ces fortes valeurs font que le coefficient de variation est de l'ordre de 132% (pour l'ensemble des valeurs y compris les valeurs nulles).

Considérant ce type d'histogramme, une transformation des données afin de ramener la variable à une distribution plus « normale » s'impose. Nous avons essayé d'utiliser le logarithme de la population afin de réduire l'influence des valeurs extrêmes, mais les coefficients de corrélation avec la concentration du dioxyde d'azote ne se sont pas améliorés. Cette possibilité a donc été écartée.

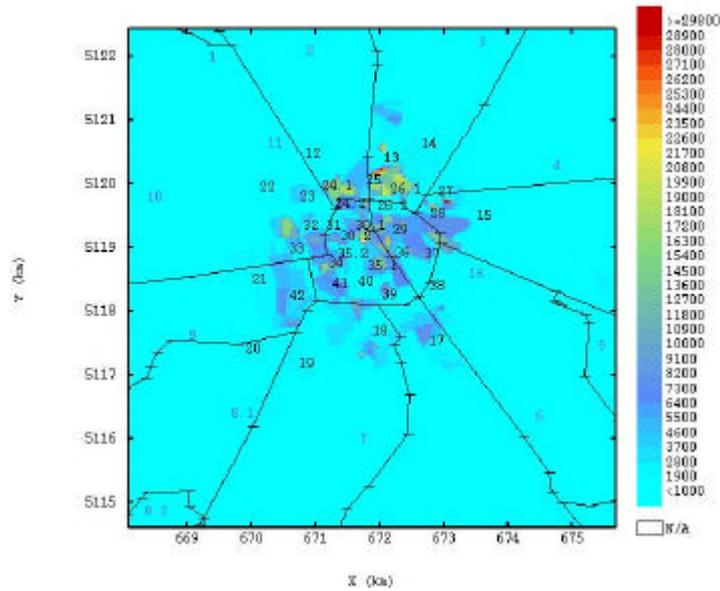
<i>Moyennes saisonnières avec densité de population nulle</i>	<i>Nb</i>	<i>V Minimale</i>	<i>V Maximale</i>	<i>Moyenne</i>	<i>Variance</i>	<i>Coefficient de variation (?/m)</i>
Moyenne de l'été 2001	13	7.9	16.4	12.5	7.9	0.22
Moyenne de l'hiver 2001-2002	13	21.9	36.7	28.2	17.2	0.15
Moyenne annuelle	13	16.4	25.1	20.4	9.6	0.15
<i>Moyennes saisonnières avec densité de population non nulle</i>	<i>Nb</i>	<i>V Minimale</i>	<i>V Maximale</i>	<i>Moyenne</i>	<i>Variance</i>	<i>Coefficient de variation (?/m)</i>
Moyenne de l'été 2001	33	9.1	29.6	16.2	16.8	0.25
Moyenne de l'hiver 2001-2002	33	24.5	47.5	34.4	24.8	0.14
Moyenne annuelle	32	16.8	38.6	25.7	20.5	0.18
Densité de population	34	42.2	15674.4	4182.1	17255632.3	0.99

Tableau 6 : Statistiques des deux sélections des tubes passifs en fonction de la valeur de la densité de population



En bleu tubes dont la densité de population est nulle

Figure 5 : Histogrammes des moyennes saisonnières et annuelle du dioxyde d’azote, et de la densité de population migrée aux tubes



Les numéros correspondent aux échantillonneurs passifs

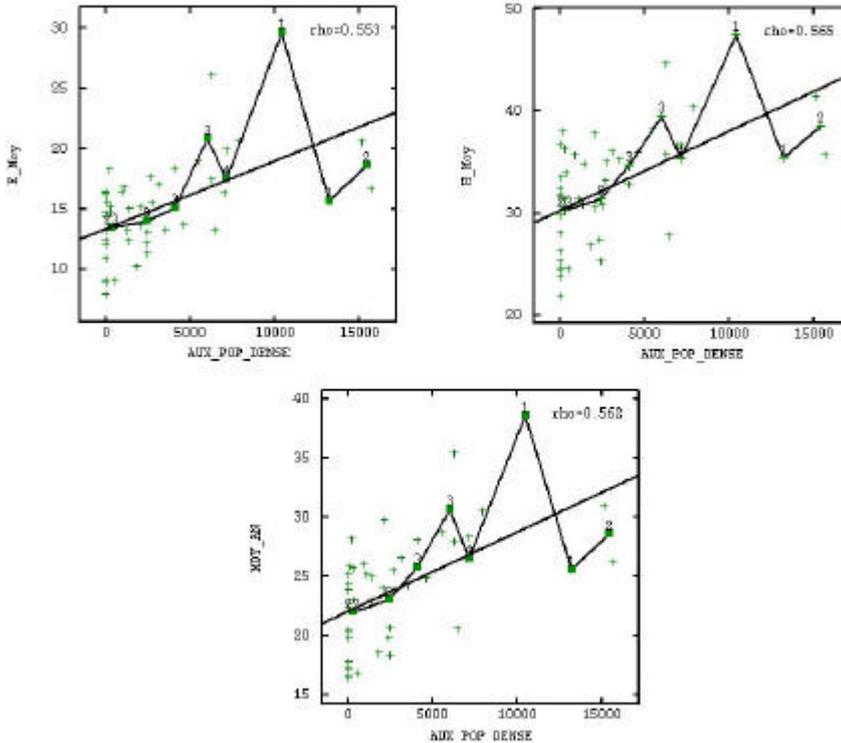
Les lignes représentent le réseau routier

Figure 6 : Carte de densité de population

Les nuages de la Figure 7 montrent la droite de régression linéaire et la courbe liant les moyennes des concentrations saisonnières et annuelles aux classes de valeurs de densité de population; les coefficients de corrélation obtenus ne sont pas très élevés mais ils sont meilleurs que pour les variables secondaires étudiées précédemment. Leur valeur oscille autour de 0.55.

La plupart des concentrations des tubes se trouve dans les deux premières classes de densité de population, c'est-à-dire jusqu'à une valeur de 5 000 hab/Km² environ (Tableau 7). Aux trois dernières classes de densité ne sont associés que quatre tubes passifs. On observe que les zones de plus forte densité de population sont soumises à des concentrations plutôt moyennes pour toutes les saisons.

Remarque : Dans le Tableau 7, pour la première classe, qui comporte 50% des tubes, la densité de population est différente pour les trois périodes, à cause de quelques mesures qui ne sont disponibles que pour une période (par exemple la moyenne annuelle comporte un tube de moins par rapport aux moyennes saisonnières).



Les lignes brisées représentent la moyenne par classe de densité de population, les numéros indiquent le nombre de tubes pris en compte dans le calcul de chaque moyenne (voir Tableau 8)

La ligne droite représente la régression linéaire de la concentration en fonction de la densité de population (voir Tableau 8)

Figure 7 : Nuages de corrélation entre les moyennes saisonnières et annuelle du dioxyde d’azote et la densité de population

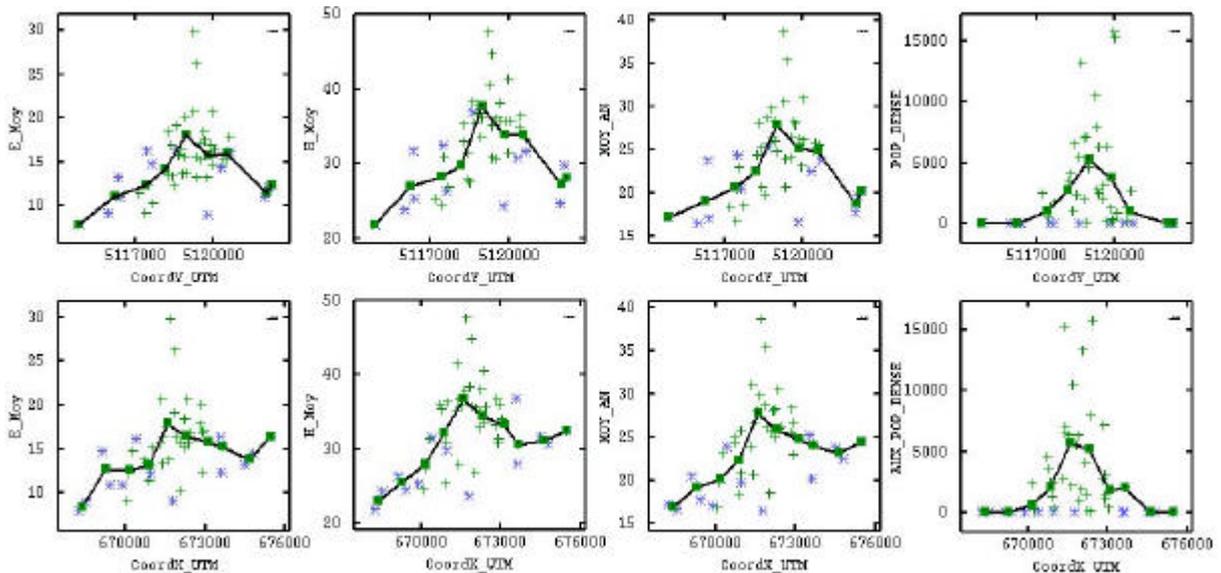
Nb de tubes par classe	Densité de population	Saison		
		Eté	Hiver	Année
23 (Année* : 22)	310 (été); 270 (hiver); 266 (Année)	13.5	30.3	22.0
9	2 401	14.0	31.4	23.0
3	4 054	15.1	34.7	25.7
3	6 010	20.9	39.5	30.6
4	7 114	17.5	35.4	26.5
1	10 434	29.6	47.5	38.6
1	13 199	15.7	35.4	25.6
2	15 414	18.7	38.5	28.6
Moyenne :	3025	15.1	32.7	24.1

Tableau 7 : Moyennes des concentrations par classe de densité de population

<i>Période</i>	<i>Coefficient de corrélation (?)</i>	<i>Pente de la droite de régression linéaire (B)</i>	<i>Ordonnée à l'origine (A)</i>
Eté	0.55	0.000565	13.38
Hiver	0.565	0.00077	30.315
Année	0.56	0.00067	22.048

Tableau 8 : Paramètres de la droite de régression entre les moyennes saisonnières du dioxyde d'azote et la densité de population

La Figure 8 montre les nuages de corrélation entre les concentrations de dioxyde d'azote et la longitude en haut (coordonnée Y, UTM) et la latitude en bas (coordonnée X, UTM). Ces nuages permettent de vérifier la tendance (ou dérive) des concentrations, dans ce cas elles révèlent la présence d'une non stationnarité qui se traduit par une augmentation de la concentration du dioxyde d'azote lorsque l'on se rapproche du centre ville (localisé à peu près au milieu du domaine).



En haut l'abscisse est la latitude, en bas l'abscisse est la longitude.

Les ordonnées sont dans un ordre de gauche à droite: moyenne estivale, moyenne hivernale, moyenne annuelle et densité de population

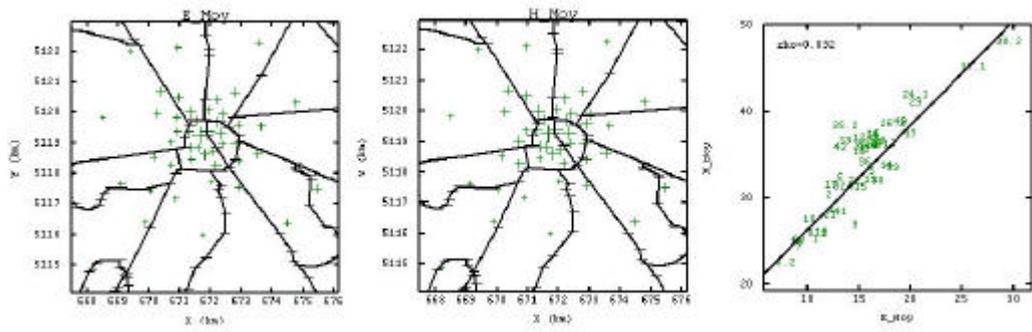
Les astérisques bleues représentent les tubes où la densité de population est nulle

Les lignes brisées représentent la moyenne de la concentration par classes de coordonnées

Figure 8 : Nuages de corrélation entre les concentrations saisonnières et annuelle du dioxyde d'azote et la densité de population le long des axes coordonnées

5.6 RELATION ENTRE MOYENNES SAISONNIERES DU DIOXYDE D'AZOTE

Dans les deux cartes de la Figure 9, les croix représentent la concentration du dioxyde d'azote pour chaque saison. Tube par tube, les tailles des croix dans les deux cartes sont très proches, en conséquence les concentrations sont proportionnelles. Cette observation est confirmée dans le nuage de corrélation où l'on repère une relation linéaire entre les concentrations saisonnières, avec un fort coefficient de corrélation (0,9).



A gauche les croix correspondent à la moyenne de la concentration estivale, à droite à celle de la concentration hivernale.

La taille des croix est proportionnelle à la concentration (l'échelle de proportionnalité n'est pas exactement la même d'une saison à l'autre).

Les numéros dans le nuage de corrélation correspondent aux échantillonneurs passifs (voir carte des figures 2 et 6).

Figure 9 : Comparaison entre les moyennes saisonnières

5.7 ANALYSE VARIOGRAPHIQUE DE L'HIVER

La Figure 10 montre la carte variographique des concentrations hivernales ; cet outil nous permet d'examiner la corrélation spatiale des données dans toutes les directions de l'espace.

Une cellule de la carte variographique représente une famille de couples de points dont les directions et les distances entre points sont égales. La valeur de la cellule représente la valeur du variogramme pour cette famille de paires.

Si la corrélation spatiale est la même dans toutes les directions de l'espace (isotropie), on obtiendrait la même couleur pour toutes les cellules à une même distance (cercle) du centre, au contraire si une direction est privilégié (anisotropie) on obtiendrait des valeurs plus fortes ou plus faibles dans la direction concernée.

La carte variographique montre des valeurs de variogrammes à peu près constantes (autour de 20 $(\mu\text{g}/\text{m}^3)^2$) dans toutes les directions et à toutes les distances sauf dans la direction N20°E où on observe un variogramme d'environ 50 $(\mu\text{g}/\text{m}^3)^2$ à une distance d'environ 1,8 Km (le pas utilisé pour le calcul des variogrammes est de 300 m avec une tolérance de 50%).

Nous avons donc calculé deux variogrammes directionnels avec une tolérance angulaire de 90° chacun, de façon à couvrir l'ensemble du domaine, les variogrammes sont affichés dans la partie gauche de la Figure 10, les directions sont N30°E (+/- 45°) et N30°W (+/- 45°).

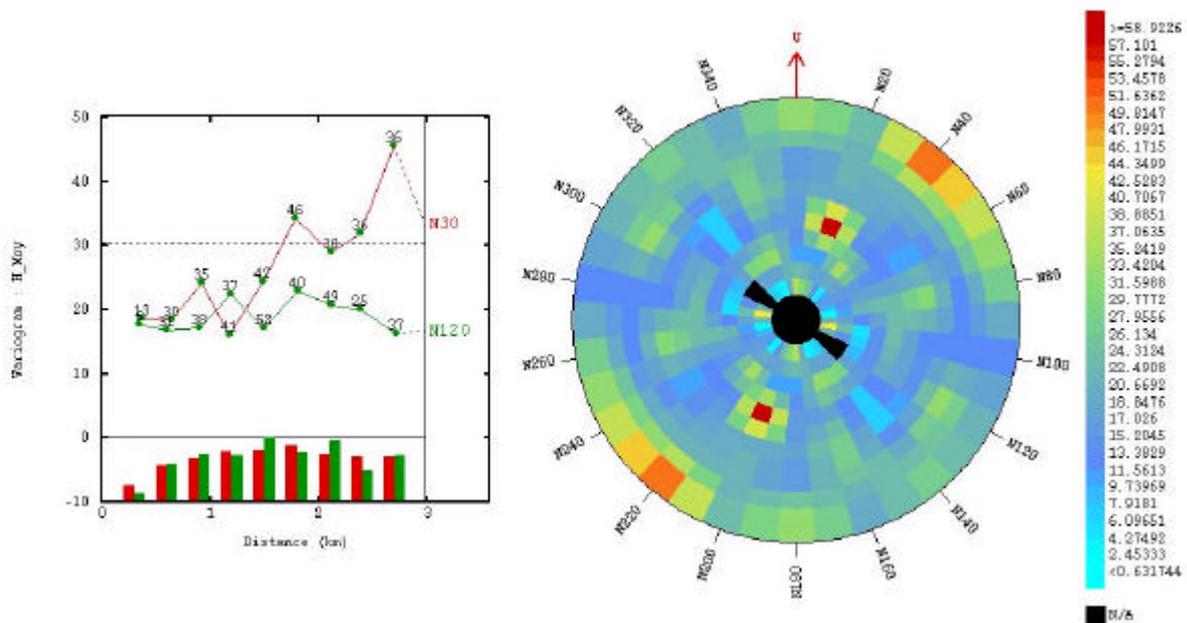


Figure 10 : Variogrammes directionnels et carte variographique des concentrations hivernales

Jusqu'à une distance de 1,5 Km le phénomène est isotrope et sans structure spatiale avec une variation presque constante autour de 20 ($\mu\text{g}/\text{m}^3$)².

Après 1,5 Km la direction N30°E s'écarte de la direction N30°W et montre une non stationnarité avec un variogramme qui croit avec la distance.

Ces observations sont vérifiées dans le variogramme omnidirectionnel de la Figure 11, qui présente une valeur constante autour de 20 ($\mu\text{g}/\text{m}^3$)² pour les distances inférieures à 1,5 km, puis une croissance tendant vers la variance des données pour une distance supérieure à 2,5 Km.

Le manque de structure spatiale pour une variable comme le dioxyde d'azote est sans doute dû aux facteurs locaux qui influencent la mesure des tubes, par exemple la typologie des rues et le type de quartiers où sont localisés les tubes « urbains ».

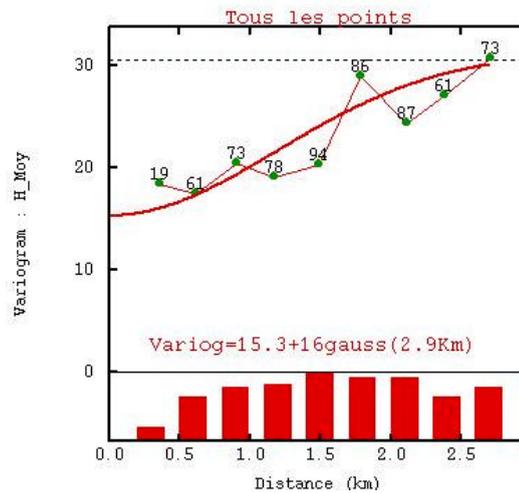


Figure 11 : Variogramme omnidirectionnel de l'hiver

En géostatistique, une méthode d'estimation non stationnaire bien connue est le krigeage avec dérive externe. Dans ce type d'approche l'estimation se fait à l'aide d'une fonction de forme connue à peu près partout dans le domaine spatial et censée mesurer le même phénomène mais d'une façon plus approximative.

On décompose le phénomène en deux parties, la première représentée par la dérive et la deuxième appelée résidu. Dans notre cas, la densité de population sera prise pour dérive externe car elle représente indirectement les sources de pollution et est corrélée avec la concentration de dioxyde d'azote.

Si l'on décide d'effectuer un krigeage en dérive externe, il convient de choisir le modèle de covariance (ou variogramme) adéquat. En théorie il faut modéliser la structure des résidus. Or, en pratique, les résidus vrais ne sont pas accessibles. Il existe alors deux manières de procéder. La première approche est de se placer dans le cadre du modèle de fonctions aléatoires intrinsèques d'ordre k (FAI-K). La covariance y est appelée covariance généralisée.

La calcul de la covariance généralisée est automatisé dans le logiciel ISATIS. Le processus consiste à proposer différentes structures dont les paliers sont ajustés numériquement dans un voisinage local puis, par validation croisée, à choisir la combinaison de structures qui donne la variance des erreurs réduites la plus proche de l'unité (cas où la variance de krigeage est égale à la variance expérimentale des données prises lors de la validation).

Une deuxième façon de déduire cette covariance de résidus est de «supposer» qu'à certaines directions et à une certaine échelle la dérive s'annule; dans ce cas la covariance des résidus est égale à la covariance de la variable. L'analyse variographique doit permettre de confirmer cette «supposition».

Nous avons vu qu'il existe une dérive car la concentration du dioxyde d'azote augmente lorsqu'on s'approche du centre ville, mais la Figure 11 montre qu'il existe aussi une certaine stationnarité pour les petites distances, qui peut être modélisée par un effet de pépite et une structure gaussienne de portée 2.9 Km.

Pour choisir le modèle, trois structures ont été comparées par validation croisée. La première est le modèle ajusté dans la Figure 11, la seconde est le modèle ajusté dans la Figure 11 en prenant en compte la densité de population comme dérive externe, la troisième est le modèle ajusté automatiquement dans le cadre des FAI-K (effet de pépite de 5.5 et structure sphérique de palier 11 et de portée 2.4 Km).

La somme des paliers du modèle des résidus (sphérique) est de 16.5 et celle du modèle gaussien est de 31.3, soit presque le double. Les résultats des validations croisées sont indiqués dans le Tableau 9, et les nuages de corrélation entre valeurs vraies et valeurs estimées dans la Figure 12.

Modèle	Erreur		Erreur relative	Erreur réduite	Coefficient de corrélation
	Moyenne	Variance	Variance	Variance	
Gaussien sans dérive externe	0.16	11.82	0.012	0.6	0.790
Gaussien avec dérive externe	0.17	12.04	0.010	0.59	0.780
Sphérique avec dérive externe	0.22	12.62	0.013	1.12	0.77

Tableau 9 : Validation croisée des modèles pour l'azote hivernale

Les résultats de la validation croisée indiquent que la prise en compte de la densité de population n'améliore pas l'estimation aux points de mesures. Les tubes sont suffisamment rapprochés (surtout en centre ville) pour que le modèle sans dérive externe fournisse une estimation acceptable ; en revanche, la variance de l'erreur réduite inférieure à 1 unité laisse penser qu'une estimation sans dérive externe donnera une variance d'estimation non représentative de l'erreur réelle.

Nous avons privilégié le modèle des résidus composé par un effet de pépité et une structure sphérique, et le voisinage unique au moment de l'estimation a été conservé. Ce modèle donne des résultats corrects, coefficient de corrélation 0,77 et variance de l'erreur réduite proche de 1 unité (1,12).

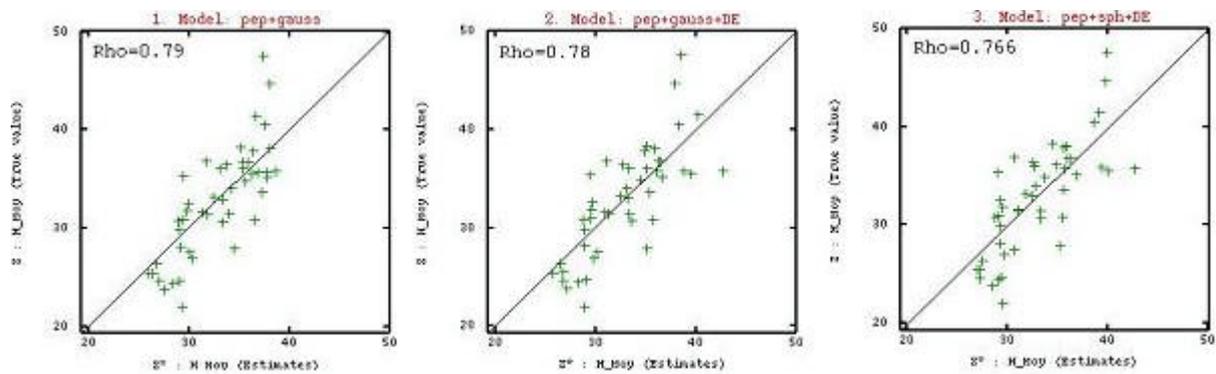


Figure 12 : Nuages des corrélation valeurs vraies - valeurs estimées, modèles hivernales

5.8 ESTIMATIONS DE L'HIVER

Deux estimations sont ici présentées pour la concentration hivernale du dioxyde d'azote. La carte de gauche de la Figure 13 est une interpolation entre les concentrations qui ne tient compte d'aucune information auxiliaire; la carte de droite montre les niveaux de concentration en prenant en compte la densité de population.

L'apport de l'information auxiliaire permet donc d'affiner la cartographie : alors que la carte sans dérive externe ne fournit pas de représentation détaillée des niveaux de concentration dans le centre ville, la carte avec dérive externe montre des variations plus réalistes de la concentration lorsque les quartiers sont plus ou moins peuplés.

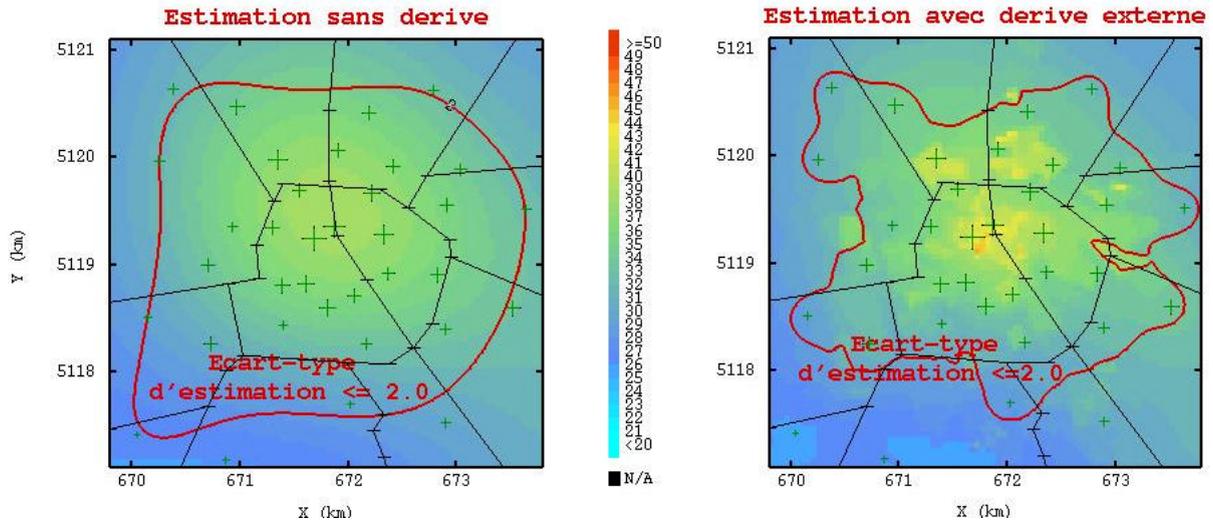


Figure 13 : Cartes des estimations de la moyenne hivernale

Sur ces cartes nous avons effectué une sélection des zones dans lesquelles les résultats sont plus fiables, ces zones correspondent aux estimations dont l'écart-type est plus faible. Un écart-type de $2 \mu\text{g}/\text{m}^3$ a été pris arbitrairement comme seuil, il est matérialisé par la courbe rouge dessinée sur les cartes de la Figure 13.

Les zones sélectionnées ont à peu près les mêmes caractéristiques, elles correspondent aux mesures urbaines, là où les prélèvements ont été plus denses.

Le Tableau 10 montre les statistiques de l'ensemble du domaine et des zones sélectionnées pour chaque type d'estimation. Les valeurs représentent la moyenne des concentrations sur une surface de 2500 m^2 .

Les zones sélectionnées représentent 16.2% des blocs pour l'estimation sans dérive externe et 12.8% des blocs pour l'estimation avec dérive externe. La moyenne sur les sélections est légèrement supérieure à la moyenne sur l'ensemble du domaine mais plus proche de la moyenne expérimentale des données ($30.5 \mu\text{g}/\text{m}^3$).

Les valeurs extrêmes des concentrations estimées sont plus petites que celles des données expérimentales. Ce résultat est dû à l'effet de lissage de l'interpolation et à l'effet de support, car les estimations représentent une surface plus grande (il y a donc plus de chance de retrouver de valeurs intermédiaires).

La variance ponctuelle qui était de $30.5 (\mu\text{g}/\text{m}^3)^2$ est passée à un peu plus de $8.0 (\mu\text{g}/\text{m}^3)^2$ pour les mailles de 250 m^2 .

Model	Nb mailles	V Minimale	V Maximale	Moyenne	Variance	Coefficient de variation (?/m)
Gaussien sans dérive externe	23 864	25.15	38.66	29.41	8.82	0.1
Sphérique avec dérive externe		24.3	45.52	29.46	8.24	0.1
Ecart- type d'estimation <= 2.0		16.2% et 12.8% des mailles				
Gaussien sans dérive externe	3 859	26.39	38.66	33.89	8.35	0.09
Sphérique avec dérive externe	3 068	27.32	44.40	34.26	8.24	0.08

Tableau 10 : Statistiques des estimations du NO2 hivernale

Les cartes des écarts- types, de la Figure 14, ont permis de sélectionner les zones où les estimations sont les plus précises. En valeurs numériques (Tableau 11) les écart- types sont très proches, mais pour l'estimation avec dérive externe une carte plus détaillée des erreurs est obtenue.

Model	Nb mailles	V Minimale	V Maximale	Moyenne	Ecart -type	Coefficient de variation (?/m)
Gaussien sans dérive externe	23864	1.15	4.21	2.75	0.69	0.25
Sphérique avec dérive externe		1.35	3.7	2.63	0.48	0.18

Tableau 11 : Statistiques des écarts- types d'estimation du NO2 hivernale

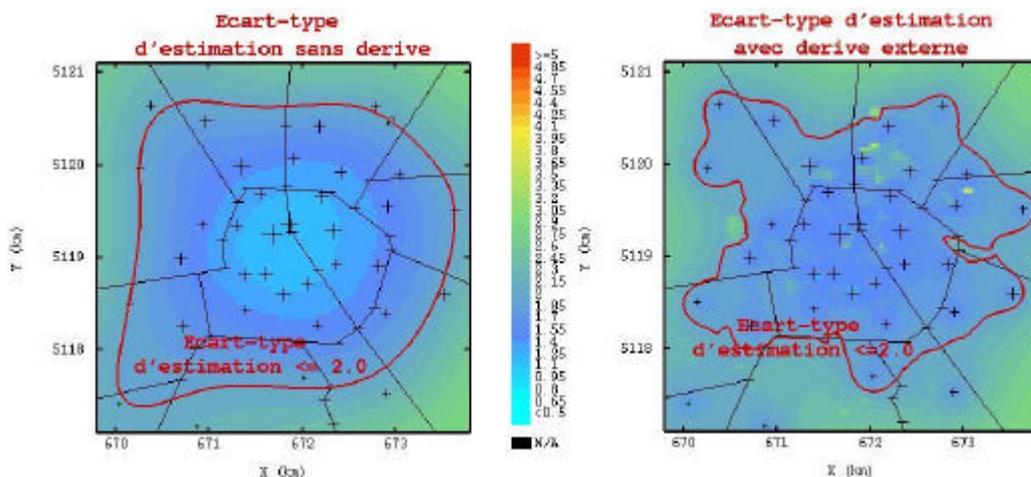


Figure 14 : Carte des écarts- types d'estimation d'une moyenne hivernale

5.9 ANALYSE VARIOGRAPHIQUE DE L'ETE

En comparaison avec l'hiver, les mesures estivales présentent un variogramme expérimental moins structuré. L'influence des deux fortes valeurs du centre ville est plus marquée (astérisques bleu de la nuée variographique de la Figure 15).

Si l'on retire ces deux valeurs, le phénomène devient plus structuré et donc plus facile à modéliser. Le variogramme omnidirectionnel des mesures estivales sans les deux valeurs maximales (courbe droite de la Figure 15) ressemble au variogramme expérimental des mesures hivernales (Figure 11).

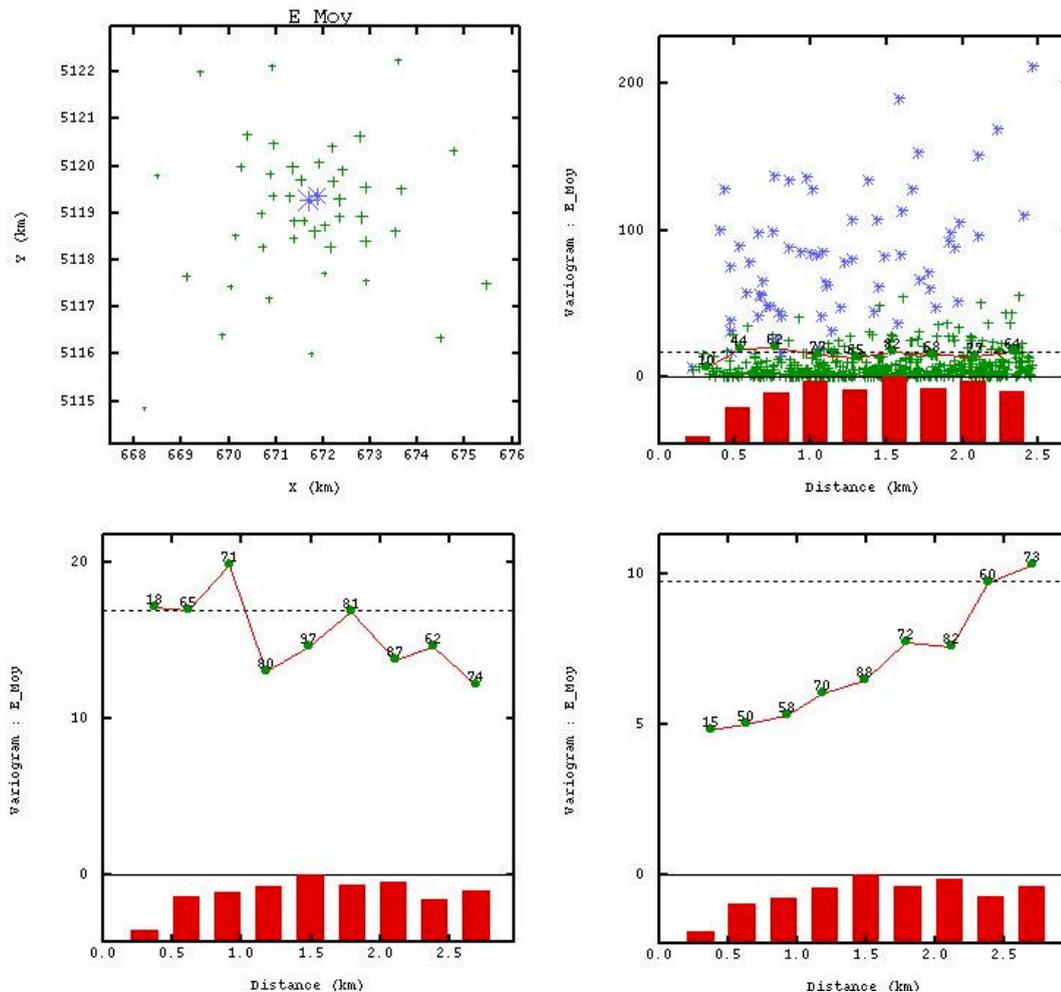


Figure 15 : variogrammes omnidirectionnels pour le NO2 estivale (avec et sans 2 points extrêmes)

La moyenne estivale présente une légère anisotropie dans la direction de N20°E (Figure 16). Cette anisotropie n'a toutefois pas été prise en compte dans les modélisations, car on considère que la dérive externe doit permettre de la représenter.

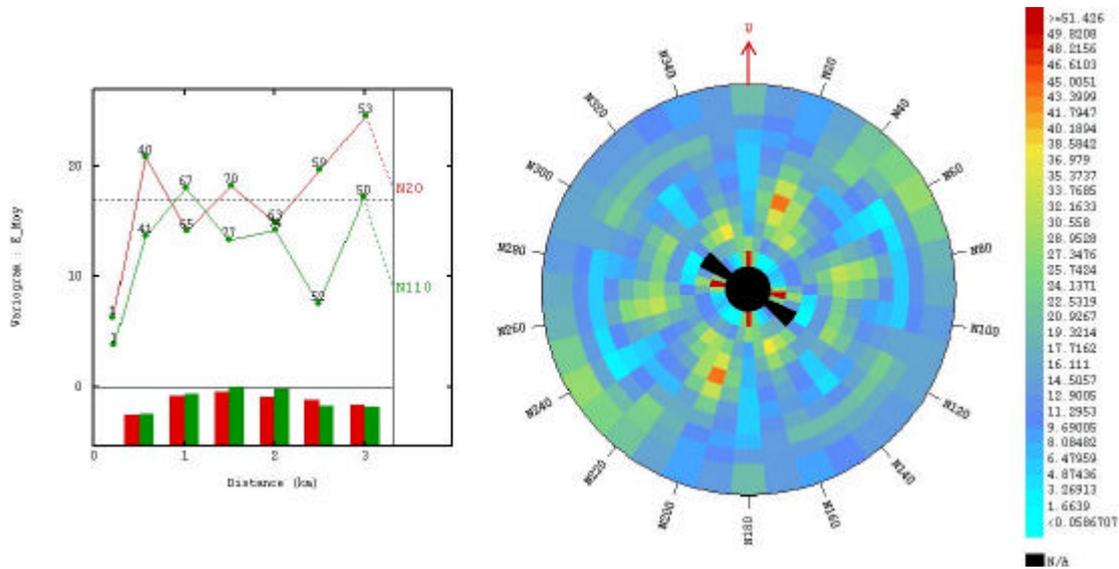


Figure 16 : variogrammes directionnels pour le NO2 estivale

Comme dans le cas des mesures hivernales, plusieurs modèles de covariance ont été testés par validation croisée, dans la perspective d'une estimation par krigeage en dérive externe.

Les modèles testés sont ceux montrés Figure 17. La courbe de gauche de cette figure dessine un phénomène presque aléatoire : l'effet de pépite représente près de 70% de la variance des données. La courbe de droite représente un phénomène plus structuré modélisé par un effet de pépite et une structure gaussienne de portée 2.93 Km.

La somme des paliers du dernier modèle est égale à 10, soit 60% de la variance des données ($16.9 (\mu\text{g}/\text{m}^3)^2$). On suppose donc que la dérive externe représente la partie de la variation du phénomène qui n'est pas prise en compte par le modèle lui-même.

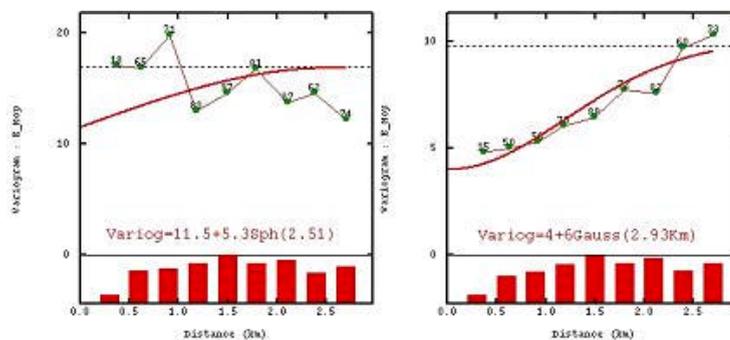


Figure 17 : Deux modèles de variogrammes testés

Trois validations ont été effectuées. La première avec le modèle sphérique sans dérive externe, la seconde avec le modèle gaussien avec dérive externe (Figure 18, Tableau 12), et la troisième avec le modèle sphérique avec dérive externe. Les résultats des validations sont très proches, le coefficient de corrélation valeurs estimées - valeurs vraies oscille entre 0.63 et 0.66. Une fois de plus, l'apport de la dérive externe n'améliore pas les estimations.

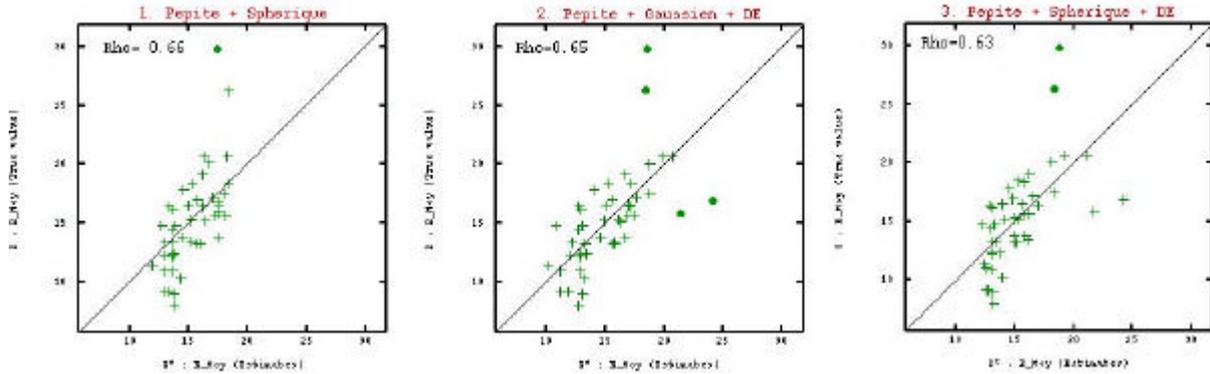


Figure 18 : Validation croisée des trois modèles de variogrammes choisis

Modèle			Erreur		Erreur relative	Erreur réduite	Coefficient de corrélation.
			Moyenne	Variance	Variance	Variance	
Sphérique externe	sans	dérive	0.09	10.3	0.048	0.70	0.66
Gaussien avec dérive externe			0.13	9.9	0.041	1.84	0.65
Sphérique externe	avec	dérive	0.10	10.3	0.046	1.42	0.63

Tableau 12 : Validation croisée plusieurs modèles de dérive externe, moyenne estivale

Le modèle choisi pour comparer les cartes avec et sans dérive externe est le modèle sphérique, lequel présente des statistiques légèrement meilleures (moyenne des erreurs plus proche de zéro et variance de l'erreur réduite plus proche de 1).

5.10 ESTIMATIONS DE L'ETE

Pour chaque type d'estimation, une sélection des concentrations les plus précises a été effectuée. Le critère de l'écart- type d'estimation a été encore utilisé.

Les écarts- types d'estimation (Figure 20 et Tableau 14) sont plus importants pour le modèle sans dérive externe. Lorsque la dérive externe est introduite dans le système de krigeage, on obtient une diminution de l'écart- type d'estimation. Il y a donc une amélioration en dehors des zones non échantillonnées et là où la densité de population est connue.

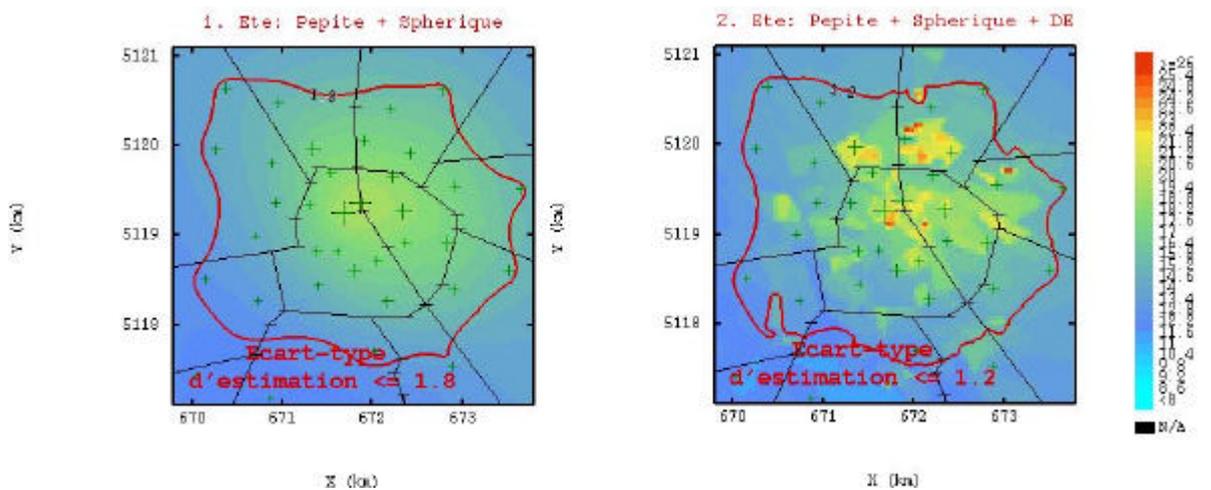


Figure 19 : Cartes d'estimation de la moyenne estivale

Dans les cartes de la Figure 19, les sélections des concentrations sont matérialisées par la courbe rouge. Pour les estimations sans dérive externe nous avons appliqué comme seuil (choisi arbitrairement) un écart- type de $1.8 \mu\text{g}/\text{m}^3$. Pour les estimations avec dérive externe, un écart- type de $1.2 \mu\text{g}/\text{m}^3$ a été adopté comme seuil.

Les zones sélectionnées sont à peu près les mêmes (environ 15% des mailles de la grille), leurs moyennes sont plus proches de la moyenne des données expérimentales ($16.9 \mu\text{g}/\text{m}^3$) mais leurs variances sont beaucoup plus faibles en raison de l'effet de lissage de l'interpolation et de l'effet de support.

L'intérêt de la dérive externe est de fournir une carte plus détaillée qui tient compte du tissu urbain représenté par la densité de population. De plus, le cokrigeage entre moyennes saisonnières permet d'augmenter la précision des estimations.

Modèle	Nb mailles	Minimale	Maximale	Moyenne	Variance	Coefficient de variation (?/m)
Sphérique sans dérive externe	23 864	11.73	19.49	13.61	1.6	0.09
Sphérique avec dérive externe		11.2	26.93	13.29	2.2	0.11
Ecart- type <=1.8		15.4% des mailles				
Sphérique sans dérive externe	3 666	12.16	19.49	15.69	2.7	0.11
Ecart- type <=1.2		13.8% des mailles				
Sphérique avec dérive externe	3 293	11.79	22.11	15.36	3.3	0.12

Tableau 13 : Statistiques des estimations de la moyenne estivale

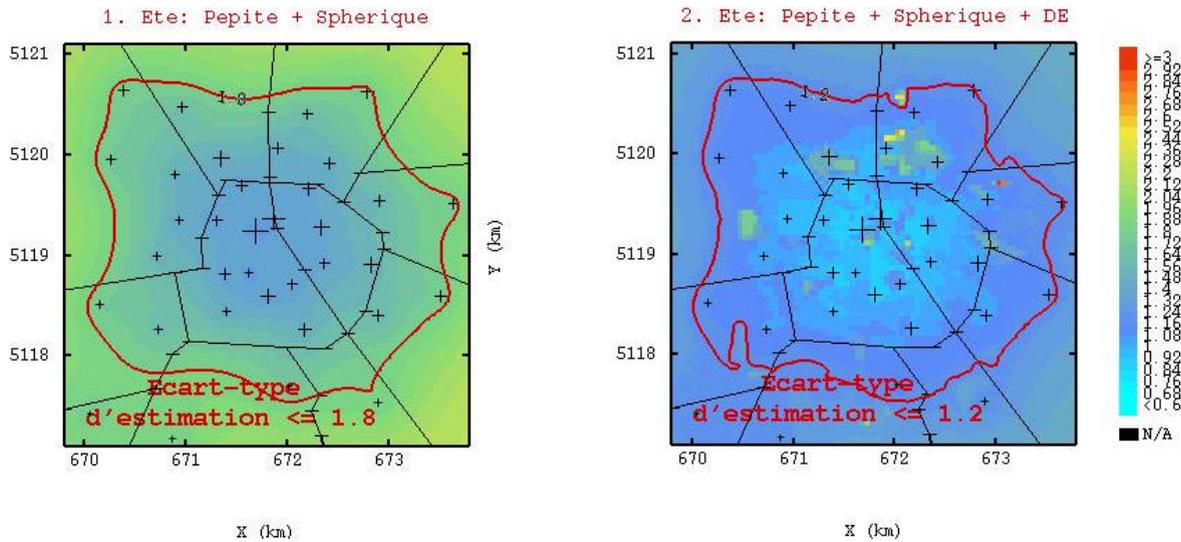


Figure 20 : Cartes des écarts-types des estimations de la moyenne estivale

Modèle	Nb mailles	Minimale	Maximale	Moyenne	Ecart- type	Coefficient de variation (?/m)
Sphérique sans dérive externe	23864	1.3	2.45	2.09	0.26	0.13
Sphérique avec dérive externe		0.87	2.94	1.38	0.15	0.11

Tableau 14 : Statistiques des écarts-types des estimations de la moyenne estivale

5.11 LE COKRIGEAGE ENTRE MOYENNES SAISONNIERES

5.11.1 Cokrigeage Ordinaire

Le principe du cokrigeage est le même que celui du krigeage, mais il permet en plus d'effectuer une estimation simultanée de plusieurs variables régionalisées.

Deux principales raisons justifient le cokrigeage :

- ? Profiter, pour estimer une variable sous-échantillonnée, de l'information apportée par une autre variable, liée à la première, et mieux échantillonnée.
- ? Dans le cas où toutes les variables sont connues aux points expérimentaux, améliorer la précision et la cohérence des résultats d'estimation en prenant en compte les liens entre variables.

Etant donné un ensemble de N variables indicées i. L'estimateur par cokrigeage ordinaire de la variable z_i en un point cible s'écrit comme une combinaison linéaire pondérée des mesures des N variables situées dans le voisinage de ce point:

$$Z_{i_0}^*(x_0) = \sum_{i=1}^N \lambda_i^i Z_i(x_i)$$

Le nombre d'échantillons n_i dépend de l'indice i des variables afin d'inclure dans la notation le cas de l'hétérotopie (une partie ou l'ensemble des points de mesures n'est pas commun aux variables).

Si l'on ne compte deux variables (Z_1 et Z_2), leurs estimateurs respectifs par cokrigeage sont les combinaisons linéaire suivantes:

$$Z_{1_0}^*(x_0) = \sum_{i=1}^{n_1} \lambda_i^1 Z_1(x_i) + \sum_{i=1}^{n_2} \lambda_i^2 Z_2(x_i)$$

$$Z_{2_0}^*(x_0) = \sum_{i=1}^{n_1} \lambda_i^1 Z_1(x_i) + \sum_{i=1}^{n_2} \lambda_i^2 Z_2(x_i)$$

Dans le cadre d'une hypothèse de stationnarité intrinsèque conjointe, on désire estimer une variable particulière d'un ensemble de N variables, en se basant sur une erreur d'estimation nulle en moyenne. Cette condition est remplie en choisissant des pondérateurs de somme unité pour la variable d'intérêt, et de somme nulle pour les variables auxiliaires :

$$\sum_{j=1}^{n_i} \lambda_{ij}^1 \delta_{ij} = \begin{cases} 1 & \text{si } i = i_0 \\ 0 & \text{si } i \neq i_0 \end{cases}$$

Le système de cokrigeage est établi de façon à minimiser la variance de l'erreur d'estimation. Comme dans le cas du krigeage, les contraintes sur les pondérateurs font appel à N paramètres de Lagrange μ_i . Dans le cas de deux variables, les équations de cokrigeage sont les suivantes :

Système de krigeage ordinaire de la variable Z₁:

$$\begin{aligned} \sum_{j=1}^{n_1} \lambda_{1j}^1 \delta_{1j}(x_j - x_j) + \sum_{j=1}^{n_2} \lambda_{2j}^2 \delta_{2j}(x_j - x_j) + \lambda_{1o} \delta_{1o}(x_o - x_o) & \text{ pour } i = 1, \dots, n_i \\ \sum_{j=1}^{n_1} \lambda_{1j}^1 & = 1 \\ \sum_{j=1}^{n_2} \lambda_{2j}^2 & = 0 \end{aligned}$$

Variance de cokrigeage de la variable Z₁:

$$\sigma_{CK^1}^2 = \sum_{j=1}^{n_1} \lambda_{1j}^1 \delta_{1j}(x_j - x_o) + \sum_{j=1}^{n_2} \lambda_{2j}^2 \delta_{2j}(x_j - x_o) + \lambda_{1o} \delta_{1o}(x_o - x_o)$$

Système de krigeage ordinaire de la variable Z₂:

$$\begin{aligned} \sum_{j=1}^{n_1} \lambda_{1j}^1 \delta_{1j}(x_j - x_j) + \sum_{j=1}^{n_2} \lambda_{2j}^2 \delta_{2j}(x_j - x_j) + \lambda_{2o} \delta_{2o}(x_o - x_o) & \text{ pour } i = 1, \dots, n_i \\ \sum_{j=1}^{n_1} \lambda_{1j}^1 & = 1 \\ \sum_{j=1}^{n_2} \lambda_{2j}^2 & = 0 \end{aligned}$$

Variance de cokrigeage de la variable Z₂ est :

$$\sigma_{CK^2}^2 = \sum_{j=1}^{n_1} \lambda_{1j}^1 \delta_{1j}(x_j - x_o) + \sum_{j=1}^{n_2} \lambda_{2j}^2 \delta_{2j}(x_j - x_o) + \lambda_{2o} \delta_{2o}(x_o - x_o)$$

Dans toutes ces formules apparaît un terme additionnel qui est le variogramme croisé entre les variables, ce variogramme est calculé expérimentalement par l'expression suivante :

$$\gamma_{1,2}(h) = \frac{1}{2 * n_h} \sum_{z_1, z_2} \sum_{x_1, x_2} (z_1(x_1) - z_2(x_2))^2$$

avec $\gamma_{1,2}(h) = \gamma_{1,2}(h)$

Si Z1 et Z2 étaient indépendantes, ce variogramme croisé serait identiquement nul. A la différence des variogrammes simples, le variogramme croisé entre deux variables peut prendre des valeurs négatives, une augmentation de l'une des variables allant alors de pair avec une diminution de l'autre.

5.11.2 Ajustement des variogrammes simples et croisés

L'ajustement des variogrammes (ou covariances) simples et croisés doit être mathématiquement cohérent, de façon à garantir des variances positives. Les modèles linéaires de corégionalisation sont les plus couramment employés.

Dans ces modèles, les variables se décomposent linéairement en composantes élémentaires, et leurs structures simples et croisées sont donc des combinaisons linéaires des schémas élémentaires.

Les composantes propres à chacune des variables disparaissent des structures croisées. Inversement, un schéma élémentaire donné (par exemple sphérique de portée donnée) ne peut apparaître dans la structure croisée entre deux variables s'il n'est pas présent dans leurs structures simples.

Une fois ajusté, un tel modèle permet le cokrigeage des variables. Dans notre cas il s'agit de tirer parti de la bonne corrélation qui existe entre les moyennes saisonnières (dans la section 5.6 une corrélation de 0.9 a été mise en évidence).

La Figure 21 présente les variogrammes simples et croisé des moyennes saisonnières ; grâce au modèle linéaire de corégionalisation, un modèle cohérent a pu être ajusté à ces variogrammes. Il est composé d'un effet de pépite et d'une structure gaussienne de portée 2.51 Km, la matrice des paliers est présentée dans le Tableau 15.

Ce modèle a un effet de pépite très prononcé, surtout pour la moyenne estivale. Comme on l'a vu dans la section 5.9, cette variable est moins bien structurée que la moyenne hivernale.

Après avoir ajusté le modèle de covariance, on peut effectuer le cokrigage de chaque variable, puis considérer la densité de population comme dérive externe, de façon à obtenir une cartographie plus détaillée des niveaux de pollution en centre ville.

Effet de Pépité	Moyenne estivale	Moyenne hivernale
Moyenne estivale	10	10
Moyenne hivernale	10	16

Structure Gaussienne (Portée=2.51 Km)	Moyenne estivale	Moyenne hivernale
Moyenne estivale	7	10
Moyenne hivernale	10	15

Tableau 15 : Paliers du modèle de corégionalisation linéaire

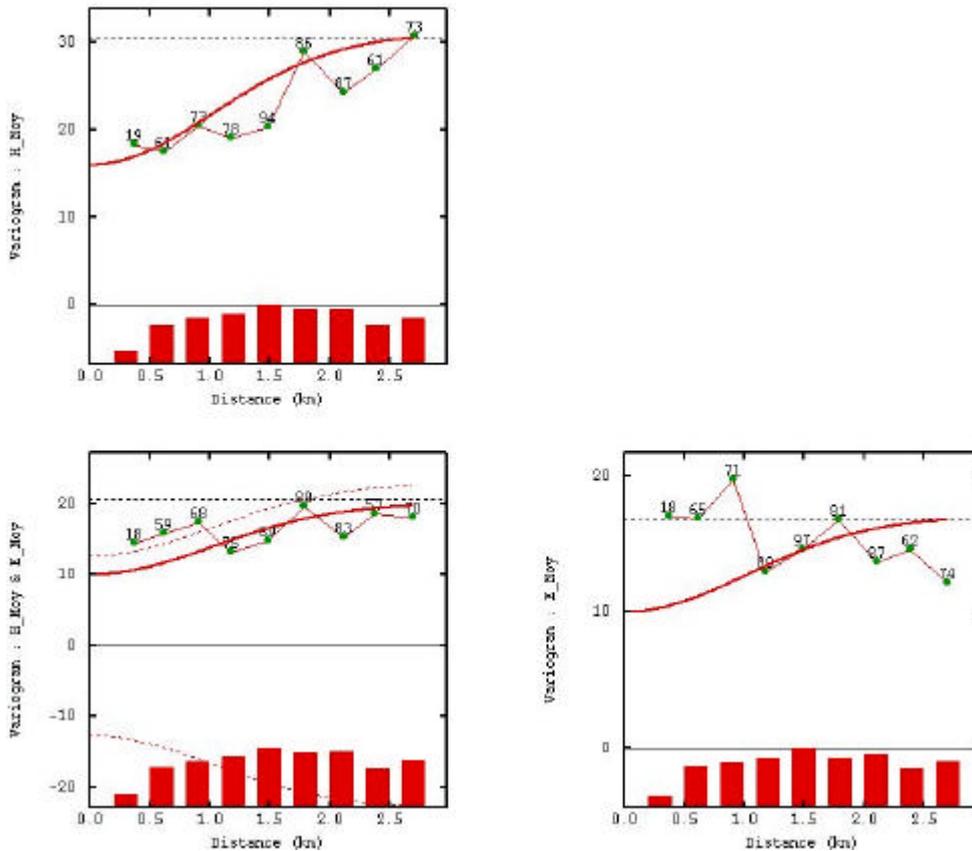


Figure 21 : Covariogrammes des variables brutes

5.11.3 Validation croisée des modèles ajustés

Les résultats de la validation croisée montrent une nette amélioration des statistiques par rapport à l'estimation monovariante : moyennes des erreurs plus proches de zéro, variances de l'erreur et de l'erreur relative plus petites.

Le coefficient de corrélation entre valeurs vraies et valeurs estimées est passé de 0.63 (Figure 18) à 0.88 (Figure 22) pour la moyenne estivale et celui de la moyenne hivernale passe de 0.77 (Figure 12) à 0.91 (Figure 22).

Grâce à l'information apportée par les données de l'autre saison, les valeurs extrêmes (maximales et minimales) sont bien réestimées en cokrigage, ce qui n'était pas le cas en krigeage monovariante.

Estimation	Erreur		Erreur relative	Erreur réduite	Coefficient de corrélation.
	Moyenne	Variance	Variance	Variance	
Moyenne estivale	0.05	3.94	0.02	0.93	0.876
Moyenne hivernale	0.00	5.31	0.01	0.77	0.909

Tableau 16 : Statistiques de la validation croisée du modèle multivariante

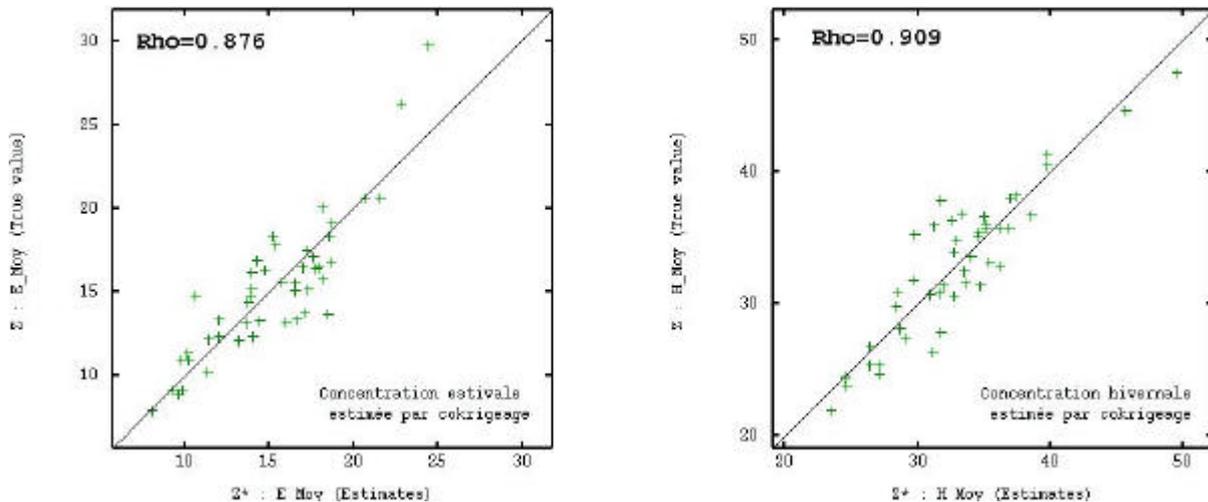


Figure 22 : Nuage de corrélation entre valeurs vraies et valeurs estimées par cokrigage en dérive externe

5.11.4 Résultats des cokrigeages saisonniers

L'écart- type d'estimation par cokrigeage de la moyenne hivernale est plus élevé que celui de la moyenne estivale. Ceci provient du fait qu'en hiver, la concentration et la variance expérimentale des données sont plus fortes. En revanche, les coefficients de variation sont très proches, indiquant une variabilité similaire pour les deux saisons (Tableau 17).

Une sélection des écarts- types les plus faibles a été effectuée afin d'identifier les zones où les estimations sont les plus précises. Un seuil de 1.5 pour le cokrigeage de la moyenne estivale et de 2.0 pour le cokrigeage de la moyenne hivernale (Figure 23 et Figure 24) a été choisi arbitrairement.

Ces sélections correspondent à 12% (2 866) des mailles pour la moyenne estivale et à 14.7% (3 512) des mailles pour la moyenne hivernale. Géographiquement on s'aperçoit que les deux sélections correspondent aux données urbaines et sont relativement similaires.

Dans les zones sélectionnées, les concentrations ont plus de variabilité (zones urbaines et périurbaines), les moyennes et les variances des concentrations sélectionnées augmentent alors par rapport au domaine entier.

Cependant ces variances sont encore beaucoup plus faibles que celles des données expérimentales (le coefficient de variation est aussi plus petit, comparer le Tableau 5 et le Tableau 18). Les effets de lissage et de support influencent là encore ces statistiques.

Si on compare les cartes et les statistiques des estimations par cokrigeage en dérive externe (Figure 23), avec celles des estimations monovariabiles par krigeage en dérive externe (Figure 13 et Figure 19), la différence n'est pas très grande (remarque : les échelles des couleurs sont légèrement différentes).

Dans notre cas, les deux variables étaient connues en tout point d'échantillonnage, ce que l'on désigne par « échantillonnage isotopique ». Si l'échantillonnage avait été « partiellement hétérotopique » (une partie des points de mesures n'est pas commune aux deux variables) alors le cokrigeage aurait amélioré l'estimation des deux variables sur l'ensemble du domaine.

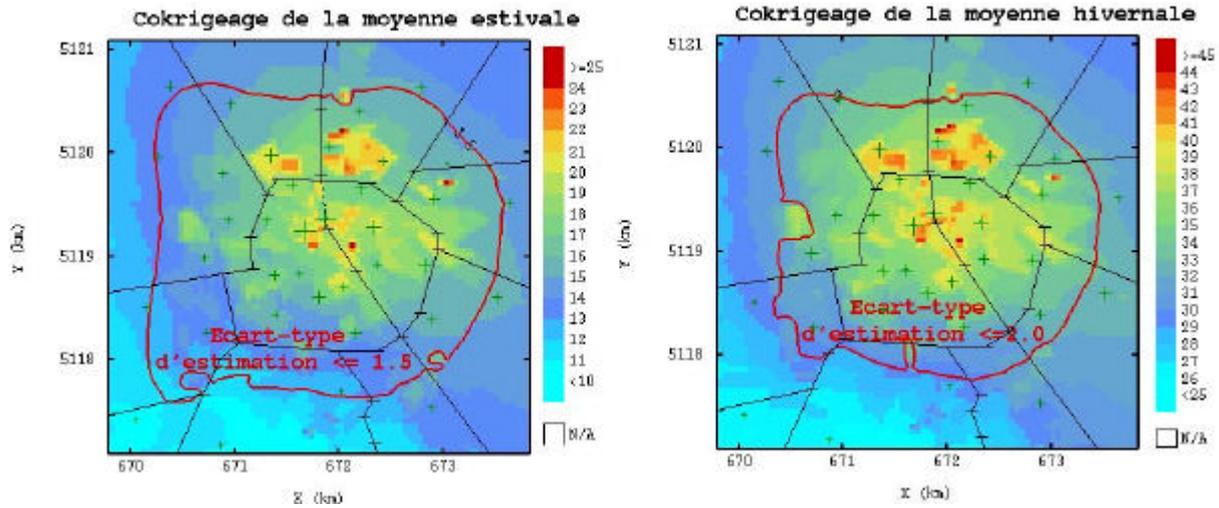


Figure 23 : Cokrigage des moyennes saisonnières

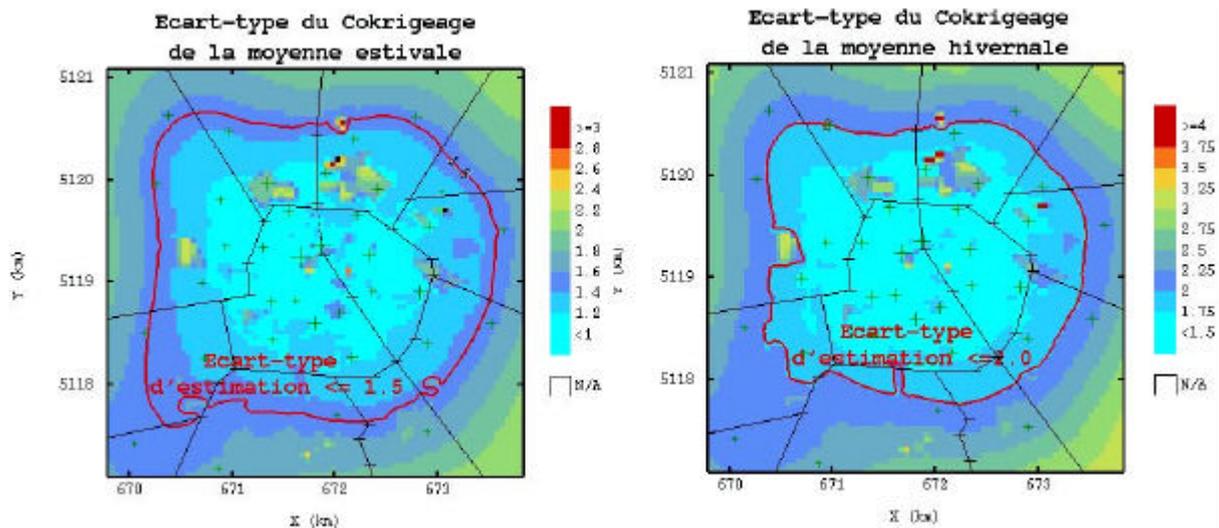


Figure 24 : Ecart-type d'estimation par cokrigage des moyennes saisonnières

Ecart- type d'estimation	Nb mailles	V Minimale	V Maximale	Moyenne	Variance	Coefficient de variation (?/m)
Moyenne estivale	23 864	0.92	3.86	2.02	0.18	0.21
Moyenne hivernale		1.26	5.08	2.90	0.41	0.22

Tableau 17 : Ecart-type d'estimation par cokrigage

<i>Estimations</i>	<i>Nb mailles</i>	<i>V Minimale</i>	<i>V Maximale</i>	<i>Moyenne</i>	<i>Variance</i>	<i>Coefficient de variation (?/m)</i>
Moyenne estivale	23 864	10.07	24.55	12.94	3.66	0.15
Moyenne hivernale		25.19	44.73	29.41	7.87	0.10

<i>Estimations</i>	<i>Nb mailles</i>	<i>V Minimale</i>	<i>V Maximale</i>	<i>Moyenne</i>	<i>Variance</i>	<i>Coefficient de variation (?/m)</i>
Moyenne estivale	3 512	11.32	24.55	16.01	4.99	0.14
Moyenne hivernale	2 866	28.83	44.73	34.63	8.06	0.08

Tableau 18 : Statistiques des estimations par cokrigage des moyennes saisonnières (ensemble du domaine et sélections des écarts- types plus petites)

5.12 KRIGEAGE DE LA MOYENNE ANNUELLE

A ce stade de l'étude, les estimations des moyennes saisonnières ont été réalisées par krigeage et par cokrigeage (avec comme dérive externe la densité de population). La suite du travail est maintenant consacrée à l'estimation de la moyenne annuelle.

Deux méthodes peuvent être proposées : la première consiste à effectuer une moyenne des estimations des moyennes saisonnières, la seconde à effectuer directement un krigeage sur la moyenne annuelle calculée à partir des périodes de mesures.

La validation croisée est un moyen de comparer ces deux méthodes.

5.12.1 Krigeage de la moyenne annuelle expérimentale

La Figure 25 montre les principales statistiques de la moyenne annuelle de la concentration du dioxyde d'azote dans l'agglomération de Bourg-en-Bresse, pour l'année 2001.

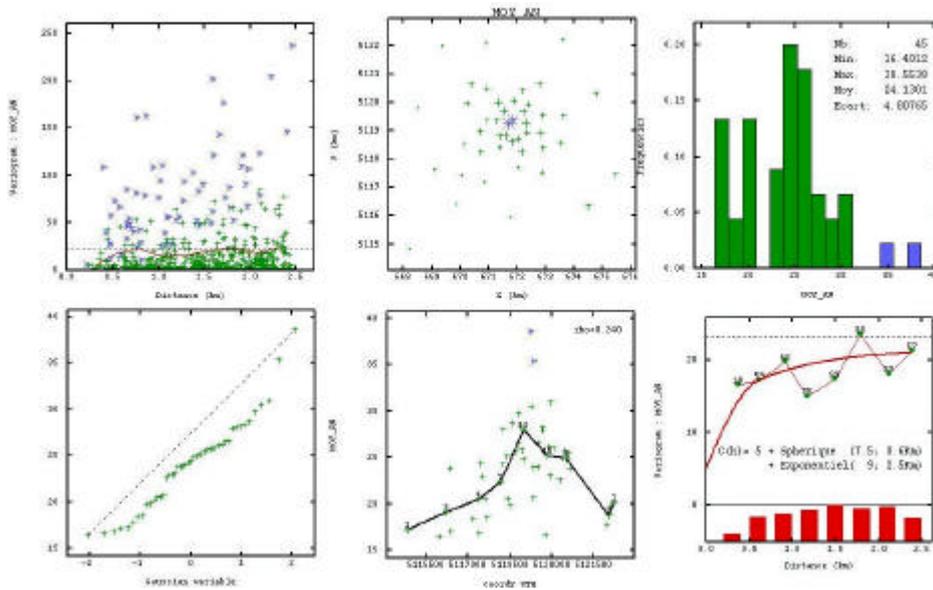


Figure 25 : Principales statistiques de la moyenne annuelle du dioxyde d'azote

Comme dans le cas des moyennes saisonnières, l'absence d'échantillons aux petites distances ne permet pas de connaître la variation dans les premiers mètres. L'allure du variogramme omnidirectionnel de la moyenne annuelle est intermédiaire entre le phénomène plus aléatoire de l'été (Figure 17) et celui plus structuré de l'hiver (Figure 11).

Ce variogramme a été modélisé par un effet de pépite de 5, une structure sphérique de petite portée 0.6 Km et de palier 7.5 et une structure gaussienne de grande portée 2.5 Km et de palier 9. La densité de population a été également prise comme dérive externe afin de modéliser la non-stationnarité due à la présence de fortes valeurs au centre ville.

Dans les résultats de krigeage, les mailles où l'écart-type d'estimation est inférieur ou égal à 3.5 ont été sélectionnées. Ce seuil, qui a été choisi arbitrairement, permet de repérer les zones où les estimations sont plus précises ; celles-ci correspondent aux zones urbaines (15% du domaine, Figure 26).

La moyenne des zones sélectionnées est plus proche de la moyenne des données expérimentales ($23.1 \mu\text{g}/\text{m}^3$), mais la variance est plus faible à cause de l'effet de lissage et de support déjà expliqués.

Tous les données	Nb de mailles	Minimum	Maximum	Moyenne	Ecart-type	Variance	Coefficient de variation (?/m)
Krigeage ordinaire en dérive externe	23 864	17.62	38.73	21.81	2.00	4.01	0.09
Ecart-type d'estimation	23 864	1.80	5.29	3.71	0.40	0.16	0.11
Ecart-type <=0.35	14.9 % des mailles						
Krigeage ordinaire en dérive externe	3 552	19.65	38.73	24.96	2.81	7.9	0.11

Tableau 19 : Statistiques du krigeage ordinaire de la moyenne annuelle

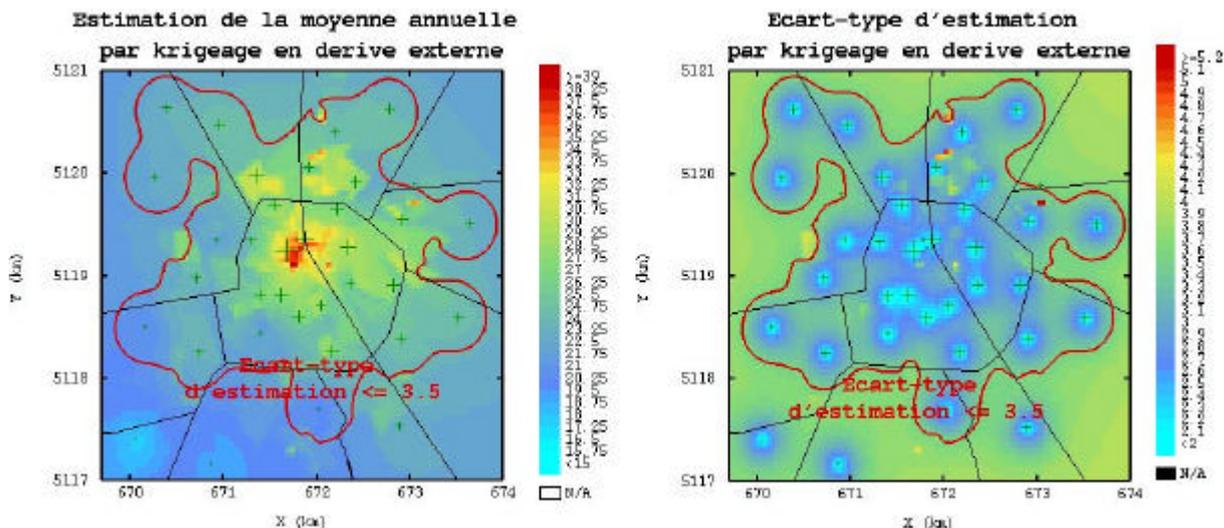


Figure 26 : Cartes du krigeage ordinaire et de l'écart-type de krigeage de la moyenne annuelle

5.12.2 Moyenne des cokrigeages saisonniers

La seconde méthode de calcul de la moyenne saisonnière consiste à calculer la moyenne arithmétique des cokrigeages saisonniers de la section 5.11. Les statistiques et cartes de cette moyenne sont montrées dans le Tableau 20 et la Figure 27.

Remarque : On a rajouté sur la Figure 27 les mêmes courbes des niveaux de la Figure 26 (sélection des estimations de la moyenne annuelle les plus fiables= écart- type d'estimation inférieure à 3.5).

Cette méthode conduit à des valeurs maximales plus faibles que le krigeage direct de la moyenne expérimentale (34.64 $\mu\text{g}/\text{m}^3$ par rapport à 38.73 $\mu\text{g}/\text{m}^3$). De ce fait la moyenne est légèrement plus faible mais reste très proche de la moyenne expérimentale (24.86 $\mu\text{g}/\text{m}^3$ par rapport à 23.1 $\mu\text{g}/\text{m}^3$).

Toutes les données	Nb de mailles	Minimum	Maximum	Moyenne	Ecart- type	Variance	Coefficient de variation (?/m)
Krigeage ordinaire en dérive externe	23 864	17.63	34.64	21.18	2.36	5.58	0.11
Ecart- type <=0.35	14.9 % des mailles						
Krigeage ordinaire en dérive externe	3552	19.48	34.64	24.86	2.66	7.06	0.11

Tableau 20 : Statistiques de la moyenne des deux cokrigeages saisonniers

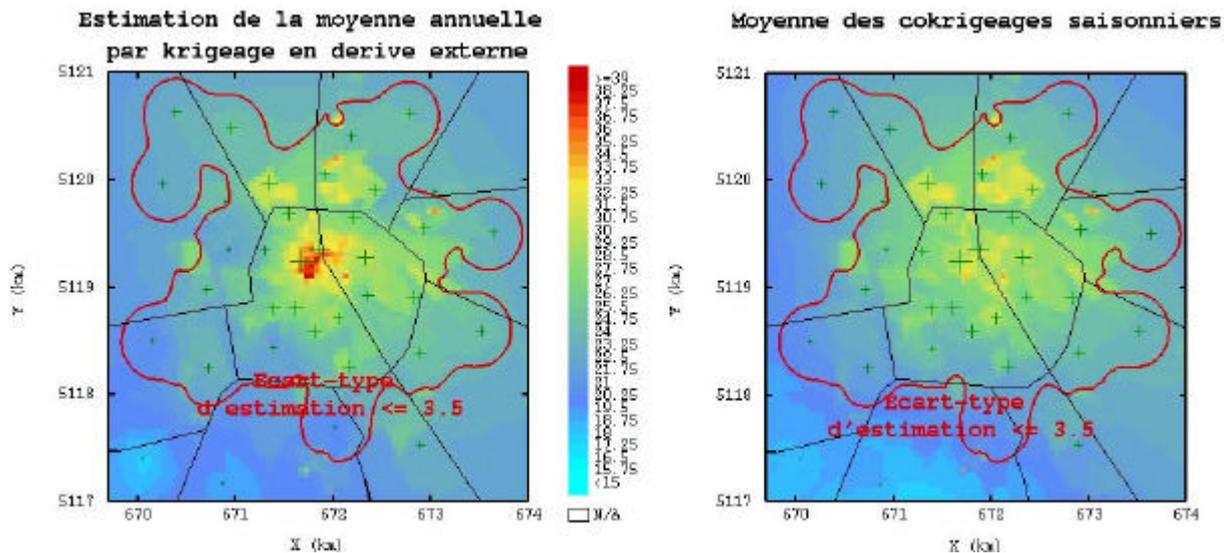


Figure 27 : Comparaison des cartes de la moyenne annuelle obtenue par deux méthodes

Rappelons que le modèle ajusté aux moyennes saisonnières pour le cokrigeage est composé d'un effet de pépite et d'une structure gaussienne (Figure 21), tandis que le modèle ajusté à la moyenne annuelle se compose d'un effet de pépite et d'une structure sphérique (Figure 25).

Les résultats obtenus dans les deux cas sont relativement proches et les cartes se ressemblent. A proximité des deux points de plus forte concentration, l'estimation directe par krigeage fournit des valeurs voisines de $38 \mu\text{g}/\text{m}^3$ tandis que la moyenne des estimations donne des valeurs de concentrations proches de $33 \mu\text{g}/\text{m}^3$.

Cependant les résultats de la validation croisée des deux méthodes sont très différents, ils plébiscitent le cokrigeage par rapport à l'estimation monovariante.

En effet, les statistiques du Tableau 21 sont meilleures pour la moyenne des cokrigeages saisonniers : le coefficient de corrélation est plus important et les variances des erreurs plus faibles.

Model	Erreur		Erreur relative	Coefficient de corrélation.
	Moyenne	Variance	Variance	
Estimation directe de la moyenne annuelle expérimentale par krigeage en dérive externe	-0.19	9.48	0.020	0.82
Moyenne des estimations saisonnières par cokrigeage en dérive externe	0.21	1.27	0.002	0.974

Tableau 21 : Statistiques de la validation croisée du modèle multivariable

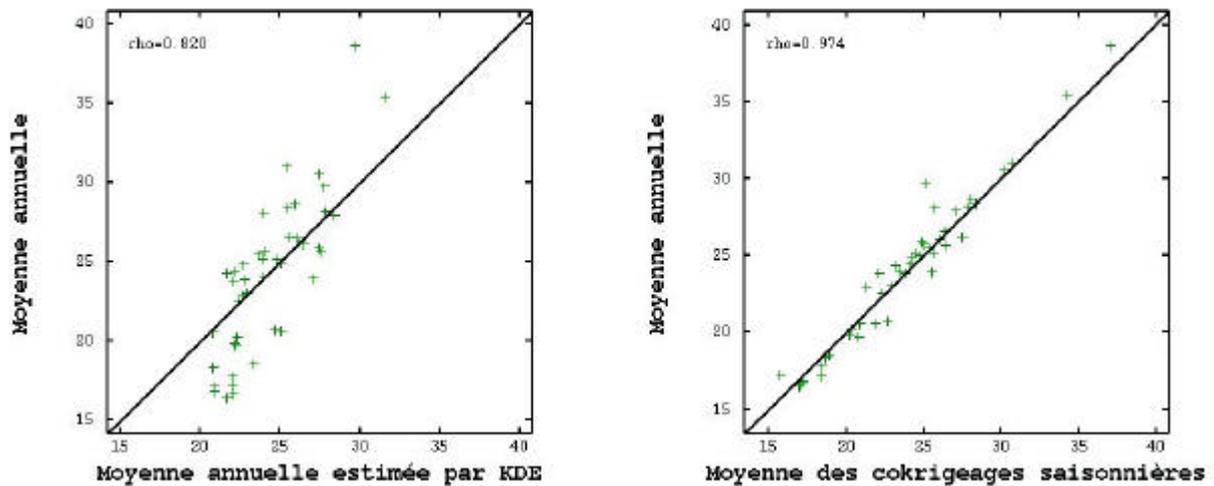


Figure 28 : Nuages de corrélation entre valeurs vraies et valeurs estimées par deux méthodes de la moyenne annuelle

Finalement les nuages de la Figure 28 montrent la nette amélioration des réestimations lorsque l'on prend la moyenne des cokrigeages.

Pour chaque cokrigeage, différentes variables auxiliaires peuvent être utilisées selon leur corrélation avec les saisons. Par conséquent, l'estimation de la moyenne annuelle peut être améliorée. Cette option n'est pas possible dans le cas d'une estimation monovariante de la moyenne annuelle expérimentale.

Un autre avantage du cokrigeage est qu'il s'avère plus adéquat dans le cas d'un échantillonnage partiellement hétérotopique. Comme il a été évoqué au paragraphe 5.11, on pourrait imaginer des plans d'échantillonnage différents pour chaque saison, en vue d'un cokrigeage des moyennes saisonnières. Ce serait une façon d'optimiser le coût des campagnes d'échantillonnage et de couvrir avec les tubes une plus grande surface.

6. CONCLUSION

Pour s'accorder avec les exigences réglementaires, la surveillance de la qualité de l'air implique de cartographier des indicateurs de pollution saisonniers ou annuels. Avec l'acquisition de logiciels spécialisés, les AASQA ont gagné en expérience dans la cartographie et y ont acquis des compétences qu'elles mettent régulièrement en pratique. Mais si les techniques de krigeage sont employées pour exploiter les campagnes de mesure et représenter spatialement les concentrations, les réflexions sur la représentativité temporelle et l'extrapolation dans la durée des cartes obtenues n'ont pas encore abouti.

Ce travail avait ainsi pour but :

- de dresser un état des lieux sur la façon dont le problème temporel est pris en compte par les AASQA dans la définition de campagnes et le traitement cartographique des données,
- d'illustrer sur un cas une des méthodologies de cartographie.

La synthèse des pratiques des associations révèle des habitudes similaires dans la stratégie adoptée : les cartographies reposent le plus souvent sur plusieurs semaines ou quinzaines consécutives de mesure en été et/ou en hiver (six à huit semaines généralement par saison). Une répartition des périodes de mesure sur toute l'année, fondée sur des recommandations européennes ou sur des considérations statistiques, est également pratiquée. Le traitement réservé aux données recueillies diffère quant à lui selon les AASQA. On distingue plusieurs cas de figure :

- Aucune extrapolation dans le temps n'est effectuée. Les moyennes des mesures sur la période étudiée (saison, année) sont introduites telles quelles dans le krigeage.
- Les données sur quelques semaines sont extrapolées sur l'année préalablement à la cartographie. A cette fin on discerne deux approches :
 - ✍ A l'aide de données de stations fixes, les moyennes des mesures sont corrigées par l'une des méthodes suivantes, à savoir l'application d'un coefficient multiplicatif, l'utilisation de la théorie des sondages, ou des méthodes de régression ;
 - ✍ Les concentrations bihebdomadaires (benzène, données de tubes) ou horaires (ozone, données de moyens mobiles) sont estimées sur l'année entière par modélisation empirique ou statistique. Le modèle, là encore, est construit à partir de mesures de stations fixes.

L'analyse géostatistique est alors conduite sur les concentrations annuelles estimées ; elle tient compte éventuellement de l'incertitude d'estimation.

- Les données d'entrée de la cartographie ne sont pas corrigées. La carte d'estimation, en revanche, est multipliée par un coefficient qui est fonction de la typologie de la maille (fond urbain, trafic) et est définie à partir des données de stations fixes.

Dans ces méthodes, la dimension temporelle est soit ignorée, soit considérée en amont ou en aval de la cartographie. Une autre méthode de cartographie saisonnière et annuelle a été récemment développée par le Centre de Géostatistique de l'Ecole des Mines de Paris et testée sur deux zones (Mulhouse et Montpellier) pour le NO₂. Elle n'effectue pas d'extrapolation préalable ou consécutive à la cartographie et suppose que les mesures représentent correctement la situation de pollution de chaque saison. A ce titre, elle se rapproche du premier cas de figure. Mais la différence importante est qu'elle :

- ne confond pas les deux saisons dans une même moyenne et tient compte d'une dynamique différente du polluant selon l'été et l'hiver ;
- tire parti des corrélations entre saisons pour estimer les concentrations saisonnières et annuelles.

La variance de l'erreur d'estimation spatiale est fournie par le krigeage ; la variance de l'erreur d'estimation temporelle, due au caractère momentané de l'échantillonnage, est approchée à l'aide des données de stations fixes.

L'intégration par le cokrigeage de la dimension temporelle dans le modèle géostatistique améliore très sensiblement la qualité de l'estimation spatiale et ouvre des perspectives intéressantes pour la définition de stratégies d'échantillonnage dans l'espace et le temps. L'application de cette méthode à une autre zone géographique a été retenue comme sujet d'étude. Elle porte sur des données de NO₂ de Bourg-en-Bresse (2002), mises à la disposition de l'INERIS par l'Air de l'Ain et des Pays de Savoie. L'absence d'un réseau de plusieurs stations fixes dans le domaine d'étude rend difficile l'évaluation de la représentativité temporelle et ne permet pas de comparaison avec les méthodes d'extrapolation précédemment citées. Des compléments sur ces points sont donc nécessaires. Cette étude a cependant le mérite d'illustrer pas à pas la méthode et d'en montrer l'efficacité. Deux logiciels ont été employés en parallèle : Isatis (Géovariances), le plus répandu dans les AASQA, et Geostatistical Analyst, module du SIG ArcView (ESRI). Pour cette application, les possibilités de Geostatistical Analyst se révèlent plus limitées.

7. REFERENCES

Barrois Q., 2004. Utilisation du module «Geostatistical Analyst » d'ARCVIEW dans le cadre de la qualité de l'air, rapport de stage INERIS.

Cottier A., 2003. Reconstitution de données pour la réalisation de cartographies d'ozone sur la région Poitou-Charentes. ATMO Poitou-Charentes, rapport de stage.

European Environment Agency (EEA), 1998. Guidance report on preliminary assessment under EC air quality directives. Technical report.

Fouquet C.(de), 2003. Méthodologie de cartographie de la concentration annuelle de NO₂ sur l'agglomération de Mulhouse, Rapport d'avancement N-6/03/G.

Fouquet C.(de), 2003. Etude sur la réalisation de cartographies de la qualité de l'air dans les zones peu/pas couvertes par les réseaux de stations fixes à l'aide de méthodes géostatistiques (complément d'étude et synthèse).

Fouquet C. (de), Gallois D., Malherbe L., Cardenas G., 2004. Mapping annual nitogene dioxide concentrations above Mulhouse urban area. GeoEnv 2004, Neuchâtel (Suisse), 13-15 octobre.

Gallois D., Fouquet C. (de), LeLoc'h, G., Malherbe L., Cardenas G. Mapping annual nitogene dioxide concentrations in urban areas. Geostats 2004, Banff (Canada), 27 septembre-1^{er} octobre.

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

Le Loc'h G., 2003. Etude exploratoire du dioxyde d'azote sur l'agglomération de Montpellier, Rapport N-8/03/G.

Christakos G., Vyas V.M., 1998. A composite space/time approach to studying ozone distribution over Eastern United States. Atmospheric Environment, Vol.32, N°16, pp 2845-2857.

Roth C., Dégardin D., 2001. Interpreting the results of diffusive sampling campaigns with respect to yearly standards. International Conference on Diffusive Smapling, Montpellier (France), 26-28 septembre 2001.

Snepvangers J.J.J.C., Heuvelink G.B.M., Huisman J.A., 2003. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, 112, pp. 253-271.

Kyriadikis Phaedon C., Journel A.G., 2001. Stochastic modeling of atmospheric pollution : a spatial time-series framework. Part I: methodology. *Atmospheric Environment*, 35, 2331-2337.

8. LISTE DES ANNEXES

Repère	Désignation précise	N° pages
1	Utilisation du logiciel Geostatistical Analyst	11
2	Quelques précisions sur le krigeage en dérive externe	1

Annexe 1

Utilisation du logiciel Geostatistical Analyst

« Geostatistical Analyst » (« GA ») est un module du logiciel SIG « ARCVIEW » de la société ESRI. L'utilisation de ce logiciel dans le contexte de la qualité de l'air a été l'objet d'un rapport de stage : Barrois Q. (annexe dans le rapport Assistance en modélisation, Ineris 2004).

Dans cette partie, nous nous limitons à la présentation et à la discussion des résultats de la méthode de cokrigage appliquée aux données de dioxyde d'azote de Bourg-en-Bresse. Pour les détails d'utilisation du logiciel il est conseillé de consulter la notice du logiciel ou le rapport de stage précédemment cité.

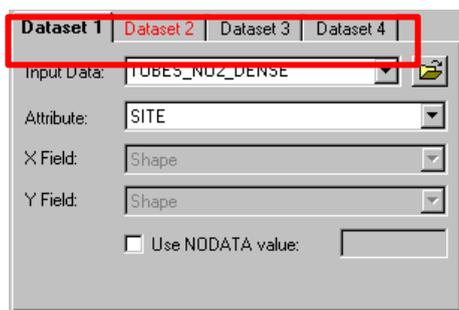
Il est important de faire remarquer que si l'étude effectuée avec Isatis sur les mêmes données (paragraphe 5) a pu considérer la densité de population comme dérive externe, en revanche, le module «GA» ne laisse pas cette possibilité.

La seule façon de modéliser une non-stationnarité dans «GA » est le filtrage de la dérive par l'ajustement d'un polynôme dont les coefficients doivent être estimés. Après plusieurs essais il a été décidé d'effectuer un cokrigage ordinaire sans considérer de dérive pour ces données.

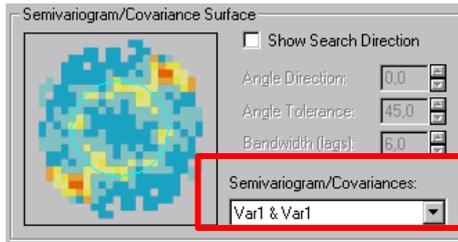
Sous «ARCVIEW» les données manquantes sont remplacées par des zéros. Le module « GA » permet de ne pas considérer ces valeurs nulles dans les estimations par l'option « NODATA Value ».

Dans ce module c'est la covariance croisée qui est ajustée (au lieu du variogramme croisée), celle-ci n'admet pas d'effet de pépite. En conséquence, il n'a pas été possible de reproduire le donc été privilégiés.

Les étapes à suivre dans le module :



Les étapes de la création d'une carte de cokrigage sous «GA» sont les mêmes que pour le krigeage. Il faut cependant préciser les autres variables à prendre en compte dans la fenêtre «Chose Input Data and Method». Dans ce module, on peut choisir jusqu'à quatre jeux de données différents, mais seul le premier peut être considéré comme la variable principale de cokrigage.



Il faut ensuite préciser les paramètres de la matrice de variance/covariance dans la fenêtre « Semivariogram /Covariance Modeling ». Pour ce faire, on choisit la variable qui nous intéresse dans la matrice avec la liste « Semivariogram/Covariance ». C’est cette liste qui nous permet de passer d’un ajustement à un autre. Plus il y a de variables, plus il y a de modèles croisés à ajuster.

Considérer trop de variables différentes peut donc rapidement conduire à un travail d’ajustement fastidieux.

Le modèle choisi est un modèle sphérique de portée 3 607 m. Son palier est de 26 pour la moyenne hivernale, de 5 pour la moyenne estivale et de 11 pour le variogramme croisé.

L’effet de pépite associé est de 11 pour la moyenne estivale et de 14 pour la moyenne hivernale.

La taille du pas utilisée est de 615m et le nuage de points est dessiné sur une distance maximale de 10 pas. On présente ci-dessous l’allure des modèles ajustés.

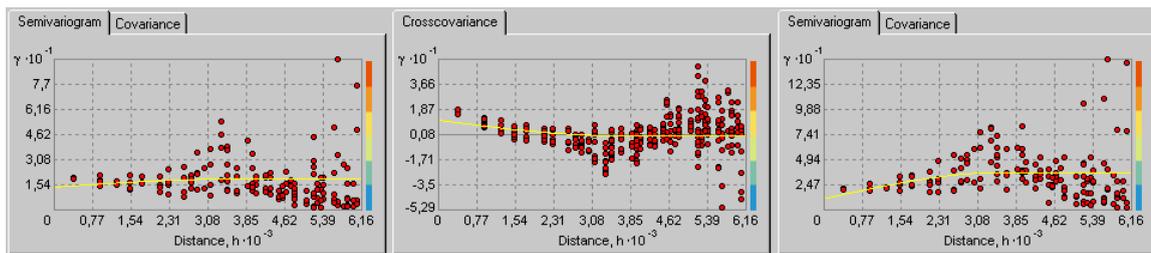


Figure 29 : Variogramme de la moyenne estivale, covariance croisée et variogramme de la moyenne hivernale

Effet de Pépite	Moyenne estivale	Moyenne hivernale
Moyenne estivale	11	0 (pas ajustable)
Moyenne hivernale	0 (pas ajustable)	14

Structure Sphérique (Portée=3.607 Km)	Moyenne estivale	Moyenne hivernale
Moyenne estivale	5	11
Moyenne hivernale	11	26

Tableau 22 : Paliers du modèle de corégionalisation linéaire sous « GA »

En ce qui concerne la limite de voisinage, il faut préciser celle-ci pour chacun des jeux de données. On considère, pour réaliser les cartes, un voisinage unique, ce qui se traduit par une limite à 10 000 m et un secteur unique.

La suite, qui est l'étape de la validation croisée, se déroule comme pour le krigeage.

Résultats du cokrigeage :

Un krigeage monovarié a été réalisé pour chacune des moyennes saisonnières en utilisant la partie « monovarié » des modèles que l'on vient d'ajuster.

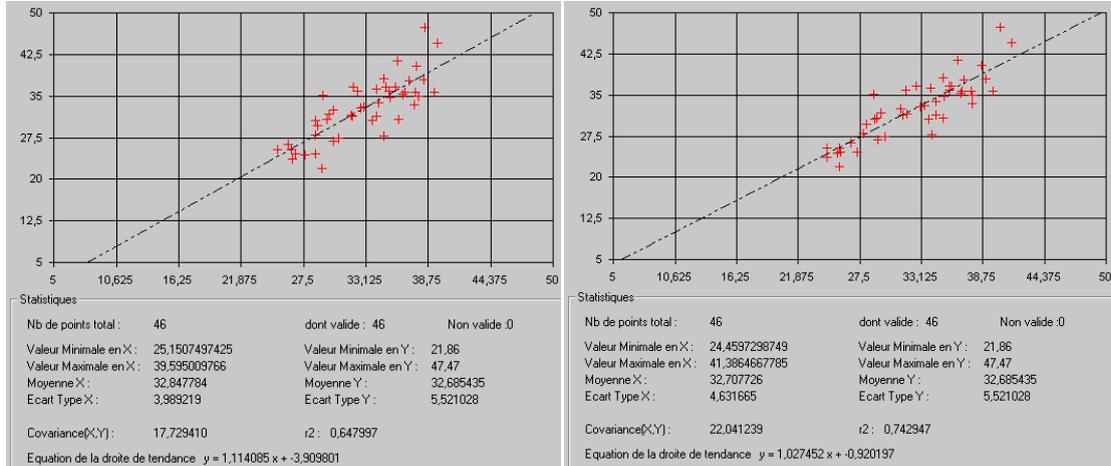
Nous allons examiner les statistiques de la validation croisée du cokrigeage et les comparer à celles du krigeage des moyennes estivales et hivernales. Le Tableau 23 récapitule les résultats.

Moyenne saisonnière	méthode	statistiques sur l'erreur			
		erreur moyenne (0)	variance de l'erreur (1)	variance de l'erreur réduite (1)	variance de l'erreur relative
Hiver	krigeage ordinaire	0,1623	11,1798	0,5704	0,0111
	cokrigeage ordinaire	0,0223	8,0261	0,4492	0,0072
Eté	krigeage ordinaire	0,0775	10,52	0,6365	0,0451
	cokrigeage ordinaire	0,0208	7,7184	0,5067	0,0283

Tableau 23 : Récapitulatif des statistiques de krigeage et de cokrigeage des données de la pollution

Les statistiques du cokrigeage sont meilleures aussi bien pour la moyenne que pour la variance. Ceci confirme que l'estimateur par cokrigeage est plus précis et moins biaisé.

Lorsque l'on observe les nuages de points correspondants, on observe une meilleure corrélation pour le cokrigage. Toutefois, cette observation est moins nette que dans l'étude effectuée sous Isatis avec la prise en compte de la dérive externe. Le nuage est néanmoins plus groupé et les valeurs extrêmes sont mieux estimées dans le cas du cokrigage.



Les valeurs vraies sont placées en ordonnée et les valeurs estimées en abscisse.

Figure 30 : Nuage de corrélation du krigeage et du cokrigeage sur la moyenne hivernale

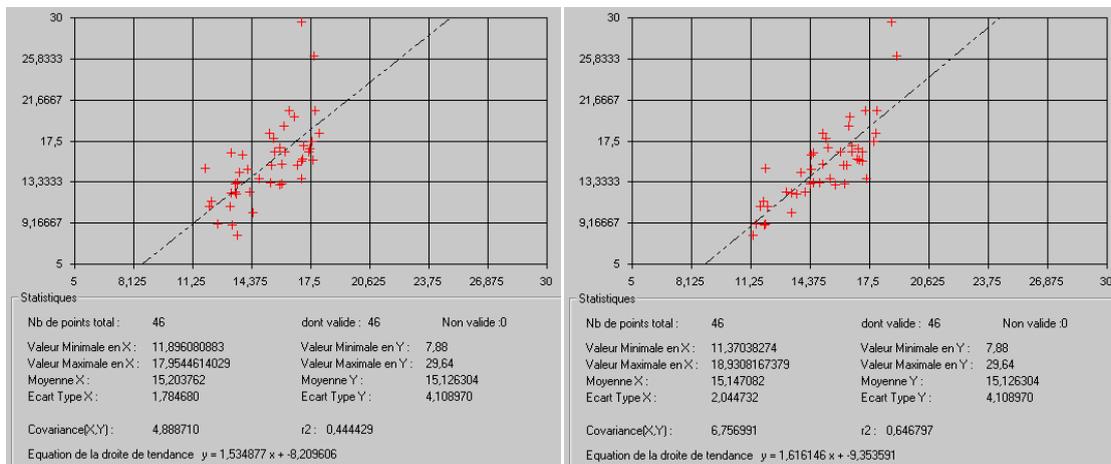


Figure 31 : Nuage de corrélation du krigeage et du cokrigeage sur la moyenne estivale

Les cartes du krigeage monovariante, puis les cartes des cokrigeages des moyennes saisonnières sont présentées ci-dessous :

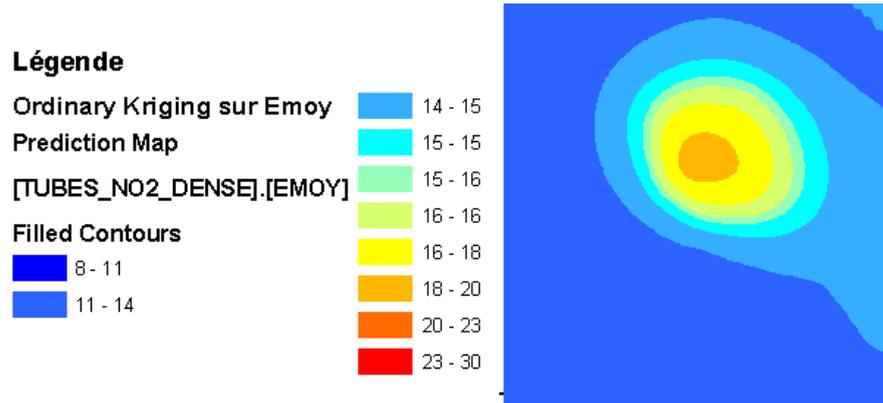


Figure 32 : Krigage Ordinaire de la moyenne estivale

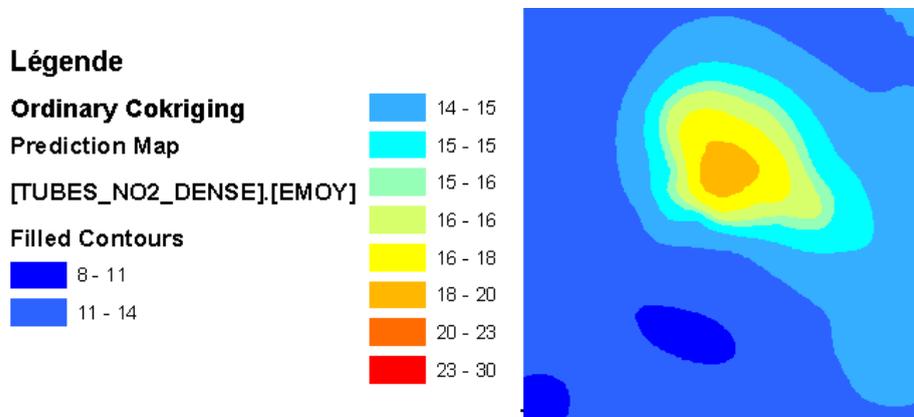


Figure 33 : Cokrigage Ordinaire de la moyenne estivale

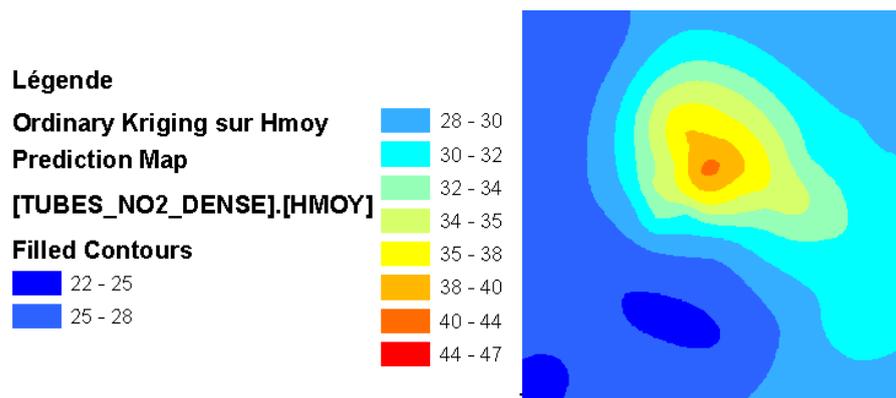


Figure 34 : Krigage Ordinaire de la moyenne hivernale

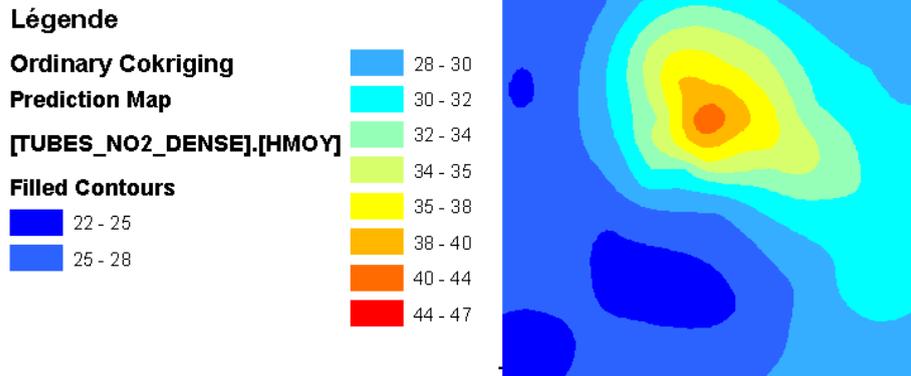
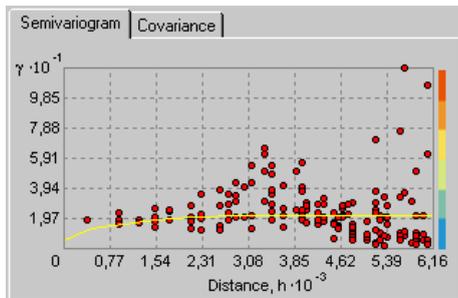


Figure 35 : Cokrigage Ordinaire de la moyenne hivernale

L’allure des cartes est très semblable pour les deux types de krigeage. Mais les résultats de la validation croisée sont différents. Cette remarque a déjà été faite dans l’étude réalisée avec Isatis.

Krigeage de la moyenne annuelle

Le krigeage de la moyenne annuelle est à présent comparé à la moyenne des cokrigeages précédemment réalisés.



La taille du pas est toujours de 615 m et on dessine le variogramme sur une distance de 10 pas. Le modèle choisi est un modèle sphérique de portée 3.8 Km et de palier 14, l’effet de pépite est de 15. Il s’agit des paramètres que le module propose par défaut. La forme du modèle ajusté est présentée ci-contre.

Les résultats de la validation croisée de ce modèle et du modèle croisé (Figure 29, Tableau 22) sont reportés ci-dessous :

Modèle	Erreur		Erreur relative	Coefficient de corrélation
	moyenne	variance	variance	
estimation directe de la moyenne annuelle expérimentale par krigeage ordinaire	0,18	8,89	0,018	0,636
moyenne des estimations saisonnières par cokrigeage ordinaire	-0,02	5,91	0,009	0,742

Tableau 24 : Récapitulatif des statistiques de la validation croisée des deux méthodes d'estimation de la pollution moyenne annuelle

Les statistiques relatives au cokrigeage sont meilleures là aussi que celles du krigeage. L'estimateur est moins biaisé et plus précis car la moyenne des erreurs et la variance sont moindres. Il est aussi mieux corrélé, comme le montre le coefficient de corrélation.

La figure ci-dessous représente les nuages de corrélation pour les deux méthodes :

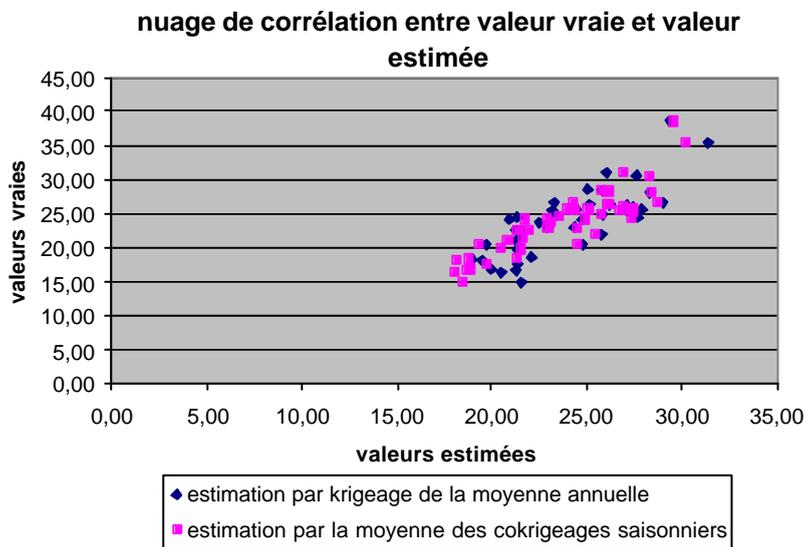


Figure 36 : Présentation des nuages de corrélation pour les deux méthodes de prise en compte de la moyenne

Le nuage de l'estimation par la moyenne des cokrigeages est légèrement mieux groupé. Il présente en effet moins de valeurs éloignées. Ces nuages illustrent donc bien ce qui a été observé sur le tableau des statistiques.

Ces deux méthodes conduisent à la production des cartes suivantes:

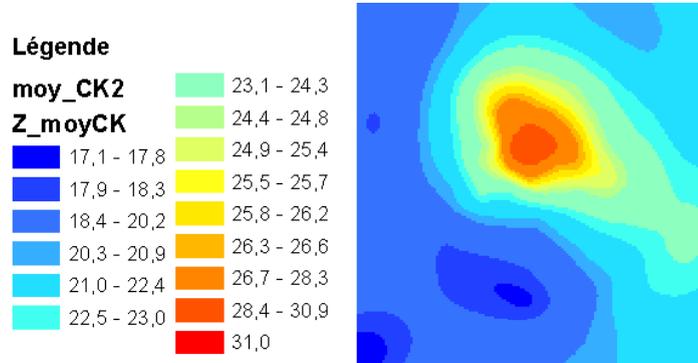


Figure 37 : Carte de la moyenne des cokrigeages saisonniers

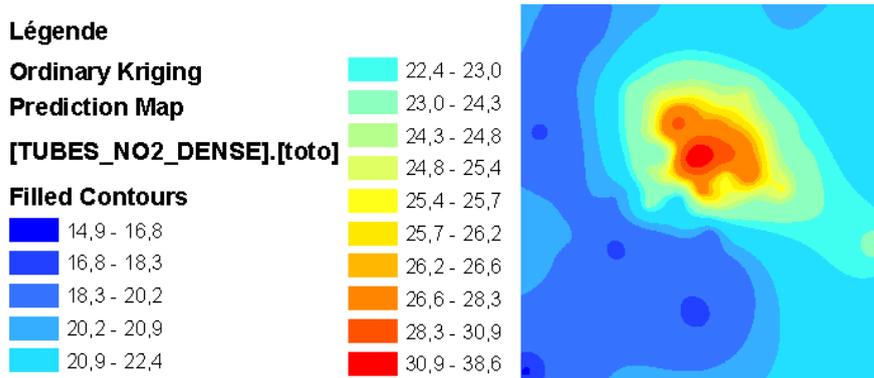


Figure 38 : Cartes du krigeage de la moyenne annuelle expérimentale

Au final, l'étude des données sous «GA» confirme les résultats obtenus avec Isatis, avec cependant moins de netteté. Les deux méthodes d'estimation se distinguent bien, mais l'étude sous Isatis montre des nuages et des statistiques mieux différenciés.

On constate que l'estimation par krigeage de la moyenne conduit à des valeurs plus fortes que la moyenne des cokrigeages. Ces zones correspondent aux sites où ont été mesurées les plus fortes concentrations. Cette étude avec le module «GA » conduit donc aux mêmes remarques qu'avec Isatis.

∞ Quelques remarques sur l'utilisation du module :

Le module présente certaines limites qui ne permettent pas d'effectuer une étude optimale : d'une part, comme il a déjà été mentionné, on ne peut pas prendre en compte la dérive externe avec une fonction auxiliaire, d'autre part d'autres outils SIG doivent être utilisés pour compléter les opérations requises.

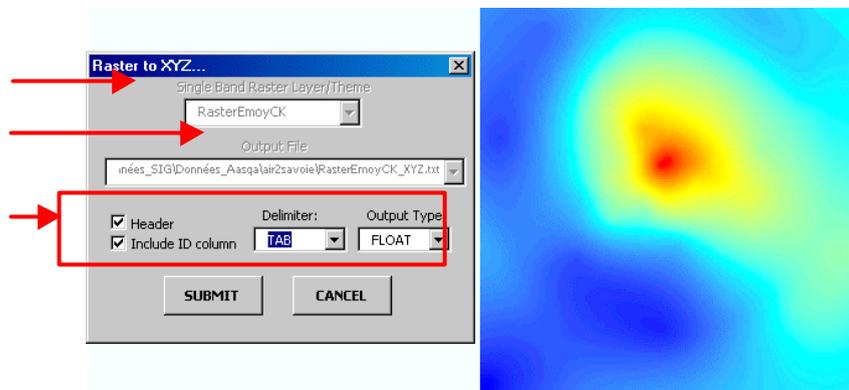
Le module suit une progression linéaire au long de l'étude qui rend son utilisation difficile lorsque l'on souhaite comparer plusieurs méthodes : il faut sans cesse revenir dans les fenêtres précédentes, modifier les paramètres, continuer la nouvelle étude, enregistrer la validation croisée, et ainsi de suite. Si on veut comparer plus de deux méthodes, il est nécessaire rassembler les statistiques de validation croisée dans un tableau à part.

Le logiciel révèle aussi des limites dans l'affichage des données. En effet, le variogramme expérimental n'est pas disponible, l'ajustement du modèle théorique se faisant uniquement avec la nuée de points. Les paramètres sont changés à la main, sans possibilité d'ajustement interactif.

L'échelle d'affichage du variogramme peut prêter à confusion pour un utilisateur peu habitué au module. Dans la fenêtre de validation croisée, les nuages de corrélation ne sont pas toujours lisibles. Nous avons en effet choisi d'utiliser une application auxiliaire, «Scatter», disponible sur le site de support d'ESRI pour afficher nos nuages de points.

Il est impossible d'effectuer directement des opérations sur les résultats du module et de les cartographier : pour obtenir la carte de la moyenne des cokrigeages plusieurs étapes ont été nécessaires, celles-ci sont commentées par la suite.

- ? La « couche » créée lors des estimations ne présente pas de table attributaire, ceci empêche l'analyse statistique directe des concentrations estimées. Pour récupérer les valeurs des estimations, une autre application auxiliaire a été utilisée. Il s'agit d'une macro (téléchargée à partir du site Internet d'ESRI : <http://support.esrifrance.fr>) appelée «Raster to XYZ». L'exécution de la macro donne la fenêtre suivante :



Présentation de l'outil « Raster to XYZ » et de l'image correspondante à convertir

? Dans ce fenêtre, on doit indiquer les informations suivantes :

- ⊗ la « couche raster » à convertir en tableau (Single Band Raster Layer/Theme),
- ⊗ le fichier d'enregistrement (Output File),
- ⊗ le séparateur de colonnes du tableau résultant (Delimiter),
- ⊗ le format des nombres: entier ou réel (Output Type),
- ⊗ l'ajout d'une colonne avec un identifiant (Include ID column).
- ⊗ L'ajout d'un en- tête au fichier résultant.

Grâce à cet outil, on obtient une table en format *.txt qui comporte (au plus) quatre colonnes : l'identifiant, la latitude et la longitude du point, et la valeur de la concentration. Cet outil a été utilisé pour récupérer les concentrations estimées par cokrigage de la moyenne estivale et hivernale.

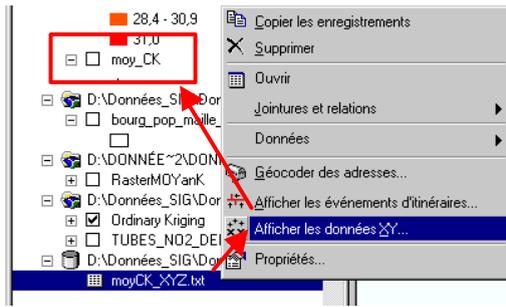
? Le processus de calcul de la moyenne arithmétique des deux estimations par cokrigage peut se faire sur un tableur comme EXCEL. Pour rassembler les deux fichiers, de façon à effectuer les calculs sur un fichier unique, on peut utiliser la colonne de l'identifiant.

? La figure ci-dessous montre le fichier résultant (format TXT) après calcul de la moyenne arithmétique des cokrigages sous un tableur.

X	Y	Z_EmoyCK	Z_HmoyCK	Z_moyCK ID
668139	5122387	12,47977	26,73139	19,60558
668189	5122387	12,45965	26,68537	19,57251
668239	5122387	12,44012	26,64078	19,54045
668289	5122387	12,42121	26,59764	19,509425
668339	5122387	12,40291	26,55599	19,47945
668389	5122387	12,38526	26,51585	19,450555
668439	5122387	12,36809	26,47691	19,42256
668489	5122387	12,35135	26,43901	19,39518
668539	5122387	12,33507	26,40223	19,36865
668589	5122387	12,31928	26,36667	19,342975
668639	5122387	12,30404	26,33244	19,31824
668689	5122387	12,28949	26,29988	19,294685
668739	5122387	12,27666	26,27137	19,274015
668789	5122387	12,26586	26,24759	19,256725
668839	5122387	12,25711	26,22865	19,24288
668889	5122387	12,2505	26,2147	19,232615
668939	5122387	12,24611	26,20596	19,226035
668989	5122387	12,24406	26,2007	19,22228

Présentation de la table de la moyenne en format *.txt, on a renommé les colonnes Z pour ne pas les confondre

? On importe ensuite la table de la moyenne dans le « bloc de données » qui nous intéresse.



A partir de celle-ci, on crée la couche des points d’estimation de la moyenne (commande « Afficher les données XY... »). A ce stade, nous disposons d’une couche de points où est enregistrée la moyenne des cokrigeages. Or pour une représentation graphique, nous devons avoir des entités surfaciques. En effet, la conversion raster vers table nous a fait passer des données

surfaciques à des données ponctuelles, les points étant les centres des carrés de la maille raster. L’étape suivante permet de revenir à des données surfaciques.

- ? Lorsqu’on dispose d’un maillage déjà créé (par exemple la grille de densité de population), il suffit de joindre géographiquement les tables des deux couches (cette option est disponible dans le menu « géoréférencement » ou dans l’option « jointures et relations »).
- ? S’il faut créer le maillage on peut utiliser la macro « AnalyseParMailles », disponible sur le site Internet d’Esri : « <http://support.esri.fr> » (l’utilisation de cette macro n’est pas expliquée ici).
- ? Il ne reste plus qu’à appliquer une symbologie en dégradé de couleur. Cependant, aucune méthode de classification n’étant commune à une « couche d’interpolation » (créée avec le krigeage de la moyenne annuelle) et à un « shapefile » (que nous venons de créer), il faut faire correspondre manuellement les classes entre elles.

Seule la succession de ces étapes a permis de créer une nouvelle carte de la moyenne des cokrigeages saisonniers.

Comme le suggèrent ces remarques, ce module de géostatistique est principalement destiné à un usage relativement superficiel. Une analyse poussée des données et un traitement rigoureux ne sont pas facilités.

Annexe 2

Quelques précisions sur le krigeage en dérive externe

La densité de population est utilisée comme dérive externe ; en faisant cela nous considérons le phénomène comme la superposition de deux composantes, une partie déterministe appelée dérive et un résidu supposé stationnaire et de moyenne nulle :

$$\text{NO}_2 = \text{RESIDU} + \text{DERIVE EXTERNE}$$

$$\text{DERIVE EXTERNE} = A + [B * (\text{Densité de Population})] : \text{régression linéaire}$$

Il reste à décider du modèle de covariance (ou variogramme) à prendre en compte dans les équations de krigeage ; en théorie il s'agit du modèle des résidus.

Le calcul des résidus peut se faire de plusieurs manières ; la plus simple est de calculer les résidus à partir des coefficients de la régression linéaire (repérés dans le Tableau 8) :

$$\text{RESIDU} = \text{NO}_2 - A - [B * (\text{Densité de Population})]$$

Ensuite, on effectue un krigeage à moyenne connue des résidus (supposés de moyenne nulle) et on obtient l'estimation du dioxyde d'azote par la relation suivante :

$$\text{NO}_2^{\text{KSR}} = B * (\text{Densité de Population}) + A + \text{RESIDU}^{\text{KS}}$$

Une autre façon est d'ajuster les coefficients de la régression linéaire dans un voisinage glissant ; dans ce cas les résidus varient d'un voisinage à un autre et la covariance des résidus n'est pas directement accessible (ajustement graphique par exemple). Dans ce type d'estimation, qui est le véritable krigeage en dérive externe, les coefficients n'interviennent que dans le calcul des résidus locaux et non pas dans le calcul des estimations.

Pour ajuster la covariance des résidus, on propose différentes structures dont les paliers sont ajustés numériquement et par validation croisée on choisit la combinaison de structures qui donne la variance des erreurs réduites la plus proche de l'unité (cas où la variance de krigeage est égale à la variance expérimentale des données prises lors de la validation).

Dans le cas particulier d'une covariance péritique et d'un voisinage unique, le krigeage en dérive externe est équivalent à la régression linéaire multiple expliquée plus haut.

Fin du rapport envoyé au client. Ce qui suit est destiné au dossier maître