



Laboratoire Central de Surveillance de la Qualité de l'Air



Traitements numériques

Application des méthodes géostatistiques pour l'exploitation conjointe des mesures de fond et de proximité

Novembre 2005
Convention: 05000051

Laure Malherbe - Giovanni Cardenas





Ministère de l'Ecologie
et du Développement Durable

PREAMBULE

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

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

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



Ministère de l'Écologie
et du Développement Durable

Application de méthodes géostatistiques pour l'exploitation conjointe de données de fond et de proximité

Laboratoire Central de Surveillance
de la Qualité de l'Air

Thème : Traitements numériques

Convention 05000051

**Financée par la Direction des Préventions des Pollutions et des Risques
(DPPR)**

Novembre 2005

Laure Malherbe – Giovanni Cardenas

Ce document comporte 32 pages (hors couverture et annexes).

	Rédaction	Vérification	Approbation
NOM	Laure MALHERBE	Laurence ROUÏL	M.RAMEL
Qualité	Ingénieur Etudes et Recherches Direction des Risques Chroniques	Ingénieur Etudes et Recherches Direction des Risques Chroniques	Responsable LCSQA/INERIS Direction des Risques Chroniques
Visa			

TABLE DES MATIERES

1. RÉSUMÉ.....	3
2. GLOSSAIRE.....	4
3. INTRODUCTION	7
3.1 Contexte et objet de l'étude.....	7
3.2 Méthode	7
3.3 Outils.....	8
4. SYNTHÈSE DE LA DEMARCHE	8
4.1 Echantillonnage du dioxyde d'azote.....	8
4.2 Cartographie de la pollution de fond à l'échelle régionale, avec prise en compte de variables auxiliaires	9
4.2.1 Etude exploratoire des données de concentration de fond – recherche de variables explicatives.....	9
4.2.1.1 Caractérisation préliminaire: statistiques monodimensionnelles	9
4.2.1.2 Création sous SIG de différentes variables auxiliaires	10
4.2.1.3 Etude des corrélations entre concentrations de fond et variables auxiliaires	11
4.2.1.4 Sélection de variables explicatives.....	13
4.2.2 Etude variographique et estimation	14
4.2.2.1 Modélisation sans et avec prise en compte des variables explicatives	14
4.2.2.2 Evaluation et comparaison des modèles par validation croisée	16
4.2.2.3 Définition d'une grille d'estimation et réalisation des cartographies ..	17
4.3 Prise en compte des mesures de proximité	20
4.3.1 Evaluation des estimations de fond dans les villes et les zones de proximité.....	20
4.3.1.1 Krigeage aux points des stations fixes urbaines ; comparaison entre valeurs estimées et mesurées.....	20
4.3.1.2 Krigeage aux points des sites de proximité ; comparaison entre valeurs estimées et mesurées.....	21
4.3.2 Explication et modélisation statistique de la différence entre estimation de fond et mesure de proximité routière	23
4.3.3 Correction des estimations en zone de proximité routière	25

4.3.3.1	Resserrement de la grille d'estimation dans les zones de proximité routière	25
4.3.3.2	Calcul des variables auxiliaires sur la grille fine	25
4.3.3.3	Estimation des concentrations de fond sur les mailles plus fines	26
4.3.3.4	Calcul et ajout du terme correctif a ces valeurs estimées	26
4.3.3.5	Représentation finale	28
5.	CONCLUSION.....	29
6.	RÉFÉRENCES	31
7.	LISTE DES ANNEXES	32

1. RESUME

Afin de fournir au public et aux gestionnaires une information la plus complète et représentative possible sur la qualité de l'air, il est important d'intégrer l'ensemble des sources de pollution, qu'elles résultent d'émissaires lointains ou proches. Cela revient à rechercher une information portant à la fois sur les niveaux de concentrations atmosphériques de fond et sur la pollution de proximité (routière, industrielle).

Inclure dans une même cartographie ces deux catégories de données représente une certaine difficulté puisqu'elles ne supposent pas le même type de traitement numérique, les échelles spatiales et temporelles concernées étant de nature différente. Considérant l'élaboration d'une cartographie représentative de situations de pollution à l'échelle régionale et la prise en compte du fort impact des infrastructures routières, une méthodologie consacrée à la cartographie régionale du dioxyde d'azote (NO₂) est développée. Elle repose sur l'exploitation de données d'échantillonnage passif fournies par l'association LIG'Air et sur l'étude des relations entre les concentrations de NO₂ et des variables auxiliaires pertinemment choisies. Elle fait largement appel aux possibilités de calcul des SIG, notamment dans le traitement de ces variables, et s'appuie exclusivement sur des outils d'analyse statistique et géostatistique.

La démarche proposée, testée sur des campagnes de l'hiver 2005, se déroule en plusieurs étapes :

1. recherche de variables explicatives des concentrations de fond ;
2. étude géostatistique et estimation par krigeage de la pollution de fond régionale, avec usage de variables auxiliaires ;
3. modélisation statistique de la différence entre valeurs estimées et mesurées aux points de proximité ;
4. raffinement du maillage dans les zones de proximité et correction de l'estimation de fond par le modèle précédemment établi.

Si les trois premières étapes ont été conduites sur toute la région Centre, la dernière, limitée par des contraintes informatiques, ne concerne qu'une partie du domaine. Malgré ces difficultés pratiques, l'expérience est prometteuse et mérite d'être affinée et poursuivie auprès des AASQA.

2. GLOSSAIRE

Analyse en composantes principales

L'analyse en composantes principales (ACP) est une méthode d'analyse statistique multidimensionnelle permettant de représenter synthétiquement un ensemble de données. Elle sert à identifier des similarités ou des oppositions entre variables et à repérer les variables les plus corrélées entre elles. Ces variables sont par exemple des concentrations hebdomadaires ou bihebdomadaires mesurées pendant plusieurs périodes successives sur un ensemble de tubes et des variables auxiliaires connues ou estimées en ces mêmes points.

L'ACP consiste à remplacer une famille de variables par de nouvelles variables de variance maximale, non corrélées deux à deux et qui sont des combinaisons linéaires des variables d'origine. Ces nouvelles variables, appelées *composantes principales*, définissent des plans factoriels qui servent de base à une représentation graphique plane des variables initiales. L'interprétation des résultats se restreint généralement aux deux premiers plans factoriels, sous réserve que ceux-ci expliquent la majeure partie de la variance du nuage des variables initiales.

En ACP normée, les variables projetées sur chaque plan factoriel se trouvent à l'intérieur d'un cercle de rayon unité. Plus une variable est projetée vers le bord du cercle, mieux elle est représentée. Par ailleurs, deux variables bien représentées et proches l'une de l'autre sont corrélées positivement tandis que deux variables qui s'opposent sont corrélées négativement. Une orthogonalité entre deux variables traduit l'absence de corrélation linéaire.

Notons que l'ACP ne mesure que des liens linéaires entre variables. Avant de conclure sur l'existence ou l'absence de relations entre variables, il est donc utile d'examiner l'allure de leurs nuages de corrélation.

Estimateurs du krigeage

Différents estimateurs dont le principe et des exemples d'application sont fournis en [1], [2] et [5], sont considérés dans cette étude. Leur expression est rappelée ci-après.

Soient s_i les points d'échantillonnage et s un point où l'on souhaite estimer la concentration de dioxyde d'azote.

On désigne par

$Z(s_i)$, les concentrations mesurées aux points s_i

$Z^*(s)$, la concentration estimée au point s

$\Phi(s_i)$ et $\Phi(s)$ la variable auxiliaire connue aux points s_i et s

$R(s_i)$, les résidus expérimentaux de la régression linéaire de Z par Φ

$R^*(s)$, le résidu estimé au point s .

Krigeage ordinaire

L'estimateur est une combinaison linéaire des données expérimentales :

$$Z^*(s) = \sum_i \lambda_i Z(s_i)$$

Les pondérateurs λ_i sont tels que Z^* est sans biais et de variance minimale.

Krigeage des résidus

1^{ère} étape : régression linéaire entre Z et Φ

$$Z(s_i) = a \cdot \Phi(s_i) + b + R(s_i)$$

Le résidu R est considéré comme une variable aléatoire.

2^e étape : modélisation du variogramme expérimental des résidus de la régression et estimation de R par krigeage ordinaire

$$R^*(s) = \sum_i \lambda_i R(s_i)$$

3^e étape : estimation finale

$$Z^*(s) = a\Phi(s) + b + \sum_i \lambda_i R(s_i)$$

Krigeage avec dérive externe

$$Z^*(s) = \sum_i \lambda_i Z(s_i)$$

Les pondérateurs λ_i satisfont aux mêmes contraintes que le krigeage ordinaire ainsi qu'à la condition sur l'espérance de la variable $Z(s)$: $E[Z(s)] = a \cdot \Phi(s) + b$ dans un certain voisinage. Les résidus n'étant pas calculés explicitement, on peut utiliser la procédure de modélisation non stationnaire d'Isatis (*non stationary modeling*), qui teste différentes structures par validations croisées successives.

En voisinage unique, le krigeage avec dérive externe correspond à la formulation du krigeage des résidus.

Cokrigeage multicolocalisé

En cokrigeage, l'estimateur de Z est une combinaison linéaire des valeurs expérimentales de Z et de Φ :

$$Z^*(s) = \sum_i \lambda_i Z(s_i) + \sum_j \mu_j \Phi(s_j)$$

Les pondérateurs vérifient les conditions de non biais et de variance minimale.

Lorsque Φ est connue sur tout le domaine, on effectue un cokrigage multicolocalisé : pour des raisons numériques, le voisinage considéré pour la variable Φ est restreint au point d'estimation et aux points de mesure de Z .

Ecart-type de krigeage :

On désigne couramment par *écart-type de krigeage* l'écart-type σ_K de l'erreur d'estimation. Il est défini par :

$$\sigma_K^2(s) = \text{var}[Z^*(s) - Z(s)] \quad (\text{var : fonction variance})$$

Validation croisée

La validation croisée a pour objet de contrôler l'adéquation entre le variogramme modélisé et les données expérimentales qui ont servi à construire ce modèle.

Elle consiste à éliminer temporairement un site de mesure et à y estimer la concentration par krigeage, à l'aide des données restantes et du modèle de variogramme. On dispose ainsi d'une concentration estimée Z_i^* et d'une concentration mesurée Z_i . Cette opération est répétée pour chacun des points d'échantillonnage, ce qui permet de calculer des statistiques sur l'erreur Z^*-Z .

3. INTRODUCTION

3.1 CONTEXTE ET OBJET DE L'ETUDE

Les outils de la géostatistique sont couramment employés par les AASQA pour cartographier la pollution de fond à l'échelle urbaine ou régionale (dioxyde d'azote, ozone, benzène) [1]. Les données exploitées sont les mesures recueillies sur les sites de fond, le plus souvent par échantillonnage passif ; la pollution de proximité fait généralement l'objet d'études séparées.

Cependant, si les cartes obtenues reproduisent correctement la répartition des concentrations dans la zone d'étude, elles n'offrent pas nécessairement une image fine des gradients de concentration près des grandes sources de pollution telles que les axes de circulation. Elles peuvent en particulier sous-estimer les niveaux de pollution au voisinage des routes ou des voies urbaines.

Pour une représentation plus fidèle de la réalité et une information plus complète du public, les AASQA s'interrogent sur les moyens de traiter conjointement des mesures relatives à différentes échelles spatiales (fond et proximité). L'enjeu de cette étude est ainsi d'imbriquer des données de fond et de proximité dans une unique cartographie. Cette expérience pilote résulte d'une collaboration avec l'association LIG'AIR qui, dans ce but, a réalisé des campagnes mensuelles de mesure du NO₂,

- au voisinage de nœuds routiers, d'avril 2003 à mars 2004,
- en situation de fond, durant l'année 2005.

L'objectif est de développer des procédures de traitement géostatistique permettant de superposer les informations issues de ces deux types de campagnes.

3.2 METHODE

Ce travail repose sur une analyse approfondie des données de concentration de fond et de proximité et des corrélations entre ces variables et différentes variables auxiliaires.

Pour faciliter la compréhension de la démarche, seule une synthèse, illustrée des principaux résultats et graphiques, est donnée dans le corps du texte. Des détails techniques sont fournis en annexe.

La méthodologie proposée se décompose en quatre étapes :

1. Recherche de variables explicatives des concentrations de fond ;
2. Cartographie régionale de la pollution de fond avec introduction de variables auxiliaires ; compte tenu de leur caractère local, les données de proximité ne sont pas utilisées à ce stade de la cartographie.
3. Comparaison entre estimations de fond et mesures de proximité et modélisation statistique de la différence ;
4. Correction statistique de la cartographie de fond dans les zones de proximité.

3.3 OUTILS

L'analyse des données et l'ensemble des traitements géostatistiques ont été effectués avec le logiciel ISATIS (Géovariances, version 5.0.2). Certains calculs statistiques (estimation de la densité de population) ont été réalisés avec R (<http://lib.stat.cmu.edu/R/CRAN/>, v2.1.1).

Les informations auxiliaires ont été traitées à l'aide du SIG Arcview (ESRI, version 9.0).

4. SYNTHÈSE DE LA DEMARCHE

4.1 ÉCHANTILLONNAGE DU DIOXYDE D'AZOTE

Les données de concentration de NO_2 proviennent de 19 sites de proximité et de 49 sites de fond répartis sur toute la région Centre (Figure 1). Des tubes à échantillonnage passif ont été exposés en ces points pendant douze périodes consécutives d'un mois : d'avril 2003 à mars 2004 pour les mesures de proximité [3] de janvier 2005 à décembre 2005 pour les mesures de fond.

Concernant l'échantillonnage de fond, les villes de Blois, Tours, Chartres et Châteauroux comptent chacune deux points de mesure (paires de points rapprochés, figure de gauche), des sites supplémentaires ayant été installés près d'analyseurs automatiques.

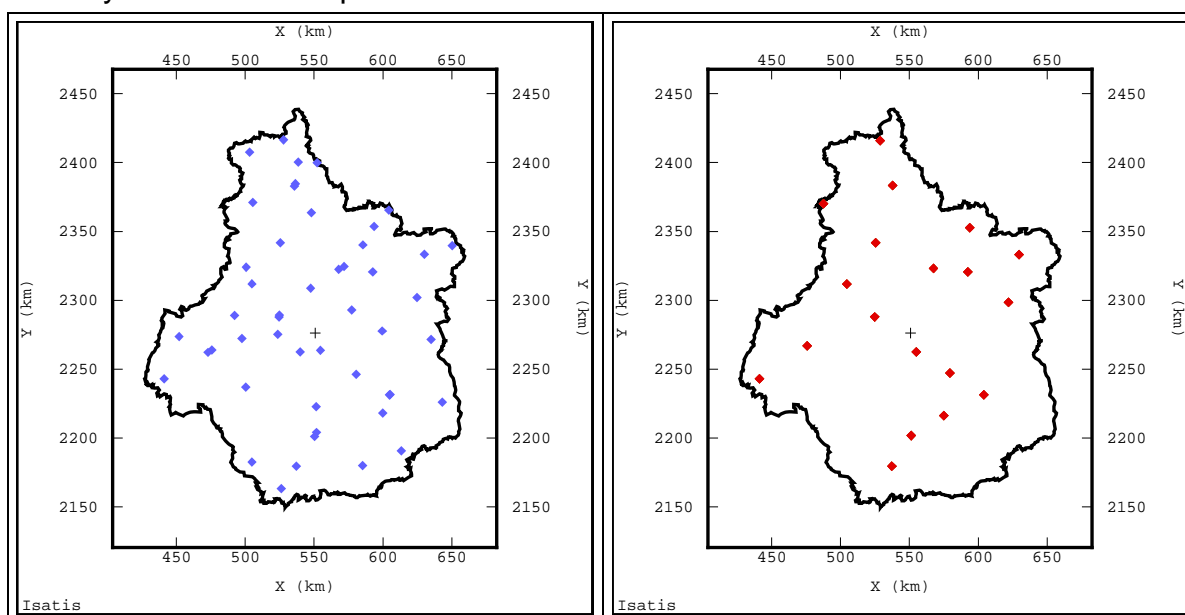


Figure 1 – Situation des mesures de fond (en bleu figures de gauche) et de proximité (en rouge, figure de droite)

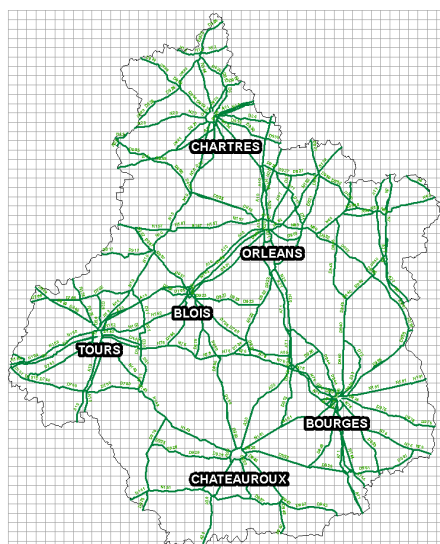


Figure 2 – Principales routes de la région Centre

4.2 CARTOGRAPHIE DE LA POLLUTION DE FOND A L'ECHELLE REGIONALE, AVEC PRISE EN COMPTE DE VARIABLES AUXILIAIRES

L'analyse porte sur la **moyenne hivernale du NO₂**, calculée sur les **données de fond** de janvier, février et mars 2005 (premier trimestre complet de mesures disponible en cours d'année). Un travail complémentaire sera effectué en 2006 sur l'été et l'année complète.

4.2.1 ÉTUDE EXPLORATOIRE DES DONNEES DE CONCENTRATION DE FOND – RECHERCHE DE VARIABLES EXPLICATIVES

4.2.1.1 CARACTERISATION PRELIMINAIRE: STATISTIQUES MONODIMENSIONNELLES

L'histogramme représente la fréquence de données par classe de concentrations. Il présente deux modalités qui ne correspondent pas à un découpage de l'espace (les données associées à chacune se répartissent sur tout le territoire) mais sont liées à l'importance des villes et des routes présentes dans l'environnement des mesures. Il montre ainsi la contribution des émissions locales à la pollution de fond. Il ne justifie pas un traitement séparé des données de chaque modalité, puisque celles-ci caractérisent un même type de pollution, mais il suggère l'utilité de recourir à des variables descriptives des sources.

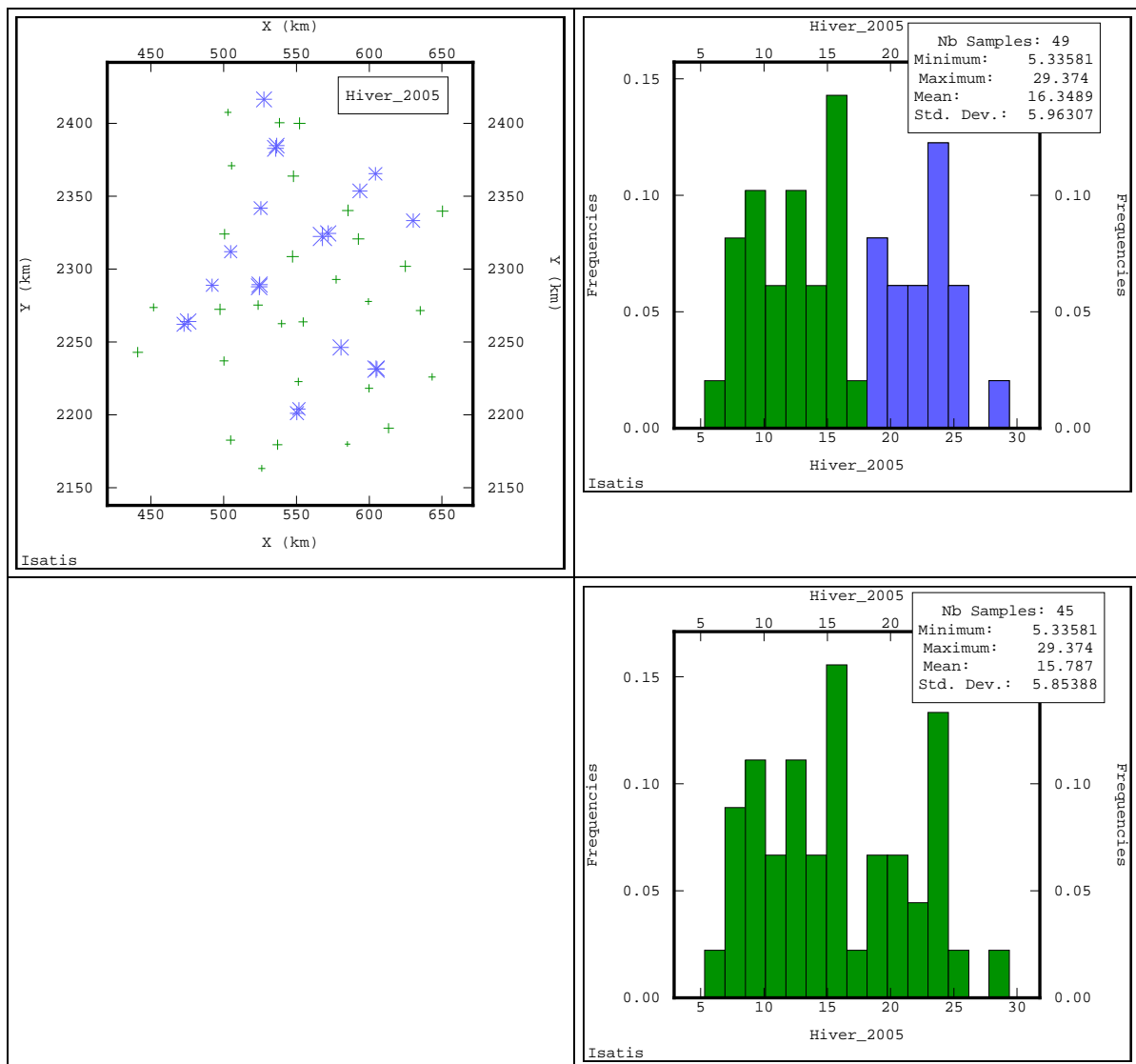


Figure 3 – Histogramme des concentrations de fond de NO₂ (moyennes de janvier, février, mars 2005). En haut, avec, en bas, sans, les mesures effectuées près des stations fixes de Blois, Tours, Chartres et Châteauroux.

4.2.1.2 CREATION SOUS SIG DE DIFFERENTES VARIABLES AUXILIAIRES

Les données auxiliaires brutes fournies par LIG'AIR (voir annexe 1) sont disponibles en tout point d'un maillage ou sous la forme de polygones lisibles par un système d'information géographique. Il s'agit :

- du cadastre kilométrique des émissions établi par LIG'AIR en 2000 et mis à jour en 2005 ;
- des données communales INSEE (frontières géographiques et caractéristiques socio-démographiques des communes) ;
- des données d'occupation du sol CORINE Land Cover (IFEN).

A partir de ces données, une programmation informatique avec ARC View permet de construire différentes variables secondaires potentiellement liées aux

concentrations. Dans un premier temps, les valeurs de ces variables sont calculées aux points de mesure.

Tableau 1 – Variables auxiliaires calculées aux points de mesure

Variable calculée	dans un rayon de
Portion de surface [entre 0 et 1] occupée par chacune des classes de la base CORINE	500 m , 1 km, 2 km, 5 km et 10 km autour des tubes
Densité d'émissions [kg/m ²]	500 m , 1 km, 2 km, 5 km et 10 km autour des tubes
Densité de population [nb d'hab./km ²]*	500 m , 1 km, 2 km, 5 km autour des tubes

* La densité de population a été déduite des données communales de population (recensement 1999). Au préalable, la population de chaque commune a été répartie spatialement en fonction de l'occupation du sol. La procédure statistique développée à cette fin est décrite en annexe 2.

4.2.1.3 ETUDE DES CORRELATIONS ENTRE CONCENTRATIONS DE FOND ET VARIABLES AUXILIAIRES

Les concentrations de dioxyde d'azote sont corrélées linéairement aux variables considérées. Le coefficient de corrélation évolue avec le rayon de calcul des densités : il augmente puis diminue (Tableau 2). **Ainsi, la distance jusqu'à laquelle l'influence d'une variable auxiliaire est prise en compte constitue un paramètre déterminant dans la recherche de facteurs explicatifs.**

La transformation logarithmique de la variable auxiliaire (de la forme $\ln(1+Y)$) a pour effet de linéariser les nuages de corrélation.

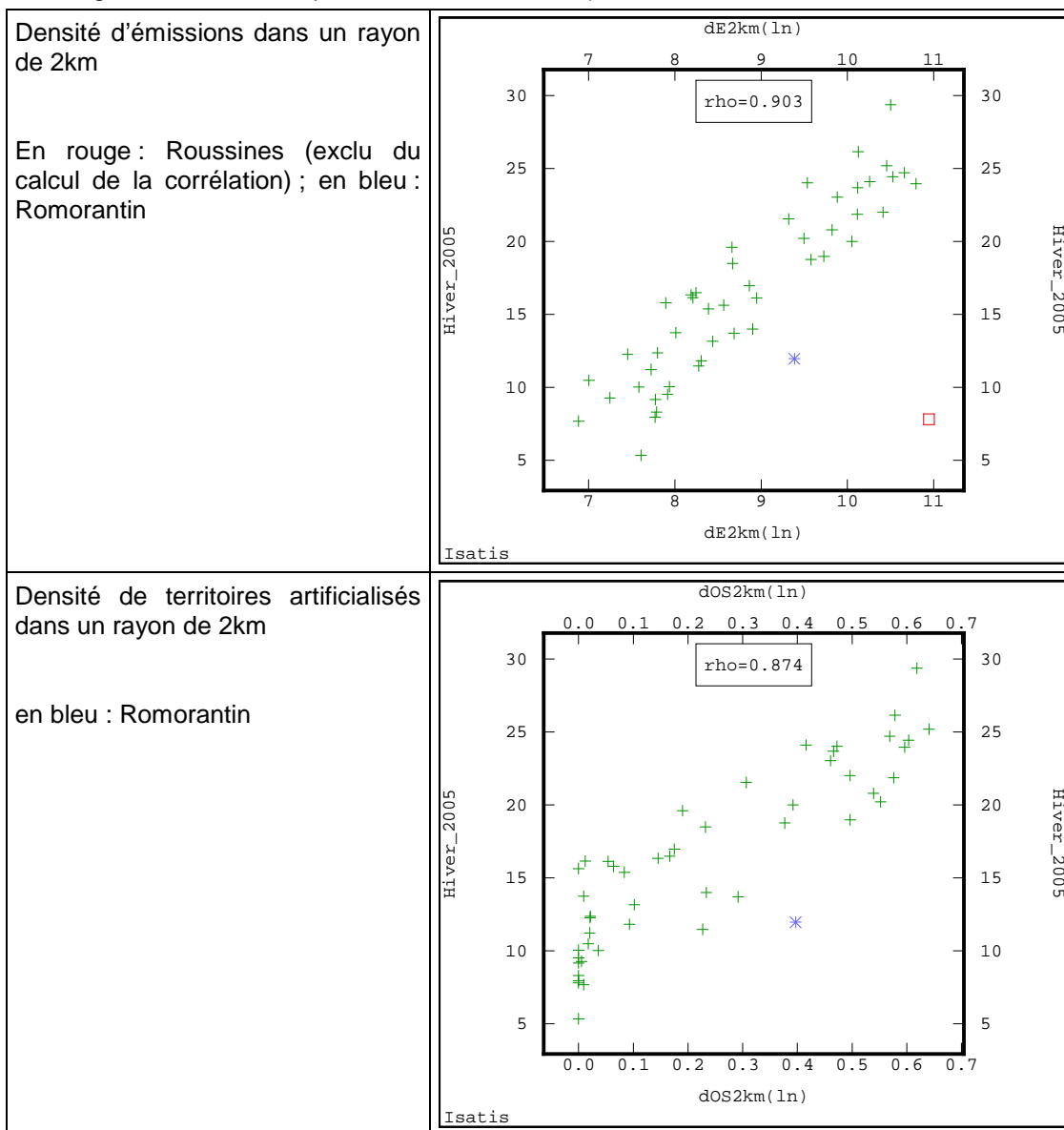
Tableau 2 – Coefficients de corrélation entre concentrations de fond de NO₂ et variables auxiliaires

Variable auxiliaire après transformation logarithmique	Remarque	Rayon de calcul de la densité				
		500 m	1 km	2 km	5 km	10 km
Densité des émissions de NOx	Sans le site de Roussines*	0.729	0.8508	0.903	0.894	0.824
Densité de « territoires artificialisés ¹ »	Classe 1**	0.781	0.824	0.874	0.856	0.766
Densité de population		0.766	0.772	0.846	0.884	Non calculée

¹ Classe de la nomenclature CORINE regroupant les zones urbanisées, les zones industrielles ou commerciales, les axes de circulation, les décharges et chantiers, les espaces verts artificialisés.

* : Un point de faible concentration, situé en zone rurale, se démarque par une densité d'émissions particulièrement élevée. Il s'agit du site de Roussines, proche d'une source industrielle de NOx. La diminution des émissions de cette source n'est que partiellement incluse dans la dernière version de l'inventaire.

** : la corrélation est maximale lorsqu'on fait la somme de toutes les sous-classes de territoires artificialisés (zones urbanisées, zones industrielles ou commerciales, axes de circulation, décharges et chantiers, espaces verts artificialisés.)



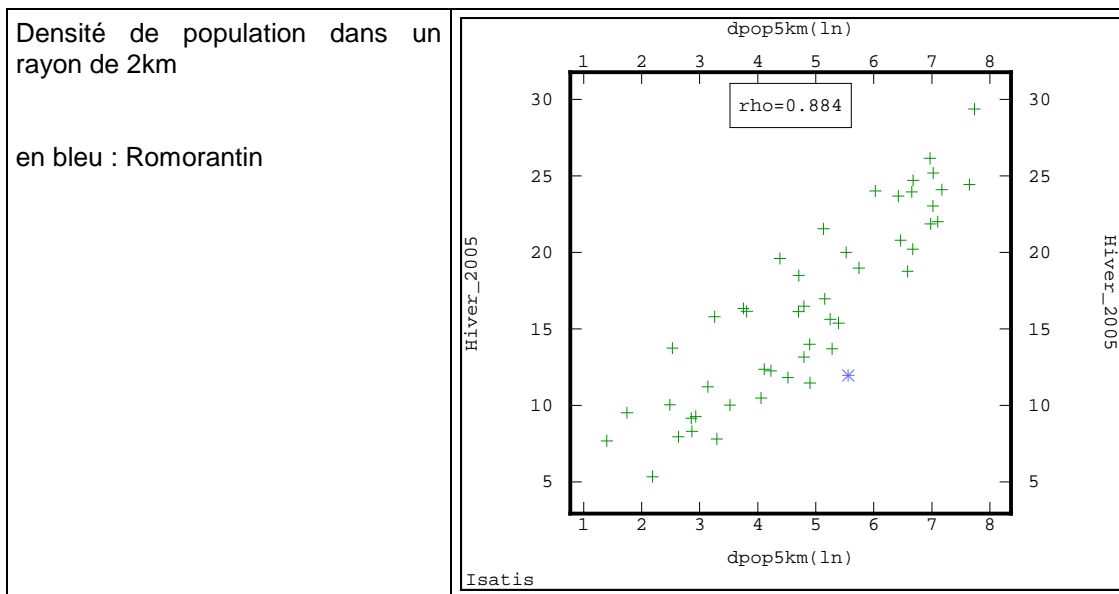


Figure 4 – Nuages de corrélation entre la concentration hivernale de NO_2 aux sites de fond et les variables auxiliaires qui lui sont le plus corrélées.

REMARQUE : Le traitement préalable des données d'émission représente un gain significatif en terme de corrélation par rapport à l'utilisation du cadastre kilométrique brut. En effet

- avant traitement, le coefficient de corrélation entre les concentrations moyennes hivernales de NO_2 et le logarithme du cadastre migré (i.e. à chaque site de mesure est affectée la valeur d'émission du point de grille le plus proche) est de 0,581. Le nuage de corrélation est dispersé.
- après traitement, le coefficient de corrélation entre les concentrations moyennes hivernales de NO_2 et logarithme de la densité d'émissions calculée par SIG dans un cercle de rayon de 2km est de 0,903. Le nuage de corrélation est plus linéaire.

4.2.1.4 SELECTION DE VARIABLES EXPLICATIVES

La concentration de NO_2 est bien corrélée à la densité d'émissions dans un rayon de 2 km, à la densité de territoires artificialisés dans un rayon de 2 km et à la densité de population dans un rayon de 5 km (après transformation logarithmique). Ces trois variables, notées respectivement dE2km(ln), dOS2km(ln), dpop5km(ln), sont sélectionnées pour la suite de l'analyse. Etant très corrélées entre elles, comme l'atteste une analyse en composantes principales – cf. glossaire- (Figure 5), elles sont prises en compte chacune séparément.

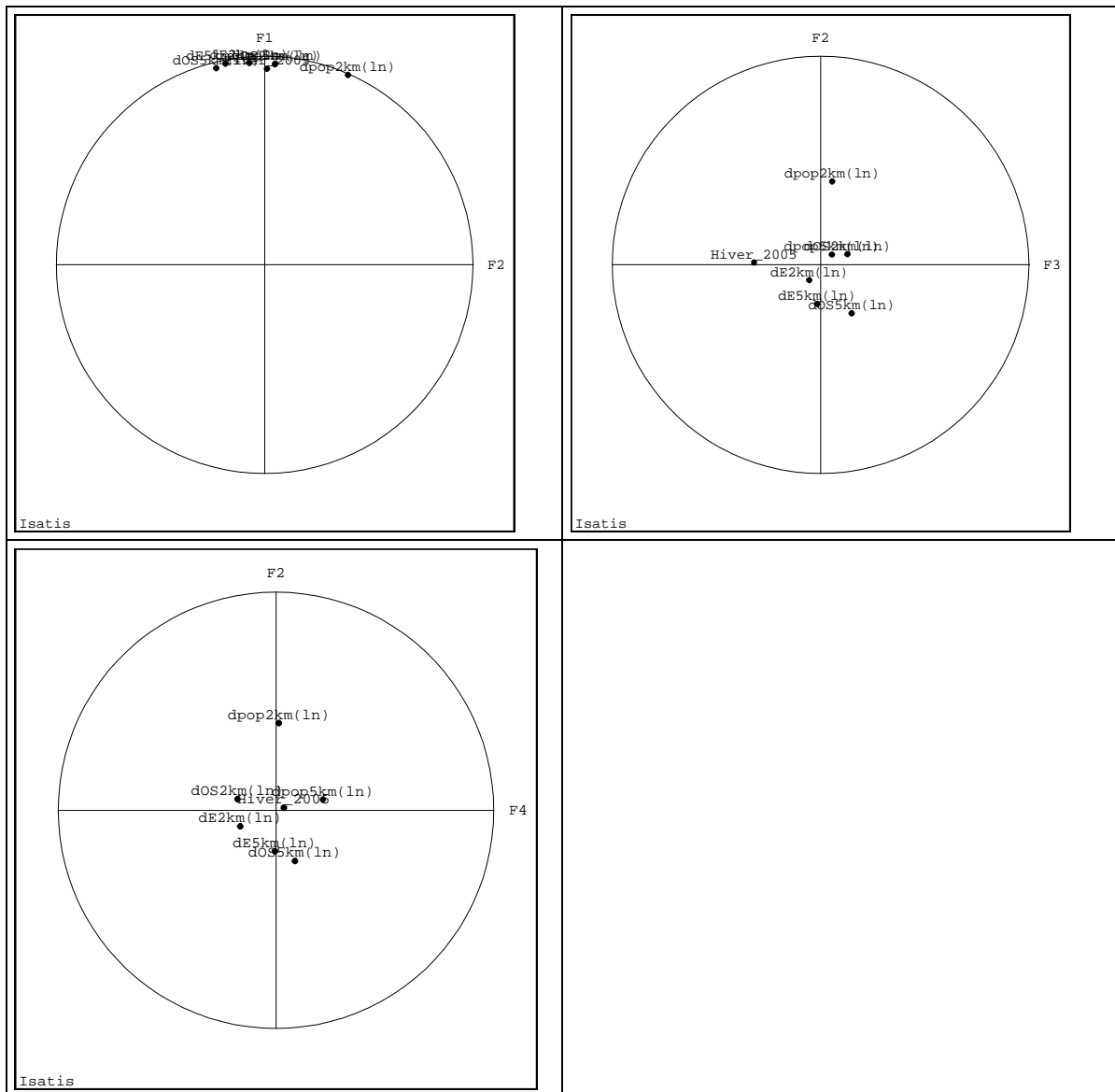


Figure 5 – Analyse en CP : projection sur les trois premiers plans factoriels des concentrations hivernales de NO₂ et des densités d'émissions, d'urbanisation et de population calculées dans un rayon de 2km et 5km.

La corrélation entre les variables de concentration, d'émissions, d'occupation du sol (territoires artificialisés) et de population est mise en évidence par l'ACP. Toutes ces variables sont très bien représentées par le premier plan factoriel -en particulier le 1^{er} axe- sur lequel elles se projettent à proximité les unes des autres. Les autres axes sont peu discriminants. Seul l'axe 2 sépare la densité de population à 2km des autres variables.

4.2.2 ETUDE VARIOGRAPHIQUE ET ESTIMATION

4.2.2.1 MODELISATION SANS ET AVEC PRISE EN COMPTE DES VARIABLES EXPLICATIVES

Afin d'apprécier l'amélioration due aux variables auxiliaires et l'efficacité des différents krigeages, plusieurs modélisations sont réalisées.

Méthode d'estimation envisagée	Variable auxiliaire	Variable dont on modélise le variogramme	Remarque	
Krigeage ordinaire	/	Hiver_2005	Cas de base permettant d'évaluer l'apport des variables auxiliaires	
Krigeage des résidus	$\ln(1+dE2km)$	Résidus de la régression Hiver_2005= $a \cdot \ln(1+dE2km)+b$	a=4.88 b=-26.67	$\bar{\varepsilon} = -0.000$ $\sigma_{\varepsilon}^2 = 6.39$
Krigeage des résidus	$\ln(1+dOS2km)$	Résidus de la régression Hiver_2005= $a \cdot \ln(1+dOS2km) +b$	a=22.92 b=10.85	$\bar{\varepsilon} = 0.000$ $\sigma_{\varepsilon}^2 = 8.37$
Krigeage des résidus	$\ln(1+dpop5km)$	Résidus de la régression Hiver_2005= $a \cdot \ln(1+dpop5km) +b$	a=3.15 b=0.99	$\bar{\varepsilon} = -0.000$ $\sigma_{\varepsilon}^2 = 7.78$
Krigeage avec dérive externe	$\ln(1+dE2km)$	Modélisation non stationnaire (option d'ISATIS) tenant compte de la dérive		
Krigeage avec dérive externe	$\ln(1+dOS2km)$	Modélisation non stationnaire (option d'ISATIS) tenant compte de la dérive		
Krigeage avec dérive externe	$\ln(1+dpop5km)$	Modélisation non stationnaire (option d'ISATIS) tenant compte de la dérive		
Cokrigeage colocalisé	$\ln(1+dE2km)$	Hiver_2005 et $\ln(1+dE2km)$	Modélisation effectuée seulement pour les émissions à titre de test	

a, b : coefficient d la régression. $\bar{\varepsilon}$, σ_{ε}^2 : moyenne et variance des résidus de la régression linéaire.

Dans le cas du krigeage ordinaire, du krigeage des résidus et du cokrigeage colocalisé, les variogrammes sont ajustés manuellement (Figure 6). Ils sont du type *effet de pépité + structure sphérique*. Pour le krigeage avec dérive externe, la procédure de modélisation non stationnaire est employée. Les paramètres des modèles sont fournis en annexe 3.

Krigeage ordinaire			
	Variogramme de la concentration de NO ₂		

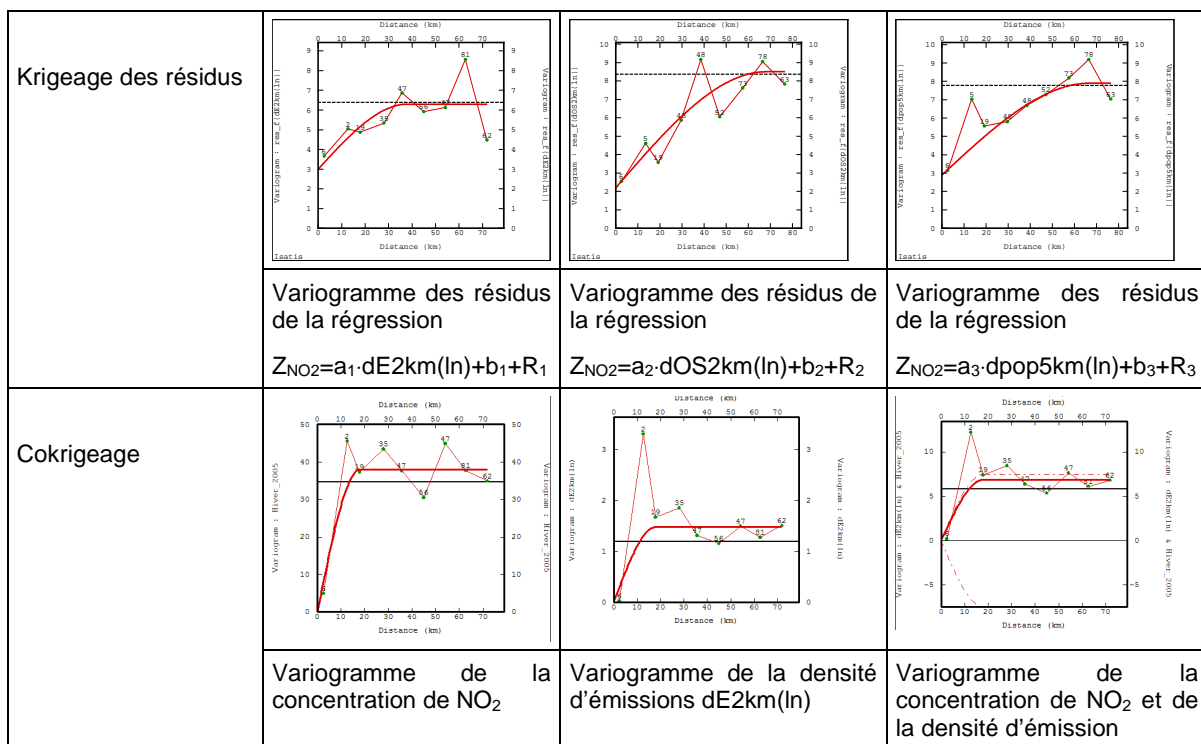


Figure 6 – Variogrammes expérimentaux et modélisés

4.2.2.2 EVALUATION ET COMPARAISON DES MODELES PAR VALIDATION CROISEE

L'adéquation des différents modèles aux données expérimentales est contrôlée par validation croisée. Les critères d'évaluation sont les suivants :

- moyenne et variance des erreurs et des erreurs standardisées ;
- statistiques des erreurs relatives ;
- corrélation entre concentrations mesurées et estimées.

Différents voisinages d'estimation sont également testés. Le krigeage avec dérive externe et le krigeage des résidus sont ici les plus efficaces. La comparaison des statistiques d'erreur (voir annexe 3) conduit à choisir :

- pour la densité d'émissions à 2km : le krigeage avec dérive externe, avec le modèle défini par Isatis et un voisinage unique (cela revient donc à effectuer un krigeage des résidus mais ici, les paramètres variographiques ajustés par Isatis diffèrent de ceux que nous avons ajustés visuellement);
- pour la densité de population : le krigeage des résidus, avec le modèle ajusté sur le variogramme expérimental des résidus et un voisinage glissant de 60 km de rayon ;
- pour la densité de territoires urbanisés : le krigeage des résidus, avec le modèle ajusté sur le variogramme expérimental des résidus et un voisinage glissant de 60 km de rayon.

Les résultats sont meilleurs avec les deux premières variables (site de Roussines exclus pour les émissions). La variance des erreurs d'estimation est relativement faible (inférieure à $7 (\mu\text{g}/\text{m}^3)^2$); celle des erreurs standardisées s'approche de la valeur théorique unitaire. En moyenne, l'erreur relative n'excède pas 15% et le

coefficient de corrélation entre estimations et mesures est supérieur à 0,9. (En krigeage ordinaire ou cokrigeage multicolocalisé, cette corrélation est inférieure à 0,7.)

Dans l'estimation finale, ces deux variables seront utilisées de préférence à l'occupation du sol.

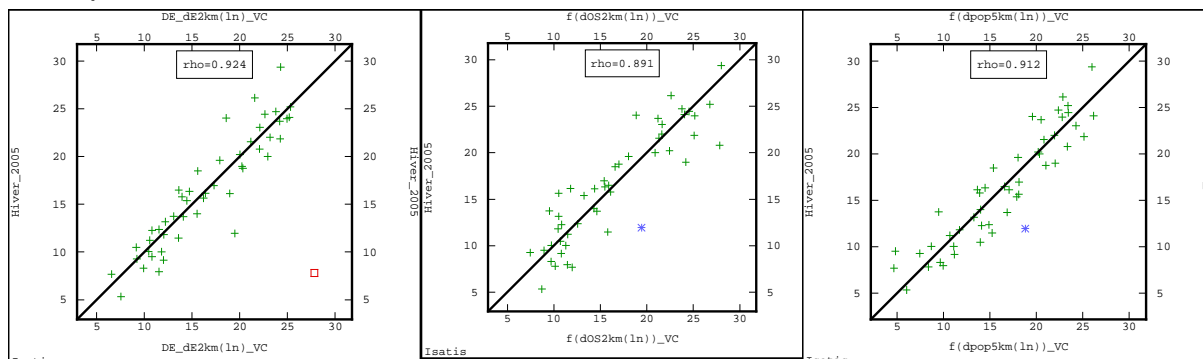


Figure 7 – Corrélation entre concentrations estimées par validation croisée (abscisses) et concentrations mesurées (ordonnées). En rouge : site de Roussines ; en bleu : Romorantin

4.2.2.3 DEFINITION D'UNE GRILLE D'ESTIMATION ET REALISATION DES CARTOGRAPHIES

Compte tenu :

- de la dimension de la zone d'étude : environ 250 km x 200 km ;
- de la distance minimale entre sites : 1 à 2 km entre points intra-urbains, environ 15km entre points interurbains ;
- du type de pollution à cartographier : pollution de fond ;

une maille d'estimation de 5km de côté est choisie.

La mise en œuvre du krigeage des résidus ou du krigeage avec dérive externe impose de calculer les variables auxiliaires en tout point de la grille d'estimation. Les programmes de calcul développés dans l'analyse exploratoire sont donc appliqués à tous les centres de maille.

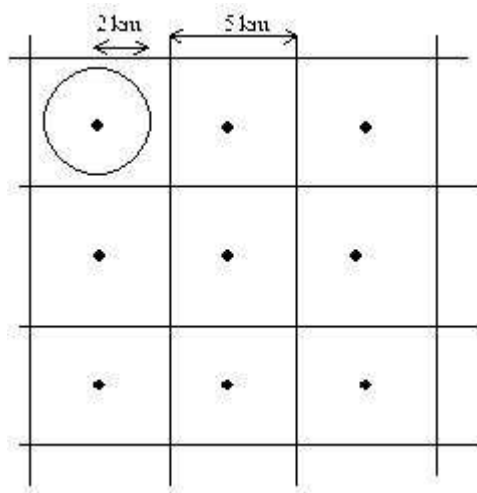
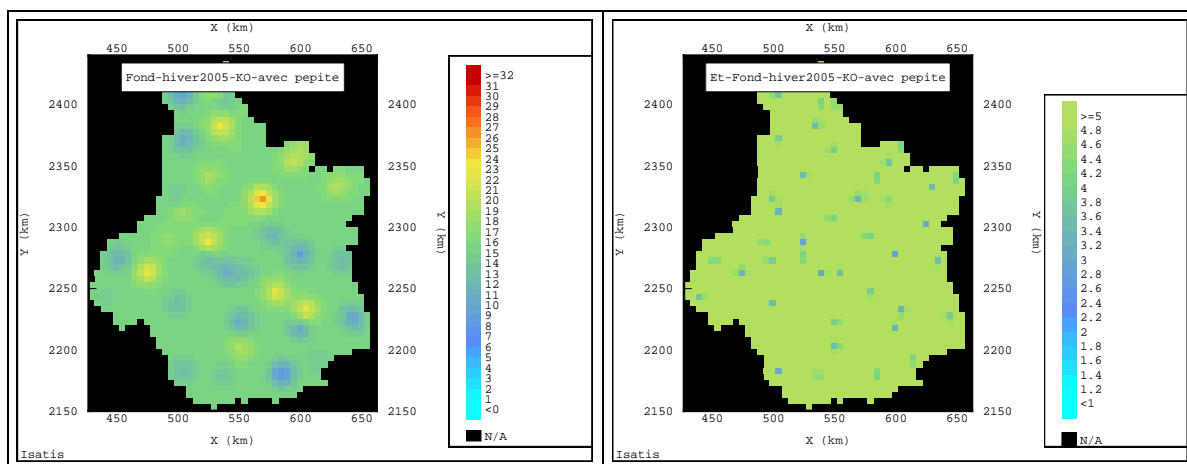


Figure 8 – Calcul des densités d'émissions et de territoires artificialisés dans un rayon de 2km autour des points d'estimation (i.e. des centres de maille). La procédure est identique pour la densité de population, mais les cercles de rayon 5km se coupent.

Les concentrations sont alors estimées par krigeage sur toute la région, aux centres de maille d'une part (krigeage ponctuel), en moyenne par maille (krigeage de blocs). Nous présentons les résultats du krigeage ponctuel. Les estimations par krigeage de blocs sont très similaires.



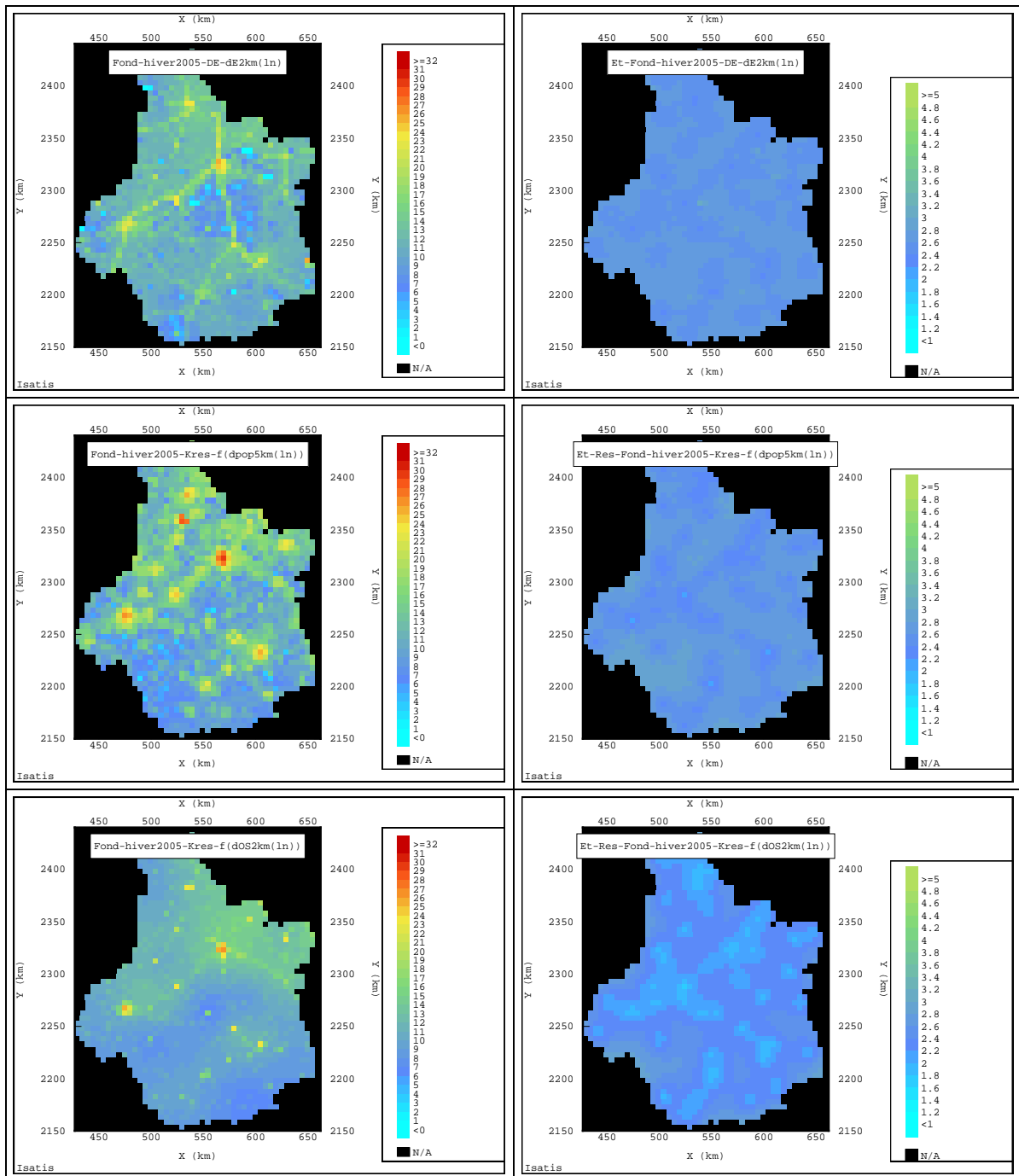


Figure 9 – Concentrations estimées par krigeage ponctuel (colonne de gauche) et écart-type de l'erreur d'estimation (colonne de droite). De haut en bas : krigeage ordinaire ; krigeage avec la densité d'émissions à 2km en dérive externe; krigeage des résidus avec la densité de population à 5km ; krigeage des résidus avec la densité de territoires artificialisés à 2km.

Dans la carte de krigeage ordinaire, les concentrations, uniformes sur la région, augmentent ou diminuent localement autour des sites de mesure, formant des taches colorées peu réalistes. L'écart-type de krigeage est relativement élevé. L'utilisation d'une variable auxiliaire modifie sensiblement la cartographie : elle abaisse les niveaux de concentration loin des villes et des routes; elle maintient des valeurs relativement élevées le long des routes.

Le choix de la variable auxiliaire est également déterminant. Avec la densité d'émissions, le tracé des routes ressort nettement. Avec la densité de population, il apparaît de manière plus diffuse alors que la pollution des villes est intensifiée. Toutefois, cette différence pourrait s'expliquer en partie par la méthode d'estimation de la densité de population et par les hypothèses faites sur la répartition spatiale des habitants. Des **données de population à plus haute résolution** (IRIS et îlots) permettraient de le vérifier. Ce résultat montre **l'importance de disposer d'une variable auxiliaire fiable et précise.**

Avec les territoires urbanisés, la variabilité des concentrations est très atténuée.

4.3 PRISE EN COMPTE DES MESURES DE PROXIMITE

Le krigeage avec variable auxiliaire permet de rendre compte des variations de concentration dans les zones urbaines ou routières mais n'exploitant que les mesures de fond, il ne restitue pas nécessairement les valeurs observées près des voies de circulation ou dans les centres-villes.

Cette partie a pour but d'évaluer, dans ces zones, l'écart entre estimation de fond et pollution réelle et si possible, de corriger cette estimation.

4.3.1 EVALUATION DES ESTIMATIONS DE FOND DANS LES VILLES ET LES ZONES DE PROXIMITE

4.3.1.1 KRIGEAGE AUX POINTS DES STATIONS FIXES URBAINES ; COMPARAISON ENTRE VALEURS ESTIMEES ET MESUREES

Les concentrations de fond de NO₂ sont estimées par krigeage aux points des stations fixes. Elles sont comparées aux valeurs mesurées par ces stations sur la même période (janvier-mars 2005).

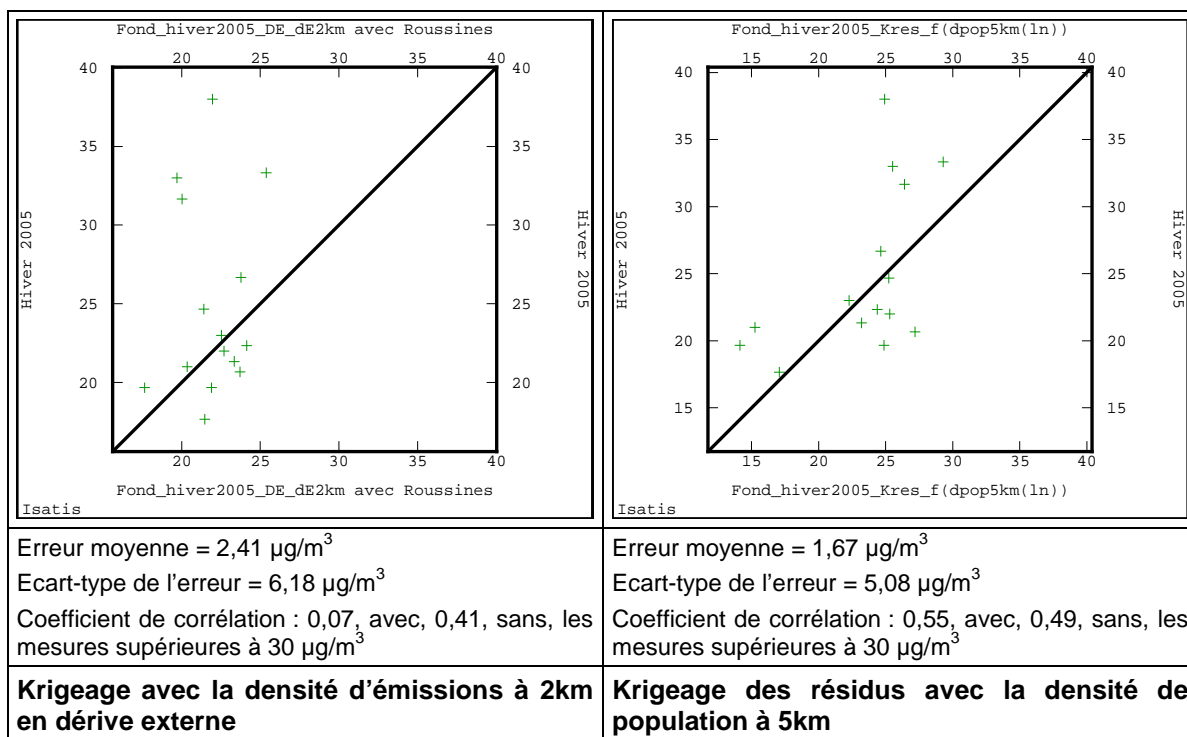


Figure 10 – Comparaison entre estimations de fond et mesures aux stations fixes urbaines

Dans l'ensemble, le krigeage avec variable auxiliaire restitue correctement les niveaux de fond urbains. A Tours et Orléans, les valeurs supérieures à 30 $\mu\text{g}/\text{m}^3$ sont sous-estimées : dans ces villes, les mesures des tubes utilisées dans le krigeage sont comprises entre 22 et 30 $\mu\text{g}/\text{m}^3$. Cette sous-estimation est moins marquée lorsque la densité de population est prise comme variable auxiliaire (un site d'Orléans fait exception, le site de La Source où la concentration enregistrée atteint 38 $\mu\text{g}/\text{m}^3$ pour une densité de population relativement faible ; cette valeur n'est donc pas bien restituée).

4.3.1.2 KRIGEAGE AUX POINTS DES SITES DE PROXIMITE ; COMPARAISON ENTRE VALEURS ESTIMEES ET MEASUREES

Les **concentrations de fond de NO₂** sont estimées par krigeage aux 19 points de proximité, selon la démarche décrite précédemment. Les données d'entrée sont exclusivement les 49 mesures hivernales de fond, et les variables auxiliaires, la densité d'émissions à 2km ou la densité de population à 5km calculées en ces 19 sites.

Les valeurs estimées sont comparées aux valeurs mesurées en ces points durant la même période de l'année 2004 (Figure 12). A défaut de mesures disponibles, **on suppose en effet que les concentrations de 2005 sont identiques à celles de 2004**. Cette hypothèse est testée sur les deux sites fixes de proximité de la région (Figure 11). Elle se vérifie pour la station d'Orléans. A Tours en revanche, les concentrations de 2005 conservent le profil temporel de l'année précédente mais augmentent d'environ 10 $\mu\text{g}/\text{m}^3$.

Malgré ces différences, l'hypothèse est maintenue afin de pouvoir mener à bien la méthodologie proposée.

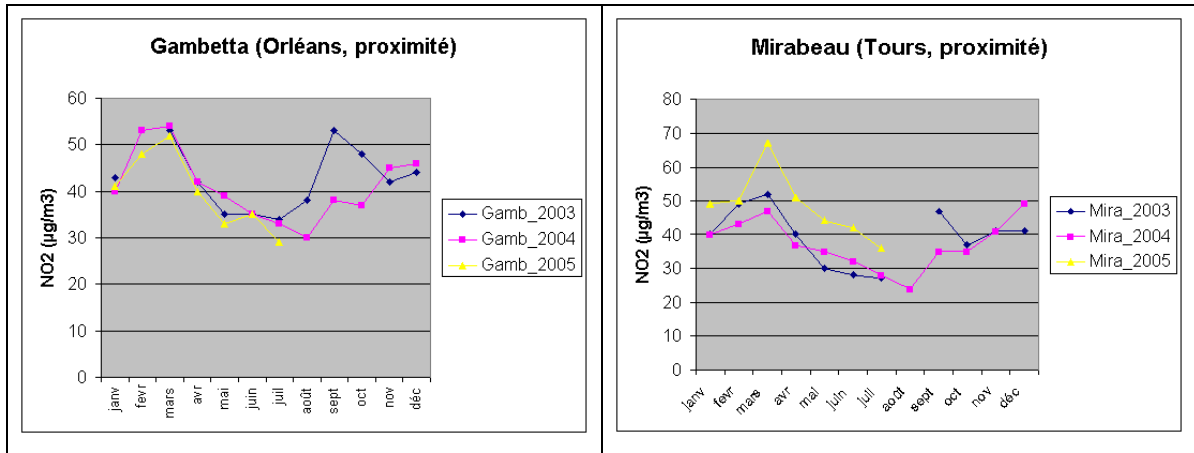


Figure 11 – Evolution interannuelle des concentrations aux stations de proximité de la région Centre

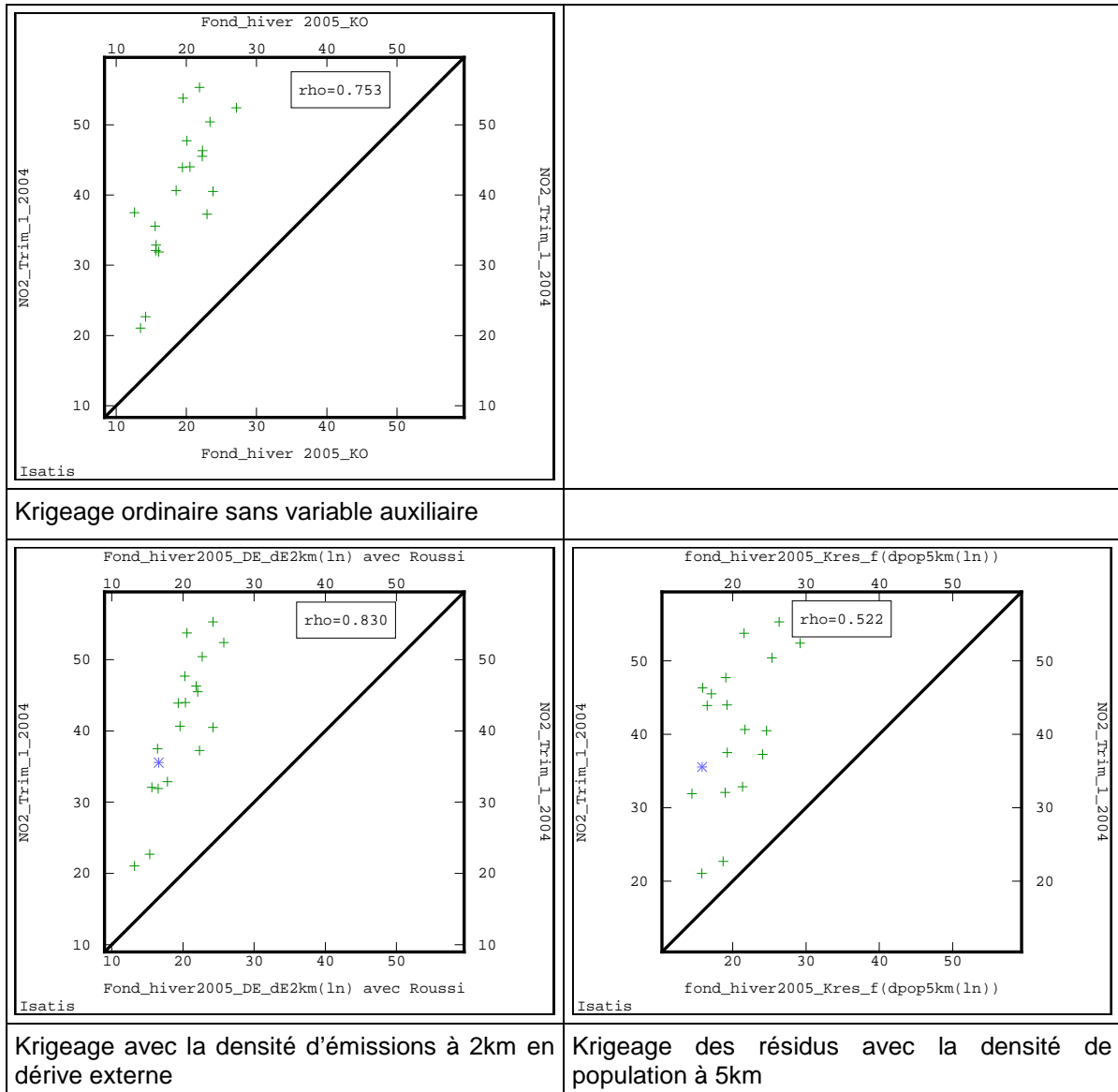


Figure 12 – Comparaison entre estimations de fond et mesures aux sites d'échantillonnage de proximité (campagne 2004)

Les concentrations estimées aux points de proximité sont corrélées linéairement aux mesures. Toutefois, malgré l'usage de variables auxiliaires, la sous-estimation est importante (entre 32% et 62%). Ce résultat ne surprend pas car si, comme les données de fond, les concentrations de proximité sont corrélées aux variables auxiliaires, les relations de régression sont différentes. Prenons l'exemple de la densité d'émissions :

$$Z_{\text{NO}_2\text{fond}} = 4,88 \cdot dE2\text{km}(\ln) - 26,67 + \varepsilon$$

$$Z_{\text{NO}_2\text{proximité}} = 12,09 \cdot dE2\text{km}(\ln) - 76,15 + \varepsilon$$

Pente et constante sont multipliées par un facteur compris entre 2 et 3.

Une étude plus complète des corrélations entre mesures de proximité et variables auxiliaires est jointe en annexe 4.

Il convient cependant de nuancer ces résultats puisque les mesures de proximité ont été effectuées à moins de 10 m des axes routiers. Or les concentrations décroissent rapidement avec la distance à la route : à environ 200 m, elles atteignent leur niveau de fond [4] [5].

Afin de quantifier plus précisément la différence entre estimations de fond et observations de proximité, il serait utile :

- de connaître les distances exactes entre les points d'échantillonnage et les routes qui influencent la mesure effectuée en ces points (non accessible au moment de l'étude, cette information sera exploitée en 2006);
- de disposer de données sur des transects perpendiculaires aux divers types de routes (autoroutes, nationales, départementales).

Dans cette étude, les données de proximité sont supposées représentatives des concentrations moyennes sur quelques dizaines de mètres de part et d'autre des routes.

4.3.2 EXPLICATION ET MODELISATION STATISTIQUE DE LA DIFFERENCE ENTRE ESTIMATION DE FOND ET MESURE DE PROXIMITE ROUTIERE

Nous voulons ici expliquer la différence (notée Δ_{P-F}) entre mesures de proximité et concentrations de fond estimées par krigeage. Cet écart, calculé expérimentalement aux 19 points de proximité, est représenté en fonction des variables auxiliaires.

La corrélation est maximale pour la densité d'émissions dans un rayon d'1km :

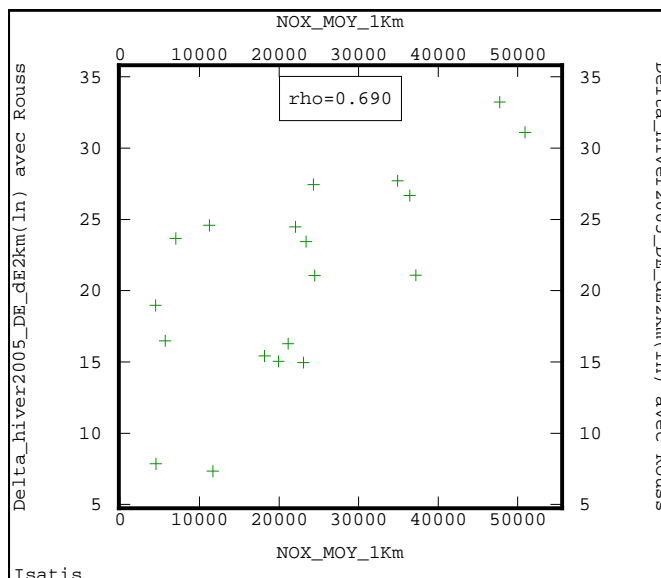


Figure 13 – Corrélation entre l'écart Δ_{P-F} (estimation fond-mesure de proximité) et la densité d'émissions à 1km

Il est donc possible de modéliser statistiquement Δ_{P-F} . Avec la variable d'émission sélectionnée, la relation de régression s'écrit :

$$\Delta_{P-F} = 3,56 \cdot 10^{-4} \cdot dE1km + 12,87 + \varepsilon$$

Une relation unique est établie pour l'ensemble des sites. Avec des données complémentaires², une modélisation plus élaborée pourrait être envisagée : distinction du type de route et de la part d'émissions d'origine automobile ; pondération en fonction de la distance à la route.

Nous proposons d'appliquer la démarche qui suit :

- raffinement de la maille dans les zones de proximité (choix d'une maille de 500m de côté) puis, aux centres des petites mailles ainsi définies,
- calcul de la ou des variables auxiliaires nécessaires ;
- estimation des niveaux de fond ;
- estimation de l'écart *proximité-fond* (Δ_{P-F});
- estimation de la concentration de proximité comme somme du niveau de fond et de l'écart *proximité-fond*.

La différence Δ_{P-F} obtenue dans notre étude peut être surestimée du fait de la position des tubes de proximité. Dans un but méthodologique, la procédure de correction proposée est néanmoins appliquée.

² Certaines de ces données ont été mises récemment à la disposition de l'INERIS et seront exploitées en 2006.

4.3.3 CORRECTION DES ESTIMATIONS EN ZONE DE PROXIMITE ROUTIERE

4.3.3.1 RESSERREMENT DE LA GRILLE D'ESTIMATION DANS LES ZONES DE PROXIMITE ROUTIERE

Le resserrement local de la maille d'estimation est **effectué avec le SIG Arcview**, à partir du fichier des routes de la région Centre.

La grille de résolution 5km est subdivisée en mailles plus fines de 250 m de côté.

- Pour tenir compte de la largeur des routes, on délimite autour de celles-ci une bande de 400 m de large.
- Toutes les petites mailles qui coupent cette bande sont sélectionnées.

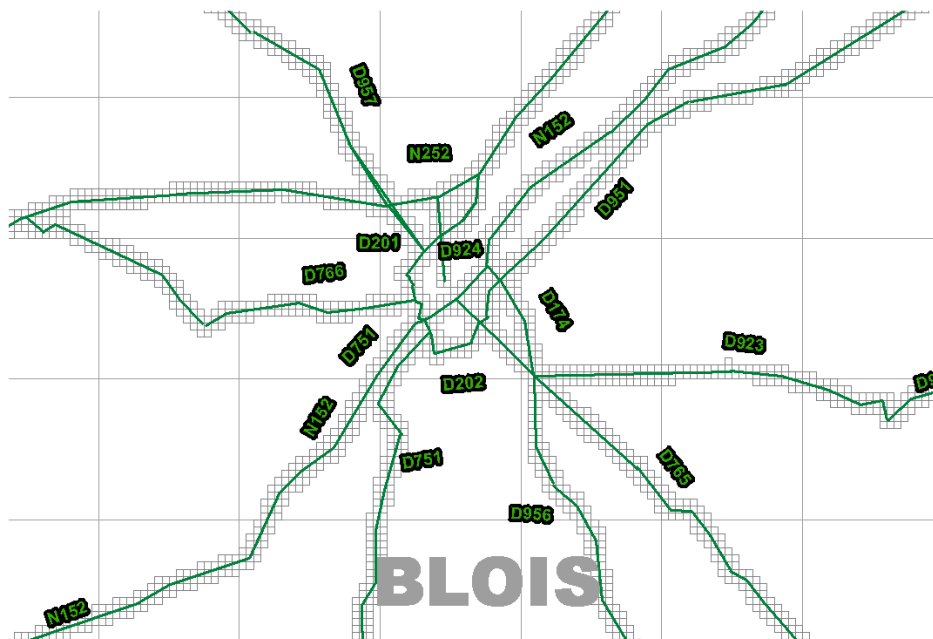


Figure 14 – Resserrement de la maille . Zoom sur la ville de Blois

4.3.3.2 CALCUL DES VARIABLES AUXILIAIRES SUR LA GRILLE FINE

En théorie, les variables auxiliaires peuvent être calculées de la même façon que pour le maillage de pas 5km. **En pratique se pose le problème du temps de calcul.**

Nous nous sommes donc restreint, à titre d'exemple :

- à une zone centrée sur la ville de Blois (Figure 14) ;
- au calcul de la densité d'émissions dans un rayon de 1 km.

Les centres des mailles kilométriques sélectionnées, caractérisés par leurs coordonnées (X,Y), et les densités d'émissions associées sont importés dans Isatis **en tant que points**. Les étapes décrites ci-après sont effectuées dans Isatis. Le maillage raffiné n'est pas traité comme un fichier *grille* mais comme un fichier *points*.

4.3.3.3 ESTIMATION DES CONCENTRATIONS DE FOND SUR LES MAILLES PLUS FINES

Faute d'avoir pu calculer les variables auxiliaires à la résolution 250m (densité d'émissions dans un rayon de 2km ou densité de population dans un rayon de 5km), les concentrations de fond ne sont pas réestimées sur le maillage fin.

Au centre de chaque petite maille est affectée par migration³ la concentration moyenne (résultant du krigeage de blocs) de la grande maille qui le contient. Ce choix se justifie puisqu'il s'agit d'une pollution de fond.

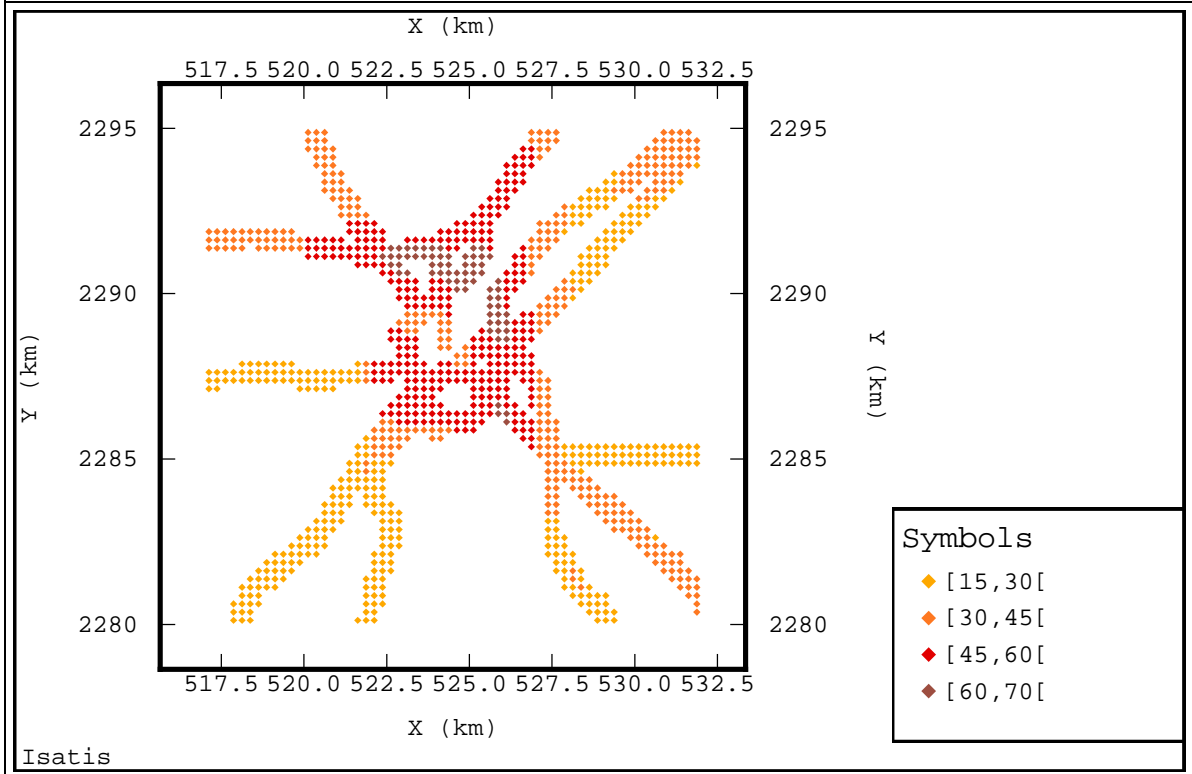
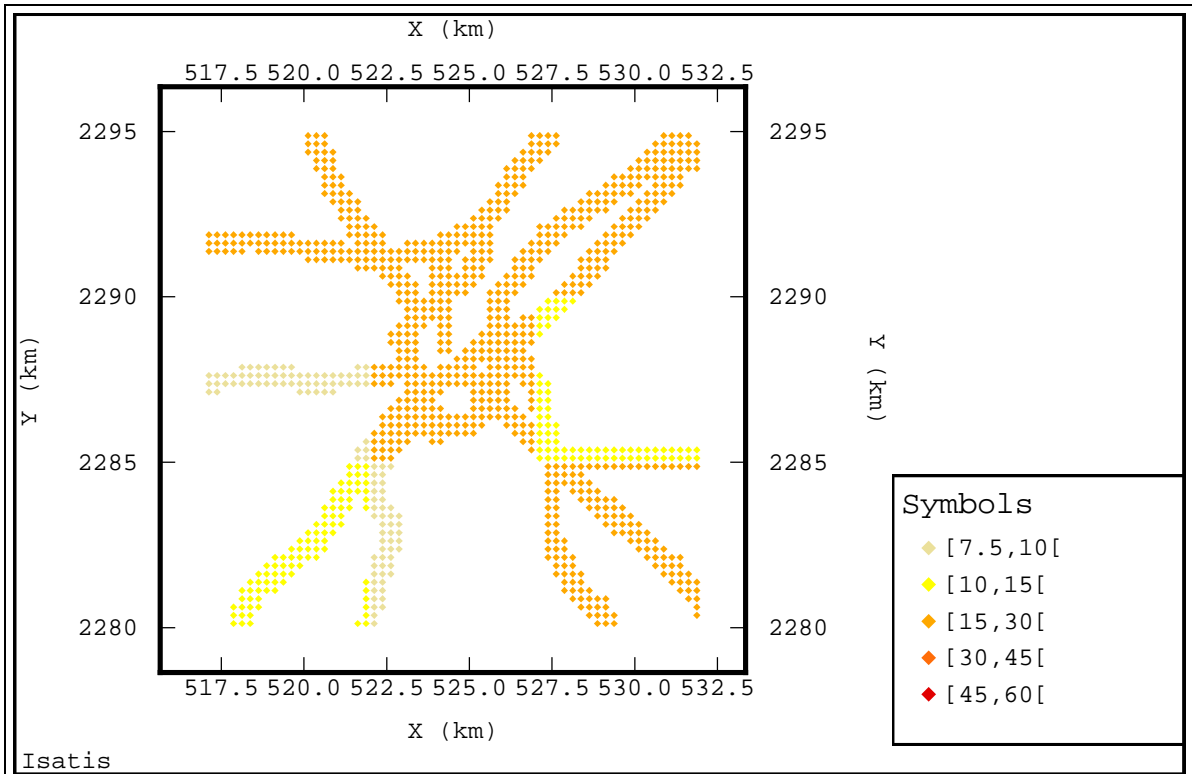
4.3.3.4 CALCUL ET AJOUT DU TERME CORRECTIF A CES VALEURS ESTIMEES

Le terme correctif associé à chaque petite maille est calculé selon la relation de régression établie en 4.3.2. puis ajouté à la valeur de fond estimée.

Plus important aux croisements des routes, il varie entre 12 et 44 $\mu\text{g}/\text{m}^3$ dans la zone considérée. Comme nous l'avons fait remarquer aux paragraphes 4.3.1.2 et 4.3.2, la correction est établie sur des mesures effectuées à quelques mètres seulement des voies et ne tient pas compte de la décroissance des concentrations avec la distance aux axes.

A titre de comparaison, nous présentons les estimations annuelles faites par LIG'Air avec le logiciel Street. Celles-ci montrent également des niveaux élevés sur la nationale 252 au Nord de la ville. Sur les autres axes, elles s'approchent plutôt du krigeage de fond.

³ Migration de grille à points



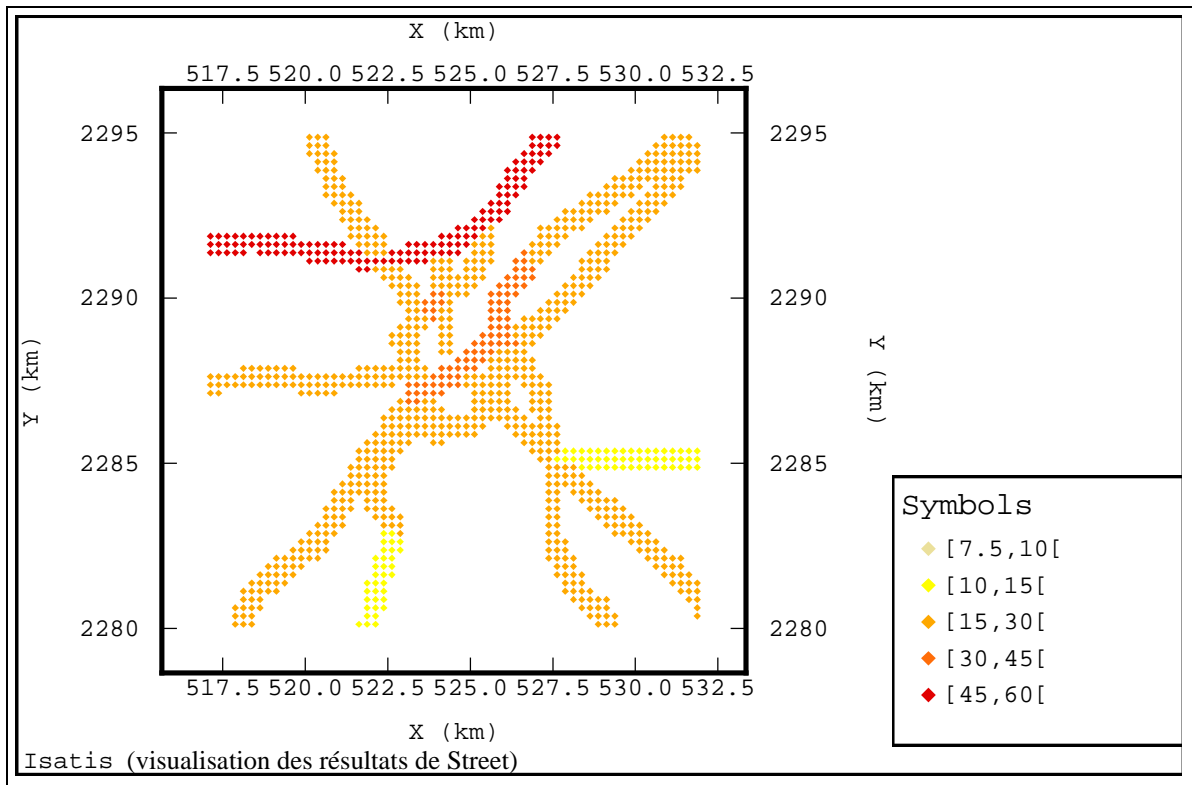


Figure 15 – Correction des niveaux de fond à proximité des routes. Exemple de la zone de Blois. De haut en bas : estimation de fond avec la densité d'émissions à 2km en dérive externe ; estimation corrigée en fonction de la densité d'émissions à 1km ; résultats du logiciel Street.

Selon le temps de calcul requis, cette procédure pourra être mise en œuvre zone par zone ou sur l'ensemble du maillage fin.

4.3.3.5 REPRESENTATION FINALE

Deux types de représentation sont proposés :

- une carte des concentrations de fond, accompagnée de zooms sur les zones corrigées ;
- une représentation simultanée de la pollution de fond et de proximité, si la correction a pu s'appliquer à toutes les zones de proximité routière de la région. On superpose à cette fin les deux grilles d'estimation (5km et 250 m) dans un SIG (Figure 16.)

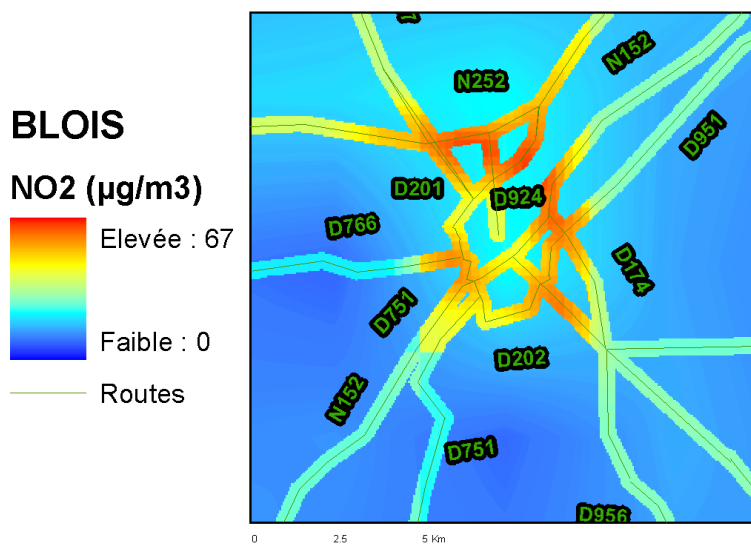


Figure 16 – Superposition des concentrations de fond et de proximité estimées dans la zone de Blois

5. CONCLUSION

L'objet de ce travail est de concevoir une méthodologie permettant de prendre en compte des données de fond et de proximité dans une même cartographie. Les données d'échantillonnage exploitées à cette fin ont été fournies par l'association LIG'Air. Elles portent sur la cartographie régionale du dioxyde d'azote.

Leur analyse nous conduit à proposer la démarche suivante, démarche dont l'application est limitée par les données et ressources informatiques disponibles.

1) Cartographie du fond

La première étape consiste à effectuer un krigeage des mesures de fond. Une représentation plus précise des gradients de concentration nécessite l'introduction d'une ou plusieurs variables explicatives. Deux techniques donnent ici satisfaction : le krigeage avec dérive externe ou la régression (multi)linéaire suivie d'un krigeage des résidus.

Remarques sur la recherche, le calcul et le choix des variables explicatives :

- Les variables auxiliaires utilisées sont l'inventaire des émissions, l'occupation du sol et le recensement par commune. A partir de ces informations, **des densités d'émissions, d'occupation du sol et de population** sont calculées dans des **cercles de différents rayons** (de 500m à 10km) autour des sites de mesure. Cette étape requiert le **développement et la mise en œuvre de programmes de calcul sous SIG**. Les résultats sont importés dans Isatis.
- Le traitement des données auxiliaires brutes permet de lisser les nuages de corrélation (effet de moyenne) et de tenir compte de la distance d'influence de ces variables. **Ainsi, le coefficient de corrélation augmente puis diminue avec le rayon considéré**, montrant un maximum entre 2km et 5km. La densité d'émissions calculée dans un rayon de 2km et la densité de population

calculée dans un rayon de 5km (après transformation logarithmique) sont ici les mieux corrélées aux concentrations de fond. Le coefficient de corrélation est d'environ 0,9. Ces variables étant très corrélées entre elles, soit l'une soit l'autre est utilisée.

- Les variables de densité sélectionnées sont calculées par SIG au centre de chaque maille de la grille d'estimation (une maille de 5km de côté a été définie).
- Le choix de la variable explicative influence sensiblement la cartographie. Il est donc important de disposer de données auxiliaires fiables et précises sur le domaine d'étude.

2) Modélisation de la différence entre fond et proximité

L'introduction de variables explicatives permet de représenter l'augmentation des concentrations près des sources de pollution mais ne rend pas compte des niveaux atteints dans les zones de proximité. Les mesures de proximité routière peuvent être exploitées afin de modéliser statistiquement, en fonction d'une ou plusieurs variables auxiliaires (préalablement calculées aux points de proximité), la part de concentration sous-estimée.

Dans cette étude, l'écart entre proximité et fond est exprimé comme une fonction linéaire de la densité d'émissions à 1km. Une modélisation plus fine de cette différence pourrait inclure le type de voie et la distance à la route . Des **mesures réalisées sur des transects perpendiculaires aux routes** seraient notamment utiles.

A l'échelle de la carte, l'estimation du fond urbain est satisfaisante, comme l'indique la comparaison avec les mesures de stations fixes. Quelques moyennes élevées sont sous-estimées au centre des plus grandes agglomérations (Tours, Orléans) mais cette différence s'explique.

3) Correction de la carte de fond

Le maillage est raffiné autour des routes. La taille de maille est de 250m, compromis entre la nécessité de représenter les variations de petite échelle et les contraintes numériques et graphiques liées aux dimensions du domaine. La densité d'émissions à 1km est calculée au centre de chaque petite maille. Malgré la capacité de calcul du SIG Arcview, **cette étape se heurte à un problème de ressources informatiques** (saturation de la mémoire). Par suite, le calcul des variables auxiliaires sur le maillage fin doit se faire morceau par morceau. A cause des délais impartis pour l'étude, la méthode de correction n'est testée que sur un domaine réduit. Une zone centrée sur Blois a été définie à titre d'exemple.

La concentration de fond d'une petite maille est supposée égale à l'estimation par krigeage de bloc de la grande maille qui l'englobe. Le terme correctif, déduit du modèle statistique établi en 2), est ajouté à cette valeur.

La possibilité d'étendre la procédure à l'ensemble de la région Centre sera évaluée en 2006.

Travaux futurs

L'étude a porté sur les données hivernales du premier trimestre 2005. Si le phénomène perd en structure spatiale, une analyse préliminaire des données

estivales indique que celles-ci sont également corrélées aux variables auxiliaires et suggère d'appliquer la même méthode. D'autre part, les données de l'année entière seront prochainement disponibles.

A l'occasion du programme LCSQA de 2006⁴, il est proposé de compléter le travail par la cartographie estivale et annuelle du NO₂. On pourra s'appuyer sur des études antérieures qui ont montré qu'à l'échelle urbaine, la prise en compte des corrélations entre saisons améliorerait la précision de l'estimation [6 à 9]. Nous chercherons également à affiner le modèle de correction dans les zones de proximité. Les nouvelles campagnes de proximité prévues par LIG'Air pour 2006 devraient fournir des données utiles à la validation. Une comparaison de la méthode avec d'autres approches (modélisation empirique et déterministe) sera également effectuée.

6. REFERENCES

- [1] Rouïl L., Malherbe L., 2003. Méthodes de représentation de la qualité de l'air dans les zones peu/pas couvertes – Guide d'utilisation de la géostatistique linéaire. Rapport LCSQA – www.lcsqa.org
- [2] Cardenas G., Malherbe L., 2002. Méthodes de représentation de la qualité de l'air dans les zones peu/pas couvertes – Application à la problématique d'une association. Rapport LCSQA – www.lcsqa.org
- [3] LIG'AIR, 2004. Qualité de l'Air. Proximité d'axes circulants. Campagne régionale. Mesure du dioxyde d'azote et du benzène. Rapport final, disponible sur le site www.ligair.fr
- [4] Séguret S., 2003. Estimation du dioxyde d'azote routier dans la vallée de la Thur, Rapport N-1/03/G.
- [5] ASPA, 2002. Observatoire de la qualité de l'air dans la vallée de la Thur. Rapport relatif à la campagne de mesure qui s'est déroulée du 12 juin 2001 au 4 septembre 2001 et du 29 novembre 2001 au 21 février 2002. Rapport 02051701-ID
- [6] Cardenas G., Malherbe L., 2004. Géostatistique et prise en compte de l'aspect temporel. Rapport LCSQA – www.lcsqa.org
- [7] 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.
- [8] 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).
- [9] Fouquet C. (de), Gallois D., Malherbe L., Cardenas G., 2004. Mapping annual nitrogen dioxide concentrations above Mulhouse urban area. GeoEnv 2004, Neuchâtel (Suisse), 13-15 octobre.

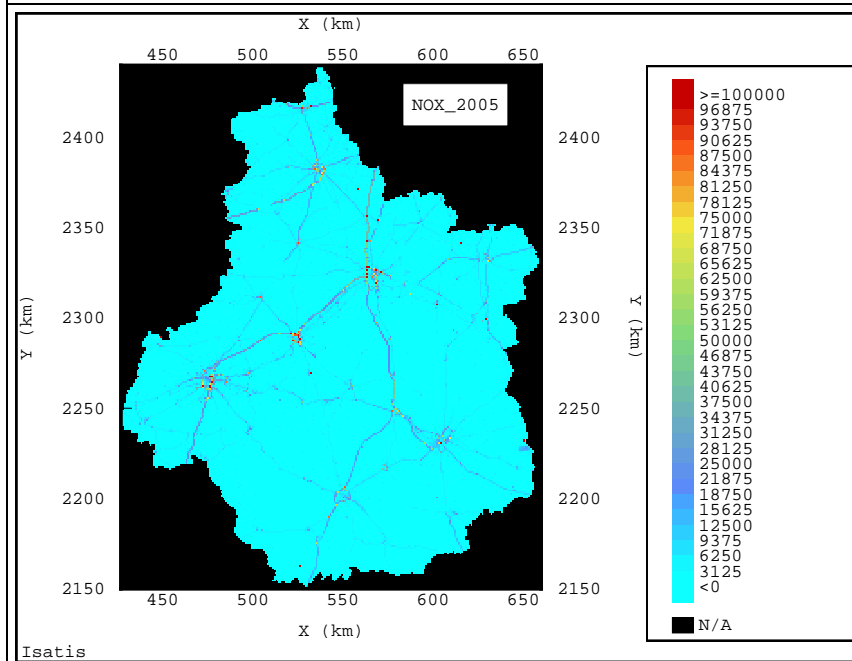
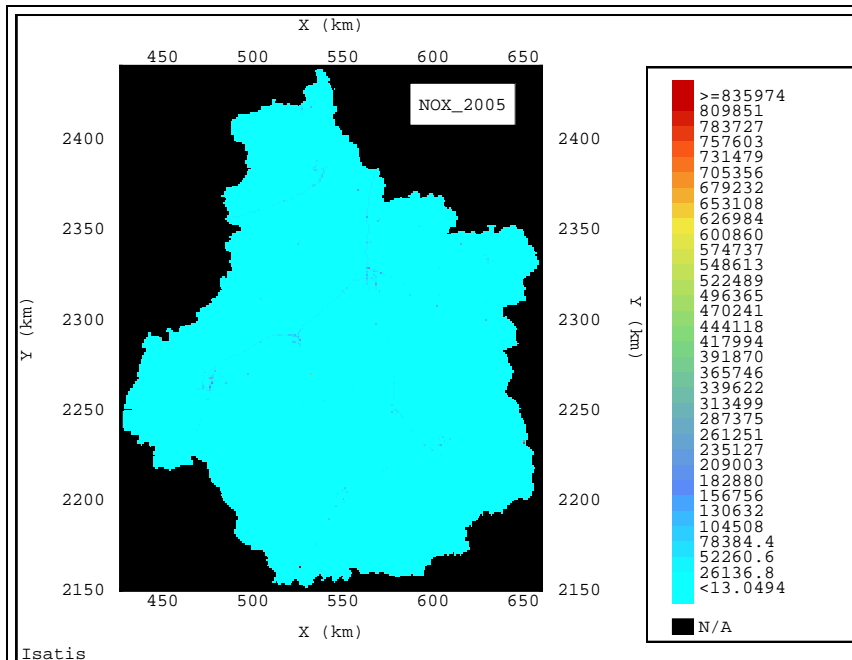
⁴ Synthèse des techniques de cartographie de NO₂ : valorisation des travaux 2003 – 2005.

7. LISTE DES ANNEXES

Repère	Désignation	Nombre de pages
Annexe 1	Variables auxiliaires	4 A4
Annexe 2	Estimation de la densité de population	2 A4
Annexe 3	Etude des données de fond	2 A4
Annexe 4	Etude des données de proximité	10 A4

Annexe 1 : variables auxiliaires

Emissions de NOx:



Même figure que précédemment avec une échelle de valeurs réduite

Figure A11 - Emissions de NOx dans la région Centre – cadastre kilométrique de 2005. Pour une meilleure lisibilité, l'échelle a été réduite dans la figure du bas

Le trafic moyen journalier est également fourni :

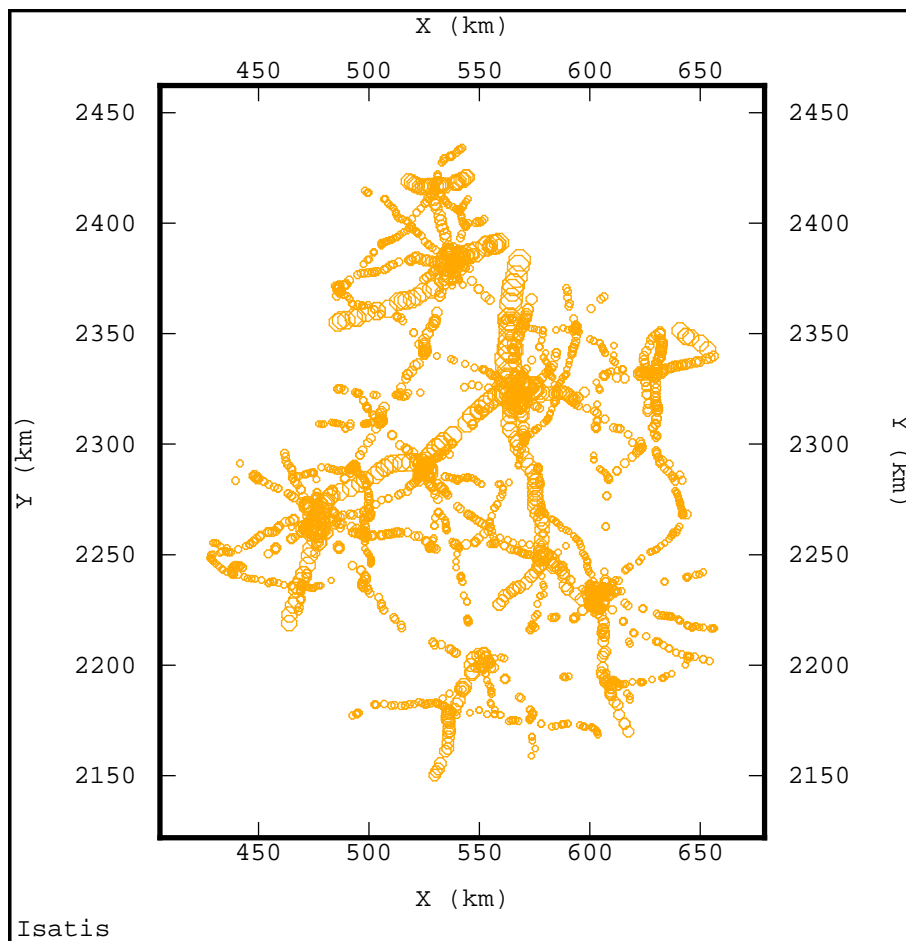


Figure A12 - Représentation du trafic moyen journalier de l'année 2003. La taille des cercles est proportionnelle au trafic.

Occupation du sol

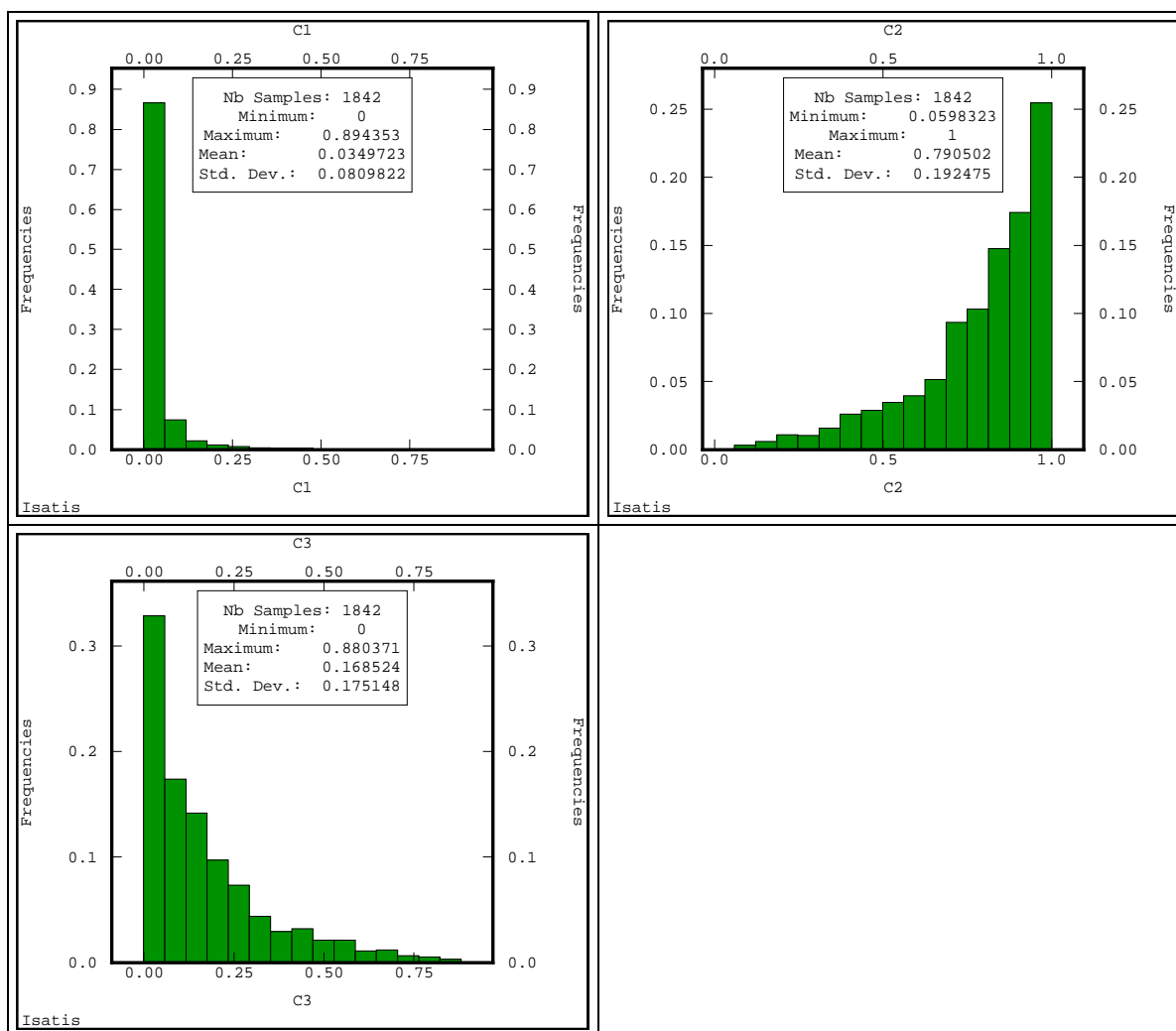


Figure A13 – Taux d'occupation des communes par les territoires artificialisés (C1), les territoires agricoles (C2) et les forêts (C3)

Données de population par commune

Tableau A11 – Statistiques des données communales

Variable	N	Min	Max	Moyenne arithmétique	Variance
Superficie (km ²)	1842	1.33	122.89	21.43	218.81
Nombre d'habitants	1842	23.0	132820.0	1325.07	27855254.48
Densité communale d'habitants (/km ²)	1842	1.53	4085.27	75.31	59328.21

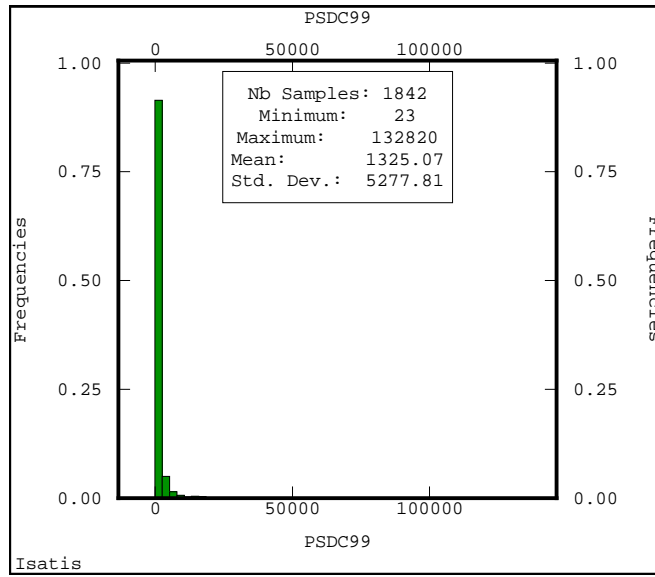


Figure A14 - Histogramme de la population par commune

Annexe 2 : estimation de la densité de population

Soit r , le rayon de calcul de la densité autour des sites de mesures.

1) On estime, à l'échelle des communes, la répartition de la population entre les classes d'occupation du sol.

La relation recherchée s'écrit :

$$P_i = \sum_j P_{ij} = \sum_j a_j S_{ij} + \varepsilon \quad i \in \{1..n\} \quad a_j \geq 0$$

n est le nombre de communes.

$\{j\}_{j \in \text{CORINE}}$ constitue une partition de l'occupation du sol. Les classes considérées sont celles de niveau 2 (1.1, 1.2...).

S_{ij} est la surface de la commune i occupée par le territoire de type j .

Les coefficients a_j sont estimés par régression multilinéaire. Seules les classes présentant une corrélation partielle positive avec la population sont prises en compte ; les autres sont supposées vides d'habitants. Sur les 1842 communes, les trois quarts (1382) sont utilisées dans la construction du modèle ; le quart restant est réservé à la validation. Les coefficients suivants sont finalement retenus :

$a_{11}=2071,355$ ($p < 2.10^{-16}$)
$a_{12}=1075,706$ ($p < 2.10^{-16}$)
$a_{14}=1356,828$ ($p < 2.10^{-16}$)
$a_{23}=7,827$ ($p = 0.079$)
$a_{24}=5,982$ ($p = 0.131$)

Les coefficients a_{23} et a_{24} , quoique non significatifs selon le critère conventionnel de la p -value, sont conservés. En effet, près de 50% de la population de la région Centre vit hors des centres urbains. Pour les autres catégories, ils sont fixés à zéro.

Etant donné une commune i de population P_i connue, la population P_{ij} présente sur l'ensemble des surfaces de type j est estimée par :

$$P_{ij} = \frac{a_j S_{ij}}{\sum_k a_k S_{ik}} \cdot P_i$$

Elle est supposée se distribuer uniformément sur ces surfaces, avec une densité égale à $\frac{P_{ij}}{S_{ij}}$.

2) On estime la population totale dans les cercles de rayon r autour des tubes.

Etant donné un site de mesure T . Pour chaque commune i interceptée par le cercle de rayon r ($i \in \{1 \dots n_T\}$), un calcul sous SIG permet d'évaluer la surface s_{ij} occupée par le territoire j ($j \in \{1 \dots 5\}$). Compte tenu de l'hypothèse précédente, la population présente sur cette surface peut être estimée par :

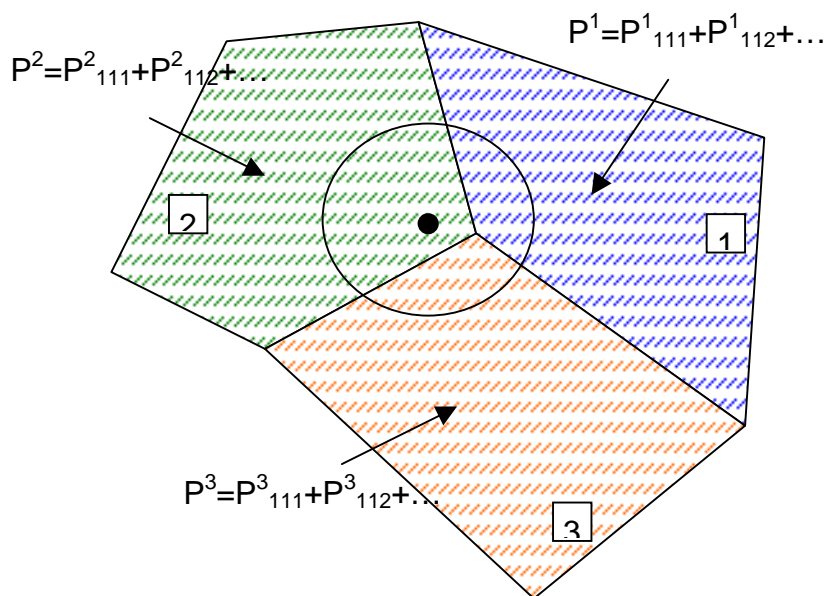
$$p_{ij}^T = s_{ij}^T \frac{P_{ij}}{S_{ij}}$$

et la population totale à l'intérieur du cercle est alors :

$$p_T = \sum_1^{n_T} \sum_1^5 s_{ij} \frac{P_{ij}}{S_{ij}}$$

La densité de population correspondante vaut :

$$d_{p_T} = \frac{p_T}{\pi \cdot r^2}$$



Annexe 3 : étude des données de fond - modélisation variographique et validation croisée

Krigeage ordinaire	Pépité 4 Sphérique (33.5, 20500 m)		
	Variogramme de la concentration de NO₂		
Krigeage des résidus	Pépité 3 Sphérique (3.3, 38000m)	Pépité 2.2 Sphérique (6.3, 68000m)	Pépité 2.9 Sphérique (5.0, 67000m)
	Variogramme des résidus de la régression $Z_{NO_2}=a_1 \cdot dE2km(ln)+b_1+R_1$	Variogramme des résidus de la régression $Z_{NO_2}=a_2 \cdot dOS2km(ln)+b_2+R_2$	Variogramme des résidus de la régression $Z_{NO_2}=a_3 \cdot dpop5km(ln)+b_3+R_3$
Krigeage avec dérive externe	<i>Dérives proposées :</i> dE2km(ln), x, y <i>Structures proposées :</i> Pépité, structures sphériques de différentes portées <i>Dérives retenues :</i> dE2km(ln) <i>Structures retenues :</i> Pépité, structure sphérique de portée 30000m	<i>Dérives proposées :</i> dE2km(ln), x, y <i>Structures proposées :</i> Pépité, structures sphériques de différentes portées <i>Dérives retenues :</i> dE2km(ln) <i>Structures retenues :</i> Pépité, structure sphérique de portée 40000m	<i>Dérives proposées :</i> dE2km(ln), x, y <i>Structures proposées :</i> Pépité, structures sphériques de différentes portées <i>Dérives retenues :</i> dE2km(ln) <i>Structures retenues :</i> Pépité, structure sphérique de portée 50000m
	Modélisation non stationnaire en voisinage unique	Modélisation non stationnaire en voisinage unique	Modélisation non stationnaire en voisinage unique
Cokrigeage	Sphérique (38.0, 18000 m)	Sphérique (1.5, 18000 m)	Sphérique (6.9, 18000 m)
	Variogramme de la concentration de NO₂	Variogramme de la densité d'émissions dE2km(ln)	Variogramme de la concentration de NO₂ et de la densité d'émission

Résultats de validation croisée

	KO_VU	Kres_V60km dE2km(ln)	Kres_V60km dOS2km(ln)	Kres_V60km dpop5km(ln)	KDE_VU dE2km(ln)	KDE_VU dOS2km(ln)	KDE_VU dpop5km(ln)	CK_VU
N	49	48	49	49	48	49	49	48
m_e	1.169	0.119	0.078	0.068	0.016	0.091	0.049	0.320
σ_e²	19.99	5.970	7.649	6.146	6.827	7.819	7.207	6.488
m_{e_std}	0.136	0.034	0.018	0.018	0.003	0.017	0.009	0.112
σ_{e_std}²	0.605	0.961	1.260	1.015	1.007	1.035	1.000	2.240
min_{e_rel}		0.34	0.22	0.05	0.43	0.23	0.62	
P25_{e_rel}		3.83	4.28	5.67	4.08	7.37	8.04	
P50_{e_rel}		10.27	10.47	12.35	7.88	12.61	14.47	
P75_{e_rel}		17.16	20.50	17.18	16.46	20.04	21.41	
max_{e_rel}		76.26	63.33	57.41	62.94	104.65	53.27	
m_{e_rel}		14.11	15.45	14.31	11.80	17.48	16.75	
σ_{e_rel}		15.10	15.73	12.16	12.11	18.17	11.92	
ρ	0.680	0.911	0.891	0.912	0.924	0.883	0.893	0.640

Légende :

N	Nombre de données
m_e	Moyenne de l'erreur
σ_e²	Variance de l'erreur
m_{e_std}	Moyenne de l'erreur standardisée (i.e. rapportée à l'écart-type de krigeage)
σ_{e_std}²	Variance de l'erreur standardisée
min_{e_rel}	Minimum de l'erreur relative
P25_{e_rel}	Quantile 25 de l'erreur relative
P50_{e_rel}	Médiane de l'erreur relative
P75_{e_rel}	Quantile 75 de l'erreur relative
max_{e_rel}	Maximum de l'erreur relative
m_{e_rel}	Moyenne de l'erreur relative
σ_{e_rel}	Ecart-type de l'erreur relative
ρ	Coefficient de corrélation entre valeurs estimées et observées

KO_VU	Krigeage ordinaire en voisinage unique
Kres_V60km dE2km(ln)	Krigeage des résidus, avec la densité d'émissions dans un rayon de 2km comme variable auxiliaire, voisinage d'estimation de 60 km
Kres_V60km dOS2km(ln)	Krigeage des résidus, avec la densité de territoires "artificialisés" dans un rayon de 2km comme variable auxiliaire, voisinage d'estimation de 60 km
Kres_V60km dpop5km(ln)	Krigeage des résidus, avec la densité de population dans un rayon de 5km comme variable auxiliaire, voisinage d'estimation de 60 km
KDE_VU dE2km(ln)	Krigeage avec la densité d'émissions dans un rayon de 2km en dérive externe, voisinage d'estimation de 60 km
KDE_VU dOS2km(ln)	Krigeage avec la densité de territoires "artificialisés" dans un rayon de 2km en dérive externe, voisinage d'estimation de 60 km
KDE_VU dpop5km(ln)	Krigeage avec la densité de population dans un rayon de 5km en dérive externe, voisinage d'estimation de 60 km
CK_VU	Cokrigeage avec la densité d'émissions dans un rayon de 2km, voisinage unique

Annexe 4 : étude des données de proximité et de leurs corrélations avec les variables auxiliaires

7.1 STATISTIQUES ELEMENTAIRES MONODIMENSIONNELLES

7.1.1 STATISTIQUES DESCRIPTIVES

Les statistiques descriptives sont calculées site par site et sur l'ensemble des points de mesure.

Tableau 3 - Campagne de proximité NO₂. Statistiques élémentaires

	Numéro du site	Commune	Moyenne annuelle (µg/m ³)	Nb de mesures	% de données valides	Min (µg/m ³)	Médiane (µg/m ³)	P80 (µg/m ³)	Max (µg/m ³)	Ecart-type (µg/m ³)	CV
a	15	AUBIGNY-SUR-NERE		12	0.0	-	-	-	-	-	-
b	10	BLOIS	42.4	12	83.3	33.8	42.4	46.9	49.4	4.9	11.6
c	21	BOURGES	51.5	12	100.0	44.8	51.2	54.7	58.2	4.2	8.2
d	3	CHARTRES	41.7	12	100.0	35.4	39.8	44.1	56.6	5.7	13.6
e	4	CHATEAUDUN	54.0	12	100.0	42.2	56.1	60.3	62.0	6.9	12.7
f	9	CHATEAUNEUF-SUR-LOIRE	30.0	12	100.0	26.3	29.0	32.4	35.5	3.0	10.0
g	19	CHATEAUROUX	41.7	12	100.0	34.9	40.5	46.2	47.5	4.2	10.1
h	16	CHINON	21.3	12	91.7	14.7	22.6	24.1	27.6	4.3	20.0
i	1	DREUX	44.5	12	100.0	37.2	43.8	45.7	53.7	4.0	8.9
j	11	GIEN	31.0	12	91.7	25.2	32.1	33.5	34.3	2.9	9.3
k	20	ISSOUDUN	31.7	12	83.3	23.2	31.6	36.8	38.7	5.2	16.4
l	23	LA CHATRE		12	0.0	-	-	-	-	-	-
m	17	LOCHES		12	0.0	-	-	-	-	-	-
n	6	MONTARGIS	46.7	12	100.0	36.4	46.8	50.5	55.8	5.4	11.6
o	13	MONTRICHARD		12	0.0	-	-	-	-	-	-
p	2	NOGENT-LE-ROU	30.5	12	58.3	23.3	29.9	34.8	41.9	6.6	21.7
q	8	ORLEANS	49.2	12	100.0	43.3	47.2	53.9	62.2	5.7	11.5
r	5	PITHIVIERS	41.7	12	100.0	36.6	40.4	44.0	48.4	3.7	8.8
s	14	ROMORANTIN-LANTHENAY	31.8	12	91.7	24.8	32.8	36.8	39.1	5.2	16.3
t	22	SAINT-MARCEL	20.9	12	100.0	15.7	20.5	24.0	25.4	3.2	15.3
u	12	TOURS	59.6	12	83.3	50.0	57.8	63.9	76.5	7.8	13.1
v	7	VENDOME	42.7	12	100.0	37.3	41.3	42.7	62.2	6.7	15.7
w	18	VIERZON	52.8	12	83.3	37.9	54.3	59.1	64.0	8.4	15.8
Moyenne (µg/m ³)			40.3								
Ecart-type (µg/m ³)			11.0								
CV (%%)			27.4								

NB : compte tenu du nombre limité de sites, les moyennes annuelles sont conservées même lorsque le pourcentage de mesures mensuelles est inférieur à 75%.

7.1.2 HISTOGRAMMES

Dans la suite on désigne par été la moyenne des concentrations mesurées d'avril à septembre 2003 et par hiver la moyenne des concentrations mesurées d'octobre 2003 à mars 2004.

Sur plus de la moitié des sites, les concentrations saisonnières et annuelles sont supérieures à 40 µg/m³. Ces sites, sauf un, sont identiques en été et en hiver (Figure 17). Les situations estivale et hivernale se ressemblent mais la variabilité spatiale est plus grande en été, comme le suggère aussi le graphique des séries temporelles.

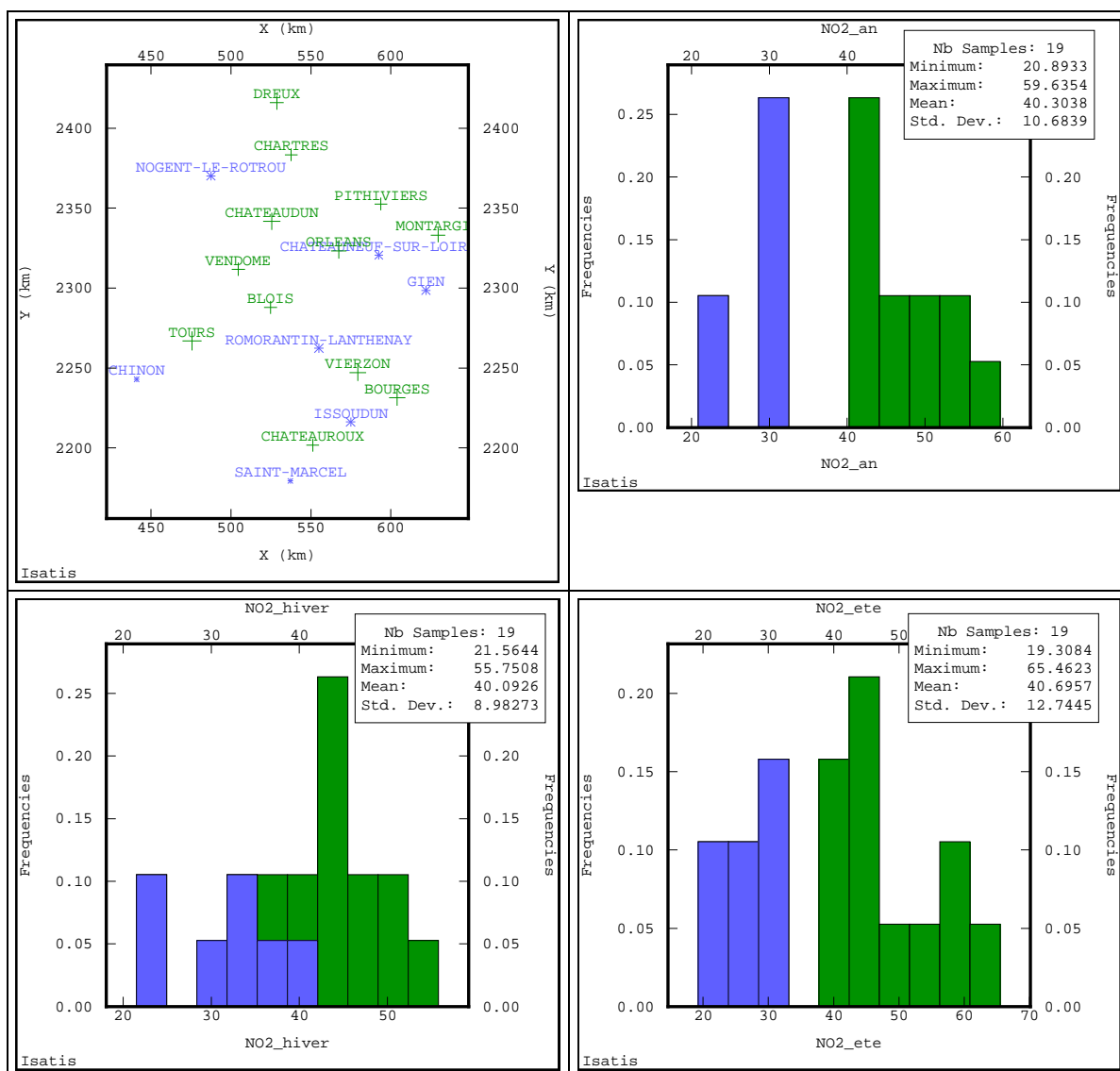


Figure 17 – Histogrammes des concentrations moyennes par saison

7.1.3 CORRELATION ETE/HIVER

Les corrélations entre trimestres sont élevées (Figure 18). Il en va de même entre la saison estivale complète (avril à septembre) et la saison hivernale qui lui succède (octobre à mars).

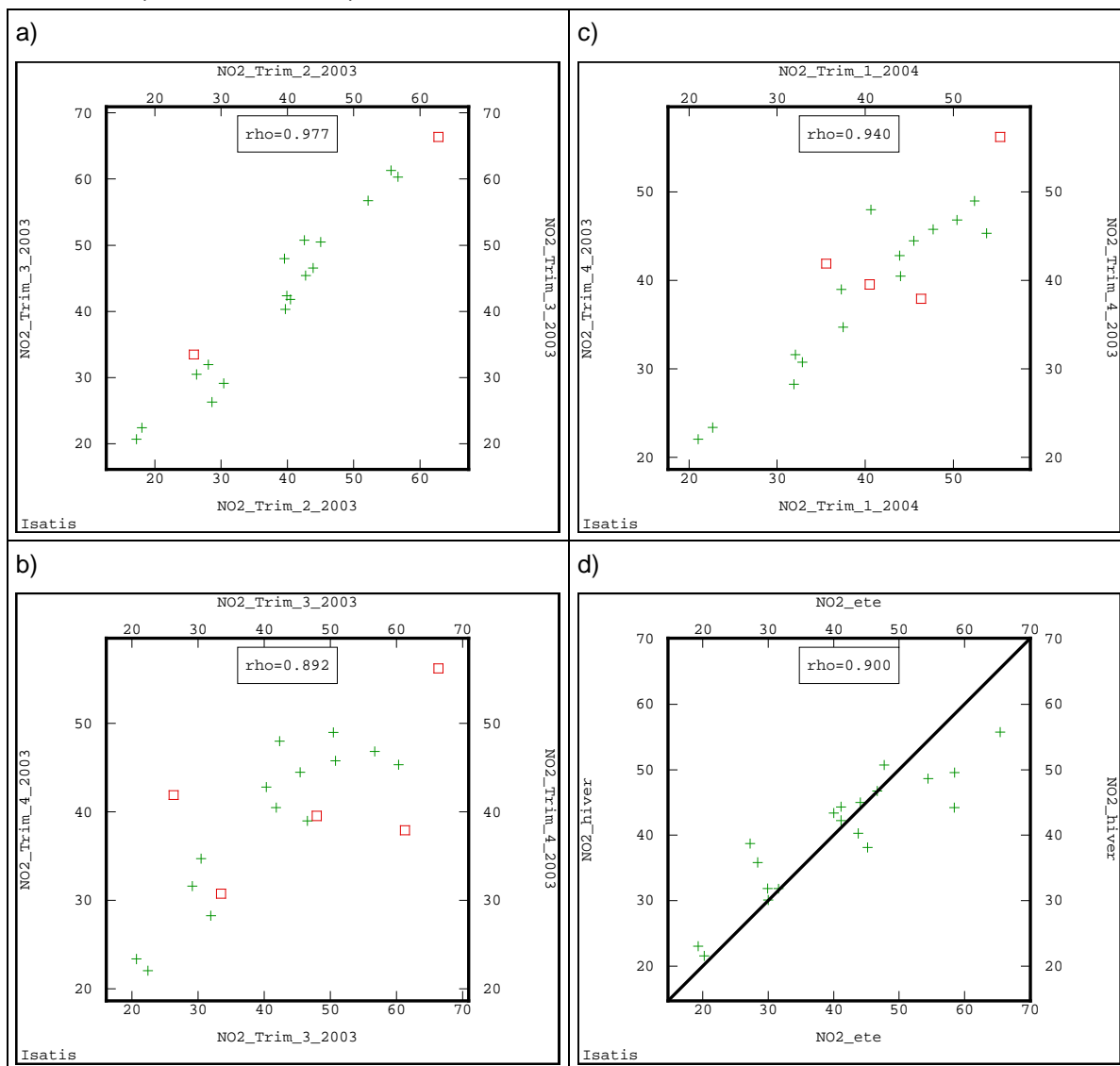


Figure 18 – a,b,c : nuages de corrélation entre trimestres. Les points masqués sont ceux pour lesquels deux mesures au moins sont manquantes.

d : nuage de corrélation entre saisons. Les points masqués sont ceux pour lesquels quatre mesures au moins sont manquantes.

7.1.4 SERIES TEMPORELLES

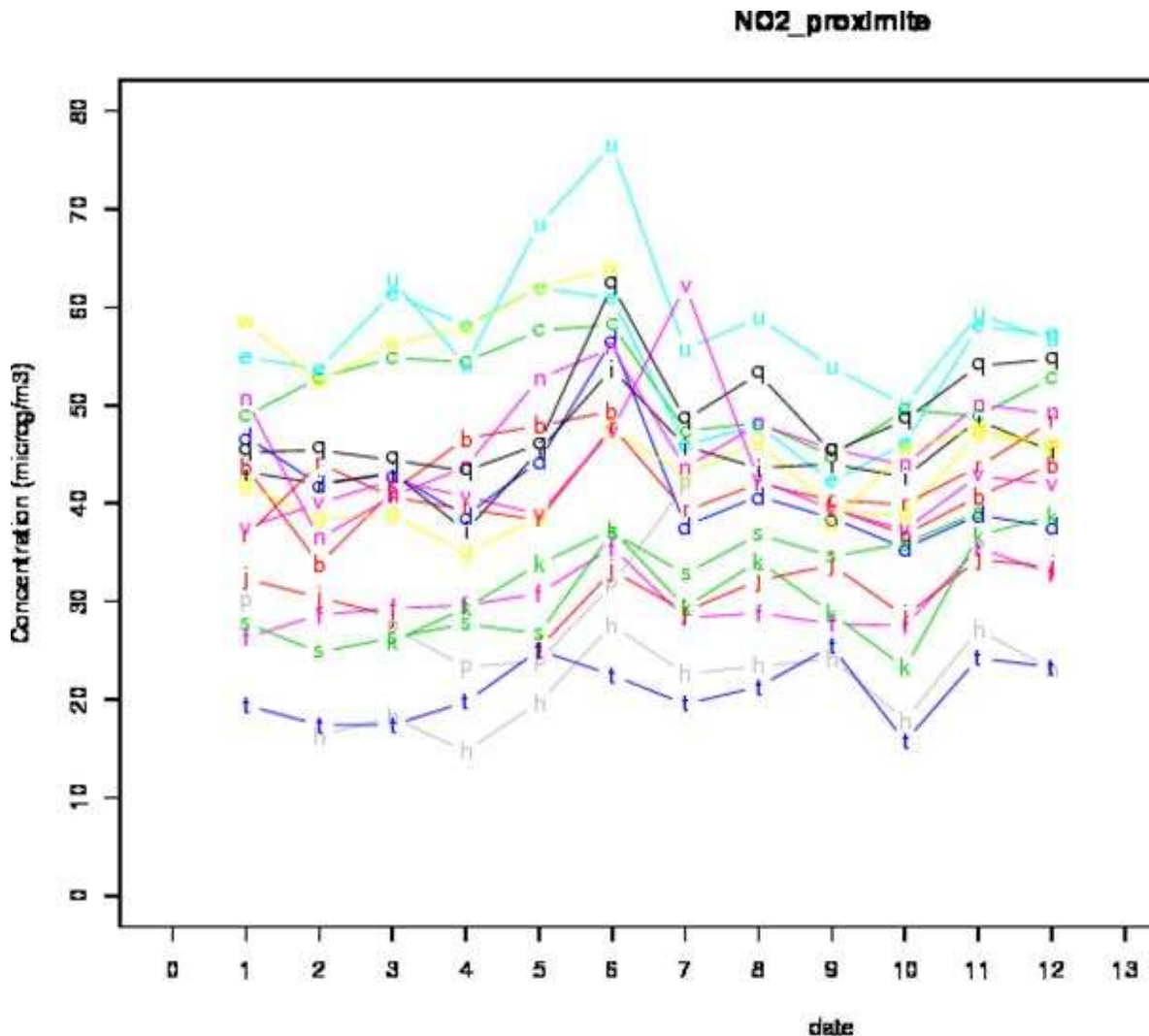


Figure 19 – Evolution des concentrations de NO₂ d'avril 2004 à mars 2003

7.2 CORRELATIONS ENTRE CONCENTRATIONS DE DIOXYDE D'AZOTE ET VARIABLES AUXILIAIRES

Remarque :

En raison de l'effectif restreint de points (19), afin de ne pas invalider un trop grand nombre de valeurs, les moyennes sont calculées sur tous les sites qui disposent d'au moins un tiers des données (soit 2 mesures par saison, 4 mesures par an). La totalité des sites de mesure est ainsi conservée dans l'étude des corrélations. On s'est assuré que les deux points comptant le plus de données manquantes ne perturbaient pas les statistiques.

7.2.1 CORRELATIONS AVEC LES EMISSIONS

Le tableau ci-après indique les corrélations entre les concentrations saisonnières et annuelles de NO₂ et les densités d'émission calculées autour des sites de mesure.

Tableau 4 – Corrélations entre les concentrations annuelles de NO₂ (sites de proximité) et les densités d'émission

	500 m	1 km	2 km	5 km	10 km
Hiver	0.324	0.752	0.703	0.601	0.629
	<i>0.356</i>	<i>0.666</i>	0.833	0.684	<i>0.651</i>
Eté	0.395	0.776	0.709	0.629	0.645
	<i>0.495</i>	<i>0.723</i>	0.820	0.726	<i>0.711</i>
Année	0.394	0.768	0.736	0.630	0.651
	<i>0.476</i>	<i>0.728</i>	0.846	<i>0.731</i>	<i>0.710</i>

Les valeurs en italique sont les coefficients de corrélation après transformation logarithmique des densités d'émission (de la forme $\ln(1+dE)$).

Sans transformation logarithmique, le maximum des corrélations est atteint pour un rayon de calcul égal à 1 km ; avec transformation il est atteint pour un rayon de 2 km. Les nuages de corrélation, en particulier ceux qui sont liés aux émissions à 500 m, font apparaître deux ensembles de points (Figure 20). A partir de 2 km, ces deux ensembles se mélangent mais les points des trois principales agglomérations de la région (Blois, Orléans, Tours) se démarquent (Figure 21). Une transformation logarithmique des densités a pour effet d'atténuer l'influence de ces villes et d'augmenter les corrélations.

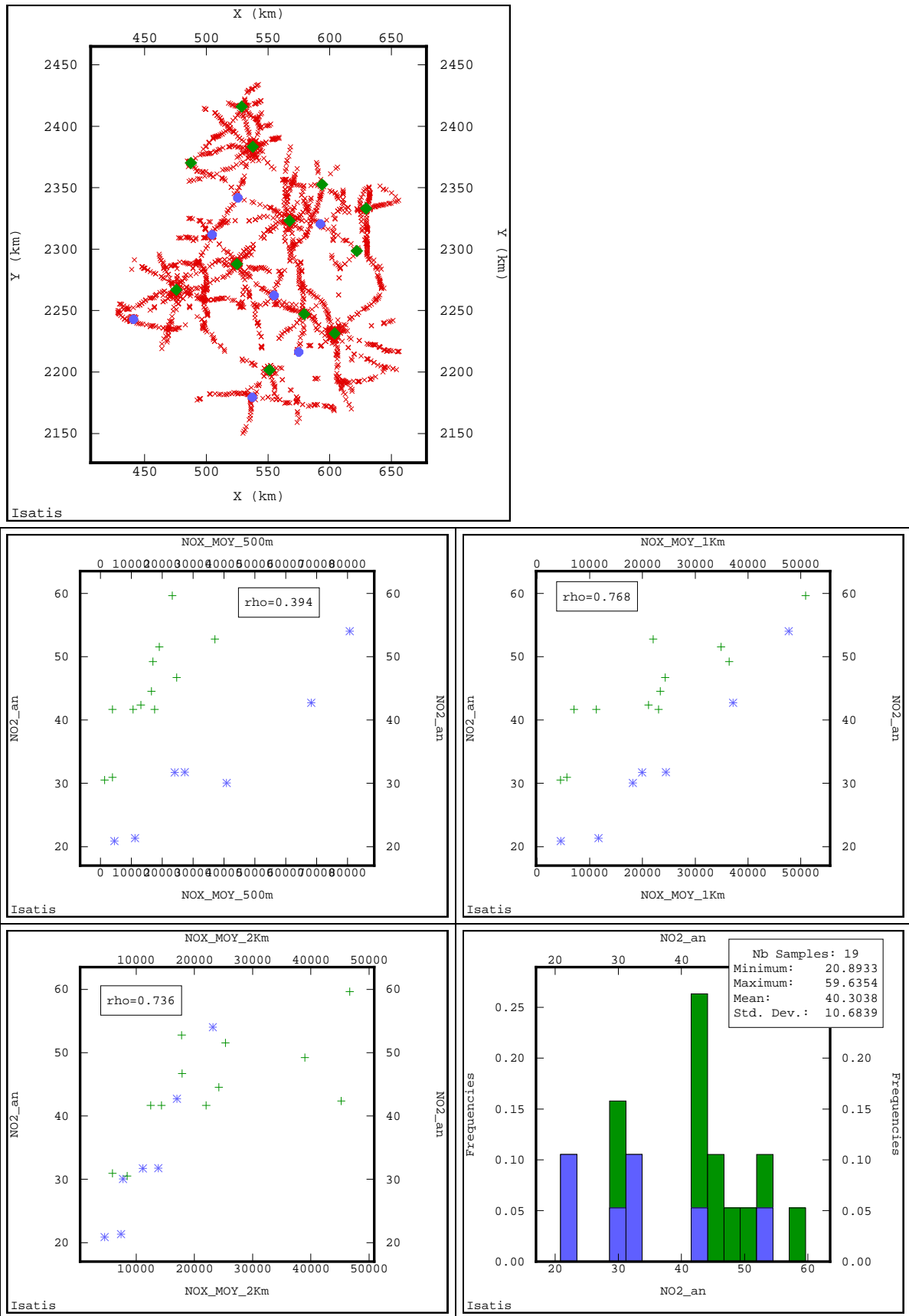


Figure 20 – Mise en évidence de deux populations de données, qui se différencient par leurs relations avec les émissions à petite distance. En haut : implantation des points. En bas : Nuages de corrélation et histogramme des concentrations

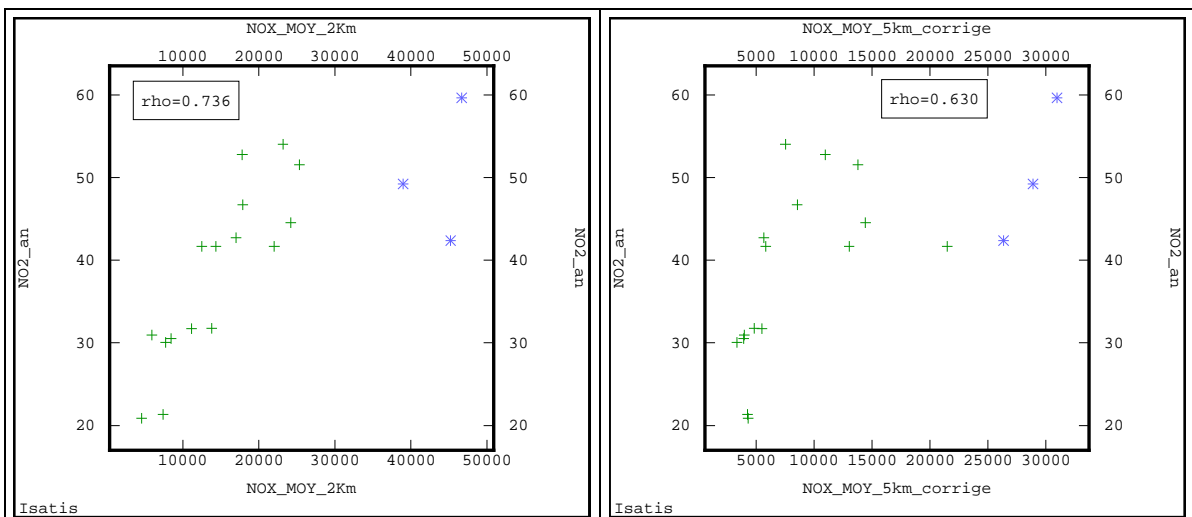
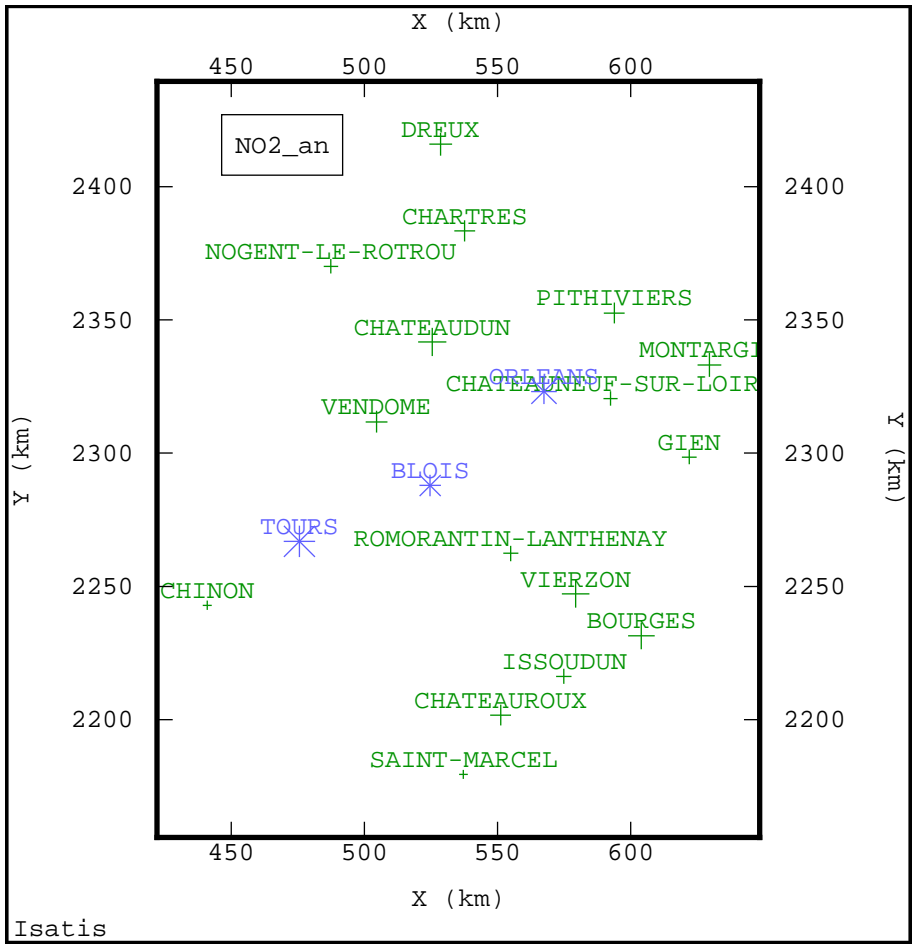


Figure 21 – Mise en évidence des trois grandes villes de la région. En haut : implantation des points. En bas : Nuages de corrélation

7.2.1.1 REMARQUE

Cette étude des corrélations révèle l'intérêt de considérer différentes variables de densité d'émission, plutôt que d'utiliser le cadastre brut. En effet, si l'on affecte à chaque site de mesure la valeur du cadastre (en g/km²) la plus proche (étant donné la résolution du cadastre, cette valeur se trouve à moins d'un kilomètre), les nuages de corrélation entre concentrations et émissions sont assez dispersés et les coefficients de corrélation relativement faibles.

Corrélation entre les concentrations annuelles de NO ₂ et....		Remarque
le cadastre migré	0.385	Ce résultat est comparable à celui qu'on obtient pour un rayon de 500m. Les deux populations sont également visible sur le nuage des corrélations
la densité d'émission dans un cercle de rayon d'1km	0.768	
Transformée logarithmique de la densité d'émission dans un cercle de rayon 2km	0,846	

7.2.2 CORRELATION AVEC LE TRAFIC MOYEN JOURNALIER

La corrélation est positive, mais peu élevée, et les nuages sont très dispersés. Si l'on masque trois points de trafic important (Chartres, Blois, Saint-Marcel), le nuage de corrélation suggère l'existence d'une relation croissante non linéaire entre les concentrations hivernales de NO₂ et le TMJA. Cette observation vaut aussi pour l'été, bien que le nuage soit plus dispersé.

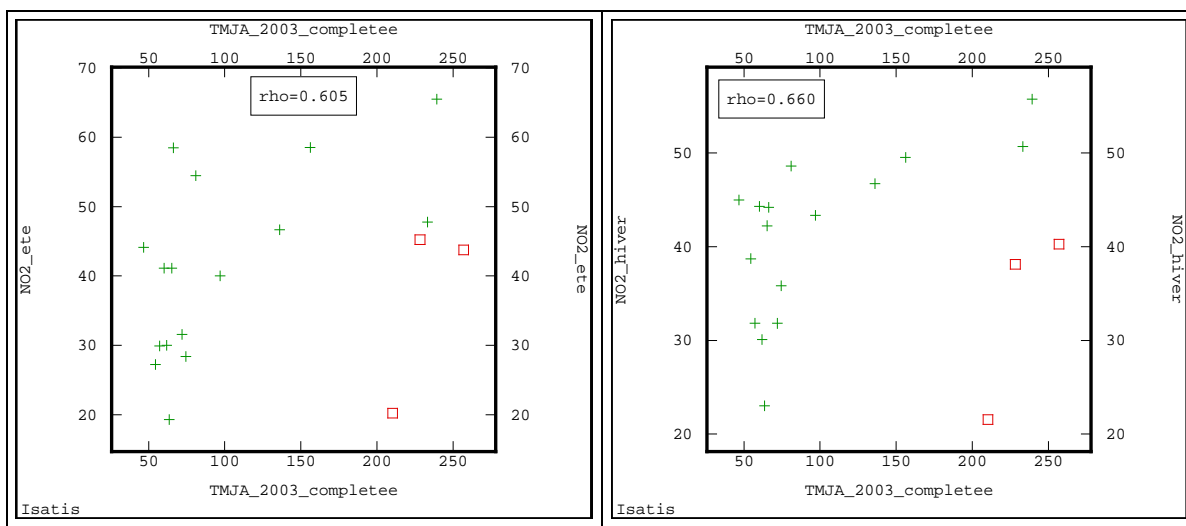


Figure 22 – Nuages de corrélation avec le TMJA. A gauche, l'été ; à droite, l'hiver.

7.2.3 CORRELATION AVEC L'OCCUPATION DES SOLS

L'étude procède de la même façon qu'avec les émissions. Pour chaque rayon, autant de variables d'occupation des sols que de classes de la nomenclature CORINE sont considérées. Le tableau ci-après fournit les valeurs de corrélation entre les concentrations annuelles de NO₂ et les territoires artificialisés

Tableau 5 – Corrélations entre les concentrations annuelles de NO₂ (sites de proximité) et les densités d'occupation des sols

	500 m	1 km	2 km	5 km	10 km
1.1.1	0.390	0.505	0.492	0.476	0.469
1.1.2	0.305	0.511	0.669	0.667	0.645
1.1	0.537	0.679	0.696	0.679	0.648
1.2	0.245	0.288	0.501	0.790	0.683
1.3	/	0.093	-0.099	0.322	0.092
1.4	0.024	-0.179	0.235	0.492	0.551
1	0.620	0.681	0.748	0.727	0.660
2	-0.574	-0.652	-0.728	-0.550	-0.180
3	-0.326	-0.419	-0.364	-0.393	-0.190

Par saison, les résultats sont de même nature :

	2 km	5 km
1.1 hiver	0.720	0.692
1.2 hiver	0.539	0.765
1 hiver	0.781	0.731
1.1 été	0.646	0.649
1.2 été	0.470	0.790
1 été	0.694	0.704

Lorsque les territoires artificialisés sont considérés dans leur ensemble (1.), les corrélations les plus élevées sont obtenues avec les densités calculées dans un rayon de 2 km ou 5 km selon la saison. La concentration de NO₂ est également corrélée avec la densité de zones industrielles ou commerciales et de réseaux de communication (1.2). Cette corrélation est maximale pour un rayon de 5km. Ce résultat peut s'expliquer par le fait que l'importance du trafic local, donc des émissions automobiles, est liée à celle, à plus grande échelle (i.e. dans un rayon de quelques kilomètres), du réseau routier. Cette hypothèse est confirmée par les corrélations suivantes :

Tableau 6 – Corrélations entre la densité locale d'émission (rayon 1 km) et la densité d'occupation du sol à différentes échelles

	Densité 1.2 - 1 km	Densité 1.2 - 2 km	Densité 1.2 - 5 km	Densité 1.2 - 10 km
Emissions - 1km	0.126	0.215	0.720	0.695

THEME : TRAITEMENTS NUMERIQUES

APPLICATION DE METHODES GEOSTATISTIQUES POUR L'EXPLOITATION CONJOINTE DE MESURES DE FOND ET DE PROXIMITE

CONTEXTE

Les outils de la géostatistique sont couramment employés par les AASQA pour produire des cartes de la pollution de fond à l'échelle urbaine ou régionale (dioxyde d'azote, ozone, benzène). Les données exploitées à cette fin sont les mesures recueillies sur les sites de fond, le plus souvent par échantillonnage passif ; la pollution de proximité fait généralement l'objet d'études séparées.

Ainsi, si les cartographies de fond reproduisent correctement la répartition des concentrations dans la zone d'étude, elles n'offrent pas nécessairement une image fine des gradients de concentration près des axes de circulation. Elles peuvent en particulier sous-estimer les niveaux de pollution au voisinage des routes ou des voies urbaines.

Pour une information plus juste et plus complète du public, les AASQA s'interrogent sur les moyens de traiter conjointement des mesures disponibles à différentes échelles spatiales (fond et proximité) afin de représenter plus fidèlement la réalité

TRAVAUX PROPOSES POUR 2005

L'enjeu de cette étude est d'imbriquer des données de fond et de proximité dans une unique cartographie. Le travail sera mené en étroite collaboration avec LIG'AIR, qui a réalisé dans cette optique des campagnes de mesure de NO₂ conjuguant la caractérisation des axes routiers, des agglomérations et du fond rural, et qui prévoit pour 2005 la conduite d'une nouvelle campagne.

L'expérience pilote qu'il est proposé de mettre en œuvre aura pour but de développer des procédures de traitement géostatistique permettant de superposer de façon adéquate les informations issues de ces campagnes.

Une attention particulière sera portée :

- à l'échantillonnage et à sa pertinence pour effectuer une analyse géostatistique satisfaisante ;
- à l'intérêt des variables auxiliaires (population, bâti, données de trafic...), aux informations à différentes échelles qu'elles peuvent apporter et à la façon de les intégrer ;
- à l'évaluation des résultats obtenus.

Selon les données disponibles, l'étude portera sur l'ensemble de la région Centre ou sera ciblée sur une ou plusieurs zones particulières.

La possibilité d'utiliser en variables auxiliaires, des résultats issus d'un modèle déterministe, typiquement PREV'AIR ou un modèle régional (pour lequel les simulations de NO₂ seront plus pertinentes) sera examinée, dans le souci de tirer parti de l'ensemble des méthodes disponibles aujourd'hui.

Enfin, en collaboration avec l'EMD, les conclusions de cette étude seront mises en perspective avec les résultats proposés par l'EMD à la suite des travaux de modélisation empirique (méthode de Stedman) ou déterministe (TRANSCHIM) réalisés en 2003 dans le cadre du LCSQA, sur la région centre précisément.

L'outil de géostatistique utilisé dans le cadre de cette étude sera le logiciel ISATIS, très répandu dans les AASQA. Cette étude de faisabilité menée sur la région Centre aboutira à un guide technique consignant des recommandations génériques à destination de l'ensemble des associations.

COLLABORATIONS

LIG'AIR , Ecole des Mines de Douai, Centre de Géostatistique de l'Ecole des Mines de Paris.

DUREE DES TRAVAUX

Cette étude prospective sera effectuée sur un an.