



Laboratoire Central de Surveillance de la Qualité de l'Air



Adaptation de méthodes géostatistiques à la cartographie automatique de NO₂

Giovanni CARDENAS et Esperanza PERDRIX

Novembre 2005





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.

ECOLE DES MINES DE DOUAI

DEPARTEMENT CHIMIE ET ENVIRONNEMENT

**ADAPTATION DE METHODES
GEOSTATISTIQUES A LA CARTOGRAPHIE
AUTOMATIQUE DE NO₂**

Esperanza PERDRIX, André WROBLEWSKI et Hervé PLAISANCE
En collaboration avec Giovanni CARDENAS (INERIS)

Novembre 2005



Ministère de l'Écologie
et du Développement Durable

ADAPTATION DE METHODES GEOSTATISTIQUES A LA CARTOGRAPHIE AUTOMATIQUE DE NO2

Laboratoire Central de Surveillance
de la Qualité de l'Air

Thème : TRAITEMENTS NUMERIQUES

Convention 05000051

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

Novembre 2005

Personnes ayant participé à l'étude :

Giovanni Cardenas
Esperanza Perdrix (EMD)

Ce document comporte 83 pages (hors couverture et annexes).

	Rédaction	Vérification	Approbation
NOM	G Cardenas	J Y Chatelier	M.RAMEL
Qualité	Ingénieur à l'unité 2IEN	Responsable Unité 2IEN	Responsable LCSQA/INERIS Direction des Risques Chroniques
Visa			

TABLE DES MATIERES

1	RÉSUMÉ DE L'ÉTUDE EMD/INERIS 2005.....	4
2	INTRODUCTION.....	5
3	APPLICATION DE LA MÉTHODE POUR L'ESTIMATION DES VALEURS JOURNALIÈRES	6
3.1	Rappel des résultats antérieurs [1] :	6
3.2	But poursuivi et methodologie :	6
3.3	Caractérisation des situations pour lesquelles les relations statistiques bi-hebdomadaires ne s'appliquent pas : cas des stations périurbaines.....	7
3.3.1	Tri selon l'erreur relative absolue	7
3.3.2	Tri par classification hiérachique	10
3.3.3	Synthese des resultats pour le cas des stations périurbaines	16
3.4	Caractérisation des situations pour lesquelles les relations statistiques bi-hebdomadaires ne s'appliquent pas : cas des stations urbaines :	17
3.4.1	Tri selon l'erreur relative absolue	17
3.4.2	Tri par classification hiérachique	20
3.4.3	Synthese des resultats pour le cas des stations urbaines	25
3.5	Synthèse des différentes situations météorologiques rencontrées, pour les stations urbaines et périurbaines.	25
4	AUTOMATISATION DE LA PROCÉDURE DE CARTOGRAPHIE DU DIOXYDE D'AZOTE.....	29
4.1	Cartographie par des méthodes géostatistiques.....	29
4.2	Automatisation de la cartographie	31
4.3	Etude de Cas.....	35
4.3.1	Données Disponibles	35
4.3.1.1	Variables auxiliaires	35
4.3.1.2	Mesures de NO ₂	36
4.3.2	Calcul de la VEM des stations virtuelles.....	37
4.3.3	Estimation des deux campagnes représentatives	42
4.3.3.1	Estimation de la campagne No. 1 (Eté)	43
4.3.3.1.1	<i>Estimation sur l'ensemble des tubes</i>	<i>43</i>
4.3.3.1.2	<i>Estimation sur la sélection des tubes corrélés aux stations de référence</i>	<i>48</i>
4.3.3.1.3	<i>Estimations faites à partir des données reestimees (stations virtuelles)</i>	<i>51</i>
4.3.3.2	Estimation de la campagne No. 18 (Hiver)	57
4.3.3.2.1	<i>Estimation sur l'ensemble des tubes</i>	<i>57</i>
4.3.3.2.2	<i>Estimation sur la sélection des tubes corrélés aux stations de référence</i>	<i>61</i>
4.3.3.2.3	<i>Estimations faites à partir des données reestimees (stations virtuelles)</i>	<i>64</i>

4.3.4	Estimation des données journalières sur les stations virtuelles.....	69
4.3.4.1	Etude des variogrammes journalieres.....	69
4.3.4.2	Validation des estimations journalières.....	71
4.3.5	Présentation de Quelques cartes obtenues.....	76
5	CONCLUSIONS.....	80
6	BIBLIOGRAPHIE.....	82
7	LISTE DES ANNEXES.....	83

1 RESUME DE L'ETUDE EMD/INERIS 2005

ADAPTATION DE METHODES GEOSTATISTIQUES A LA CARTOGRAPHIE AUTOMATIQUE DE NO2

Esperanza PERDRIX (EMD)

perdrix@ensm-douai.fr ; tel : + 33 (0)3 27 71 26 15

Giovanni CARDENAS (INERIS)

Giovanni.Cardenas@ineris.fr; tel : + 33 (0)3 44 55 64 81

Ces travaux constituent une étude de faisabilité d'une cartographie automatique du dioxyde d'azote dans une agglomération. Ces cartes doivent être réalisées à partir des données des stations fixes, en prenant en compte le nombre potentiellement insuffisant de capteurs fixes permettant une correcte interpolation spatiale des niveaux de pollution.

Dans la première partie de cette étude nous proposons une méthode pour générer des données supplémentaires, cette méthode est basée sur l'existence d'un historique des campagnes par échantillonneurs passifs.

Ces campagnes sont exploitées de façon à pouvoir grouper les sites d'échantillonnage passif (appelés désormais stations virtuelles). Ces sites doivent être corrélés temporellement à une station fixe particulière. Celle-ci est appelée station de référence.

L'influence des conditions météorologiques sur la qualité des résultats issus des corrélations, entre stations virtuelles et station de référence, a été testée. Nous montrons que dans le cas du dioxyde d'azote, les conditions météorologiques ainsi que les niveaux d'ozone doivent être pris en compte. Pour certaines conditions météorologiques précises et peu fréquentes (environ 10% des cas), nous proposons l'utilisation d'un facteur correctif.

Dans la seconde partie de cette étude nous avons testé, sur un cas d'étude, la pertinence d'une méthode de génération automatique de variogrammes et la qualité des estimations effectuées à partir de données ainsi produites.

Pour la cartographie, des méthodes géostatistiques ont été mises en œuvre, en particulier la méthode de cokrigage. Cette méthode multivariable permet d'exploiter l'information additionnelle apportée par des variables auxiliaires ; dans le cas du dioxyde d'azote, des variables décrivant le milieu, comme la densité de population ou les émissions peuvent donc être utilisées.

Pour prendre en compte l'incertitude des données générées aux stations virtuelles, nous avons inclus dans la matrice variance - covariance du krigeage une composante additionnelle, appelée la variance de l'erreur de mesure (VEM) ; une méthodologie pour la calculer est proposée.

Finalement les cartes obtenues sont bien détaillées, elles reconstituent les principales caractéristiques de la pollution par le dioxyde d'azote dans le domaine considéré.

2 INTRODUCTION

Depuis 2004, le LCSQA développe une méthode originale d'estimation et de cartographie des concentrations en dioxyde d'azote, à partir d'un nombre limité de mesures effectuées en stations et en utilisant des corrélations statistiques [1].

Les objectifs de qualité [3] requis pour les méthodes de modélisation et d'analyse objective sont rappelés dans le tableau 1.

Tableau 1 : Objectifs de qualité pour l'évaluation des concentrations en dioxyde d'azote [3].

Méthode	Type de moyenne estimée	Exactitude ou précision requise	Valeur limite (en 2010)
Modélisation	horaire	50%-60%	200 $\mu\text{g}/\text{m}^3$ (< 18 fois par an)
	journalière	50%	
	annuelle	30%	40 $\mu\text{g}/\text{m}^3$
Estimation objective	non précisée	75%	

Il avait été montré qu'il existait des corrélations statistiques significatives entre les concentrations moyennes bi-hebdomadaires (moyenne sur 14 jours) en dioxyde d'azote, mesurées par une station automatique (par chimiluminescence), et celles, mesurées (par spectrophotométrie ou chromatographie ionique) à partir de prélèvements effectués par échantillonnage passif [1].

L'étude d'un jeu de données, relatif à une zone présentant des caractéristiques essentiellement urbaines et périurbaines (conurbation Lille – Roubaix - Tourcoing-Villeneuve d'Ascq), a montré que plus de la moitié des sites étudiés étaient ainsi corrélés à une station de mesure automatique. Plus précisément, deux stations automatiques ont servi de référence, l'une urbaine et l'autre périurbaine, en relation avec les deux types de typologies rencontrés sur cette zone.

Nous avons montré également, dans cette précédente étude, que le nombre de sites corrélés et leur répartition spatiale permettait de réaliser des cartographies de répartition spatiale du dioxyde d'azote.

Les nouveaux travaux présentés dans ce rapport, visent à mieux cerner les limites de la méthode et à faciliter son application au sein des AASQA, notamment au travers d'une automatisation de la procédure de cartographie.

L'étude se scinde donc en deux parties :

- la première se rapporte à l'estimation des moyennes journalières. On souhaite savoir si les corrélations statistiques, établies sur une base temporelle bihebdomadaire, permettent d'estimer des moyennes journalières avec une précision suffisante vis-à-vis des objectifs de qualité (cf. tableau 1).
- la seconde vise à faciliter la production de cartes de répartition spatiale du dioxyde d'azote, par automatisation de la procédure d'analyse géostatistique et de krigeage. Le logiciel Isatis® version 5, de Geovariances est utilisé.

3 APPLICATION DE LA METHODE POUR L'ESTIMATION DES VALEURS JOURNALIERES

3.1 RAPPEL DES RESULTATS ANTERIEURS [1] :

Des relations mathématiques simples (régressions linaires) ont été établies entre les concentrations bi-hebdomadaires en NO₂ mesurées par une station fixe automatique et celles mesurées par échantillonnage passif dans d'autres localisations. Nous avons montré que ces mêmes relations mathématiques pouvaient servir à estimer des concentrations journalières en NO₂, en divers endroits, en se basant sur la concentration journalière mesurée par la station fixe automatique. Néanmoins, nous avons constaté que l'erreur relative absolue¹ pouvait dépasser 50% dans 5 à 7% des cas, voire 75% dans environ 2% des cas, notamment lors de l'estimation de faibles niveaux de concentrations.

3.2 BUT POURSUIVI ET METHODOLOGIE :

Il s'agit de comprendre et de déterminer dans quels cas les relations qui permettent d'estimer les concentrations journalières en NO₂ ne fonctionnent pas, c'est-à-dire ne remplissent pas les objectifs de qualité (Tableau 1). L'idée est alors de trouver des critères d'identification de ces cas, peu nombreux, où les relations ne s'appliquent pas et, a contrario, déterminer avec précision tous les cas où les relations s'appliquent en respectant bien les objectifs de qualité.

La méthodologie suivie repose sur l'utilisation conjointe de paramètres météorologiques et de mesures de pollution. A l'aide d'outils d'analyse statistique multivariable, les caractéristiques météorologiques et de pollution des jours où les relations ne s'appliquaient pas ont été recherchées.

Les situations urbaines et périurbaines ont été étudiées séparément, en suivant dans les deux cas la même méthodologie. Pour chaque typologie de sites, deux stations automatiques ont été considérées : l'une est la station de référence, fournissant des concentrations mesurées, et l'autre est la station « virtuelle » à laquelle on calcule, par relation statistique, des concentrations estimées. Comme la véritable concentration mesurée à la station « virtuelle » est connue, il est possible de calculer l'erreur commise lors de l'estimation par la relation statistique.

¹ L'erreur relative absolue est définie comme la valeur absolue de l'écart entre valeur estimée et valeur mesurée par la station automatique de référence, divisé par la valeur mesurée par la station automatique de référence.

3.3 CARACTERISATION DES SITUATIONS POUR LESQUELLES LES RELATIONS STATISTIQUES BI-HEBDOMADAIRES NE S'APPLIQUENT PAS : CAS DES STATIONS PERIURBAINES

Les deux stations automatiques périurbaines considérées sont « Armentières » et « Halluin cailloux ». Nous avons estimé les concentrations journalières de la station d'« Armentières » à partir des concentrations journalières de la station de référence d'« Halluin cailloux », en utilisant la régression linéaire établie entre les deux stations pour des concentrations bi-hebdomadaires :

$$\text{Conc_jour_Armentières } (\mu\text{g}/\text{m}^3) = 0,71 \times \text{Conc_jour_Halluin } (\mu\text{g}/\text{m}^3) + 11,30$$

(relation 1)

Nous considérons les années 1998 à 2003, à l'exception de l'année 2002 qui est gardée pour validation. Le nombre de données disponibles (correspondant aux jours où les concentrations moyennes des deux stations ont pu être mesurées simultanément) s'élève à 1420 données.

3.3.1 TRI SELON L'ERREUR RELATIVE ABSOLUE

L'erreur relative absolue est définie comme la différence, en valeur absolue, entre la valeur mesurée et la valeur estimée par calcul, divisée par la valeur mesurée. En prenant comme critère de sélection l'erreur relative en valeur absolue, on peut distinguer les jours, où la relation statistique entre les deux stations s'applique bien, de ceux où elle s'applique mal. Nous avons considéré comme « mauvais » les jours pour lesquels l'erreur relative en valeur absolue était supérieure ou égale à 50% (Tableau 2).

Tableau 2 – Estimation des concentrations journalières par régression linéaire : cas des stations périurbaines.

Stations périurbaines	Tous les jours	Jours où la relation s'applique bien (a)	Jours où la relation s'applique mal (b)
Nombre de données	1420	1304 (92%)	116 (8%)
Coefficient de corrélation (r^2) entre estimations et mesures	0,82	0,85	0,52
Concentration moyenne des mesures à la station « Armentières » ($\mu\text{g}/\text{m}^3$)	30,1	31,3	16,8
Erreur moyenne ($\mu\text{g}/\text{m}^3$)	0,74	0,02	8,83
Erreur relative moyenne en valeur absolue (%)	20,6	15,9	73,2

(a) Les jours pour lesquels la relation statistique s'applique bien sont les jours pour lesquels l'erreur relative, en valeur absolue, est strictement inférieure à 50%.

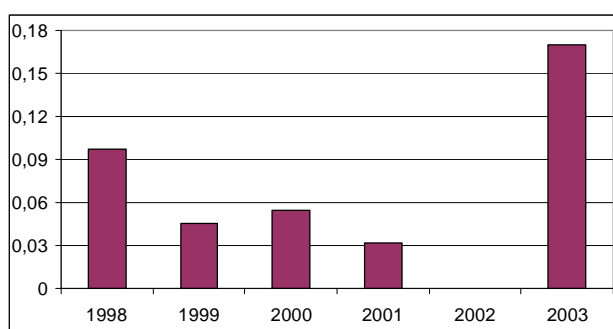
(b) Les jours pour lesquels la relation statistique s'applique mal sont les jours pour lesquels l'erreur relative, en valeur absolue, est supérieure ou égale à 50%.

La séparation des jours en deux classes, selon l'erreur relative supérieure ou non à 50%, montre que les « mauvais » jours constituent 8% des données, soit 29 j/an. Ils correspondent à des journées pour lesquelles la concentration en NO₂ est faible (16,8 µg/m³ en moyenne).

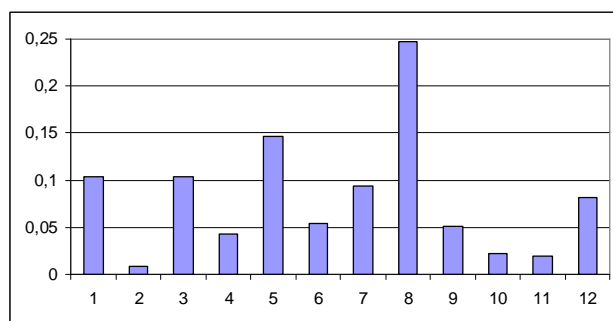
Un examen plus détaillé de ces « mauvais » jours montrent que leur fréquence est plus élevée :

- l'année 2003 (Fig. 1a) ;
- au mois d'août (Fig. 1b), surtout le mois d'août 2003 ;
- le dimanche (Fig. 1c).

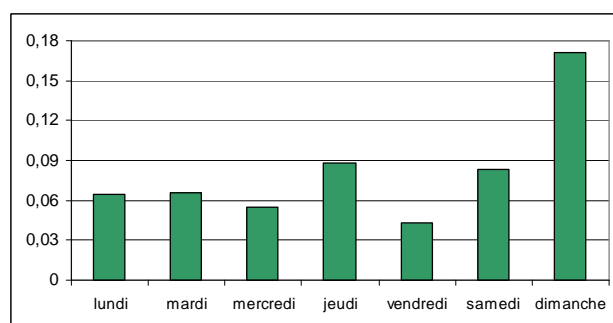
Il est à signaler que le mois d'août de l'année 2003 a été le siège d'un phénomène météorologique exceptionnel (canicule), caractérisé par de très fortes chaleurs. Le dimanche est, quant à lui, le jour de la semaine présentant les plus faibles niveaux de concentration en NO₂, en relation avec la diminution du trafic le week-end.



(a)



(b)



(c)

Figure 1 – Pourcentage de « mauvais » jours en fonction de l'année (a), du mois (b), du jour de la semaine (c) dans le cas des stations périurbaines.

Afin de mieux comprendre les liens entre les « mauvais » jours et les conditions météorologiques, nous avons mis en relation les données journalières de concentration en NO₂ avec les données météorologiques mesurées quotidiennement par la station Météo France la plus proche (Lille Lesquin).

Les paramètres météorologiques considérés sont : pression, température à 13h, température minimale, température maximale, direction et force du vent, pluviométrie.

Nous avons également considéré, pour l'analyse, les concentrations journalières en ozone mesurées à la station périurbaine d'Halluin-Cailloux (la même que pour le NO₂).

Les résultats montrent que les « mauvais » jours sont caractérisés par des secteurs de vent très différents des secteurs normalement rencontrés (Fig. 2a et 2b), en particulier provenant du nord et du nord-est.

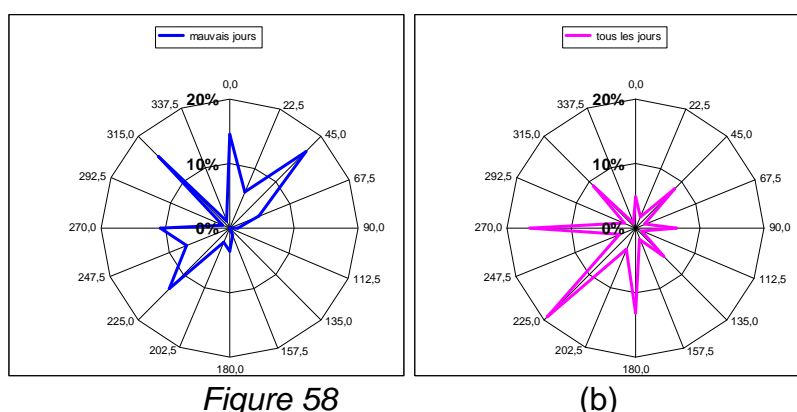


Figure 2 – Provenance du vent dans le cas des stations périurbaines
Figure 58 lors des « mauvaises » journées
Figure 58 pour la totalité des jours.

On constate également que les « mauvais » jours se caractérisent par une augmentation de la concentration en ozone et des valeurs des paramètres météorologiques suivants : pression, température, vitesse de vent, cumul des précipitations (Tableau 3).

Tableau 3 – Comparaisons des paramètres météorologiques pour les bons et mauvais jours (cas des stations périurbaines).

Stations périurbaines	Tous les jours	Jours où la relation s'applique bien ^(a)	Jours où la relation s'applique mal ^(b)
Moyenne journalière NO ₂ à Halluin (µg/m ³)	27,5	28,2	19,7
Moyenne journalière NO ₂ à Armentières (µg/m ³)	30,1	31,2	16,6
Erreur relative absolue moyenne (%)	20,3	15,9	73,6
Pression (hPa)	1015,6	1015,5	1016,6
Température à 13h (°C)	12,9	12,8	14,8
Température minimale (°C)	-7,9	-7,9	-7,3
Température maximale (°C)	35,0	35,0	29,4
Vitesse du vent (km/h)	19,5	19,2	22,2
Précipitations journalières (L/m ²)	1,98	1,97	2,05
Moyenne journalière O ₃ à Halluin (µg/m ³)	37,7	38,4	49,3

(a) Jours pour lesquels l'erreur relative, en valeur absolue, est strictement inférieure à 50%.

(b) Jours pour lesquels l'erreur relative, en valeur absolue, est supérieure ou égale à 50%.

Ces nouvelles données confirment, en partie, l'hypothèse avancée précédemment. Les jours, où la relation statistique entre stations ne s'applique pas, correspondent à des situations anticycloniques (pression élevée), à des journées plutôt chaudes (moyenne des températures plus élevée), avec une activité photochimique prononcée (concentration en ozone élevée, concentration en NO₂ basse), un vent plus fort (moyenne des vitesses de vent plus élevée, favorise la dispersion des polluants) et des précipitations plus importantes.

On constate, par ailleurs, que les concentrations en NO₂ mesurées lors des « mauvais » jours sont sur-estimées d'environ 66% par la relation 1.

Il est possible d'établir une nouvelle relation statistique pour ces « mauvais » jours (relation 2, Fig. 3) :

$$\text{Conc_jour_Armentières } (\mu\text{g}/\text{m}^3) = 0,4459 \times \text{Conc_jour_Halluin } (\mu\text{g}/\text{m}^3) + 6,05$$

(relation 2)

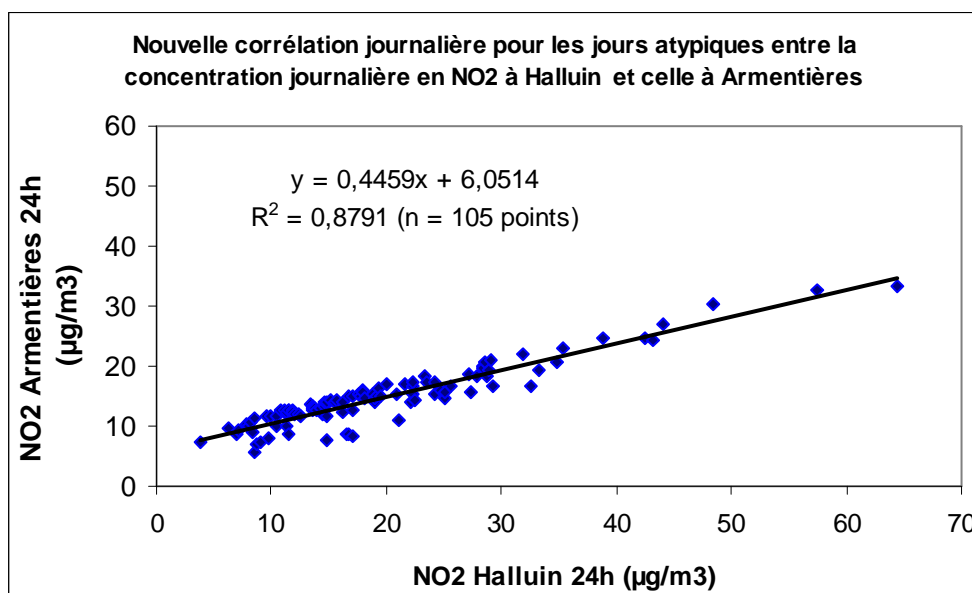


Figure 3 – Nouvelle régression linéaire entre les concentrations journalières en NO₂ mesurées à Armentières en fonction de celles mesurées à Halluin.

3.3.2 TRI PAR CLASSIFICATION HIERARCHIQUE

Afin de mieux préciser les diverses situations météorologiques rencontrées sur la zone d'étude, nous avons mené une analyse des données en utilisant une méthode d'analyse statistique multivariable (la classification hiérarchique ascendante selon la méthode de Ward), permettant de classer les jours selon leurs caractéristiques météorologiques et de pollution.

L'analyse a porté sur 698 jours pour lesquels les 10 variables suivantes étaient renseignées : erreur relative absolue, concentration en NO₂ à Armentières, concentration en NO₂ à Halluin, concentration en ozone à Halluin, pression, cumul journalier des précipitations, température maximale, température minimale, vitesse du vent et visibilité.

Six classes différentes ont été distinguées (Tableau 4).

La première classe (classe 1) se distingue des cinq autres par une erreur relative moyenne très élevée (55%).

Si l'on compare la fréquence des jours appartenant à la classe 1 en fonction de l'année, du mois ou du jour de la semaine (Fig. 4 a,b,c), on remarque que la classe 1 présente des caractéristiques similaires à celles du groupe des « mauvais » jours (Fig. 2 a,b,c), notamment une fréquence élevée l'année 2003 et le dimanche.

En revanche, la classe 1 ne comporte pas beaucoup de jours provenant du mois d'août, car les 698 valeurs étudiées présentent très peu de valeurs au mois d'août 2003. C'est plutôt le mois de mai, qui ressort.

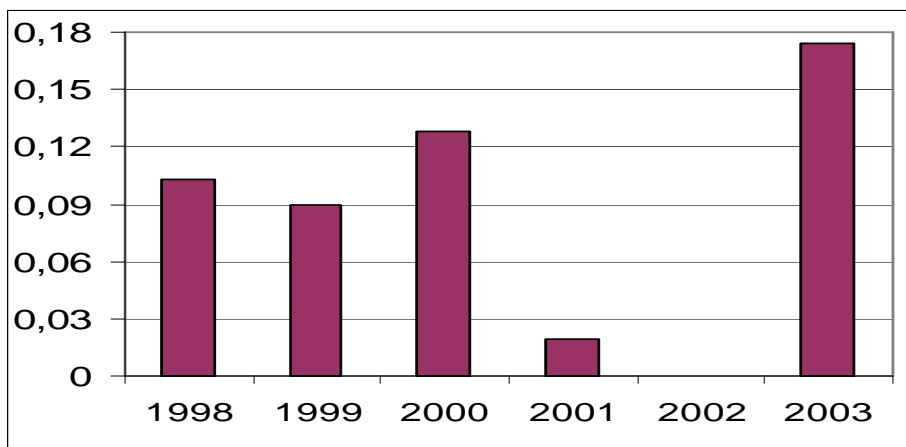
On peut noter également que les jours de la classe 1 se caractérisent par un cumul des précipitations journalier plus faible que les autres jours (temps plutôt sec), alors que les « mauvais » jours (Tableau 3) présentent plutôt des précipitations plus élevées que la moyenne (temps plus humide). Il est probable que ce critère (cumul journalier des précipitations) ne soit donc pas un critère discriminant pour notre étude.

Tableau 4 – Résultats de la classification hiérarchique menée sur les données journalières de pollution et de météorologie : cas des stations périurbaines.

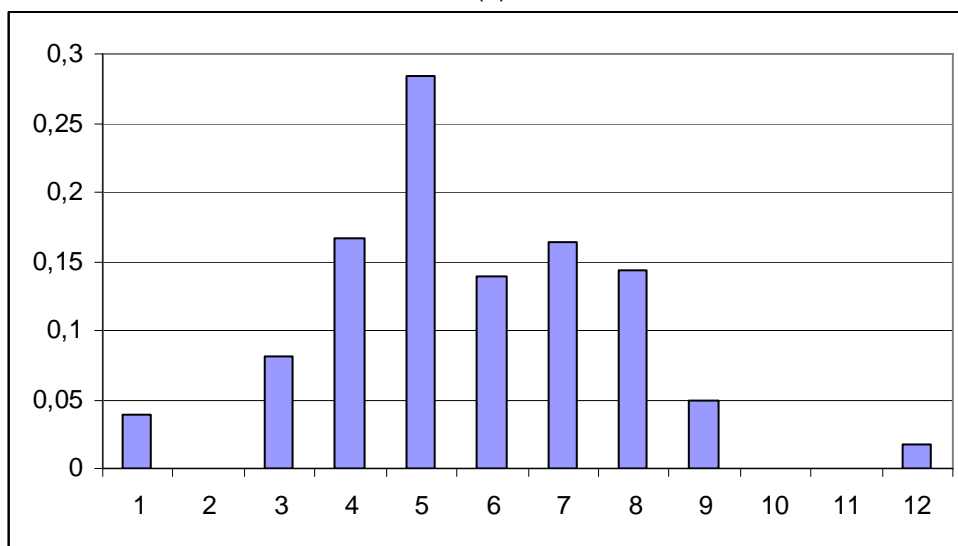
Erreur : erreur relative absolue (%)
 NO₂ A : concentration moyenne journalière en NO₂ à Armentières (µg/m³)
 NO₂ H : concentration moyenne journalière en NO₂ à Halluin (µg/m³)
 O₃ H : concentration moyenne journalière en O₃ à Halluin (µg/m³)
 P_hPa : pression moyenne journalière (hPa)
 Précip. : précipitations (L/m²)
 Tmax : moyenne des températures maximales (°C)
 Tmin : moyenne des températures minimales (°C)
 Vis_km : visibilité (km)
 VV : vitesse du vent (km/h).

classe	effectif	Erreur %	NO ₂ A	NO ₂ H	O ₃ H	P_hPa	Précip.	T_max	T_min	Vis_km	VV	Commentaire
1	75 (11%)	55	17	20	50	1020	0,61	17,4	10,1	18	22	NO ₂ faible, O ₃ élevé Régime anticyclonique, vent de NE, très sec (précipitations minimales), visibilité modérée, mois de mai
2	182 (26%)	12	22	17	51	1013	1,84	17,8	11,0	26	27	NO ₂ minimum, O ₃ maximum Pression ≅ 1 atm, visibilité maximale, mois de juin
3	67 (10%)	22	26	23	38	1009	10,2	15,4	10,1	14	19	NO ₂ faible, O ₃ moyen Régime dépressionnaire (pression minimale), vent de SO, pluviométrie maximale, températures douces
4	133 (19%)	17	30	31	47	1016	0,63	22,8	12,1	21	12	NO ₂ moyen, O ₃ élevé Régime anticyclonique, très sec, force de vent minimale, températures maximales, saison estivale
5	151 (22%)	14	35	31	28	1015	1,17	10,1	4,4	17	20	NO ₂ moyen, O ₃ faible Pression ≅ 1 atm, visibilité modérée, températures froides
6	90 (19%)	19	53	48	10	1026	0,72	7,0	0,8	8	12	NO ₂ maximum, O ₃ minimum Régime anticyclonique (pression maximale), sec, force de vent minimale, visibilité minimale, températures minimales, période hivernale

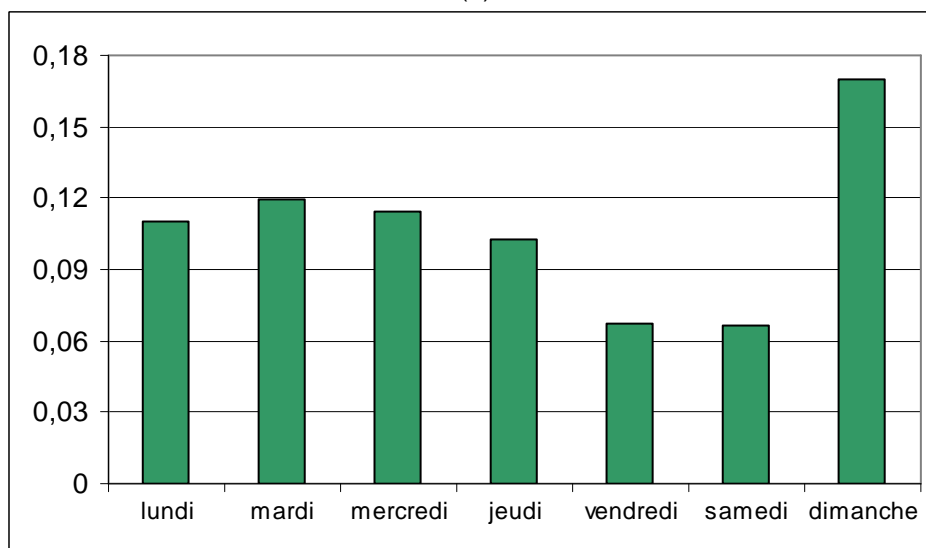
Total : 698



(a)



(b)



(c)

Figure 4 – Proportion des jours appartenant à la classe 1 en fonction de l'année (a), du mois (b), du jour de la semaine (c), pour les stations périurbaines.

Pour les jours de la classe 1, la régression linéaire entre les concentrations journalières en NO₂ à Halluin et celles à Armentières (Relation 3, Fig. 5) est proche de celle trouvée pour les « mauvais » jours (Relation 2).

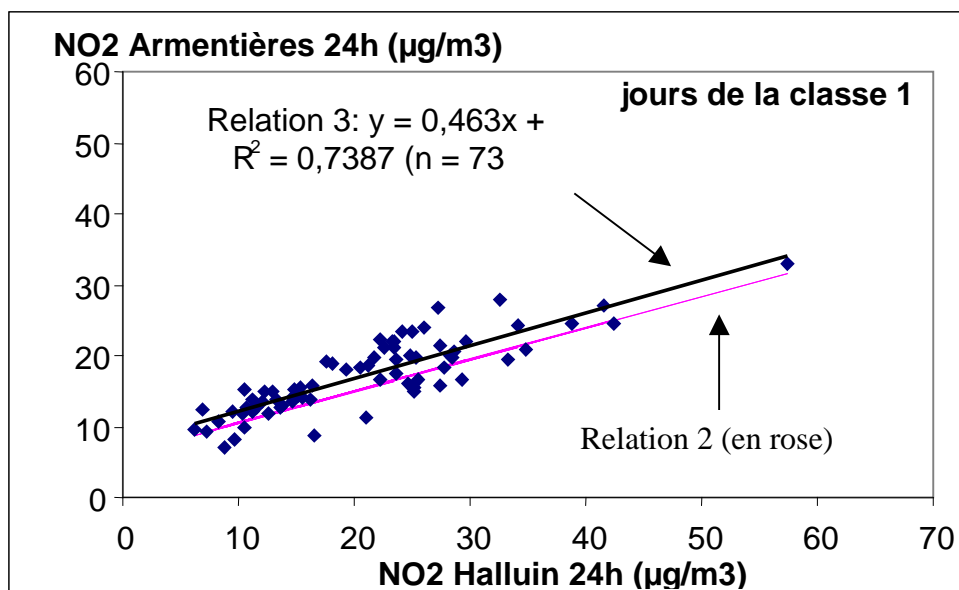


Figure 5 – Régression linéaire entre les concentrations journalières en NO₂ à Halluin et à Armentières, établie à partir des jours de la classe 1.

D'après les données du Tableau 4 et de la Figure 6, relatives aux résultats de classification hiérarchique, la classe 1 correspond à des conditions anticycloniques prononcées (pression très élevée), avec un vent de secteur nord-est, apportant un temps plutôt sec et des températures douces, à la fin du printemps (mois de mai), et générant une activité photochimique prononcée (concentration en ozone élevée).

La classe 2 correspond, elle aussi, à des concentrations en NO₂ faibles et en O₃ élevées ; mais elle se distingue de la classe 1 par la direction de vent et une pression moins élevée.

Quant à la classe 4, qui comporte aussi des vents de secteur nord-est, elle se distingue de la classe 2 par des concentrations en NO₂ plus élevées (30-31 µg/m³ en moyenne pour la classe 4 contre 17-20 µg/m³ en moyenne pour la classe 1).

Sur la Figure 6 suivante, on distinguera :

- Graphes de la colonne 1 : concentrations en NO₂ (µg/m³) calculées en fonction des concentrations vraies mesurées, à la station Halluin, pour une classe donnée.
- Graphes de la colonne 2 : fréquence des jours de la semaine pour une classe donnée.
- Graphes de la colonne 3 : rose des vents des jours appartenant à une classe donnée.

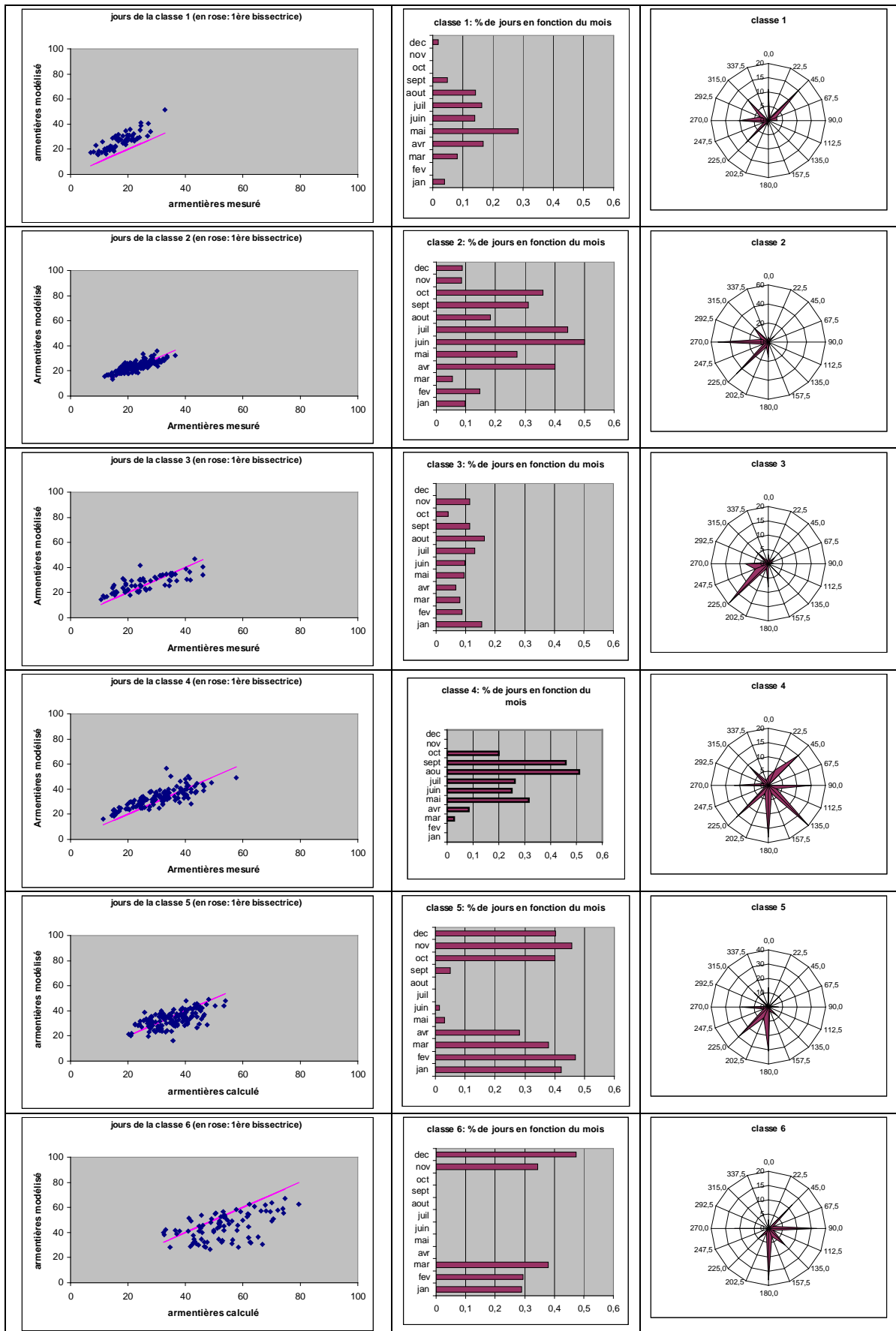


Figure 6 – Classification hiérarchique des données journalières : cas des stations périurbaines.

3.3.3 SYNTHÈSE DES RESULTATS POUR LE CAS DES STATIONS PERIURBAINES

Les résultats précédents permettent de mettre en évidence six paramètres potentiels, permettant de caractériser les jours où la relation statistique s'applique mal :

- pression (en moyenne supérieure à 1013 hPa) ;
- direction de vent particulière (nord-est) ;
- température (en moyenne supérieure à 10°C) ;
- mois de l'année (mai-août) ou saison (printemps-été) ;
- concentration en ozone (en moyenne supérieure à 50 $\mu\text{g}/\text{m}^3$) ;
- concentration en dioxyde d'azote (en moyenne faible entre 15 et 20 $\mu\text{g}/\text{m}^3$).

Nous avons donc essayé d'utiliser ces paramètres pour reconnaître les jours où la relation statistique s'applique mal.

Nous avons considéré chaque paramètre, un à un, et avons cherché à lui attribuer une valeur discriminante, permettant d'identifier le plus grand nombre de « mauvais » jours (exemple : pression > 1013 hPa). Nous avons constaté qu'aucun paramètre, à lui seul, ne permet de reconnaître au moins les deux tiers des « mauvais » jours.

Nous avons alors considéré les paramètres, deux à deux. Nous avons choisi une valeur peu restrictive pour le premier paramètre, permettant d'opérer une première sélection de jours, puis nous avons appliqué sur cette sélection un critère avec le second paramètre.

Cette méthode à deux paramètres n'a pas conduit à de meilleurs résultats, c'est-à-dire qu'aucune combinaison de deux paramètres, quel que soit l'ordre dans lequel ces deux paramètres sont utilisés, n'a permis de reconnaître au moins les deux tiers des « mauvais jours ».

Face aux mauvais résultats de ces essais, nous avons décidé d'abandonner cette voie.

3.4 CARACTERISATION DES SITUATIONS POUR LESQUELLES LES RELATIONS STATISTIQUES BI-HEBDOMADAIRES NE S'APPLIQUENT PAS : CAS DES STATIONS URBAINES :

La même démarche a été appliquée. Les stations urbaines « Lille Fives » et « Roubaix château » ont été considérées. On estime les concentrations journalières de la station urbaine « Lille Fives » à partir des concentrations journalières de la station urbaine de référence « Roubaix château », en utilisant la régression linéaire (Relation 4) établie entre les deux stations pour des concentrations bi-hebdomadaires :

$$\text{Con_jour_Fives } (\mu\text{g}/\text{m}^3) = 1,00 \times \text{Conc_jour_Roubaixchâteau } (\mu\text{g}/\text{m}^3) + 6,00$$

(Relation 4)

Comme pour le cas précédent les années 1998 à 2003 sont considérées, à l'exception de l'année 2002 qui est gardée pour validation. Le nombre de données disponibles (correspondant aux jours où les concentrations moyennes des deux stations ont pu être mesurées simultanément) s'élève à 1683 données.

3.4.1 TRI SELON L'ERREUR RELATIVE ABSOLUE

Comme pour les stations périurbaines, nous avons trié les « bons » et les « mauvais » jours, en prenant comme critère de sélection l'erreur relative en valeur absolue. Compte-tenu des objectifs de qualité (Tableau 1), nous avons considéré comme « mauvais » les jours pour lesquels l'erreur relative en valeur absolue était supérieure ou égale à 50% (Tableau 5).

Tableau 5 – Estimation des concentrations journalières par régression linéaire : cas des stations périurbaines.

Stations périurbaines	Tous les jours	Jours où la relation s'applique bien ^(a)	Jours où la relation s'applique mal ^(b)
Nombre de données	1683	1583 (94%)	100 (6%)
Coefficient de corrélation (r^2) entre estimations et mesures	0,70	0,73	0,56
Concentration moyenne des mesures à la station « Lille Fives » ($\mu\text{g}/\text{m}^3$)	39,6	40,7	23,1
Erreur moyenne ($\mu\text{g}/\text{m}^3$)	1,45	0,64	14,32
Erreur relative moyenne en valeur absolue (%)	19,0	15,8	69,4

(a) Les jours pour lesquels la relation statistique s'applique bien sont les jours pour lesquels l'erreur relative, en valeur absolue, est strictement inférieure à 50%.

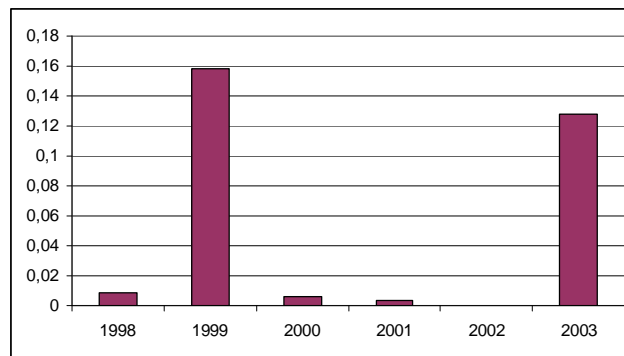
(b) Les jours pour lesquels la relation statistique s'applique mal sont les jours pour lesquels l'erreur relative, en valeur absolue, est supérieure ou égale à 50%.

La séparation des jours en deux classes, selon l'erreur relative supérieure ou non à 50%, montre que les « mauvais » jours constituent 6% des données, soit 22 j/an. Ils correspondent à des journées pour lesquelles la concentration en NO_2 est faible (23,1 $\mu\text{g}/\text{m}^3$ en moyenne).

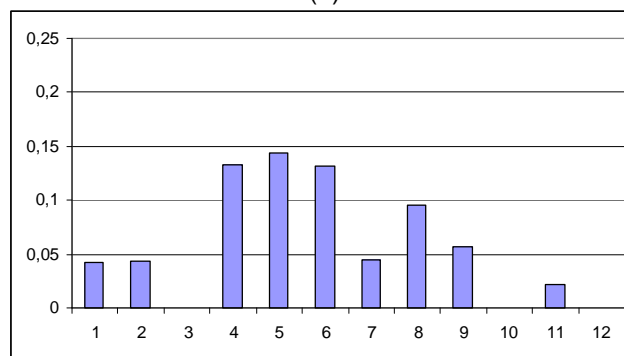
Un examen plus détaillé de ces « mauvais » jours montrent que leur fréquence est plus élevée :

- les années 1999 et 2003 (Fig. 7a) ;
- aux mois d'avril, mai, juin 1999 (Fig. 7b), et au mois d'août 2003 ;
- le dimanche (Fig. 7c).

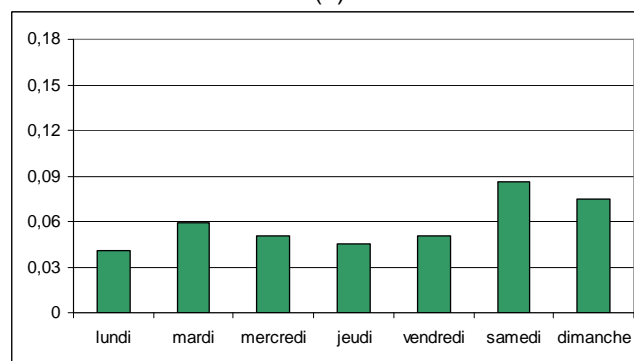
Ces caractéristiques sont parfaitement cohérentes avec celles mises en évidence pour les stations périurbaines, mis à part la fréquence élevée de « mauvais » jours à la fin du printemps 1999.



(a)



(b)



(c)

Figure 7 – Proportion de mauvais jours en fonction de l'année (a), du mois (b), du jour de la semaine (c), dans le cas des stations urbaines.

Comme pour les stations périurbaines, nous avons étudié les caractéristiques météorologiques des « mauvais jours », au travers de paramètres mesurés à la station Météo France de Lille Lesquin (pression, température à 13h, température minimale, température maximale, direction et force du vent, pluviométrie) et des concentrations journalières en ozone mesurées à la station périurbaine d'Halluin-Cailloux.

Les résultats montrent que les « mauvais » jours sont caractérisés par des secteurs de vent très différents (Fig. 8a) des secteurs normalement rencontrés (Fig. 8b), en particulier provenant du nord-nord-ouest et du quart est-nord-est.

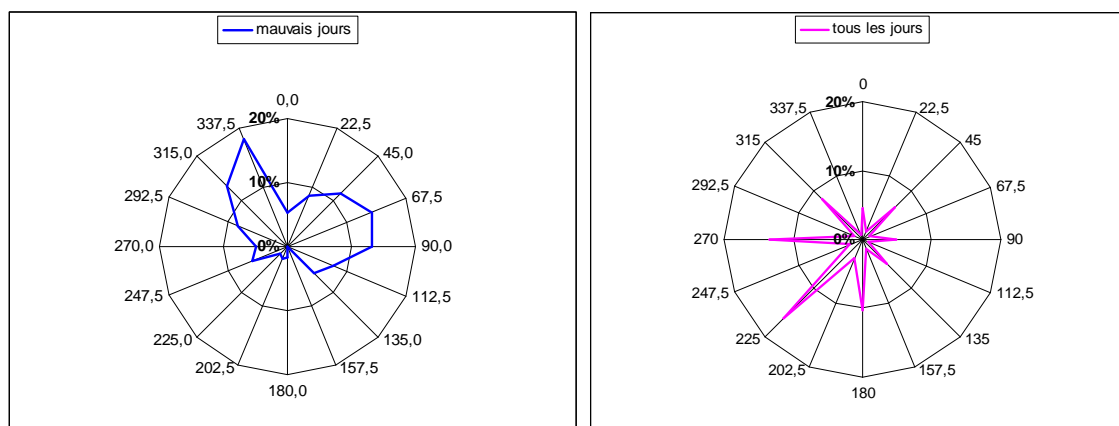


Figure 8 – Provenance du vent dans le cas des stations urbaines
(a) lors des « mauvaises » journées
(b) pour la totalité des jours.

Comme pour les stations périurbaines, on constate que les « mauvais » jours se caractérisent par une augmentation de la concentration en ozone et des valeurs des paramètres météorologiques suivants : pression, température, vitesse de vent (Tableau 6).

En revanche, contrairement au cas des stations périurbaines (Tableau 3), le cumul des précipitations diminue ; confirmant le fait que ce paramètre n'est pas un facteur déterminant.

Tableau 6 – Comparaisons des paramètres météorologiques pour les bons et mauvais jours (cas des stations urbaines).

Stations urbaines	Tous les jours	Jours où la relation s'applique bien ^(a)	Jours où la relation s'applique mal ^(b)
Moyenne journalière NO ₂ à Roubaix ch. (µg/m ³)	35,1	35,3	31,4
Moyenne journalière NO ₂ à Lille Fives (µg/m ³)	39,6	40,7	23,1
Erreur relative absolue moyenne (%)	19,0%	15,8%	69,4
Pression (hPa)	1015,5	1015,4	1017,4
Température à 13h (°C)	13,6	13,3	18,4
Température minimale (°C)	-7,9	-7,9	-3,1
Température maximale (°C)	36,6	36,6	32,3
Vitesse du vent (km/h)	19,4	19,4	19,8
Précipitations journalières (L/m ²)	2,02	2,05	1,58
Moyenne journalière O ₃ à Halluin (µg/m ³)	38,6	38,0	46,0

(a) Jours pour lesquels l'erreur relative, en valeur absolue, est strictement inférieure à 50%.

(b) Jours pour lesquels l'erreur relative, en valeur absolue, est supérieure ou égale à 50%.

On peut donc formuler la même hypothèse que pour les stations périurbaines, à savoir que les jours, où la relation statistique entre stations ne s'applique pas, correspondent à des situations anticycloniques (pression élevée), plutôt printanières ou estivales (température élevée), avec une activité photochimique prononcée (concentration en ozone élevée, concentration en NO₂ basse). Le phénomène incriminé exerce donc une influence sur toute la zone d'étude, en milieu urbain comme en milieu périurbain.

On constate, par ailleurs, comme dans le cas des stations périurbaines, que les concentrations en NO₂ mesurées lors des « mauvais » jours sont surestimées d'environ 66% par la Relation 4. Il est possible d'établir une nouvelle relation pour ces « mauvais » jours (Relation 5, Fig. 9) :

$$\text{Conc_jour_Armentières } (\mu\text{g}/\text{m}^3) = 0,6282 \times \text{Conc_jour_Halluin } (\mu\text{g}/\text{m}^3) + 2,5$$

(Relation 5)

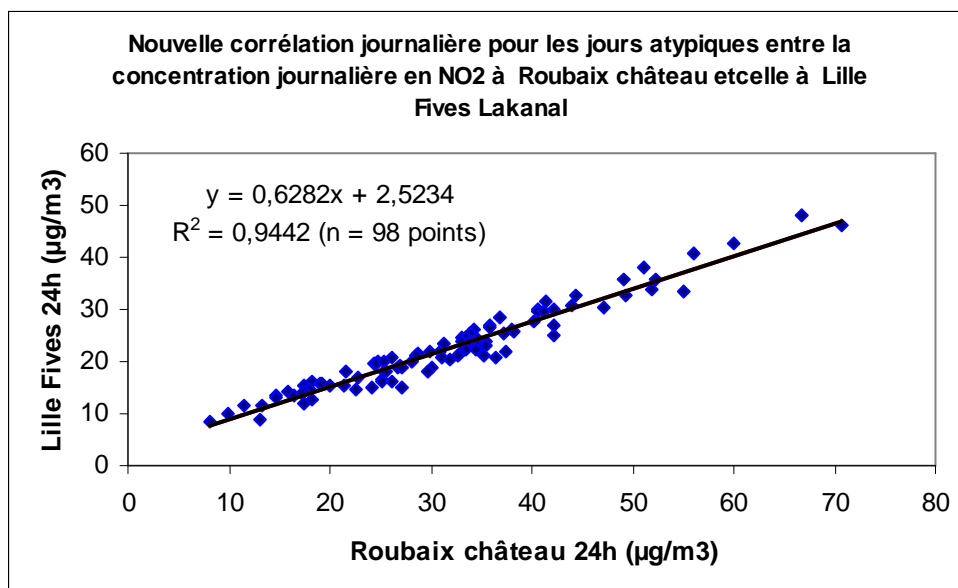


Figure 9 – Nouvelle régression linéaire entre les concentrations journalières en NO₂ mesurées à Lille Fives en fonction de celles mesurées à Roubaix château.

3.4.2 TRI PAR CLASSIFICATION HIERACHIQUE

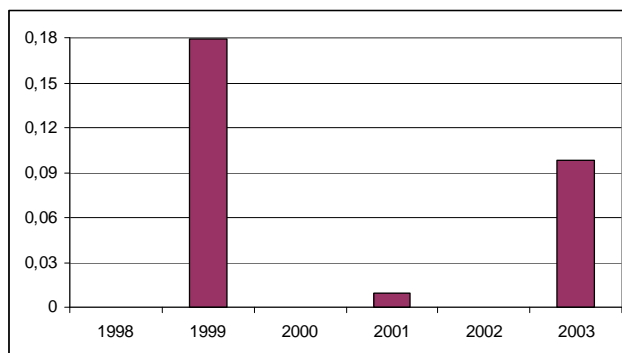
Comme pour le cas des stations périurbaines, la classification hiérarchique ascendante a été utilisée, selon la méthode de Ward, pour classer les jours selon leurs caractéristiques météorologiques et de pollution.

L'analyse a porté sur 821 jours pour lesquels les 10 variables suivantes étaient renseignées : erreur relative absolue, concentration en NO₂ à Lille Fives, concentration en NO₂ à Roubaix château, concentration en ozone à Halluin, pression atmosphérique, cumul journalier des précipitations, température maximale, température minimale, vitesse du vent et visibilité.

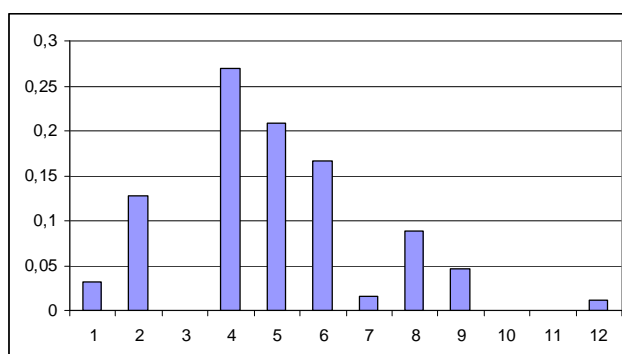
Cinq classes différentes ont été distinguées (Tableau 7).

La cinquième classe (classe 5) se distingue des quatre autres par une erreur relative moyenne très élevée (67%).

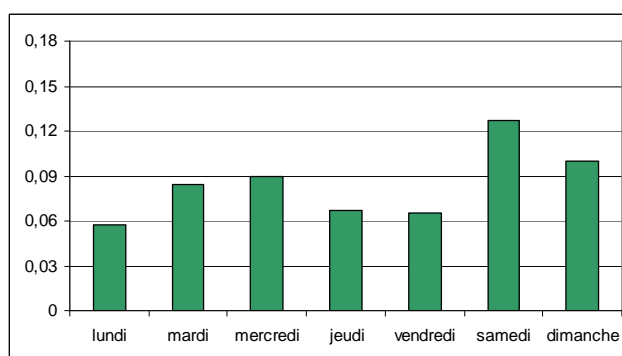
Si l'on compare la fréquence des jours appartenant à la classe 5 en fonction de l'année, du mois ou du jour de la semaine (Fig. 10 a,b,c), on remarque que la classe 5 présente des caractéristiques similaires à celles du groupe des « mauvais » jours (Fig. 7 a,b,c), notamment une fréquence élevée les années 1999 et 2003, les mois d'avril, mai et juin et les journées du samedi et dimanche.



(a)



(b)



(c)

Figure 10 – Proportion de jours appartenant à la classe 5 en fonction de l'année (a), du mois (b), du jour de la semaine (c), pour les stations urbaines.

Pour les jours de la classe 5, la régression linéaire entre les concentrations journalières en NO₂ à Roubaix château et celles à Lille Fives (Relation 6, Fig. 11) est proche de celle trouvée pour les « mauvais jours » (Relation 5).

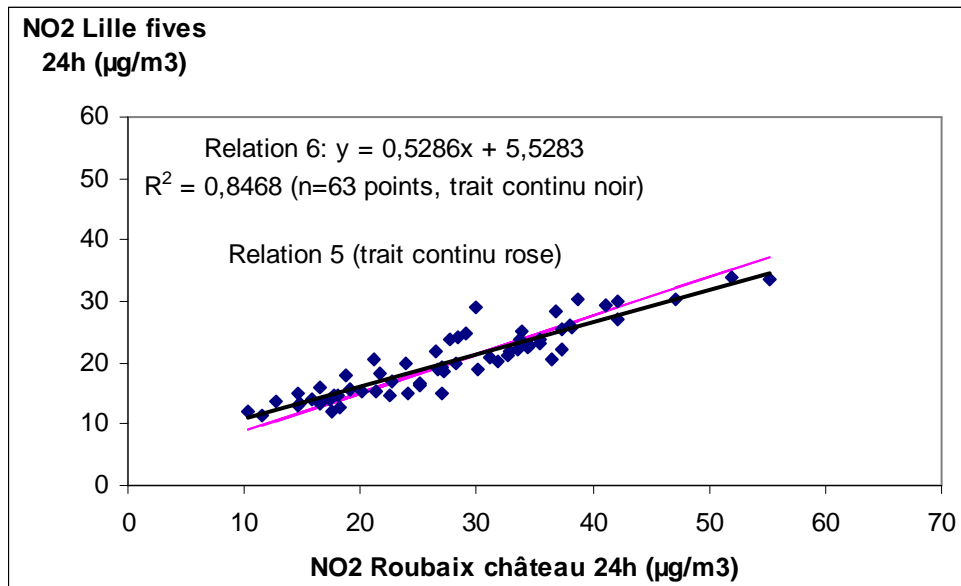


Figure 11 – Régression linéaire entre les concentrations journalières en NO₂ à Roubaix château et à Lille Fives, établie à partir des jours de la classe 5.

D'après les données du Tableau 7 et de la Figure 12, la classe 5 correspond à des conditions anticycloniques (pression élevée), avec un vent de secteur nord-est, apportant un temps très sec et des températures douces, au printemps (mois d'avril), et générant une activité photochimique prononcée (concentration en ozone élevée).

Tableau 7 – Résultats de la classification hiérarchique menée sur les données journalières de pollution et de météorologie : cas des stations urbaines.

Erreur : erreur relative absolue (%)
 NO₂ F : concentration moyenne journalière en NO₂ à Lille Fives (µg/m³)
 NO₂ R : concentration moyenne journalière en NO₂ à Roubaix Château (µg/m³)
 O₃ H : concentration moyenne journalière en O₃ à Halluin (µg/m³)
 P_hPa : pression moyenne journalière (hPa)
 Précip. : précipitations (L/m²)
 Tmax : moyenne des températures maximales (°C)
 Tmin : moyenne des températures minimales (°C)
 Vis_km : visibilité (km)
 VV : vitesse du vent (km/h).

classe	effectif	Erreur %	NO ₂ F	NO ₂ R	O ₃ H	P_hPa	Précip.	T_max	T_min	Vis_km	VV	Commentaire
1	294 (36%)	15	44	39	35	1015	1,50	14,2	7,3	17	19	NO ₂ moyen, O ₃ moyen Pression ≅ 1 atm, vents du quart SO
2	210 (25%)	21	30	25	51	1016	0,69	21,7	12,9	25	20	NO ₂ faible, O ₃ maximum Temps sec, températures maximales, visibilité maximale, saison estivale
3	106 (13%)	19	33	26	46	1004	8,36	14,8	9,5	22	28	NO ₂ faible, O ₃ moyen Régime dépressionnaire (pression minimale), vent de SO, pluviométrie maximale, températures douces
4	146 (18%)	16	58	55	12	1025	1,19	6,5	1,5	9	15	NO ₂ élevé (maximum), O ₃ très faible (minimum) Régime anticyclonique, visibilité faible, vent minimum, températures froides minimales, période hivernale
5	65 (8%)	67	21	27	46	1017	0,67	16,8	8,8	21	21	NO ₂ faible, O ₃ élevé Régime plutôt anticyclonique, vent de NE, très sec (précipitations minimales), visibilité modérée, mois d'avril

Total : 821

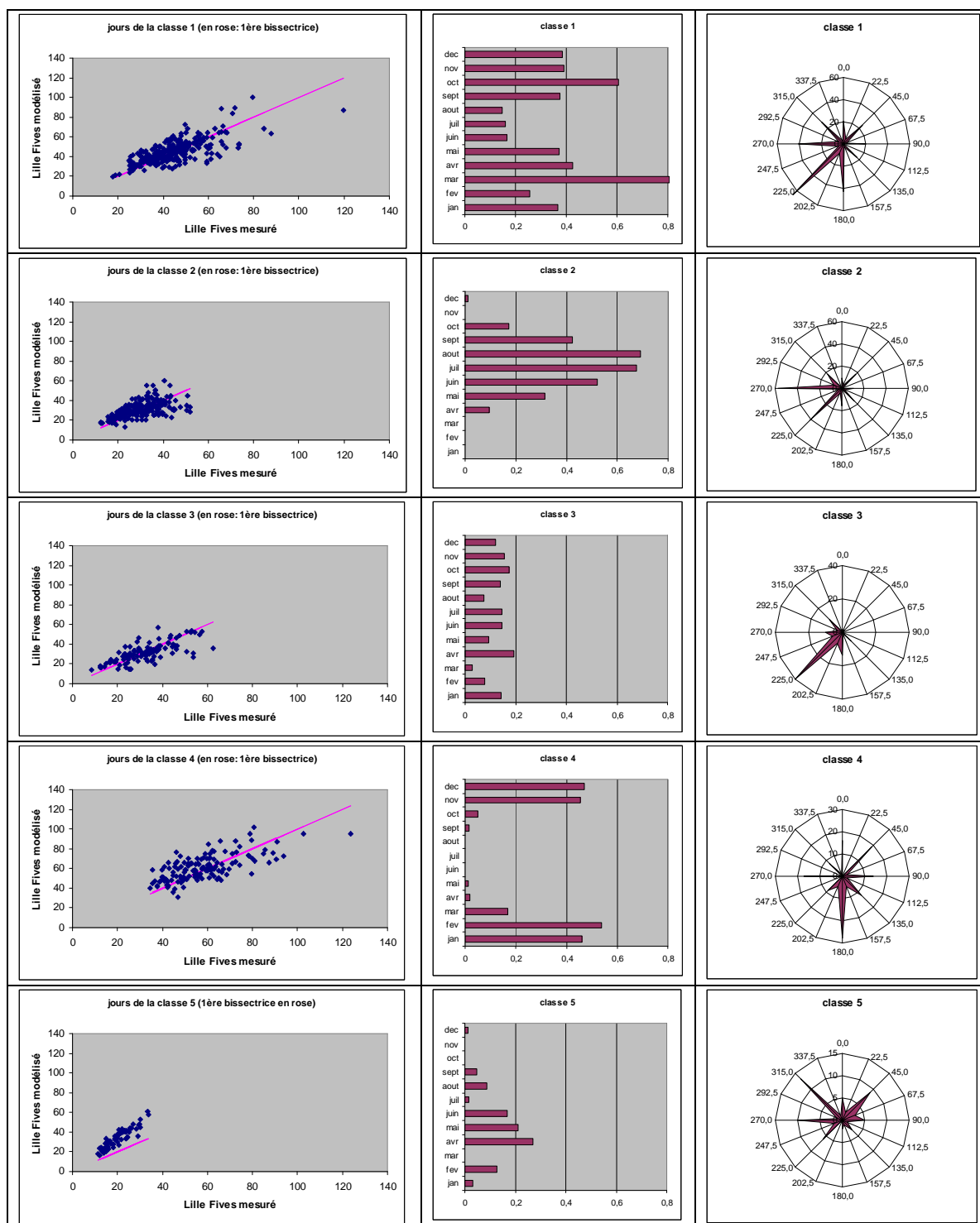


Figure 12 – Classification hiérarchique des données journalières : cas des stations urbaines.

- Graphes de la colonne 1 : concentrations en NO₂ (µg/m³) calculées en fonction des concentrations vraies mesurées, à la station Lille Fives, pour une classe donnée.
- Graphes de la colonne 2 : fréquence des jours de la semaine pour une classe donnée.
- Graphes de la colonne 3 : rose des vents des jours appartenant à une classe donnée.

3.4.3 SYNTHÈSE DES RÉSULTATS POUR LE CAS DES STATIONS URBAINES

Les résultats pour les stations urbaines montrent que les conditions météorologiques et de pollution qui règnent, lorsque la relation statistique bihebdomadaire s'applique mal, sont grosso modo les mêmes que pour les stations périurbaines. Elles correspondent à des situations météorologiques plutôt printanières ou estivales, accompagnées de concentrations en ozone plus élevées que la moyenne et de concentrations en dioxyde d'azote plus faibles que la moyenne.

Comme pour le cas des stations périurbaines, des valeurs discriminantes ont été recherchées pour les paramètres influents, pris un à un ou deux à deux. Mais aucune combinaison de paramètres ne s'est révélée satisfaisante.

3.5 SYNTHÈSE DES DIFFÉRENTES SITUATIONS MÉTÉOROLOGIQUES RENCONTRÉES, POUR LES STATIONS URBAINES ET PÉRIURBAINES.

Il est montré que, pour une même typologie de stations, les deux méthodes employées (tri selon les erreurs relatives ou classification hiérarchique) conduisent à des résultats similaires et cohérents. Il y a cohérence entre les résultats obtenus pour les différentes typologies de stations.

Ainsi, pour la méthode du « tri selon l'erreur relative », il existe 13 dates communes aux stations urbaines et périurbaines, pour lesquelles l'erreur relative absolue est strictement supérieure à 0,5.

L'erreur commise est toujours une surestimation de la concentration, en moyenne de 68%.

Les concentrations en NO₂ sont relativement faibles (Fig. 13a), la concentration en ozone plutôt élevée, des conditions météorologiques anticycloniques, printanières ou estivales, avec un vent de provenance du nord-est (Fig. 13b).

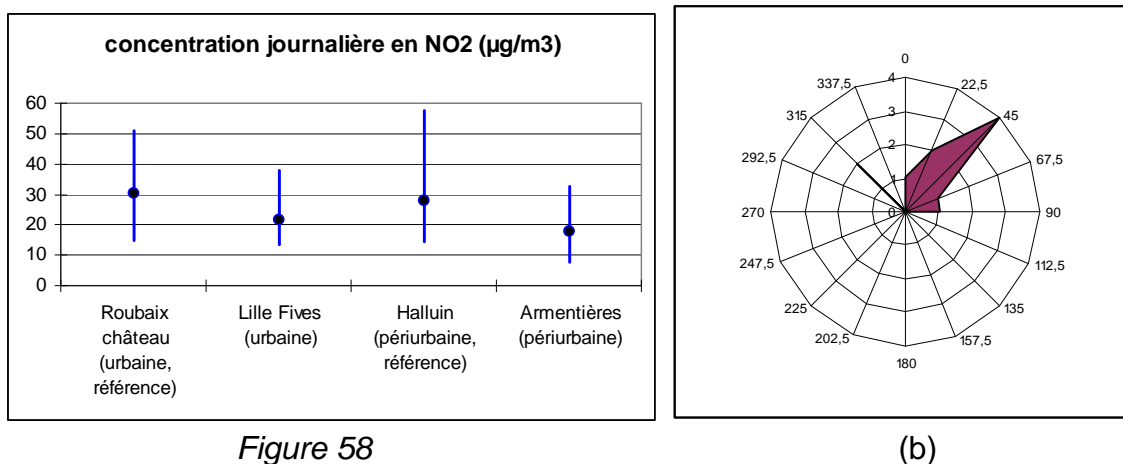


Figure 13 – Concentrations journalières moyennes en NO₂ et provenance du vent pour les 13 « mauvais » jours, triés selon l'erreur relative et communs aux stations périurbaines et urbaines.

De même, pour la méthode par classification hiérarchique, il existe 14 dates communes aux stations urbaines et périurbaines, appartenant aux classes caractérisées par une erreur relative élevée. Elles présentent les mêmes caractéristiques que précédemment (Fig. 14 a et b).

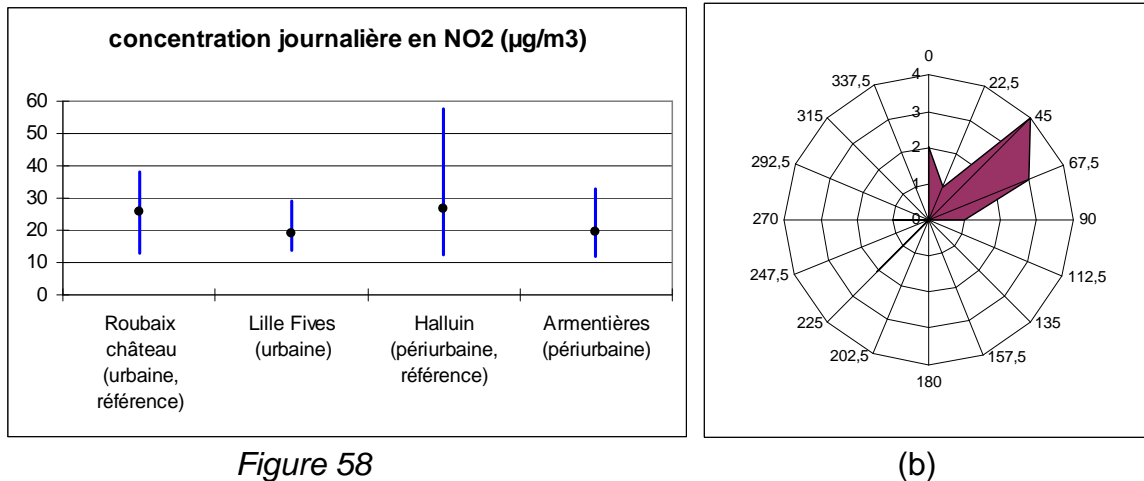


Figure 14 – Concentrations journalières moyennes en NO₂ et provenance du vent pour les 14 « mauvais » jours, triés par classification hiérarchique communs aux stations périurbaines et urbaines.

La méthode par classification hiérarchique permet en outre de distinguer les différentes conditions météorologiques de la zone sous surveillance. Si l'on procède à une classification à partir des seules variables météorologiques, on peut distinguer six grandes situations météorologiques (Annexe 1) :

- Les jours froids (38%), qui se séparent selon les précipitations en jours froids et secs (classe 3, 13%) et jours froids et humides (classe 1, 25%) ;
- Les jours chauds (62%), qui se séparent en deux groupes selon la pression atmosphérique en jours de dépressions (24%, dont 4% très pluvieux constituant la classe 5, les 20 % restant constituant la classe 2) et jours d'anticyclones (38%, dont 21% en saison estivale constituant la classe 6, et les 17% restant constituant la classe 4 plutôt au printemps et à l'automne avec peu de précipitations).

En revanche, lorsque l'on procède à la classification en ajoutant des variables de pollution (concentrations en dioxyde d'azote et en ozone) et en ajoutant l'erreur relative commise lors de l'estimation des concentrations, on oriente la classification différemment (Figure 15) :

On retrouve bien une classe correspondant à la saison hivernale et une autre à la saison estivale, ainsi qu'une classe correspondant aux jours très pluvieux (dépression, vent de SO). Cependant, l'effectif de chaque classe est alors limité. La majorité des jours se retrouve en effet dans une seule classe (à effectif élevé), c'est-à-dire que la classification est, dans ce cas, moins discriminante (elle différencie moins de classes). A noter que le paramètre « erreur relative » permet tout de même de distinguer les jours, caractérisés par une erreur relative d'estimation très importante, qui forment une classe particulière.

Cette classe particulière (cf. Figure 15 : « vent de NE printanier ») est caractérisée par des concentrations en NO₂ plus faibles qu'à l'accoutumée, un vent de nord-est, apportant un temps doux et sec de printemps (Fig. 15 a, b,c).

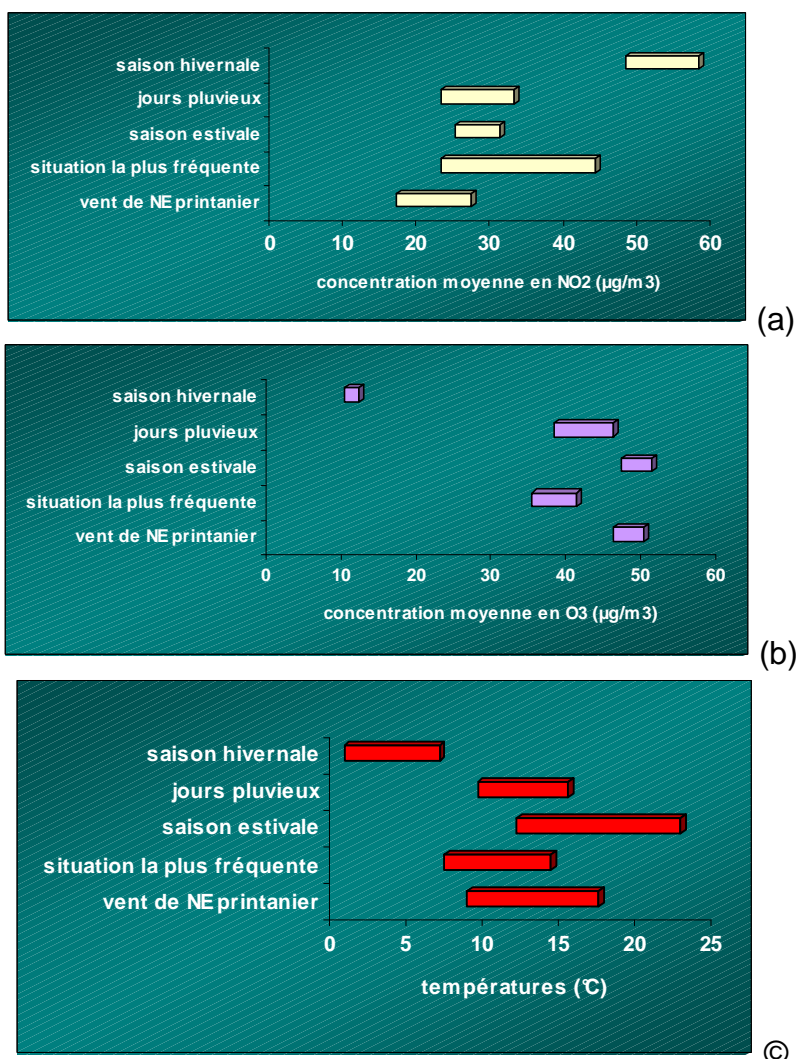


Figure 15 – Résultats de la classification selon la météorologie, la pollution et la qualité de l'estimation (erreur relative d'estimation).

Le fait que la technique de classification hiérarchique conduise à moins de classes distinctes, lorsque l'on considère à la fois les paramètres météorologiques et les paramètres de pollution, semble indiquer que, pour la zone considérée, il n'existerait pas autant de situations de pollution différentes que de situations météorologiques.

En conclusion de cette partie, nous avons étudié les conditions d'application des relations statistiques bi-hebdomadaires, entre stations, et montré que ces relations pouvaient s'appliquer avec succès à l'estimation des valeurs des concentrations journalières, sauf pour certains jours particuliers, pour lesquels une autre relation statistique doit être appliquée. Nous avons remarqué que ces jours particuliers se caractérisaient par une concentration mesurée environ 60% plus faible que celle estimée par la relation statistique classique. Une façon simple d'estimer la concentration en NO₂ de ces jours particuliers serait donc de multiplier la concentration estimée par la relation statistique classique par un facteur donné (0,6).

Pour tester cette hypothèse, nous avons considéré les données de l'année 2002 (année gardée pour la validation). Dans un premier temps, nous avons appliqué la relation statistique bi-hebdomadaire « classique ». Puis, après calcul de l'erreur relative absolue commise et identification des jours pour lesquels la relation bi-hebdomadaire s'applique mal (erreur relative absolue > 50%), nous avons ré-estimé les « mauvaises » concentrations en les multipliant par un facteur 0,6.

Appliquée au couple de stations périurbaines « Armentières » et « Halluin cailloux », la méthode précitée conduit aux résultats suivants :

Sur 365 valeurs de concentrations journalières estimées (concentrations à Armentières en 2002 estimées à partir de celles d'Halluin-cailloux en 2002), 25 sont jugées mauvaises (7%), avec une erreur relative absolue moyenne de 109%. En revanche lorsque l'on multiplie les valeurs estimées de ces 25 « mauvaises » concentrations par un facteur correctif égal à 0,6, on constate que l'erreur relative absolue moyenne diminue fortement jusqu'à 33%. Qui plus est, un examen des données montre que seules deux données présentent alors une erreur relative absolue supérieure à 50% : il s'agit de deux journées pour lesquelles les concentrations à estimer sont très faibles, comprises entre 0 et 5 µg/m³. Or cette gamme de concentrations est difficilement mesurable par les analyseurs usuels. On peut donc raisonnablement penser que ces deux mesures sont entachées d'une incertitude plus grande et qu'elles ne conviennent pas pour le calcul d'erreur. Si l'on écarte ces deux valeurs trop faibles, on constate que l'erreur relative moyenne, calculée après application du facteur correctif sur les 23 valeurs restantes, n'est plus que de 7%, ce qui constitue un excellent résultat.

Finalement, pour parachever complètement notre méthode et pouvoir l'appliquer à l'estimation de données journalières, il reste à trouver un moyen sûr d'identifier les jours pour lesquels la relation statistique s'applique bien et ceux, minoritaires, pour lesquels un facteur correctif doit être appliqué. L'analyse des situations météorologiques a permis de trouver des paramètres de sélection potentiels. Mais aucune combinaison simple de ces paramètres n'a permis de constituer un test de sélection efficace. La prochaine étape, non encore atteinte, serait donc de définir un test de sélection fiable, vraisemblablement plus complexe, à partir des variables significatives mises en évidence dans cette étude.

4 AUTOMATISATION DE LA PROCEDURE DE CARTOGRAPHIE DU DIOXYDE D'AZOTE

Dans cette partie, l'automatisation de la procédure de cartographie des concentrations en dioxyde d'azote par des méthodes géostatistiques est étudiée.

4.1 CARTOGRAPHIE PAR DES METHODES GEOSTATISTIQUES

Pour rappel, parmi les méthodes géostatistiques les plus utilisées dans le contexte de la qualité de l'air, nous pouvons mentionner : les méthodes monovariées, composées par le krigeage ordinaire et le krigeage simple ou à moyenne connue, et les méthodes multivariées dont le but est d'exploiter au mieux une information additionnelle apportée par une variable auxiliaire.

Les méthodes multivariées sont une adaptation du krigeage à différentes hypothèses de façon à prendre en compte la variable auxiliaire ; plus particulièrement, pour l'estimation du dioxyde d'azote, la prise en compte des variables auxiliaires décrivant le milieu, comme le relief, la densité de population ou le cadastre d'émissions permet d'obtenir des cartes plus détaillées de la répartition spatiale du dioxyde d'azote.

En qualité de l'air les trois méthodes multivariées qui ont été les plus utilisées, sont le cokrigeage, le krigeage en dérive externe et une variante de cette dernière qui est le krigeage des résidus.

Le krigeage repose sur la connaissance de la corrélation spatiale des données, la mesure de cette corrélation est donnée par le calcul du variogramme spatial expérimental et l'ajustement manuel d'un modèle (ou structure) prédéfini de covariance.

Le cokrigeage permet l'estimation simultanée de deux variables qui sont corrélées. L'estimation de chaque variable prend en compte sa propre corrélation spatiale mais aussi la corrélation spatiale de l'autre variable. Pour appliquer le cokrigeage, il faut donc calculer les variogrammes expérimentaux simples et croisés puis ajuster manuellement un modèle prédéfini de covariance.

La méthode de la dérive externe repose sur la décomposition de la concentration du dioxyde d'azote en deux variables : une dérive qui traduit la tendance générale, et un résidu qui représente plutôt les fluctuations, les variations à petites échelles :

$$Z(x) = D(x) + R(x)$$

Dans notre cas, la dérive peut être une fonction linéaire des variables auxiliaires (densité, émissions, relief, etc.), cela signifie qu'elles sont censées représenter l'allure générale de la concentration, et non ses variations à petites échelles.

Si la fonction de dérive est connue dans le domaine, on peut calculer les résidus dans les points expérimentaux et les estimer par krigeage en ajustant un modèle à leur variogramme expérimental. Une fois les résidus estimés, on obtient la valeur de la concentration, cette méthode est appelée le krigeage des résidus (KR):

$$D(x) = \alpha * f(x) + \beta$$

$$R(x) = Z(x) - D(x) \quad \text{et} \quad R(x) = Z(x) - [\alpha * f(x) + \beta]$$

$$R(x)^K = \sum_i \lambda_i R(x_i) \quad \text{et} \quad \text{donc} \quad Z(x)^{KR} = D(x) + R(x)^K$$

Dans le krigeage en dérive externe, on suppose que les coefficients α et β de la fonction linéaire des variables auxiliaires ne sont pas fixés; ils sont ajustés pour chaque point lors du krigeage. Le variogramme expérimental des résidus n'est donc pas accessible, on le calcule indirectement par essais successifs et de façon automatique. Dans cette méthode on estime directement la concentration du dioxyde d'azote sans passer par les résidus.

Pour choisir l'approche la plus adaptée, la démarche classique consiste à comparer les résultats obtenus par les différentes méthodes. Par ailleurs, on peut les comparer avec le krigeage usuel en l'absence de variable auxiliaire, pour justifier ou non, l'apport de celle-ci.

Voici les étapes d'une étude géostatistique ayant comme objet la cartographie du dioxyde d'azote :

1. Dépouillement des données.
2. Corrélation avec les variables auxiliaires : Il s'agit d'une phase itérative où l'on teste plusieurs configurations possibles des variables explicatives de façon à trouver une fonction très bien corrélée à la concentration.
3. Calcul du ou des variogrammes expérimentaux et ajustement du ou des modèles respectifs (en fonction de la méthode choisie):
 - . KO : Calculer le variogramme expérimental du dioxyde d'azote et ajuster un modèle.
 - . CK : Calculer les variogrammes simples et croisés du dioxyde d'azote et de la fonction explicative, et ajuster un modèle bivariable.
 - . KR : Trouver les résidus, calculer le variogramme expérimental des résidus, et ajuster un modèle.
 - . KDE : Ajuster automatiquement le modèle de variogramme approprié.
4. Choix du meilleur modèle proposé : pour choisir la méthode qui donne les meilleures estimations du dioxyde d'azote, la méthode de la validation croisée est employée. Cette méthode consiste à réestimer la concentration en tous points de mesure à l'aide du modèle de variogramme et de la méthode choisie, ensuite les vraies concentrations et les concentrations ainsi réestimées sont comparées.

4.2 AUTOMATISATION DE LA CARTOGRAPHIE

Le but de cette étude est d'étudier la faisabilité d'une automatisation de la procédure de calcul des variogrammes expérimentaux, afin d'obtenir la carte la plus précise des niveaux de pollution correspondants.

Dans un premier temps, nous nous sommes intéressés à la faisabilité d'une cartographie automatique journalière à partir des données des capteurs dans une agglomération.

Mais comme le nombre de capteurs est généralement insuffisant pour générer une carte représentative de la variabilité des concentrations, ces données doivent être complétées par des données supplémentaires.

Dans la première partie de cette étude, une méthode qui peut être utilisée pour générer ces données supplémentaires a été présentée. Elle est basée sur l'existence d'un historique de campagnes par échantillonneurs passifs dans l'agglomération.

Celles-ci sont exploitées de façon à pouvoir grouper les tubes (appelées désormais stations virtuelles), qui doivent être corrélés temporellement à une station fixe particulière appelée station de référence.

Dans les « stations virtuelles » ainsi déterminées il est donc possible, en appliquant cette méthode d'estimer la concentration du dioxyde d'azote, à n'importe quel moment, à partir de la valeur mesurée au point de référence correspondant.

Cette méthode est une alternative pour la génération des données supplémentaires. Cependant, pour les agglomérations où il n'y a pas un nombre suffisant de campagnes par tubes passifs, d'autres méthodes doivent être employées. Une autre possibilité consisterait à exploiter des campagnes mobiles et des modèles de prévision (modèles de type ANACOVA), etc.

Les données provenant des stations virtuelles sont, par construction, entachées d'erreurs. La géostatistique permet de prendre en compte leur incertitude grâce à la variance de l'erreur de mesure (VEM), terme qui peut s'ajouter à la matrice variance - covariance lors du krigeage. Le calcul de la variance de l'erreur de mesure pour les stations virtuelles sera discuté dans l'étude de cas.

Le même jeu de données que pour la première partie de l'étude a été utilisé. Il est constitué de plusieurs campagnes de mesure, de 14 jours chacune, répétées sur une même zone géographique (l'agglomération de Lille), au cours des années 1998/1999 et 2003/2004.

Les étapes essentielles de la procédure sont celles relatives à la recherche de la fonction auxiliaire corrélée spatialement avec la pollution et (par conséquent) celles relatives au calcul des variogrammes et à l'ajustement des modèles respectifs.

Dans un premier temps par souci de robustesse, des configurations préétablies ont été testées (de cette façon nous exploitons les connaissances disponibles du phénomène) :

1. Recherche de la fonction auxiliaire : le but de cette étape est de trouver une fonction connue dans l'ensemble du domaine corrélé avec la concentration du dioxyde d'azote. Cette fonction auxiliaire peut-être composée des variables décrivant le milieu comme la densité de population, le relief, ou le cadastre d'émissions. Nous proposons d'effectuer à l'avance une étude des toutes les configurations possibles de ces variables auxiliaires et de laisser celle qui est la mieux corrélée aux concentrations. Après l'avoir choisi, cette fonction sera utilisée tous les jours dans l'algorithme de cartographie.

Note : Cependant la fonction peut être différente selon la saison (une pour l'été et une autre pour l'hiver). En effet, plusieurs études ont démontré qu'en été la corrélation entre le dioxyde d'azote et les variables auxiliaires décrivant le milieu est moins bonne que pour les autres saisons. Par ailleurs nous savons qu'il existe une relation décroissante entre la température et la concentration en NO_2 , c'est pourquoi d'un point de vue de surveillance, la cartographie du NO_2 est plus intéressante en hiver qu'en été où les concentrations sont faibles.

2. Pour l'ajustement des variogrammes, les structures de base (effet de pépite, choix du modèle : sphérique, exponentiel ou gaussien, etc) ont été choisies à l'avance, à partir des informations disponibles telles que l'analyse des campagnes des tubes existants.
3. La meilleure méthode d'estimation, parmi les quatre disponibles (krigeage ordinaire, cokrigeage, krigeage en dérive externe et krigeage des résidus) a été sélectionnée à partir des conclusions que l'on peut tirer des estimations sur les campagnes des tubes existants.

Le Tableau 8 présente une liste des étapes préliminaires nécessaires pour déterminer ces paramètres préétablis. Ensuite, dans le Tableau 9 sont présentées les procédures qui peuvent être automatisées sur le logiciel Isatis.

L'étude des données de l'agglomération de Lille est présentée. A partir de cette étude de cas, la méthodologie préconisée est appliquée et la qualité des résultats obtenus est analysée dans les paragraphes suivants.

<i>Etapes préliminaires à effectuer</i>	
<i>Action</i>	<i>Résultat</i>
1. Sur les campagnes par tubes passifs existantes et les données des stations fixes : application de la méthode des corrélations temporelles.	Choix des stations de référence et des stations virtuelles. Détermination des paramètres de corrélation : la pente et l'interception de l'axe Y.
2. Calcul de la VEM des stations virtuelles.	VEM.
3. Etude des variables auxiliaires disponibles : relief, émissions, densité de population, données météo, etc.	Choix de la fonction auxiliaire la mieux corrélée aux concentrations.
4. Calcul des variogrammes anisotropes et isotropes de la concentration et de la variable auxiliaire.	Recherche de possibles anisotropies, analyse de la qualité de la variable auxiliaire.
5. Ajustement du ou des modèles de variogrammes.	Choix des structures de base et des portées des modèles.
6. Test des modèles par validation croisée.	Validation des modèles et choix de la méthode d'estimation : krigeage ordinaire, cokrigeage, krigeage en dérive externe ou krigeage des résidus.

Tableau 8 : Etapes préliminaires

Automatisation de la procédure	
<i>Action</i>	<i>Résultat</i>
1. Création du jeu de données à l'aide des données de stations de référence et des paramètres des corrélations temporelles (la pente et l'interception de l'axe Y).	Jeu de données des concentrations journalières de dioxyde d'azote.
2. Sur le logiciel Isatis : Création de l'algorithme permettant l'obtention automatique des cartes d'estimation journalières du dioxyde d'azote (travail en mode Batch).	L'algorithme doit effectuer les tâches suivantes : <ol style="list-style-type: none"> 1. Calcul du nuage de corrélation avec la variable auxiliaire préétablie. 2. Calcul des variogrammes expérimentaux. 3. A l'aide de ces variogrammes expérimentaux : ajustement des paliers de structures de base choisis à l'avance. 4. Validation croisée. 5. Par la méthode de krigeage choisie à l'avance : interpolation de concentrations sur une maille. 6. Création des cartes estimées et de la carte de la variance d'estimation à l'aide d'une échelle des couleur préétablie. 7. Exportation des résultats : <ul style="list-style-type: none"> • Exportation des cartes résultantes en format d'image bmp ou jpeg. • Exportation des fichiers résultants en format ascii ou grid d'ARCVIEW pour une exploitation ultérieure dans un SIG.
3 Analyse des résultats	Voici quelques informations nécessaires pour déterminer la qualité des estimations obtenues : <ol style="list-style-type: none"> 1. Nuage de corrélation entre la concentration et la variable auxiliaire (coefficient de corrélation). 2. Variogrammes expérimentaux et modèles ajustés. 3. Statistiques de la validation croisée : variance de l'erreur et erreur relative. 4. Nuage de corrélation : concentrations des stations virtuelles /concentrations estimées (coefficient de corrélation). 5. Statistiques des estimations résultantes et de la variance d'estimation.

Tableau 9: Automatisation de la procédure

4.3 ETUDE DE CAS

4.3.1 DONNEES DISPONIBLES

4.3.1.1 VARIABLES AUXILIAIRES

La Figure 16 montre les courbes de niveaux de l'inventaire d'émissions (NO_x) interpolés dans une maille de 250 m (à l'origine l'inventaire était disponible avec une maille de résolution de 1 Km). Comme on pouvait s'y attendre, les niveaux d'émissions sont très forts dans la zone urbaine de Lille et près des routes.

Lille: Emissions et Capteurs

Emissions NO_x (g/Km²)

— 0 - 30

— 31 - 70

— 71 - 120

Capteurs NO_2 (Nb=16)

Typologie

◆ Industriel (Nb=2)

● Périurbain (Nb=3)

▲ Trafic (Nb=5)

■ Urbain (Nb=6)

0 2.5 5 Km

Fond de Carte: Agglomération de Lille
(Source Internet: Mappy)

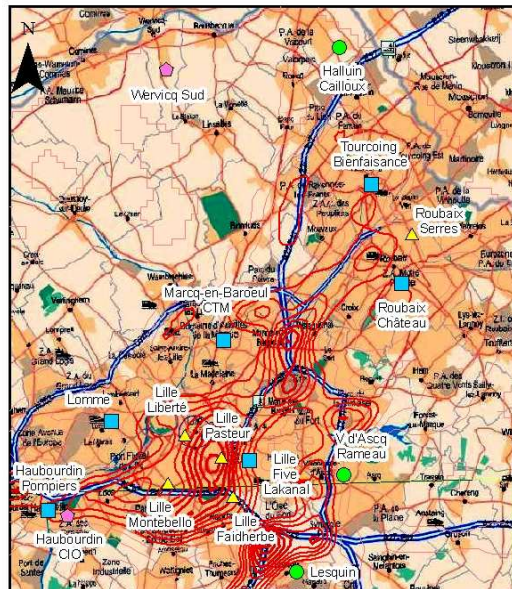


Figure 16: Cadastre d'Emissions (Isolignes) et localisation des 16 capteurs NO_2

La densité de population a été calculée à partir du nombre d'habitants par îlot (données du recensement 1999 de l'INSEE).

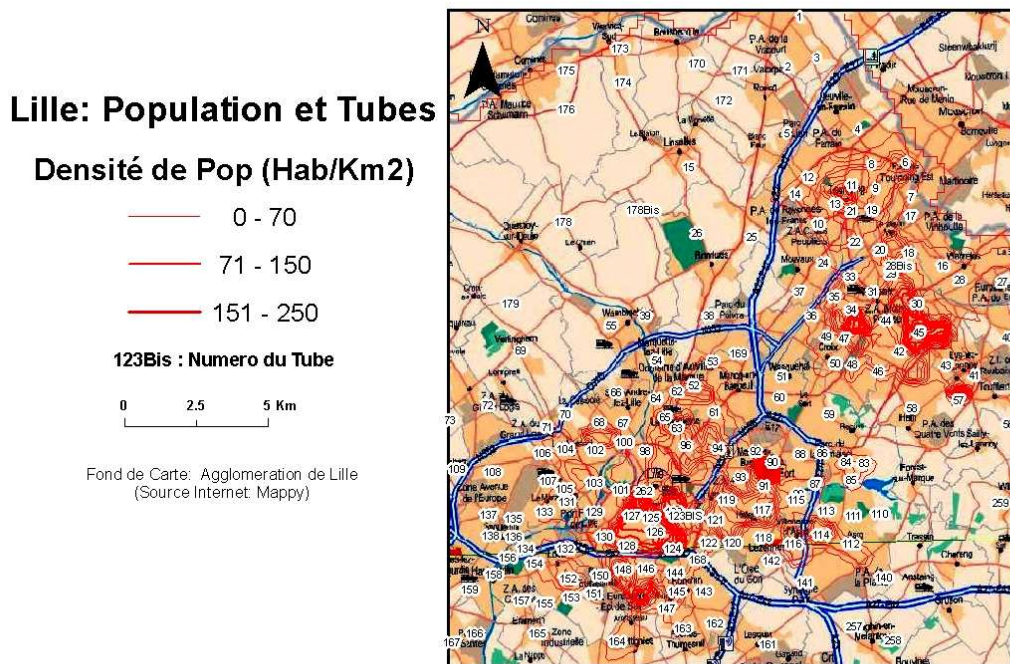


Figure 17 : Densité de Population (Isolignes) et localisation des 168 Tubes NO₂

La Figure 17 montre les courbes de niveaux correspondant à cette variable qui a été construite à l'aide d'un SIG et directement sur la maille d'estimation finale (250 m).

Les valeurs fortes d'émissions ne sont présentes que dans la zone urbaine de Lille, en revanche la variable densité de population présente également des valeurs fortes dans la zone urbaine de Roubaix et Tourcoing.

4.3.1.2 MESURES DE NO₂

Nous comptons 16 capteurs fixes de NO₂ sur la zone d'estimation et 15 campagnes de mesures par échantillonneurs passifs. Le Tableau 10 montre les périodes de prélèvements de ces campagnes (la Figure 16 montre la localisation géographique de 16 capteurs et dans la Figure 17 celle de 168 tubes NO₂).

Dans l'emplacement correspondant à 14 capteurs des tubes passifs ont également été installés. Ils sont souvent utilisés pour vérifier les mesures des stations fixes. Les seules stations où des tubes n'ont jamais été installés sont la station Urbaine Lomme et la station Industriel Haubourdin CIO.

Dans les campagnes, le NO₂ a été mesuré dans 168 sites différents. Cependant, le nombre maximal de tubes validés simultanément dans une seule campagne est de 134 (campagne 18 en hiver 2004) et le nombre minimal est de 114 (campagne 2 en été 1998).

On observe que la plupart des sites échantillonnés sont principalement localisés dans deux endroits : la zone urbaine de Lille et les communes de Roubaix et Tourcoing, la densité des tubes étant très faible au nord-ouest de la zone d'estimation.

Nb Q	Saison	Date Début	Date Fin	Nb Tubes	Min	Max	Moy	σ	σ^2 ($\mu\text{g}/\text{m}^3$) ²	CV
1	été 1998	27-mai-98	10-juin-98	119	15.1	40.7	26.48	5.3	28.06	0.2
2		10-juin-98	24-juin-98	114	9.3	32	18.27	5.12	26.18	0.28
3		24-juin-98	8-juil-98	115	9.3	52.6	18.9	6.28	39.42	0.33
4		8-juil-98	22-juil-98	124	8	40.6	17.03	5.02	25.23	0.3
5		22-juil-98	5-août-98	126	8	32.3	16.97	4.63	21.42	0.27
6		5-août-98	19-août-98	122	11	36	20.6	5.34	28.54	0.26
7	hiver 1999	3-déc-98	16-déc-98	128	11.2	68.6	38.98	8.09	65.45	0.21
10		20-janv-99	3-févr-99	129	17.6	68.3	37.57	7.82	61.16	0.21
11		3-mars-99	17-mars-99	122	26.6	61.7	43.23	6.26	39.24	0.14
12	été 2003	22-mai-03	5-juin-03	131	8	63.4	27.66	8.31	69.06	0.3
13		5-juin-03	19-juin-03	126	15.1	53.9	25.1	6.33	40.05	0.25
14		19-juin-03	3-juil-03	133	14.4	60.7	25.25	7.04	49.62	0.28
16	hiver 2004	8-janv-04	22-janv-04	131	17.46	52.33	31.18	5.87	34.41	0.19
17		22-janv-04	5-févr-04	131	19.04	47.73	33.15	5.9	34.86	0.18
18		5-févr-04	18-févr-04	134	22.89	52.04	37.87	6.39	40.85	0.17

Tableau 10: Campagnes par échantillonneurs passifs disponibles

4.3.2 CALCUL DE LA VEM DES STATIONS VIRTUELLES

Les stations virtuelles ont été sélectionnées en appliquant la méthode expliquée dans la première partie de cette étude. Elles représentent 65 mesures par échantillonneur passif qui sont très bien corrélées temporellement aux mesures de deux stations fixes de référence : la station périurbaine Halluin Caillou et la station urbaine Roubaix Châteaux (Figure 18).

Tubes sélectionnés (Nb=65)

- 88 TP liés à la SF P Halluin Cailloux (Nb= 31)
- 45: TP liés à la SF U Roubaix Château (Nb=34)
- SF de Référence

88 ou 45 : Numero du Tube
H Ca : Nom SF

0 2.5 5 Km

Fond de Carte: Agglomération de Lille
(Source Internet Mappy)

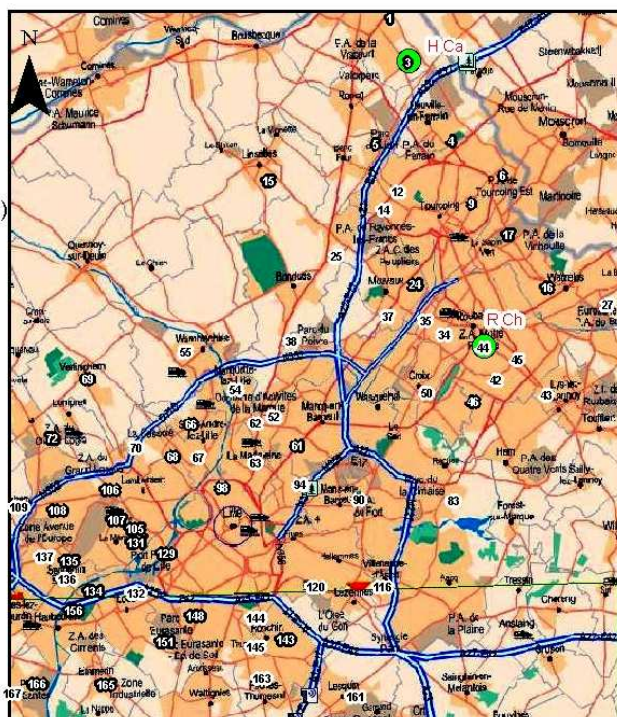
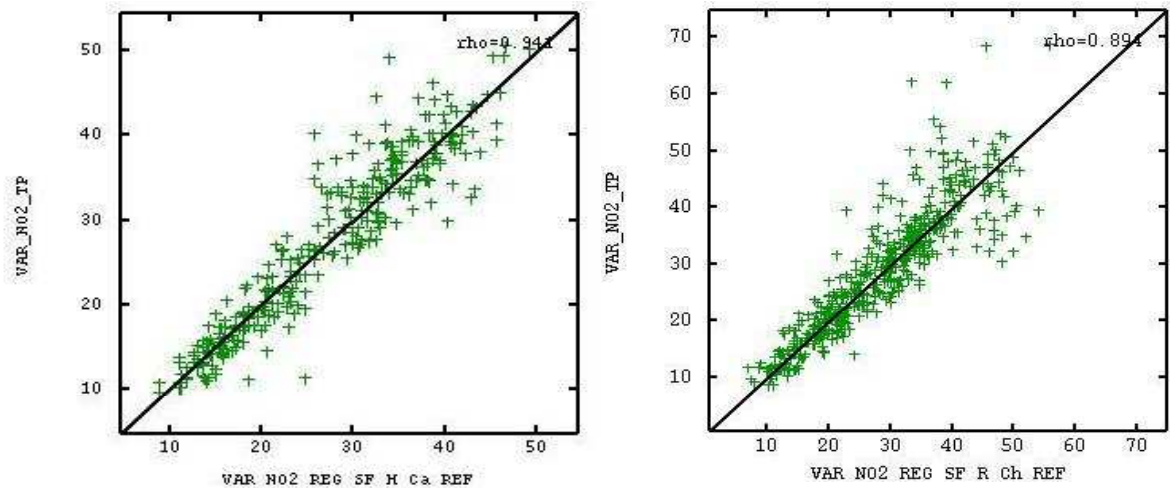


Figure 18: 65 Stations Virtuelles

Cette figure montre l'existence de points très bien corrélés temporellement aux stations, mais très éloignés de celles-ci. Ceci peut s'expliquer par une ressemblance au niveau de la typologie des sites.

Pour prendre en compte les valeurs estimées dans ces 65 stations virtuelles, nous devons calculer la variance de l'erreur d'estimation (cette variance est introduite dans Isatis en tant qu'erreur de mesure ou VEM).

Si l'on compare les valeurs estimées par rapport aux valeurs mesurées pendant les 15 campagnes on obtient les deux nuages de corrélation de la Figure 19. Dans ces nuages, les points représentant des couples de concentrations mesurées/estimées sont très proches de la bissectrice. Ces résultats sont un bon indicateur de la qualité de la méthode employée pour générer les données supplémentaires.



Nuage à gauche : Stations virtuelles liées à la SF Périurbaine Halluin Cailloux
 Nuage à droite : Stations virtuelles liées à la SF Urbaine Roubaix Château
 Axe Y : Concentrations mesurées dans les tubes lors des 15 campagnes
 Axe X : Concentrations estimées (à partir des corrélations temporelles)
 Ligne droite : Première bissectrice

Figure 19: Nuage de corrélation des concentrations estimées et concentrations mesurées lors des campagnes par tubes dans les 65 « stations virtuelles »

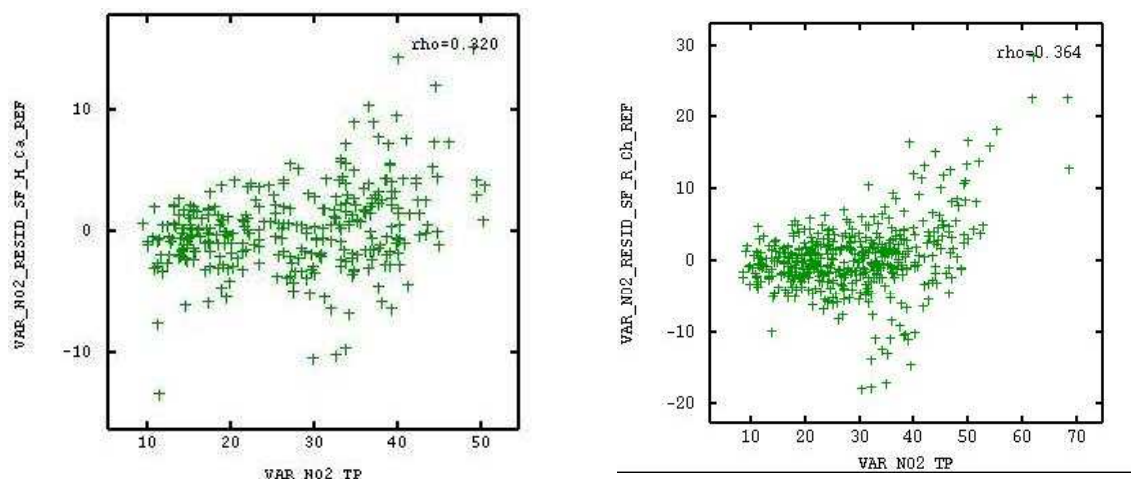
Afin de trouver une mesure de l'incertitude liée aux données estimées, nous avons calculé les résidus des estimations :

$$R(x) = NO_2 \text{ Estimé} - NO_2 \text{ Mesuré}$$

Chaque point des nuages de la Figure 20 représente la concentration mesurée et le résidu respectif pour un point de mesure (station virtuelle) et une quinzaine en particulier.

Dans ces nuages on s'aperçoit que les écarts entre concentrations estimées et concentrations mesurées peuvent aller jusqu'à 20 µg/m³. Les écarts sont en moyenne plus importants pour les concentrations supérieures à 30 µg/m³; la plupart de ces valeurs sont mesurées en hiver.

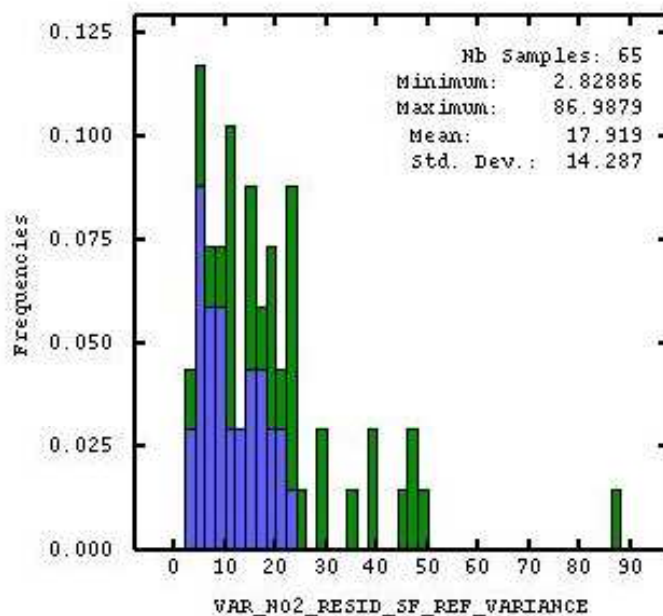
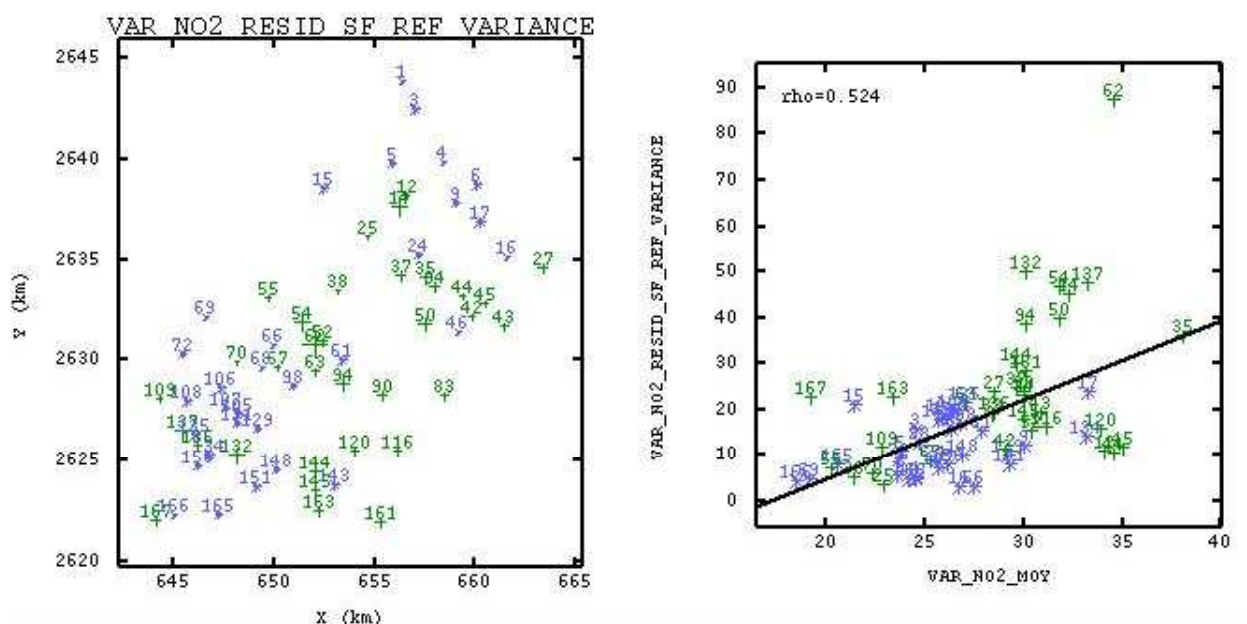
Finalement pour chaque station virtuelle on calcule la variance de résidus (ou VEM des stations virtuelles), celle-ci représente la « dispersion » des résidus autour de sa valeur moyenne sur chaque station virtuelle. La Figure 21 montre les principales statistiques de cette variable.



Nuage à gauche : Stations virtuelles liées à la SF Périurbaine Halluin Cailloux
 Nuage à droite : Stations virtuelles liées à la SF Urbaine Roubaix Château
 Axe Y : Résidus calculés dans les stations virtuelles pendant les 15 campagnes
 Axe X : Concentrations mesurées dans les stations virtuelles pendant les 15 campagnes

Figure 20: Nuage de corrélation des résidus et concentrations mesurées pendant les campagnes par tubes dans les 65 « stations virtuelles »

L'analyse détaillée de ces figures permet de conclure qu'il existe un groupe de 10 points de mesure (stations virtuelles) liés à un « environnement urbain » où il est plus difficile d'obtenir des estimations précises ; ce sont les stations 62, 132, 137, 54, 14, 356, 94, 144, 161 et 50 où la moyenne des 15 campagnes du dioxyde d'azote est supérieure à $30 \mu\text{g}/\text{m}^3$. Ces stations peuvent être représentatives d'un environnement plus local influencé par exemple par le trafic automoteur.



Carte en haut à gauche : Localisation géographique des stations virtuelles (la taille des croix est fonction de la variance des résidus).

Nuage en haut à droite: Nuage de corrélation entre la variance des résidus et la moyenne des 15 campagnes du dioxyde d'azote (La Ligne droite est la régression linéaire).

Les nombres à cote des croix correspondent aux numéros des tubes.

Figure en bas: Histogramme de la variance des résidus des stations virtuelles.

Les symboles en couleur bleue représentent les stations virtuelles liées à la station fixe périurbaine Halluin Caillou.

Les symboles en couleur verte représentent les stations virtuelles liées à la station fixe urbaine Roubaix Château.

Figure 21: Statistiques de la Variance des Résidus ou VEM

4.3.3 ESTIMATION DES DEUX CAMPAGNES REPRESENTATIVES

Les résultats de la cartographie du NO₂ sont présentés, pour deux des quinze campagnes par tubes disponibles : la première campagne (Q1) est menée au printemps 1999 (entre le 27 mai et le 10 juin), et la deuxième campagne (Q18) est menée en hiver 2004 (entre le 5 et 18 février).

L'objectif est d'essayer d'en extraire le plus d'informations en observant les caractéristiques du phénomène de pollution dans deux saisons opposées.

Pour ne pas accorder arbitrairement un poids plus fort à l'une ou l'autre des variables explicatives, et étant donné la différence d'ordre de grandeur entre elles (les émissions varient entre 0 et 127.6 g/Km² et la densité de population entre 0 et 253.2 Hab/Km²), les deux variables ont été transformées ainsi :

$$Z_{transf} = \frac{Z_{orig} - Z_{min}}{Z_{max} - Z_{min}}$$

Les nouvelles variables obtenues, qui sont comprises dans un rang entre 0 et 1, ont ensuite été migrées aux tubes (à chaque tube on assigne la valeur de la maille la plus proche).

Plusieurs essais ont été réalisés de manière à trouver une fonction de ces variables transformées la mieux corrélée aux concentrations des campagnes : il s'est avéré que la somme des deux variables (nouvelle fonction notée FK1) donne les meilleurs coefficients de corrélation avec la concentration des périodes estivales et hivernales.

Les deux nouvelles variables explicatives transformées et la fonction auxiliaire sélectionné FK1 sont comparées dans la Figure 22.

La fonction FK1 est comprise dans un intervalle allant entre 0 et 1.5.

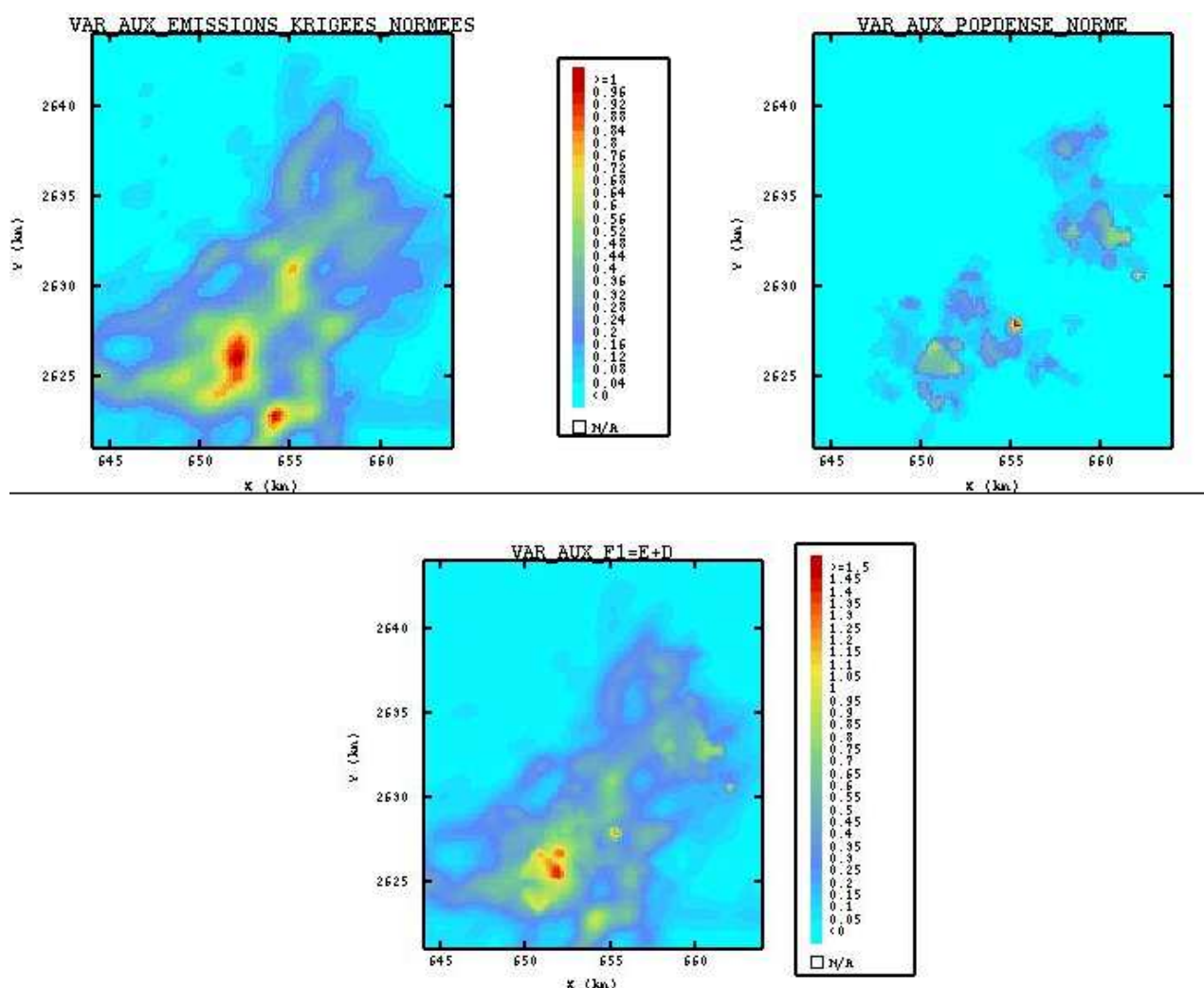


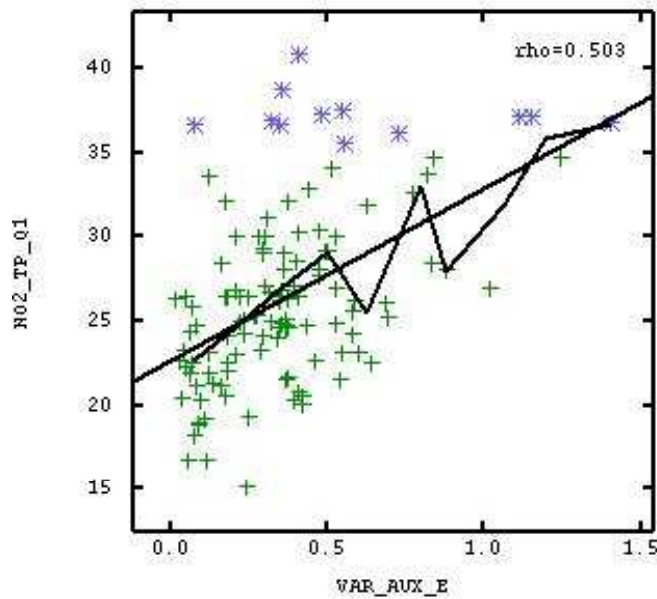
Figure en haut à gauche : Emissions estimées par krigeage et transformées entre zéro et un.
 Figure en haut à droite: Densité de Population transformée entre zéro et un.
 Figure en bas: La somme des deux variables précédentes (fonction F1K)

Figure 22 : Cartes des variables explicatives

4.3.3.1 ESTIMATION DE LA CAMPAGNE NO. 1 (ETE)

4.3.3.1.1 ESTIMATION SUR L'ENSEMBLE DES TUBES

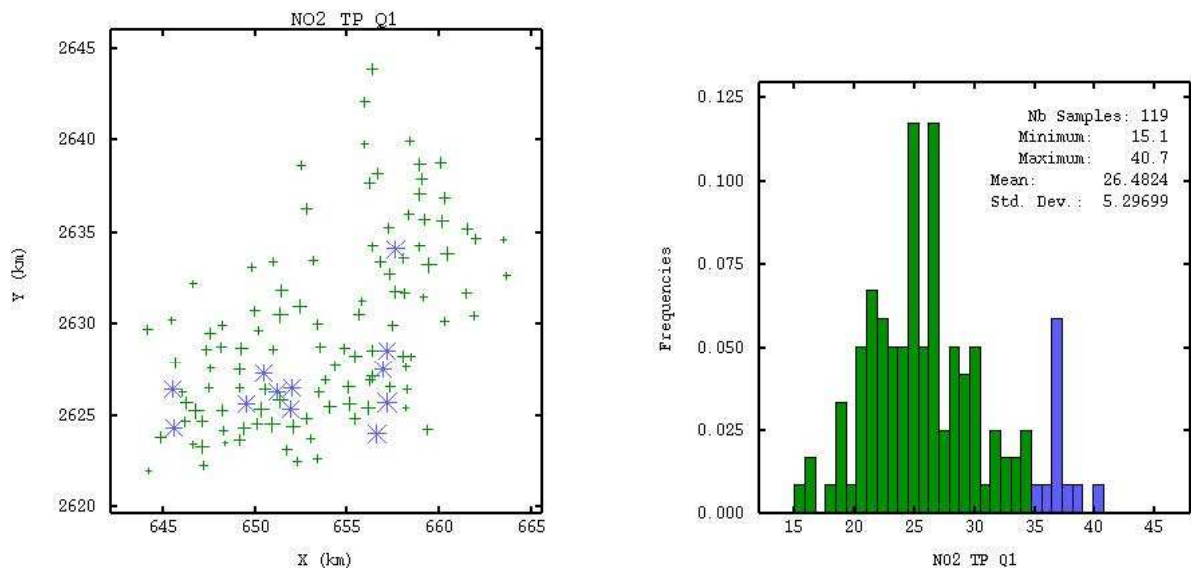
La première campagne est composée de 119 échantillons (tubes passifs) qui ont pour moyenne $26.5 \mu\text{g}/\text{m}^3$, la corrélation de la concentration avec la fonction F1K est examinée. Le nuage correspondant est présenté sur la Figure 23, l'histogramme et la localisation des points sur la Figure 24.



Symboles en bleu : Concentrations supérieures à $35 \mu\text{g}/\text{m}^3$
 Deux lignes : Régression linéaire et espérance conditionnelle

Figure 23 : Campagne No. 1: Nuages de corrélation avec la fonction auxiliaire F1K

Les émissions et la fonction auxiliaire F1K présentent une corrélation de 0.5. Comme cette valeur n'est pas très élevée, normalement le cokrigeage doit être plus adapté que le krigeage en dérive externe ou le krigeage des résidus, ces trois méthodes seront comparées par la suite.



Symboles en bleu : concentrations supérieures à $35 \mu\text{g}/\text{m}^3$

Figure 24 : Campagne No. 1: Localisation tubes et Histogramme

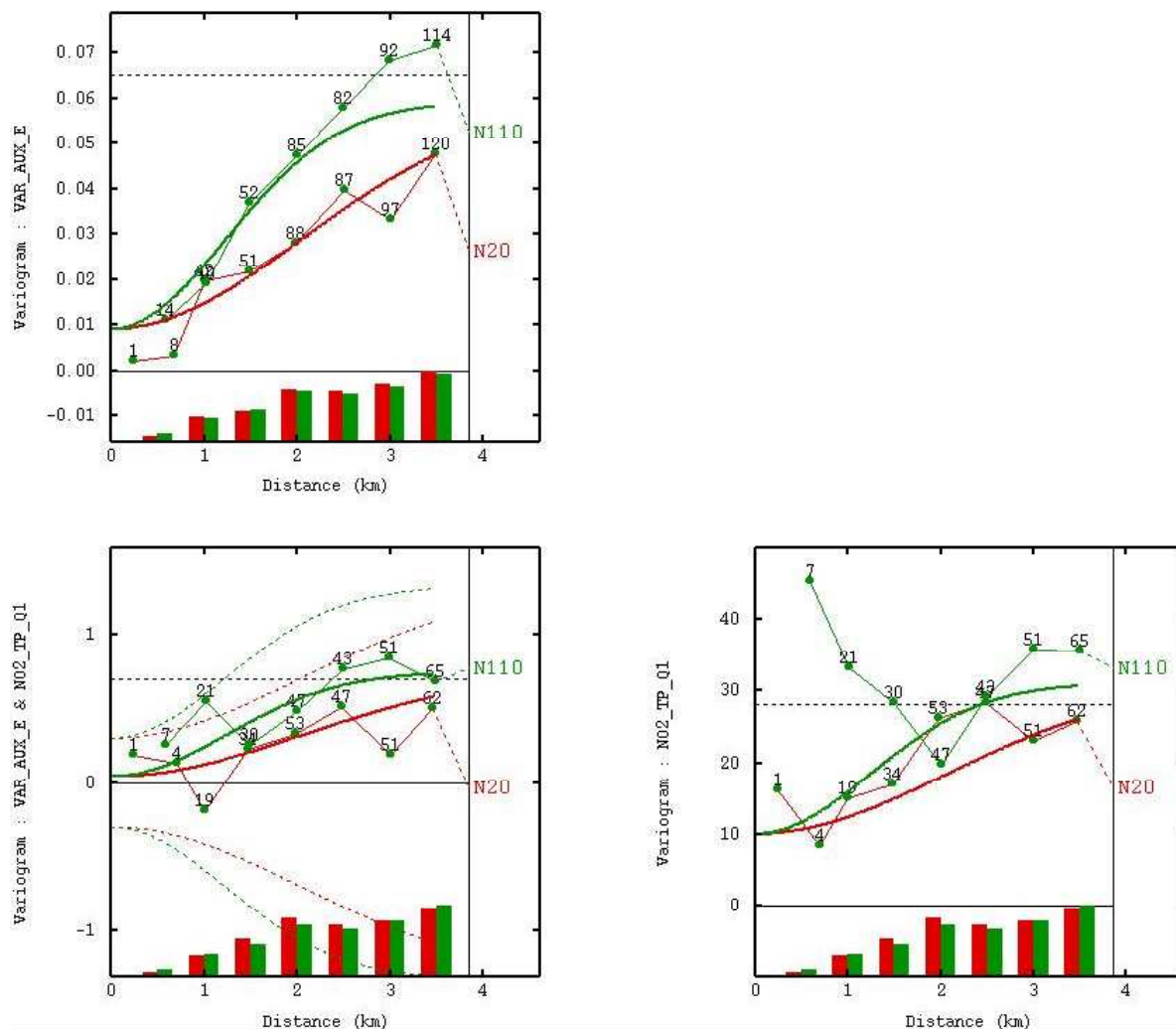
A l'aide du menu de validation croisée du logiciel Isatis les quatre méthodes d'estimation suivantes ont été testées: krigeage ordinaire, krigeage de résidus, krigeage en dérive externe et cokrigeage colocalisé, pour toutes les estimations, la VEM a été calculée comme dans la section précédente.

Dans notre cas, c'est la méthode du cokrigeage colocalisé qui a donné les meilleures statistiques. Le cokrigeage permet d'exploiter la corrélation spatiale bivariable, mais l'on doit modéliser les structures simples et croisées des variables (dans notre cas nous serons obligés de modéliser trois variogrammes tous les jours : un pour le dioxyde d'azote, un pour la fonction F1K et le variogramme croisé entre ces deux variables).

La Figure 25 montre les trois variogrammes expérimentaux et les modèles respectifs ajustés manuellement. Notons que la variance de la concentration est de $28 (\mu\text{g}/\text{m}^3)^2$ et que la somme des paliers du modèle ajusté est de $31 (\mu\text{g}/\text{m}^3)^2$, l'effet de pépité correspond à un tiers de la variabilité ($10 (\mu\text{g}/\text{m}^3)^2$).

La fonction F1K liée aux variables décrivant le milieu présente une portée de 5Km à une direction de N 20°E et une portée plus courte de 3Km dans la direction perpendiculaire à celle-ci (N70°O).

Cette anisotropie peut être expliquée par l'influence des principaux axes routiers de l'agglomération (voir Figure 16). Par ailleurs, on observe que la concentration en dioxyde d'azote est mieux structurée dans la direction parallèle aux axes routiers (courbe rouge des variogrammes). En revanche, pour la direction perpendiculaire (N 70 O), le phénomène est beaucoup plus erratique et très variable à petite distance.

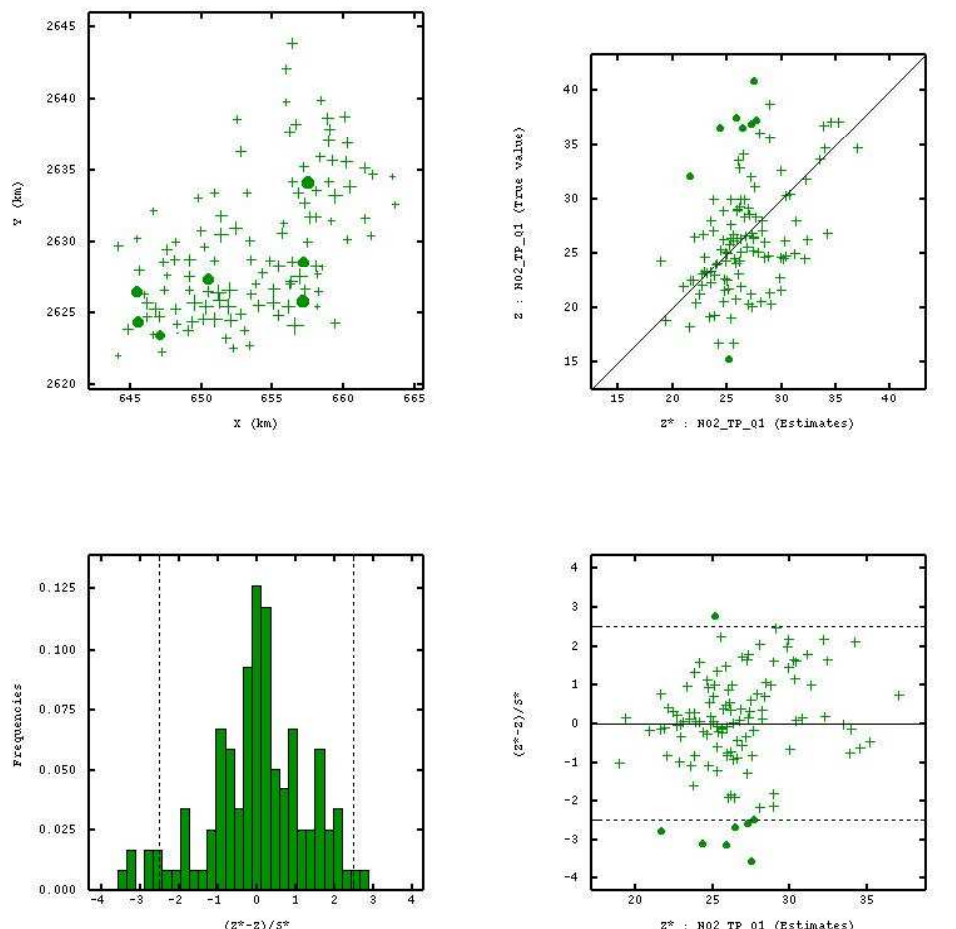


Pas de calcul variogramme expérimental : 500m, Nombre de pas= 8

Variable	Direction	Structure	Palier	Portée (Km)
NO2_Q1	N20°E	Pénite	10	----
		Gaussienne	21	5
	N70°O	Pénite	10	----
		Gaussienne	21	3
F1K	N20°E	Pénite	0.009	----
		Gaussienne	0.05	5
	N70°O	Pénite	0.009	----
		Gaussienne	0.05	3
NO2_Q1 et F1K	N20°E	Pénite	0.04	----
		Gaussienne	0.7	5
	N70°O	Pénite	0.04	----
		Gaussienne	0.7	3

Figure 25 : Campagne 1 : Modèle bivariable avec la fonction F1K

Dans les résultats de validation croisée (Figure 26) on observe un lissage des estimations autour des valeurs moyennes, ceci est dû au fort effet de pépité (un tiers de la variabilité du phénomène). Par conséquent, on ne parvient pas à estimer correctement quelques valeurs extrêmes ; dans les nuages de cette figure, ces points sont marqués par des boules (voir autour de 15 $\mu\text{g}/\text{m}^3$ et de 35 $\mu\text{g}/\text{m}^3$).



Variable (Nb=119)	Moyenne	Variance
Erreur	0.1	23.4
Erreur Réduite	0.015	1.6

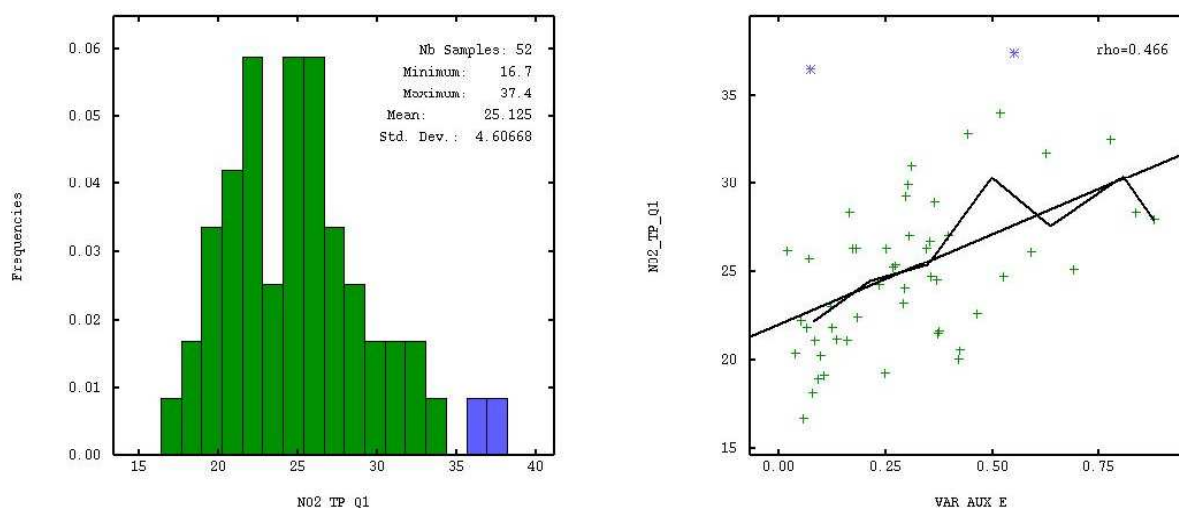
Figure 26 : Campagne 1 : Test de validation croisée du modèle bivariable

4.3.3.1.2 ESTIMATION SUR LA SELECTION DES TUBES CORRELES AUX STATIONS DE REFERENCE

La sélection des tubes corrélés aux stations fixes de référence et qui deviendront stations virtuelles, cause une forte diminution des effectifs ; par exemple pour la première campagne (Q1) on passe de 119 à 52 tubes (une diminution de 52% des effectifs). En effet, des 65 tubes qui sont corrélés aux deux stations de référence, 13 tubes ne présentent pas de mesure sur la première campagne, d'où une sélection finale de 52 tubes.

Afin d'observer l'impact occasionné par la réduction du nombre des points, nous avons examiné les principales statistiques sur la sélection des 52 tubes.

Pour cette sélection de données, les statistiques ne sont guère différentes. Le coefficient de corrélation avec la fonction F1K baisse légèrement de 0.5 à 0.47, la moyenne de 26.5 à 25.1 $\mu\text{g}/\text{m}^3$ et la variance de 28 à 21 $(\mu\text{g}/\text{m}^3)^2$. On observe que la fonction explicative F1K a perdu les valeurs supérieures à 0.8.

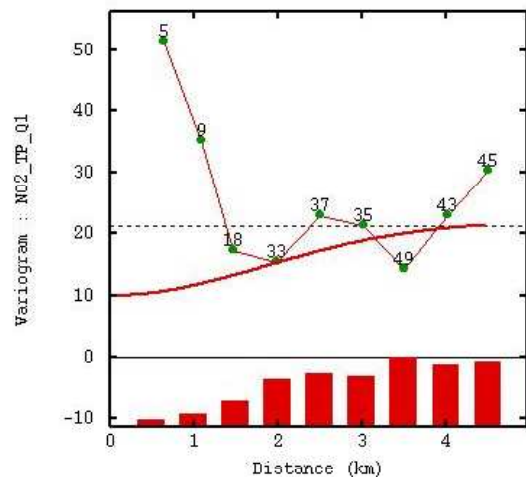
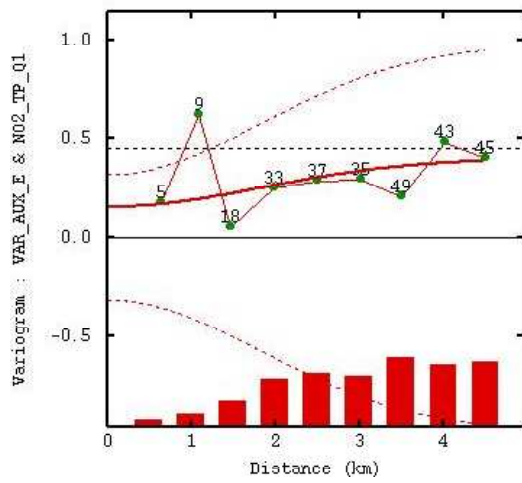
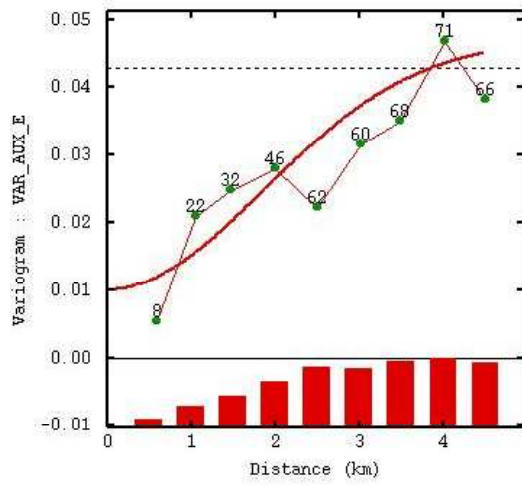


Symboles en bleu : Concentrations supérieures à 35 $\mu\text{g}/\text{m}^3$

Figure 27 : Campagne 1: Histogramme et Nuage de corrélation avec la variable explicative F1K (52 tubes)

Le nombre limité de données ne permet pas de calculer des variogrammes directionnels crédibles, c'est pourquoi nous nous sommes contentés de les calculer sous l'hypothèse d'isotropie. En revanche nous avons conservé la même largeur du pas (500 m) et le même nombre de pas (8).

Comme la variance de la concentration des 52 données est de 21 $(\mu\text{g}/\text{m}^3)^2$, la somme des paliers du nouveau modèle de variogramme du dioxyde d'azote est de 22 $(\mu\text{g}/\text{m}^3)^2$; l'effet de pépite représente maintenant 45% de la variabilité (Figure 28). La structure gaussienne est conservée, avec une portée de 4.5Km, distance proche à la plus grande portée du modèle anisotrope (5Km).



Variable (Nb=52)	Structure	Palier	Portée (Km)
NO2_Q1	Pépité	10	----
	Gaussienne	12	4.5
F1K	Pépité	0.01	----
	Gaussienne	0.037	4.5
NO2_Q1 et F1K	Pépité	0.15	----
	Gaussienne	0.25	4.5

Figure 28 : Campagne 1 : Modèle bivarié avec la fonction F1K (52 Tubes)

De manière à pouvoir évaluer le modèle sur l'ensemble du domaine, nous l'avons testé par validation croisée sur deux jeux de données : tout d'abord sur la sélection de 52 tubes corrélés aux stations fixes de référence, ensuite sur la sélection des 67 tubes restants.

Les estimations du modèle anisotrope ajusté sur l'ensemble des tubes et du modèle isotrope ajusté sur la sélection de 52 tubes sont de qualité très proche : la variance des erreurs n'a augmenté que de 8% (elle est passée de 23.4 à 25.2 $(\mu\text{g}/\text{m}^3)^2$) et les nuages (concentrations mesurées/concentrations estimées) présentent une allure semblable.

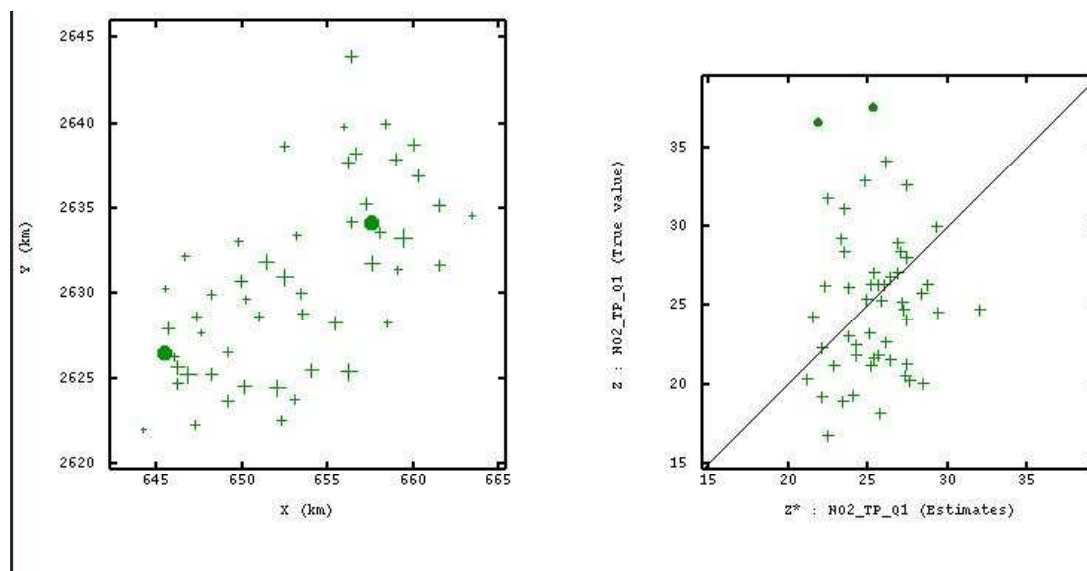


Figure à gauche: Localisation des 52 points de mesure (stations virtuelles)
 Figure à droite: Nuage de corrélation. Abscisse: valeurs estimées
 Ordonnée : valeurs mesurées
 Ligne droite : Première bissectrice

Variable (Nb=52)	Moyenne	Variance
Erreur	0.31	25.2
Erreur Réduite	0.05	1.4

Statistiques de la validation croisée

Figure 29 : Campagne 1 : Validation croisée du modèle bivariable sur 52 tubes

Le second test de validation croisée a été mené sur les données qui n'ont pas participé au calcul des variogrammes, dans ce test on estime la concentration de 67 tubes à partir des 52 tubes sélectionnés.

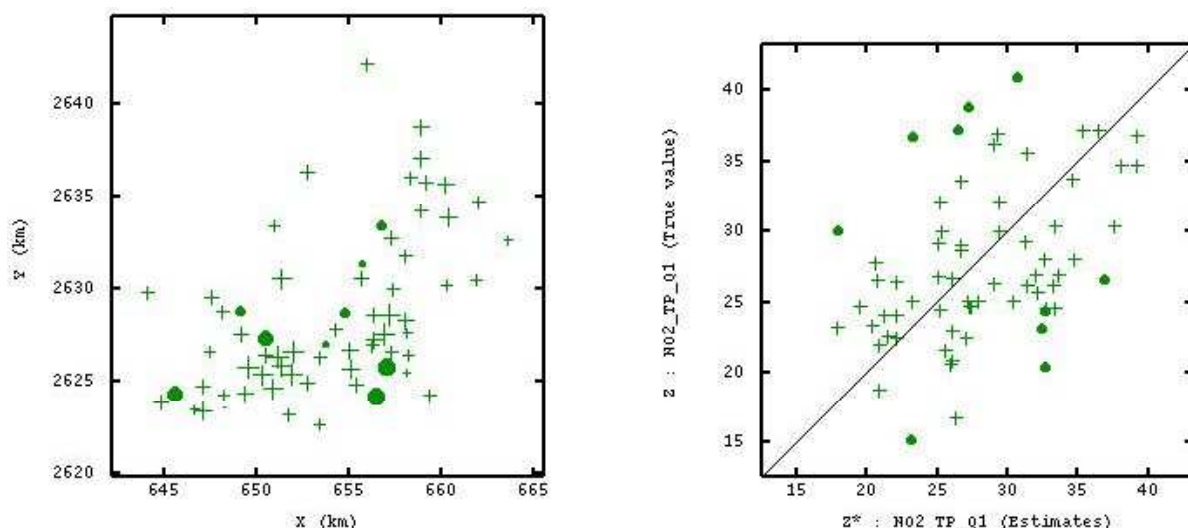


Figure à gauche : Localisation des 67 points de mesure (Hors stations virtuelles)
 Figure à droite: Nuage de corrélation. Abscisse: valeurs estimées
 Ordonnée : valeurs mesurées
 Ligne droite : Première bissectrice

Variable (Nb=67)	Moyenne	Variance
Erreur	0.57	35.2
Erreur Réduite	0.16	2.9

Statistiques de la validation croisée

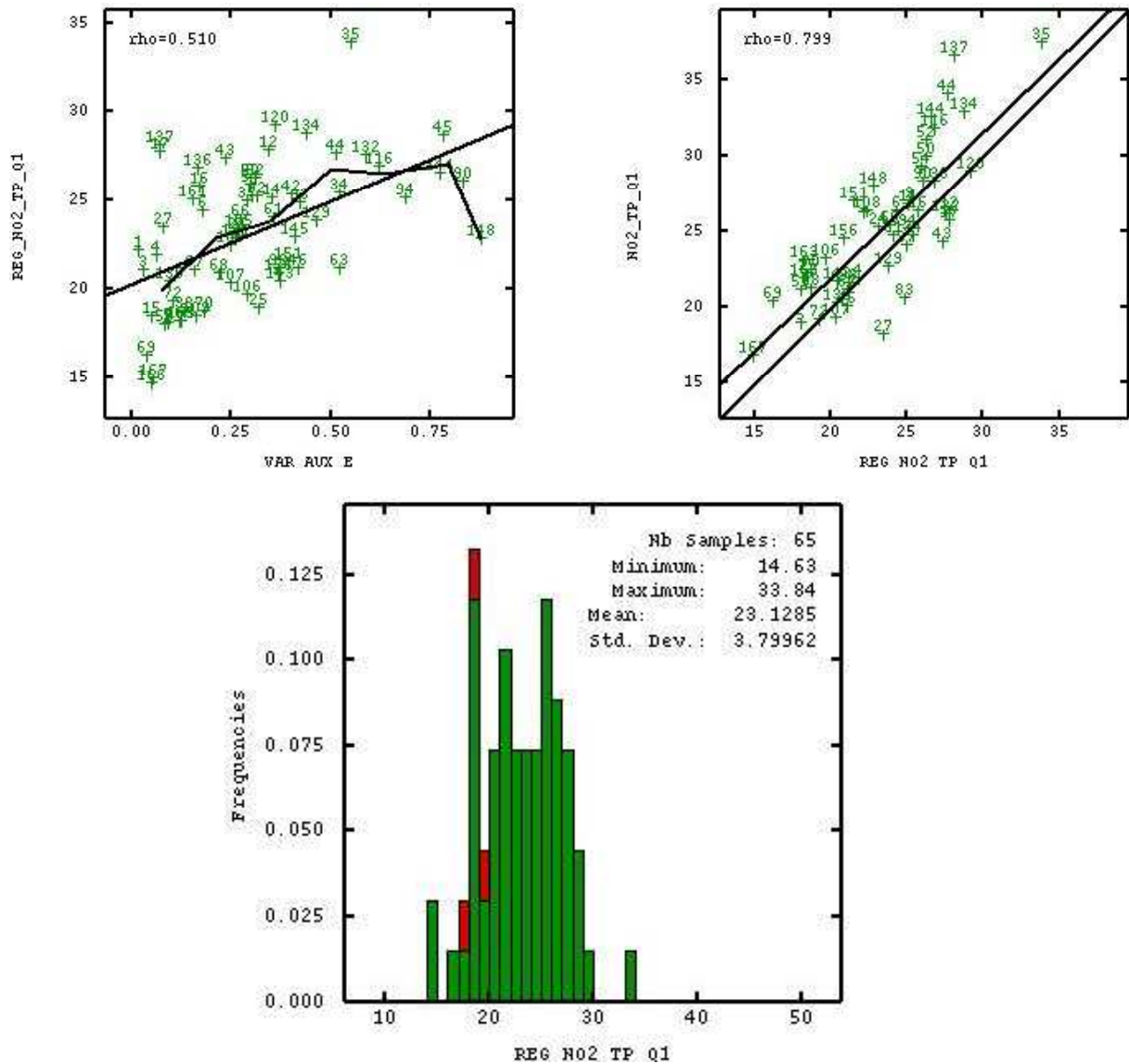
Figure 30 : Campagne 1 : Validation croisée du modèle bivariable sur 67 tubes

La variance des erreurs a beaucoup augmenté (une augmentation de 40% en passant de 25.2 à 35.2 $(\mu\text{g}/\text{m}^3)^2$). Il y a 10 tubes (15% des effectifs) dont les estimations sont nettement moins précises, et quand on les retire, la variance des erreurs diminue pour atteindre 21 $(\mu\text{g}/\text{m}^3)^2$.

Ces tests menés sur les vraies mesures, donne un aperçu de l'ampleur de la diminution de la qualité des estimations lorsque l'on réduit le nombre de données.

4.3.3.1.3 ESTIMATIONS FAITES A PARTIR DES DONNEES REESTIMEES (STATIONS VIRTUELLES)

Nous allons maintenant recalculer les concentrations aux tubes sur les 65 stations virtuelles, pour cela les paramètres des corrélations temporelles (présentées dans la première partie de cette étude) ont été utilisés. Pour la première campagne, la Figure 31 montre les principales caractéristiques des concentrations réestimées.



Nuage en haut à gauche : Corrélation avec la variable explicative F1K (Axe X) :

Les lignes correspondent à la régression linéaire et à l'espérance conditionnelle

Nuage en haut à droite: Corrélation avec les vraies concentrations de 52 tubes (Axe Y)

Les lignes correspondent à la première bissectrice et à la régression linéaire

Les nombres sur les nuages correspondent aux numéros des tubes

Figure en bas: Histogramme des concentrations réestimées

Figure 31 : Campagne 1, Stations Virtuelles: Principales caractéristiques des concentrations

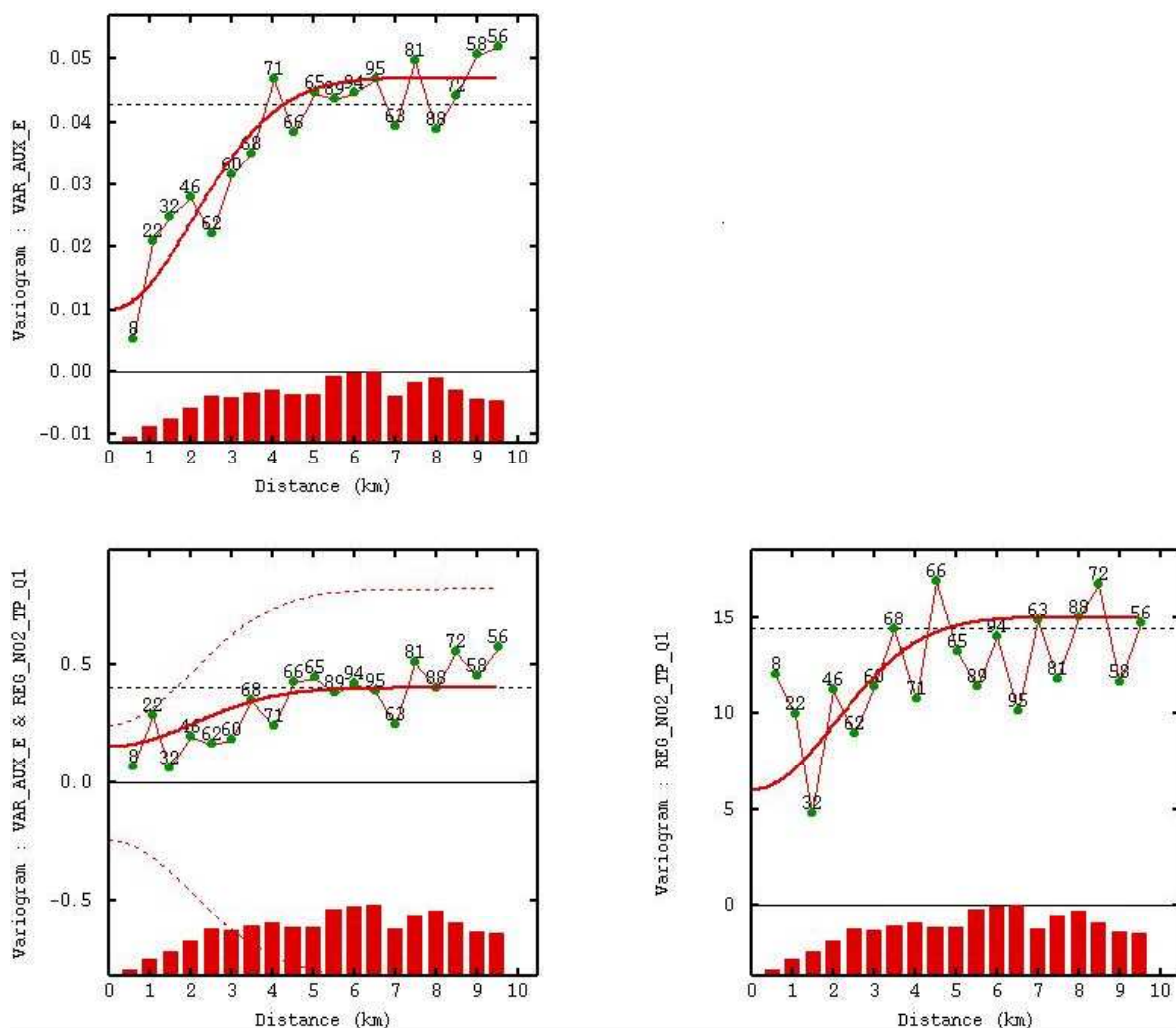
Pour la plupart des tubes de la première campagne, les concentrations réestimées sont plus faibles que les concentrations mesurées. Néanmoins le niveau de corrélation avec la variable explicative se conserve (coefficient de corrélation égale à 0.5) et les histogrammes sont semblables.

La Figure 32 affiche pour les 65 stations virtuelles, les variogrammes simples et croisés ajustés aux concentrations réestimées de la première campagne et de la variable auxiliaire F1K.

Le variogramme expérimental a été calculé sur une distance correspondant à 20 pas de 500 m, le modèle ajusté n'est guère différent de celui ajusté sur les vraies concentrations. **Ce résultat indique que les principales caractéristiques spatiales du phénomène sont bien conservées.**

Comme la variance des données réestimées a diminué, passant de 21 à 14 ($\mu\text{g}/\text{m}^3$)², les paliers des modèles sont rajustés en conséquence, maintenant la somme des paliers est de 15 ($\mu\text{g}/\text{m}^3$)² et l'effet de pépité représente le 40% de la variabilité (diminution de 5%, par rapport aux vraies concentrations).

La portée de la structure gaussienne est de 5 Km, distance égale à la plus grande portée du modèle ajusté sur les 119 tubes.



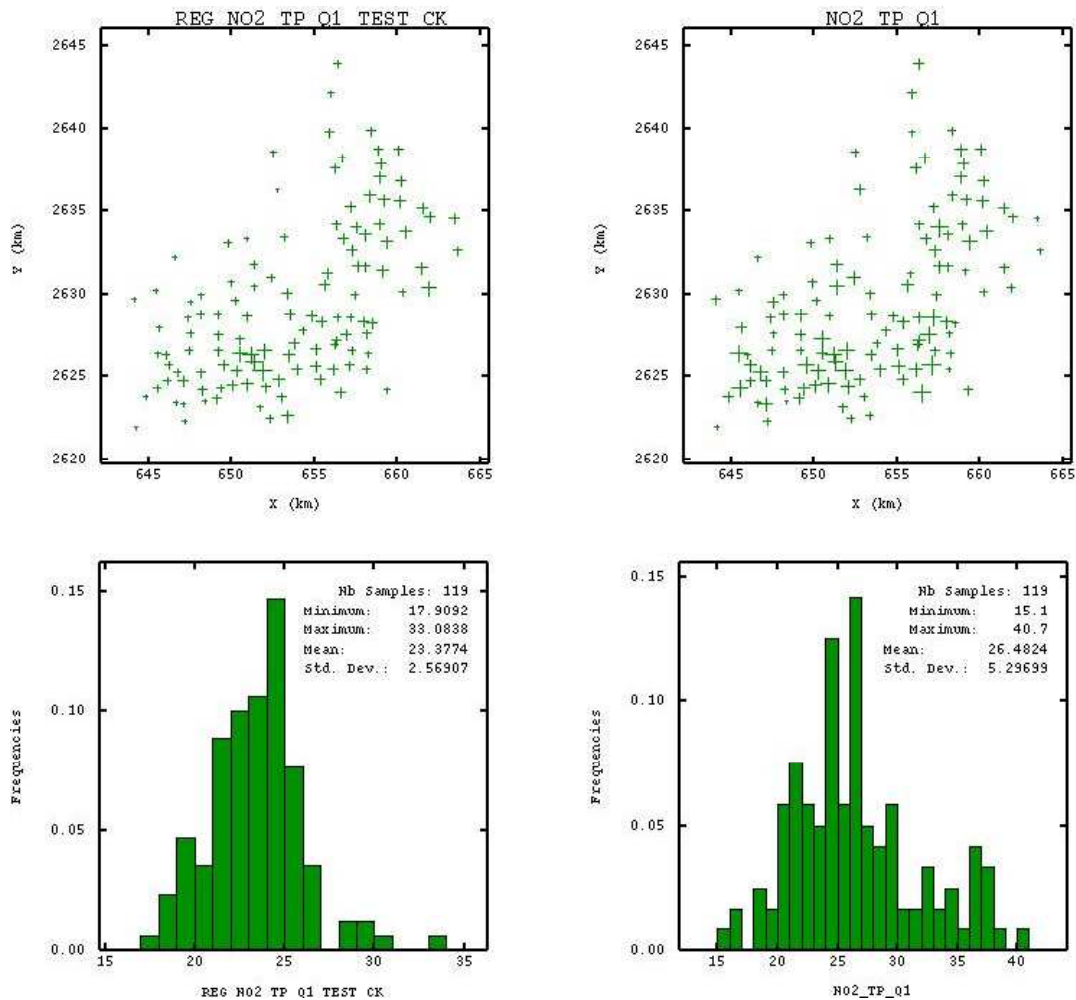
Variable (Nb=65)	Structure	Palier	Portée (Km)
NO2_Q1	Pépite	6	-----
	Gaussienne	9	5
F1K	Pépite	0.01	-----
	Gaussienne	0.037	5
NO2_Q1 et F1K	Pépite	0.15	-----
	Gaussienne	0.25	5

Figure 32 : Campagne 1, Stations Virtuelles : Modèle bivariable

Afin de valider ce modèle nous avons mené une estimation ponctuelle sur l'ensemble des échantillons (119 tubes), la méthode retenue est le cokrigage colocalisé avec la fonction F1K.

Par ailleurs, la VEM calculée dans la section 4.3.2 a été utilisée lors des estimations (celle-ci est rajoutée dans la diagonal de la matrice gauche de variance - covariance du système de krigeage, option disponible dans le logiciel Isatis).

Les données utilisées pour l'estimation sont les 65 concentrations estimées à partir des stations de référence (stations virtuelles). Ces résultats ont été ensuite comparés avec les concentrations mesurées sur les 119 tubes (Figure 33).

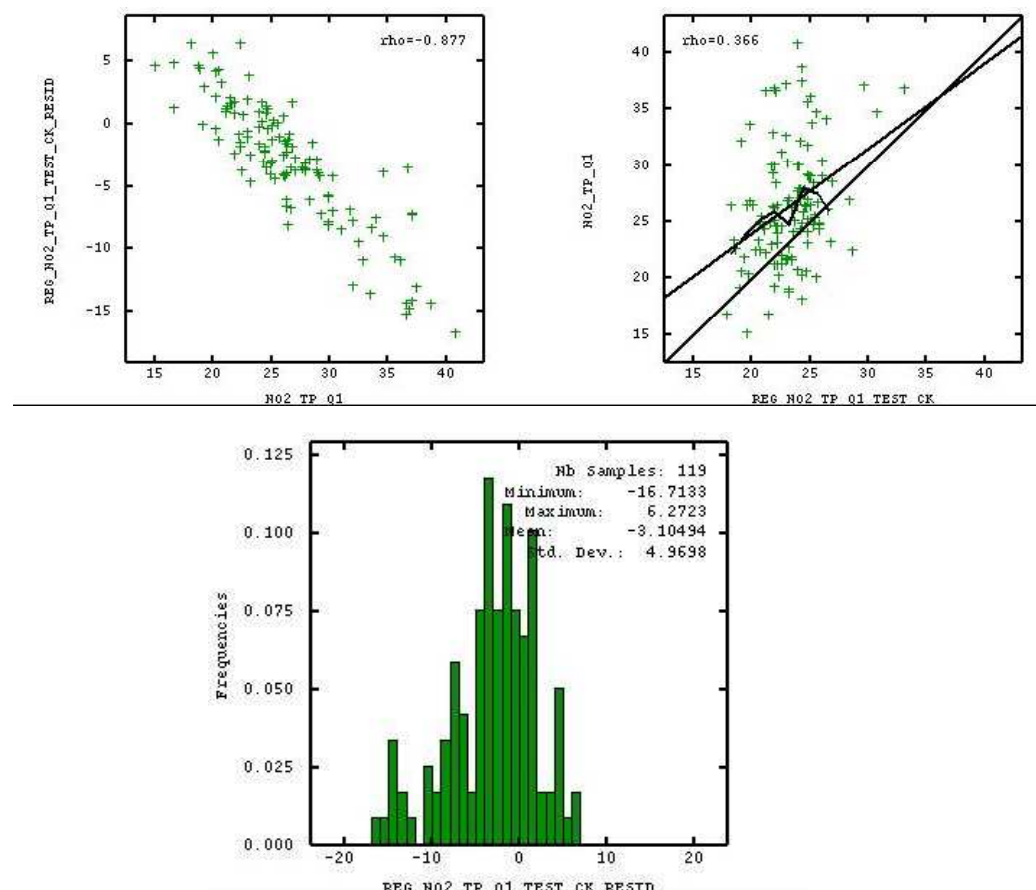


Deux figures à gauche : Concentrations estimées
 Deux figures à droite : Concentrations mesurées
 Deux cartes en haut : Localisation géographique des tubes
 Deux figures en bas : Histogrammes des concentrations de NO₂
 La taille des croix est fonction des concentrations respectives

Figure 33 : Campagne 1 : Carte et histogramme des estimations obtenues sur l'ensemble du domaine et comparaison avec les concentrations mesurées

Le manque de structure de la pollution estivale produit des valeurs estimées qui oscillent autour de la moyenne des 65 échantillons ($23.1 \mu\text{g}/\text{m}^3$), on obtient donc un lissage des concentrations (les faibles concentrations sont surestimées, les fortes sont sous-estimées : voir la Figure 34).

Par exemple, quelques concentrations supérieures à $30 \mu\text{g}/\text{m}^3$ ont été très largement sous-estimées (entre 10 et $16.7 \mu\text{g}/\text{m}^3$ d'écart), ces tubes sont localisés dans la zone de Roubaix et Tourcoing et dans la périphérie de Lille.



Nuage en haut à gauche: Corrélation entre erreurs et mesures

L'axe X correspond aux erreurs d'estimations ($Z'-Z$)

L'axe Y correspond aux concentrations mesurées sur les 119 tubes

Nuage en haut à droite : Corrélation entre valeurs estimées et mesures

L'axe Y correspond aux vraies concentrations de 119 tubes

L'axe X correspond aux estimations de la concentration par cokrigage

Les lignes correspondent à la première bissectrice, à la régression linéaire et à l'espérance conditionnelle

Figure en bas: Histogramme et statistiques des erreurs

Figure 34 : Campagne 1 : Histogramme des erreurs et nuages de corrélation avec les concentrations mesurées

Notons tout de même que la variance des erreurs est de $24.7 (\mu\text{g}/\text{m}^3)^2$, cette valeur est très proche de la variance des erreurs obtenue pour les estimations menées à partir des vraies concentrations ($23.4(\mu\text{g}/\text{m}^3)^2$, voir la Figure 26). Ceci atteste de la qualité de la démarche.

4.3.3.2 ESTIMATION DE LA CAMPAGNE NO. 18 (HIVER)

4.3.3.2.1 ESTIMATION SUR L'ENSEMBLE DES TUBES

Cette campagne a été menée en février 2004, il s'agit de la dernière campagne effectuée par échantillonneurs passifs et disponible dans notre jeu de données.

Cette campagne hivernale compte le plus grand nombre de mesures effectuées (134), elle présente une moyenne de $37.8 \mu\text{g}/\text{m}^3$ et une variance de $41 (\mu\text{g}/\text{m}^3)^2$.

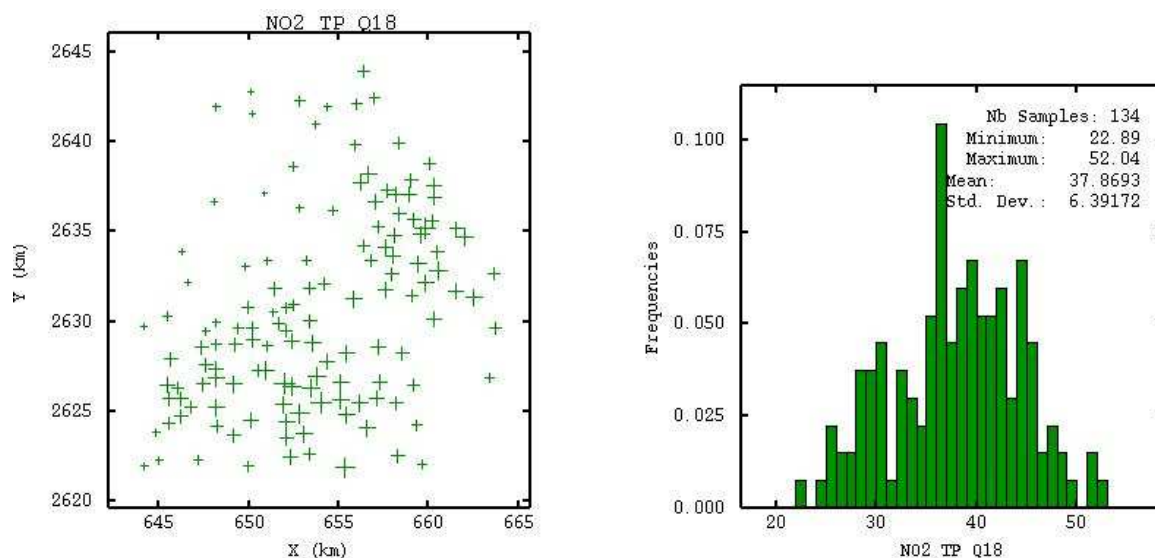
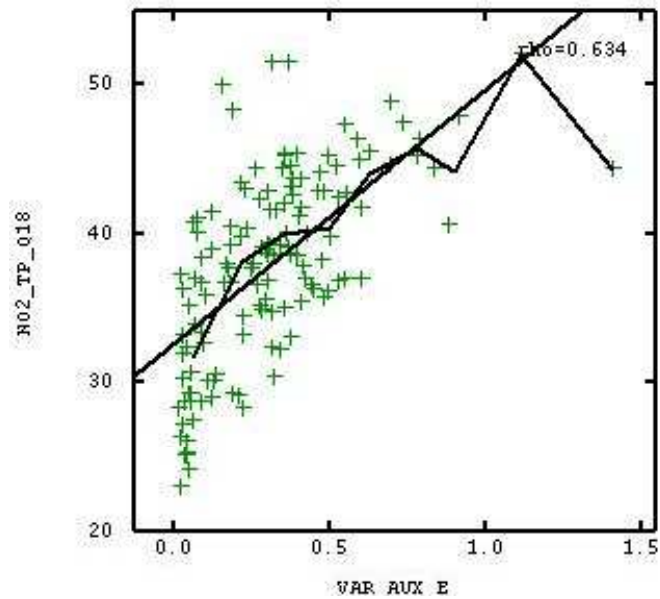


Figure 35 : Campagne No. 18: Localisation tubes et Histogramme

Pour cette campagne, le coefficient de corrélation avec la fonction auxiliaire F1K est légèrement supérieur par rapport à la campagne estivale (corrélation de 0.63).



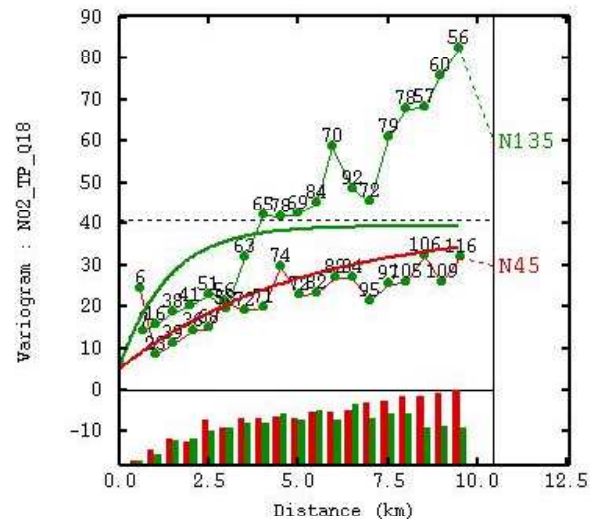
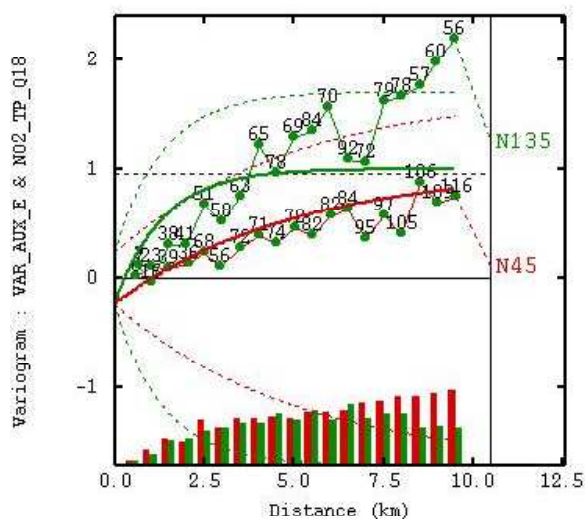
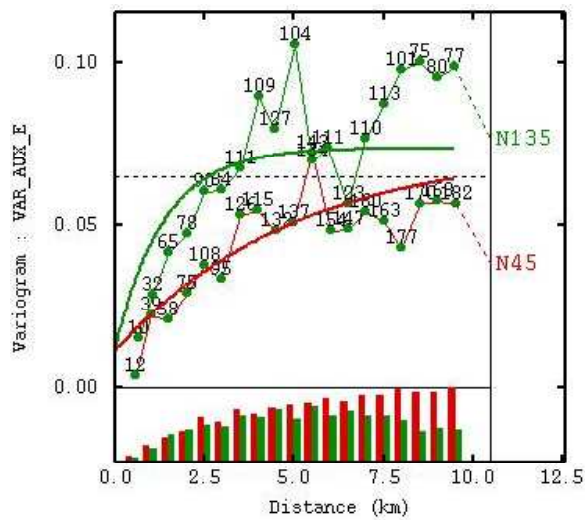
Nuage à gauche : Corrélation avec la fonction auxiliaire F1K
 Deux lignes : Régression linéaire et espérance conditionnelle

Figure 36 : Campagne No. 18: Nuages de corrélation avec les variables auxiliaires

Après plusieurs essais, on a choisi des variogrammes expérimentaux calculés sur 20 pas de 500m (10Km), et sur deux directions : N45°E et N45°O.

Pour la modélisation de ces variogrammes, une structure exponentielle avec un effet de pépite a été préférée. L'ajustement des portées de la structure exponentielle est de meilleure qualité pour la variable auxiliaire F1K. Cependant on observe, dans la Figure 37, que le résultat obtenu pour la concentration de dioxyde d'azote est satisfaisant dans la direction N45°E et acceptable jusqu'à 5 Km dans la direction de plus grand variabilité (N45°O).

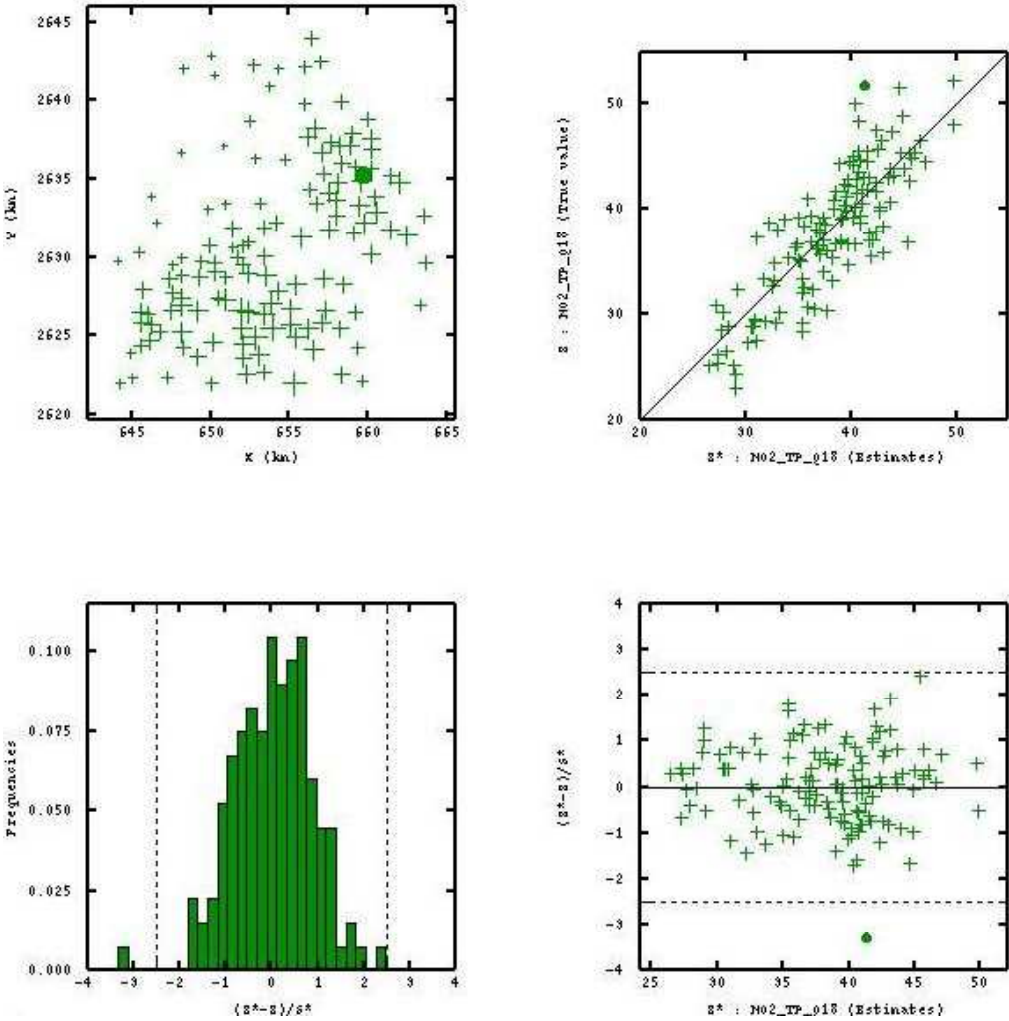
Les concentrations hivernales présentent un phénomène plus structuré, dans cette saison, l'influence du réseau routier est plus marquée : la corrélation spatiale dans la direction N45°E peut aller jusqu'à une distance de 15 Km. En été cette corrélation est nulle au-delà d'environ 5 Km; la corrélation spatiale dans la direction perpendiculaire est plus faible, elle n'atteint que 4 Km.



Variable (Nb=134)	Direction	Structure	Palier	Portée (Km)
NO2_Q18	N45°E	Pépité	5.03	-----
		Exponentiel	34.4	15
	N45°O	Pépité	5.03	-----
		Exponentiel	34.4	4
F1K	N45°E	Pépité	0.01	-----
		Exponentiel	0.06	15
	N45°O	Pépité	0.01	-----
		Exponentiel	0.06	4
NO2_Q18 et F1K	N45°E	Pépité	-0.2	-----
		Exponentiel	1.2	15
	N45°O	Pépité	-0.2	-----
		Exponentiel	1.2	4

Figure 37 : Campagne 18 : Modèle bivariable avec la fonction F1K

Comme l'effet de pépite n'est que 5 (15% de la somme de paliers du modèle) et que la variable auxiliaire est mieux corrélée avec la concentration hivernale, on a obtenu par validation croisée des estimations très proches des vraies concentrations.



Variable (Nb=134)	Moyenne	Variance
Erreur	0.19	12.76
Erreur Réduite	0.03	0.73

Figure 38 : Campagne 18 : Validation croisée du modèle bivariable

4.3.3.2 ESTIMATION SUR LA SELECTION DES TUBES CORRELES AUX STATIONS DE REFERENCE

Parmi les tubes corrélés aux stations de référence, il y a 63 tubes qui ont été mesurés dans cette campagne, ceci implique une diminution de 53% des effectifs disponibles. Malgré cette réduction des tubes, les statistiques, la corrélation avec la fonction F1K et les variogrammes expérimentaux conservent les principales caractéristiques du phénomène (Figure 39).

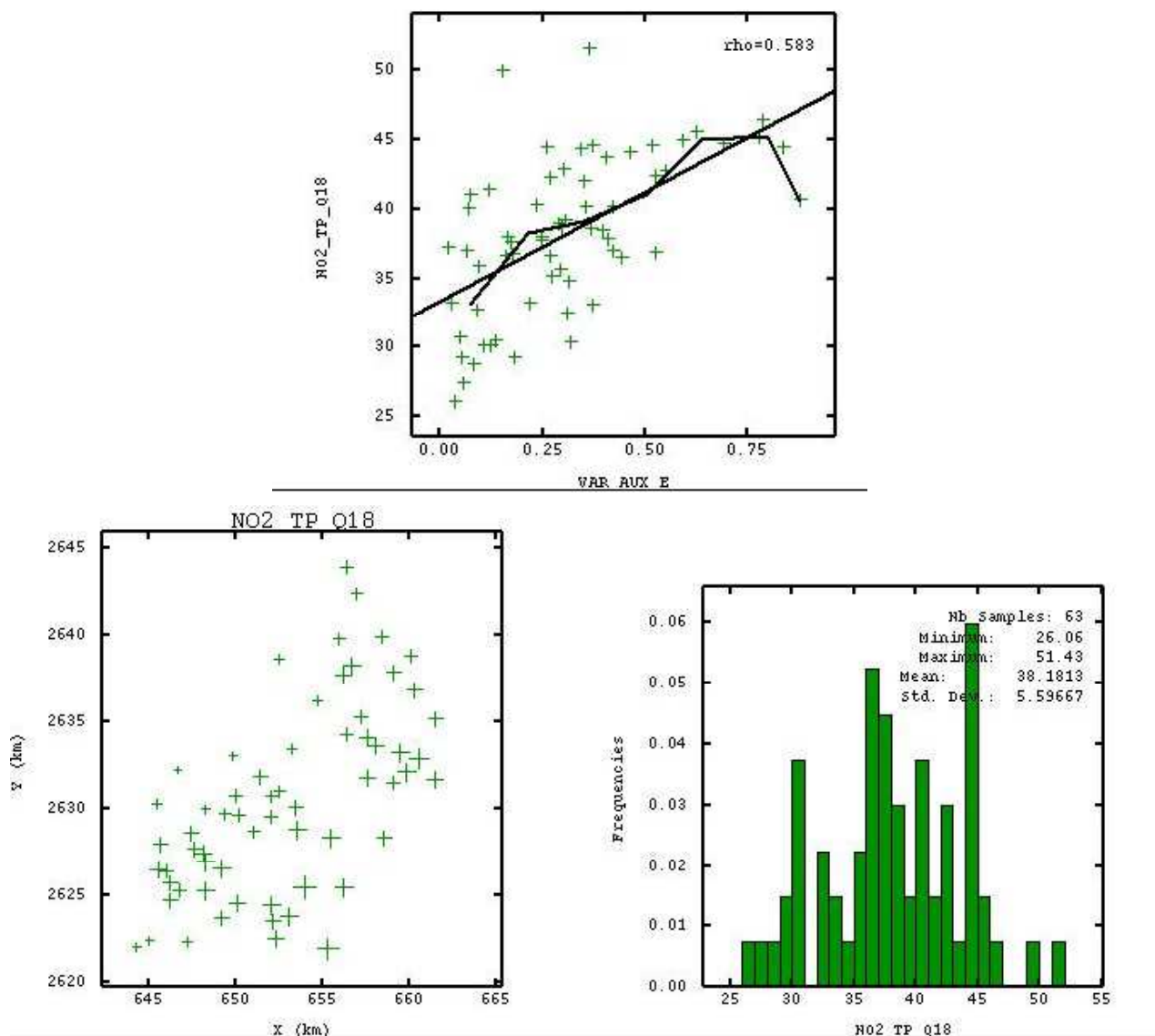
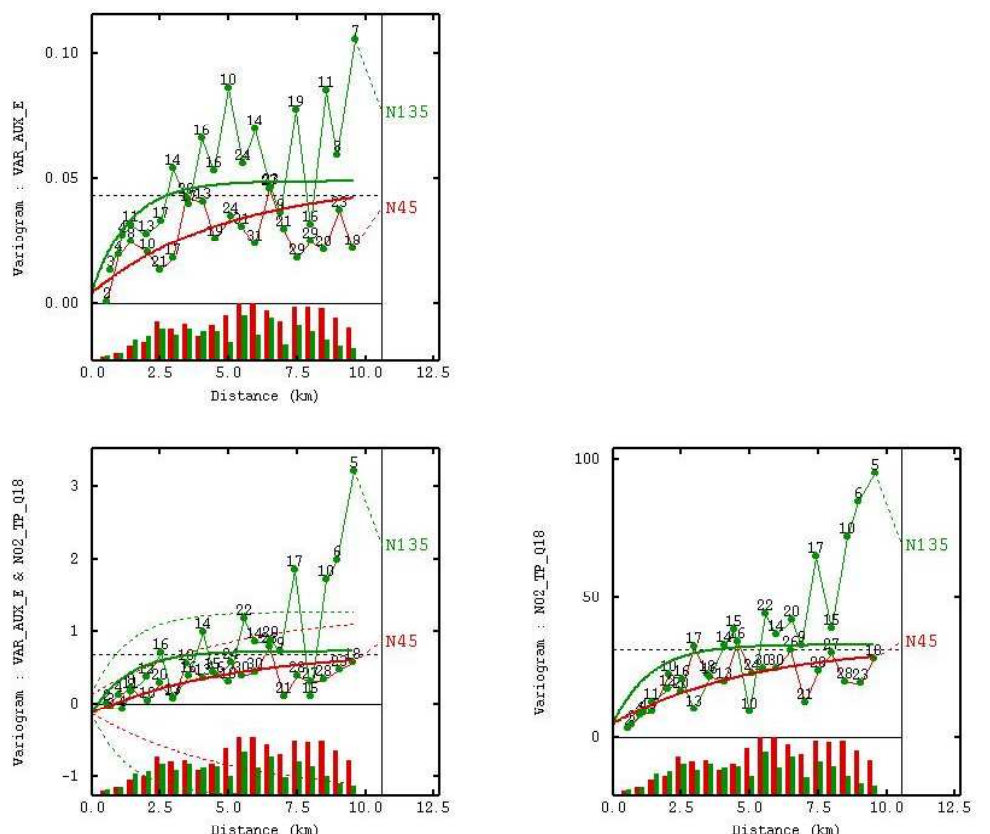


Figure en haut : Nuage de corrélation entre la concentration et la fonction auxiliaire F1K
 Deux lignes : Régression linéaire et espérance conditionnelle
 Figure en bas à gauche : implantation géographique des tubes
 La taille des croix est proportionnelle aux concentrations
 Figure en bas à droite : Histogramme

Figure 39 : Campagne No. 18: Localisation tubes, Histogramme et Nuage de Corrélation avec la variable F1K (63 tubes)

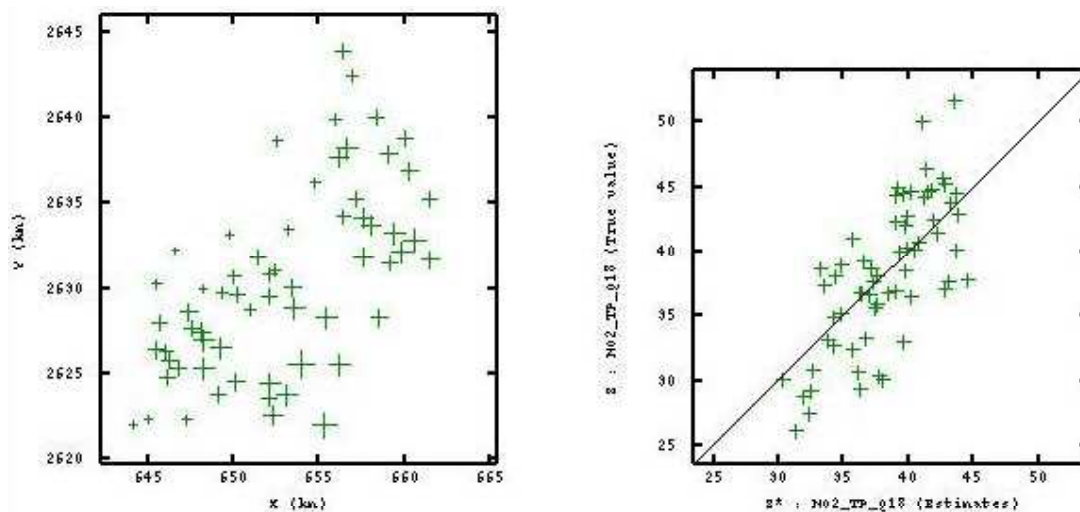
Le même modèle de variogramme ajusté précédemment est donc conservé pour l'ensemble des tubes, le seul changement consiste à réajuster les paliers du modèle à la nouvelle variance des données ($31.4 (\mu\text{g}/\text{m}^3)^2$). La somme des paliers est maintenant de $33.1 (\mu\text{g}/\text{m}^3)^2$ et la proportion de l'effet de pépite a diminué à 10% de la variabilité totale (Figure 40).



Variable (Nb=63)	Direction	Structure	Palier	Portée (Km)
NO2_Q18	N45°E	Pépite	4.6	-----
		Exponentiel	28.5	15
	N45°O	Pépite	4.6	-----
		Exponentiel	28.5	4
F1K	N45°E	Pépite	0.004	-----
		Exponentiel	0.045	15
	N45°O	Pépite	0.004	-----
		Exponentiel	0.045	4
NO2_Q18 et F1K	N45°E	Pépite	-0.13	-----
		Exponentiel	0.86	15
	N45°O	Pépite	-0.13	-----
		Exponentiel	0.86	4

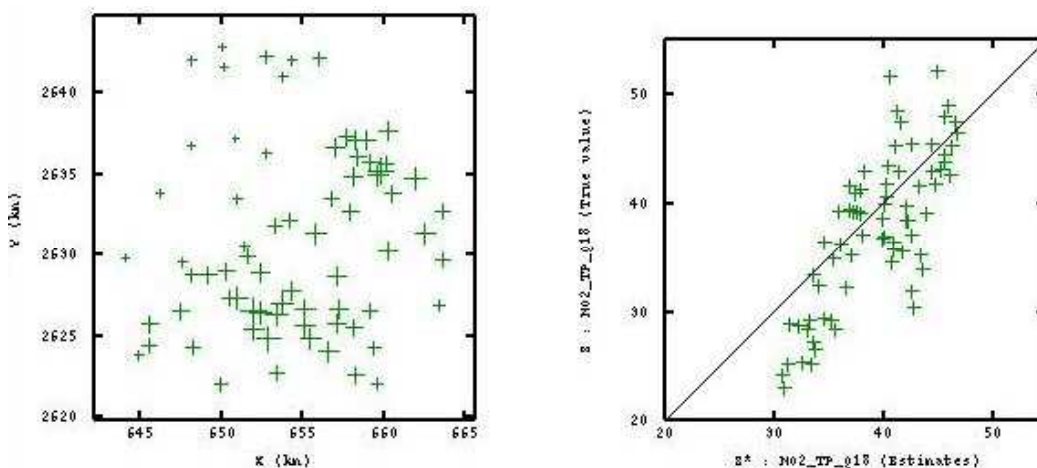
Figure 40 : Campagne No. 18 : Modèle bivariante avec la fonction F1K (63 tubes)

Pour la validation de ce modèle, nous avons procédé de la même façon que pour la campagne estivale : deux validations croisées ont été menées, la première sur la sélection de 63 tubes, l'autre sur les tubes qui n'ont pas participé à la construction du modèle. Les résultats sont illustrés dans les deux figures ci-dessous :



Variable (Nb=63)	Moyenne	Variance
Erreur	0.09	14.9
Erreur Réduite	0.016	0.8

Figure 41 : Campagne 18 Validation croisée du modèle bivariable (63 tubes)



Variable (Nb=71)	Moyenne	Variance
Erreur	1.73	20.6
Erreur Réduite	0.32	0.9

Figure 42 : Campagne 18 Validation croisée du modèle bivariable (71 tubes)

La qualité des estimations sur la sélection de 63 échantillons est très proche de celle effectuée sur l'ensemble du domaine : la variance des erreurs n'a augmenté que de 15%, elle est passée de 12 à 15 ($\mu\text{g}/\text{m}^3$)².

Les résultats de la Figure 42 sont une mesure de la qualité des estimations que l'on obtiendra si l'on estime à partir de la sélection de 63 échantillons. Dans le nuage de corrélation de cette figure on compare les mesures qui n'ont pas été sélectionnées avec leurs estimations.

Par rapport aux résultats précédents, on observe une détérioration des estimations, la variance des erreurs a augmenté de 40%, passant de 15 à 21 ($\mu\text{g}/\text{m}^3$)² (pour la première campagne cette variance avait aussi augmentée de 40%, passant de 25.2 à 35.2 ($\mu\text{g}/\text{m}^3$)²).

4.3.3.2.3 ESTIMATIONS FAITES A PARTIR DES DONNEES REESTIMEES (STATIONS VIRTUELLES)

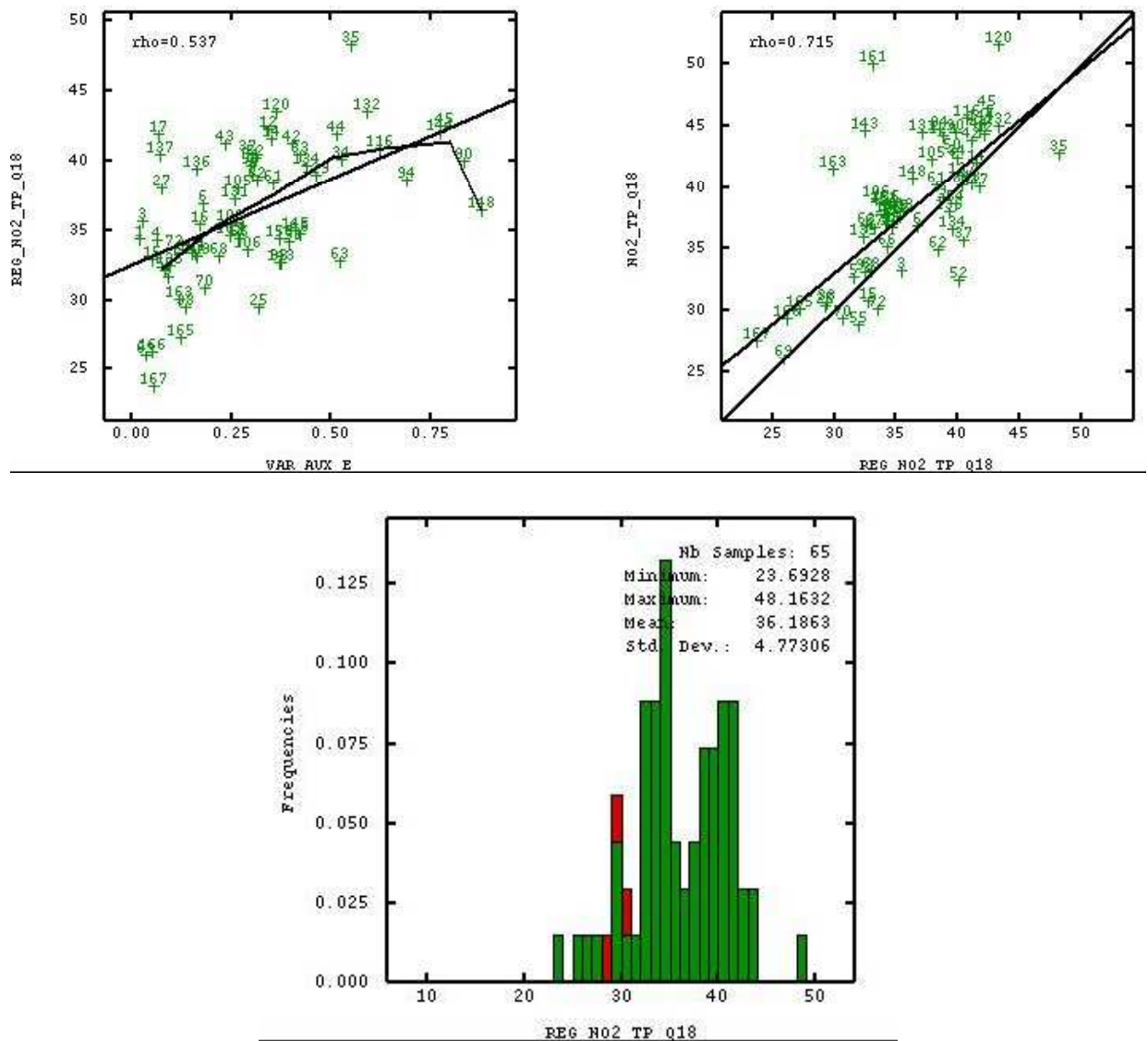
Par construction les concentrations sur les 65 stations virtuelles sont fonction de la moyenne des concentrations mesurées dans chaque station de référence.

L'histogramme des concentrations de la campagne 18 des stations virtuelles présente deux modes, la première, autour de 34 $\mu\text{g}/\text{m}^3$, représente les tubes liés à la station périurbaine et la seconde, autour de 42 $\mu\text{g}/\text{m}^3$, ceux liés à la station urbaine.

Par ailleurs, la moyenne des stations virtuelles qui est de 36.2 $\mu\text{g}/\text{m}^3$ est très proche de la moyenne des concentrations mesurées sur les deux stations de référence, qui est de 36.7 $\mu\text{g}/\text{m}^3$.

Par conséquent, les concentrations réestimées sont en moyenne plus faibles que les vraies concentrations (36.2 $\mu\text{g}/\text{m}^3$ par rapport à 38.2 $\mu\text{g}/\text{m}^3$).

Le coefficient de corrélation entre valeurs réestimées et vraies concentrations n'est que de 0.7 (nous rappelons que celui correspondant à la première campagne estivale était de 0.8), les points les moins bien réestimés correspondent aux stations 120, 161, 143, 163, 35 et 52, pour ces stations les écarts peuvent aller jusqu'à 17 $\mu\text{g}/\text{m}^3$.



Nuage en haut à gauche : Corrélation avec la variable explicative F1K (Axe X) :

Les lignes correspondent à la régression linéaire et à l'espérance conditionnelle

Nuage en haut à droite: Corrélation avec les 63 vraies concentrations (Axe Y) :

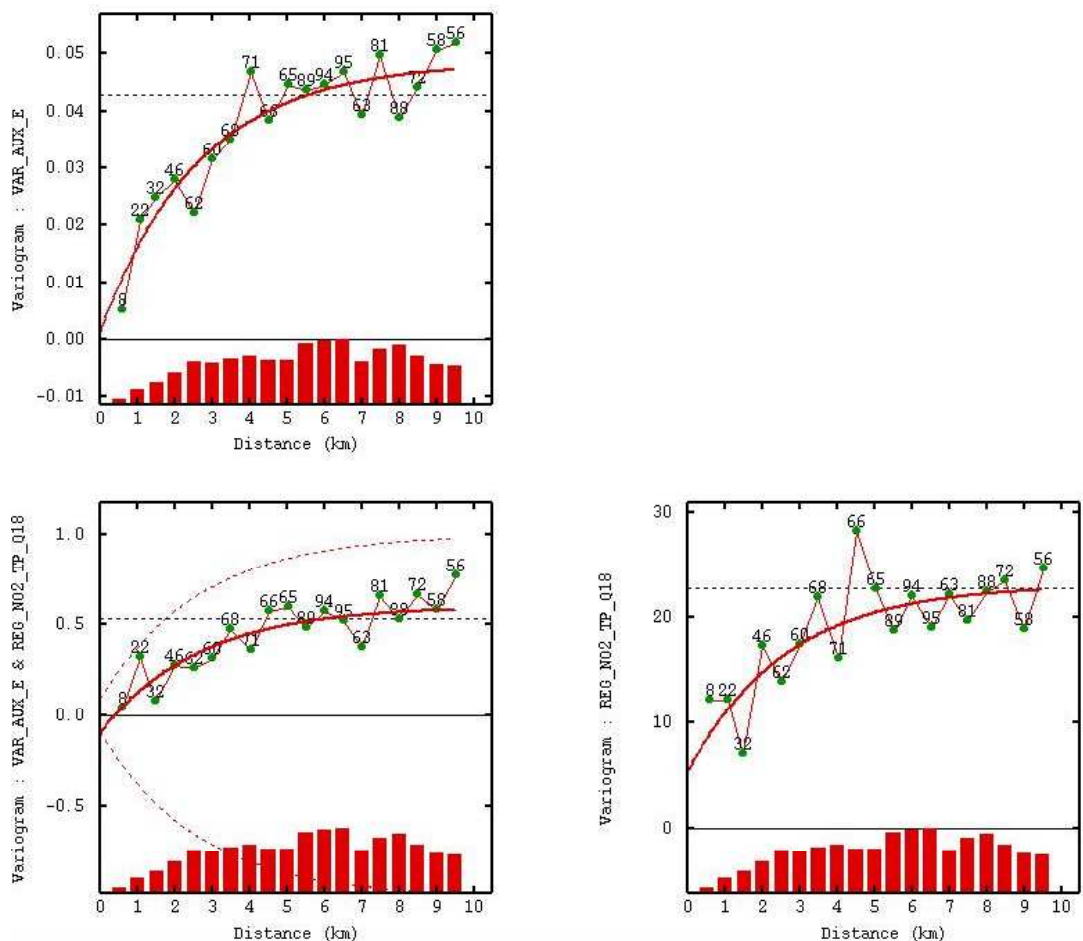
Les lignes correspondent à la première bissectrice et à la régression linéaire

Les nombres sur les nuages correspondent aux numéros des tubes

Figure en bas: Histogramme des concentrations réestimées

Figure 43 : Campagne 18: Principales caractéristiques des concentrations réestimées

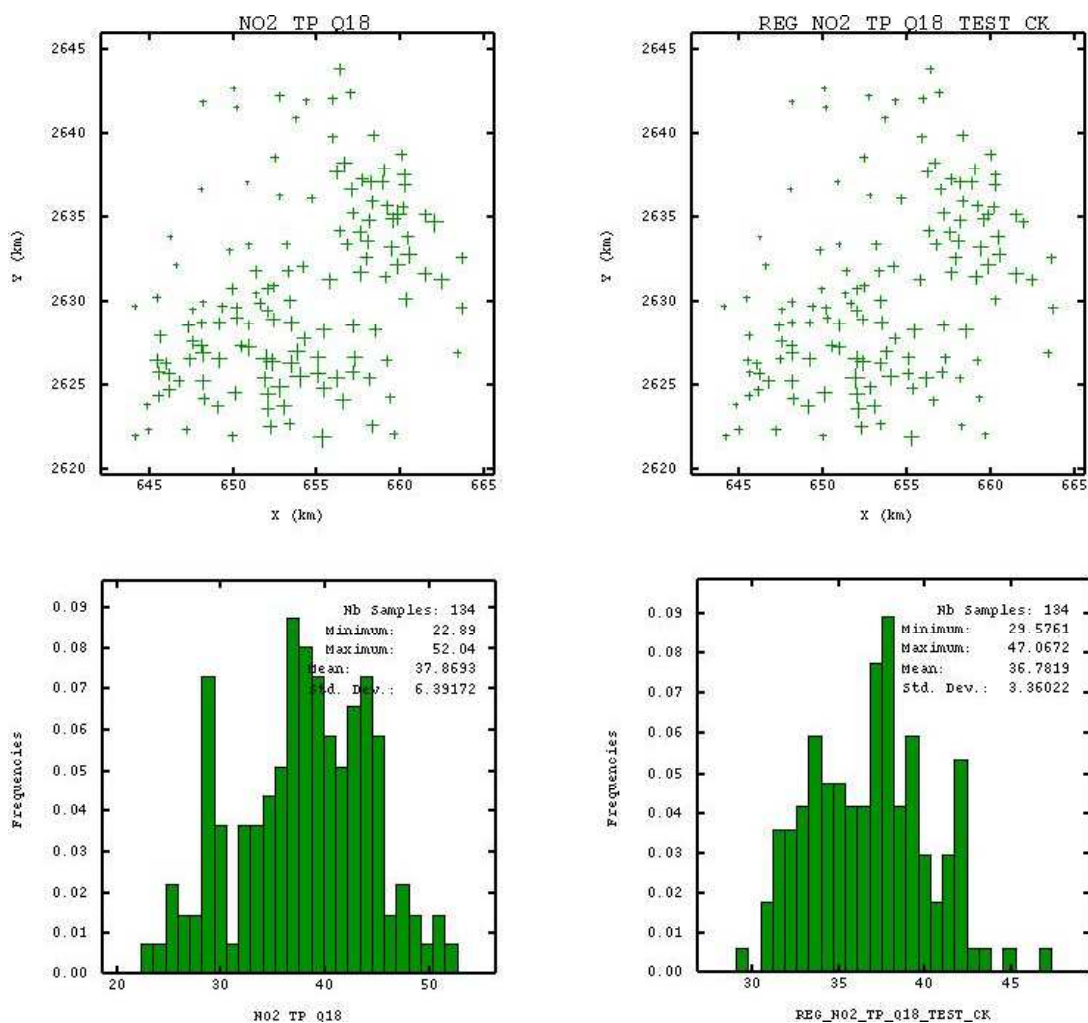
Nous avons choisi un variogramme expérimental isotrope calculé sur 20 pas de 500 m (10Km), le modèle ajusté est un effet de pépite et une structure exponentielle de 8 Km. La composante pépitique représente maintenant 23% de la variabilité totale.



Variable (Nb=63)	Structure	Palier	Portée (Km)
NO2_Q1	Pépite	5.43	----
	Exponentiel	17.7	8
F1K	Pépite	0.0015	----
	Exponentiel	0.047	8
NO2_Q1 et F1K	Pépite	-0.091	----
	Exponentiel	0.69	8

Figure 44 : Campagne 18 : Modèle calculé à partir de 63 stations virtuelles

Nous avons procédé de la même façon que pour l'analyse de la campagne estivale : une estimation par cokrigage colocalisé avec VEM a été menée sur les 134 échantillons à partir de 65 stations virtuelles, les résultats sont présentés dans les figures ci-après.



Deux figures à droite : Concentrations estimées
 Deux figures à gauche : Concentrations mesurées
 Deux cartes en haut : Localisation géographique des tubes
 La taille des croix est fonction des concentrations respectives
 Deux figures en bas : Histogrammes des concentrations

Figure 45 : Campagne 18 : Carte et histogramme des estimations obtenues sur l'ensemble du domaine et comparaison avec les concentrations mesurées

Pour cette campagne la réestimation des fortes concentrations est aussi moins précise, les écarts peuvent aller jusqu'à $13.4 \mu\text{g}/\text{m}^3$, aussi quelques tubes où l'on a mesuré une concentration d'environ $50 \mu\text{g}/\text{m}^3$, ont été réestimés à $37 \mu\text{g}/\text{m}^3$.

Par rapport aux estimations menées à l'aide de l'ensemble des mesures, la variance des erreurs a augmenté de 47%, passant de 12.8 à $18.7 (\mu\text{g}/\text{m}^3)^2$.

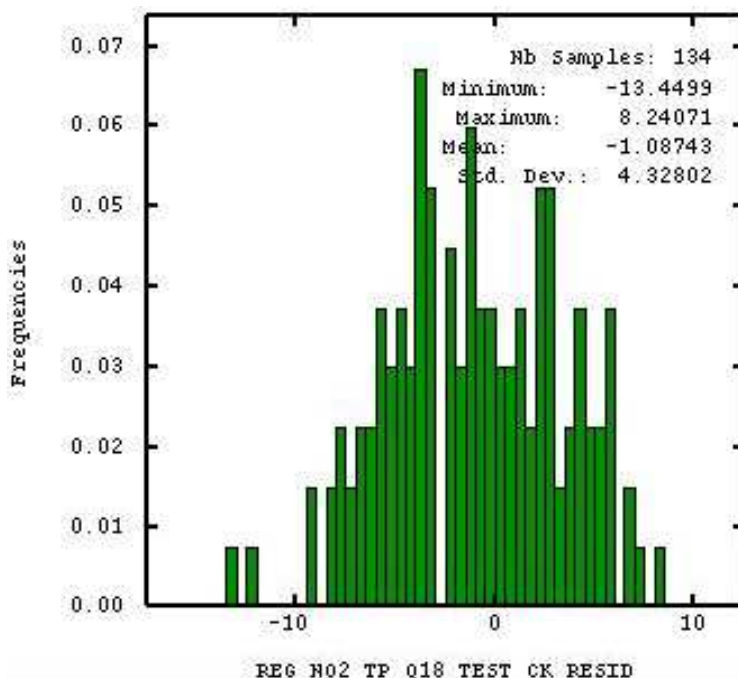
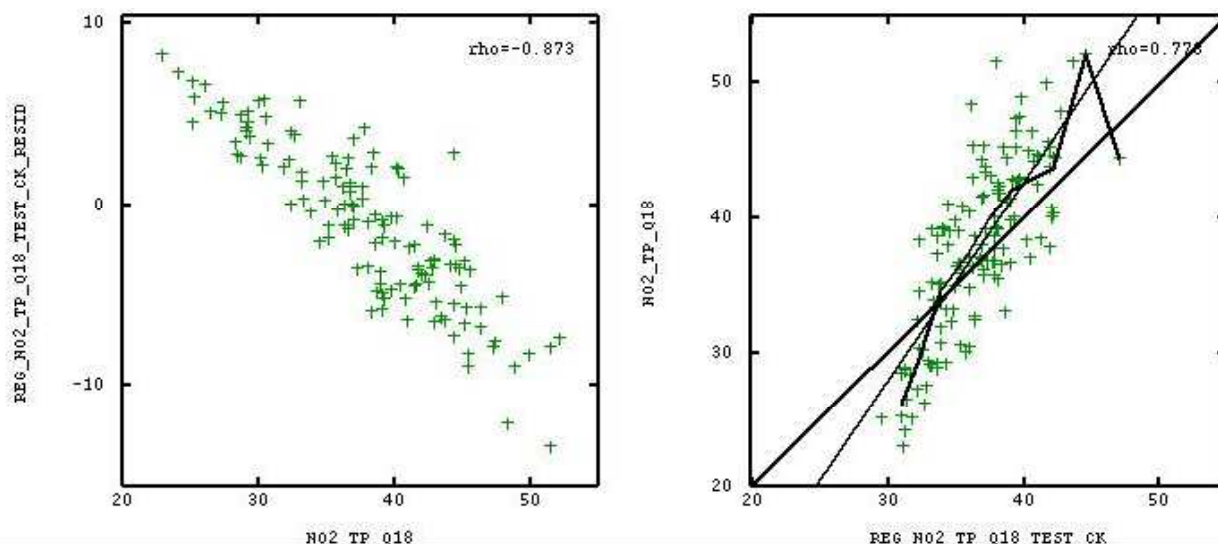


Figure en haut à gauche : Histogramme et statistiques des erreurs

Nuage en haut au milieu : Corrélation entre erreurs et mesures

L'axe X correspond aux erreurs d'estimations ($Z'-Z$)

L'axe Y correspond aux concentrations mesurées sur les 134 tubes

Nuage en bas à droite : Corrélation entre valeurs estimées et mesures

L'axe Y correspond aux vraies concentrations de 134 tubes

L'axe X correspond aux estimations de la concentration par cokrigage

Les lignes correspondent à la première bissectrice, à la régression linéaire et à l'espérance conditionnelle

Figure 46 : Campagne 18 : Histogramme des erreurs et nuages de corrélation avec les concentrations mesurées

4.3.4 ESTIMATION DES DONNEES JOURNALIERES SUR LES STATIONS VIRTUELLES

La Figure 18 décrit l'emplacement des 65 stations virtuelles corrélées aux stations de référence, en ces points il est donc possible d'estimer la concentration journalière du dioxyde d'azote.

Les paramètres ajustés à partir des campagnes bihebdomadaires (la pente et l'interception de l'axe Y) ont directement été appliqués pour l'estimation de données journalières.

En plus des données provenant des deux stations de référence, nous avons aussi à notre disposition les données de six autres stations fixes, il s'agit des données des stations urbaines Tourcoing-Bienfaisance, Marc-en-Baroueil CTM et Five-Lakanal, et de stations de trafic localisées dans la zone de Lille : Pasteur, Liberté et Faidherbe (voir la Figure 16).

Parmi les 65 tubes sélectionnés, trois tubes sont localisés au même emplacement que les stations fixes, il s'agit des tubes 3 (Caillou), 44 (Château) et 52 (Marc-en-Baroueil). Pour cette raison, ces trois tubes ont été retirés du jeu de données.

Le jeu de données final est composé de 70 points : 62 tubes ou stations virtuelles et 8 stations fixes.

La méthode consiste donc à créer ce jeu de données, puis à effectuer une carte journalière par cokrigage colocalisé avec la fonction F1K en intégrant la VEM des stations virtuelles.

Une seule fonction explicative (F1K) est maintenue pour toutes les saisons, car les coefficients de corrélation ne varient pas nettement d'une saison à une autre.

En revanche, un modèle de variogramme différent est ajusté pour chaque saison : pour l'été le modèle isotrope constitué par un effet de pépite et une structure gaussienne de portée 5 Km et pour l'hiver un effet de pépite et une structure gaussienne de portée 8 Km.

Le logiciel Isatis, utilisé pour cette étude, offre la possibilité d'ajuster les paliers des structures automatiquement, cette option a été utilisée dans un premier temps.

4.3.4.1 ETUDE DES VARIOGRAMMES JOURNALIERES

De façon à apprécier le comportement par saison des variogrammes expérimentaux journaliers, ceux-ci ont été calculés pour une période allant de début 1998 à fin 2003 ; le pas choisi pour tous les variogrammes était de 500 m, ils ont été calculés à partir de 62 stations virtuelles et des données provenant des stations fixes, quand elles sont disponibles. Ensuite, la moyenne saisonnière des variogrammes pour chaque classe de distance a été calculée.

Dans les figures suivantes, les variogrammes moyennés et les modèles ajustés sont montrés : effet de pépite et structure gaussienne de portée 5 Km pour les périodes printanières et estivales, et, effet de pépite et structure exponentiel de portée 8 Km pour les périodes automnales et hivernales. Ceux-ci ont été établis à partir des études effectuées sur les deux quinzaines des mesures par tube étudiées dans les sections précédentes, la taille de l'effet de pépite et du palier des modèles est ajustée chaque fois automatiquement en fonction des variogrammes expérimentaux.

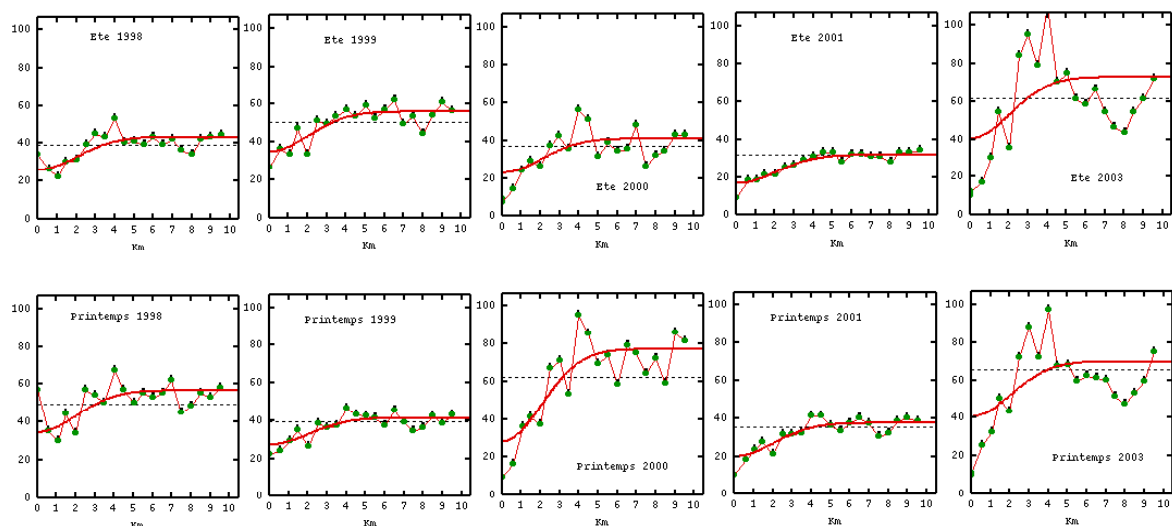


Figure 47 : Moyennes des variogrammes journalières pour les périodes printanières et estivales

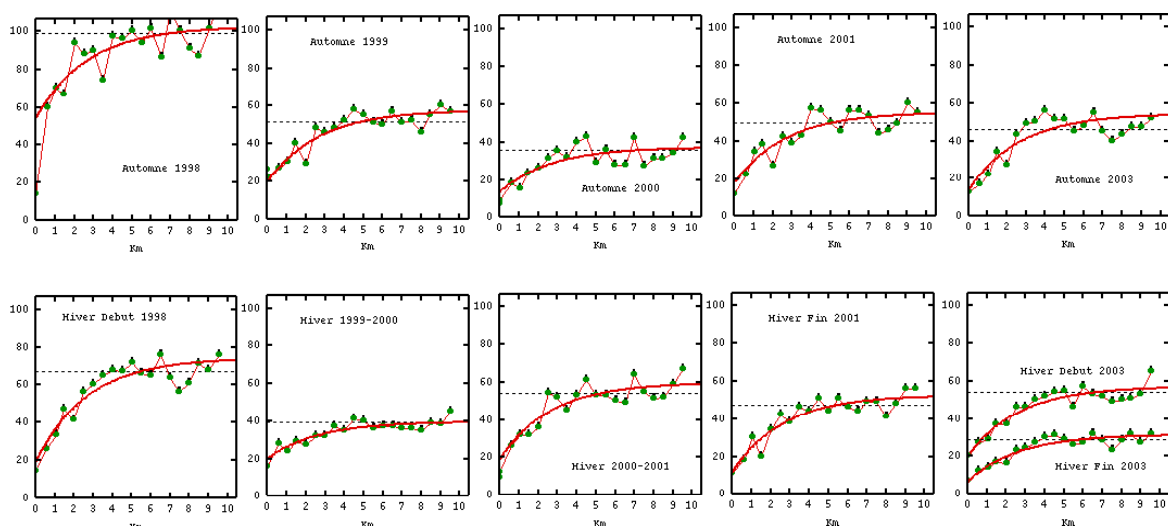


Figure 48 : Moyennes des variogrammes journalières pour les périodes automnales et hivernales

Les figures précédentes confirment que les structures choisies peuvent représenter correctement la corrélation spatiale du dioxyde d'azote.

Les saisons avec une plus grande variabilité des concentrations (variance supérieure à $60 (\mu\text{g}/\text{m}^3)^2$) sont : l'hiver du début 1998, l'automne 1998, le printemps 2000, le printemps 2003 et l'été 2003. Nous observons également que les variogrammes automnaux et hivernaux sont plus structurées (plus faible proportion de l'effet de pépite par rapport à ceux des saisons printanières et estivales).

Par ailleurs, nous pouvons conclure que la prise en compte des 3 stations de trafic ne perturbe pas l'allure des variogrammes.

4.3.4.2 VALIDATION DES ESTIMATIONS JOURNALIÈRES

Pour les estimations journalières il est difficile de valider les résultats, car nous ne disposons pas d'autres mesures journalières que celles évaluées aux huit stations fixes et utilisées pour le calage de la méthode. La seule possibilité est de comparer pour une même période de 14 jours la moyenne des estimations journalières par rapport aux concentrations mesurées par échantillonneurs passifs.

Nous avons effectué cette comparaison sur les deux quinzaines étudiées dans les sections précédentes : la première (Q1) et la dernière campagne (Q18).

Tout d'abord deux modèles ajustés automatiquement pour deux dates prises au hasard, le 9 juin 1998 et le 14 février 2004, sont étudiés.

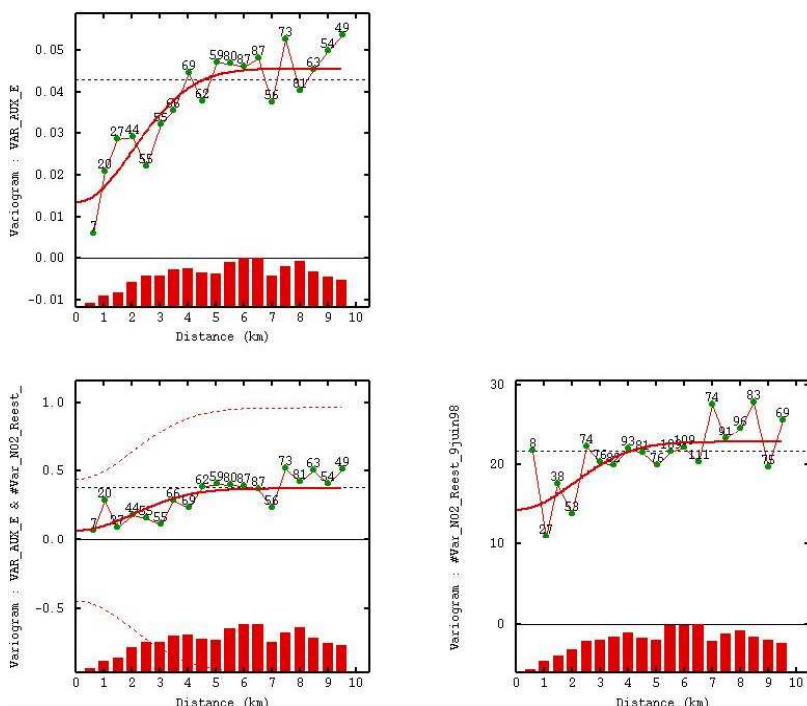


Figure 49 : Modèle ajusté automatiquement pour le 9 juin 1998

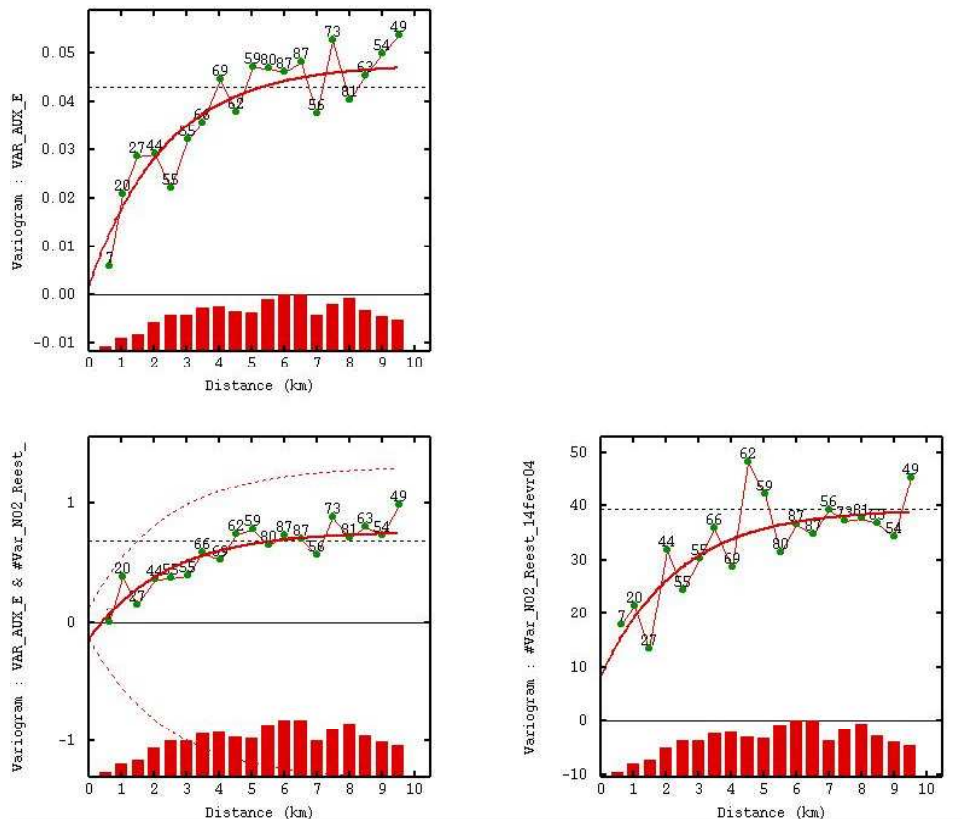


Figure 50 : Modèle ajusté automatiquement pour le 14 février 2004

Ces deux modèles montrent les caractéristiques de la corrélation spatiale de chaque saison, en été le phénomène est plus aléatoire et l'effet de pépète représente plus du 50% de la variabilité; en revanche en hiver les concentrations sont plus structurées, l'effet de pépète ne représentant que 25% de la variabilité.

Nous rappelons que ces variogrammes ont été calculés à partir des données réestimées, cependant les traits caractéristiques du phénomène sont conservés.

Sur la Figure 51, nous comparons la moyenne des 14 estimations journalières avec les mesures d'échantillonneurs passifs ; les statistiques des concentrations estimées et les deux nuages de corrélation sont très similaires à la Figure 34.

En effet, les mêmes résultats ont été obtenus lors des estimations menées à partir du calcul direct des concentrations bihebdomadaires sur les stations virtuelles (section 4.3.3.1.3). Ceci s'explique par le fait que dans les deux cas nous avons appliqué les mêmes corrélations temporelles sur les stations virtuelles et des modèles spatiaux très proches (structure gaussienne de portée 5 Km).

Pour la campagne 18, représentée en Figure 52, on obtient également un lissage des valeurs de concentration (résultats très proches à ceux de la Figure 46).

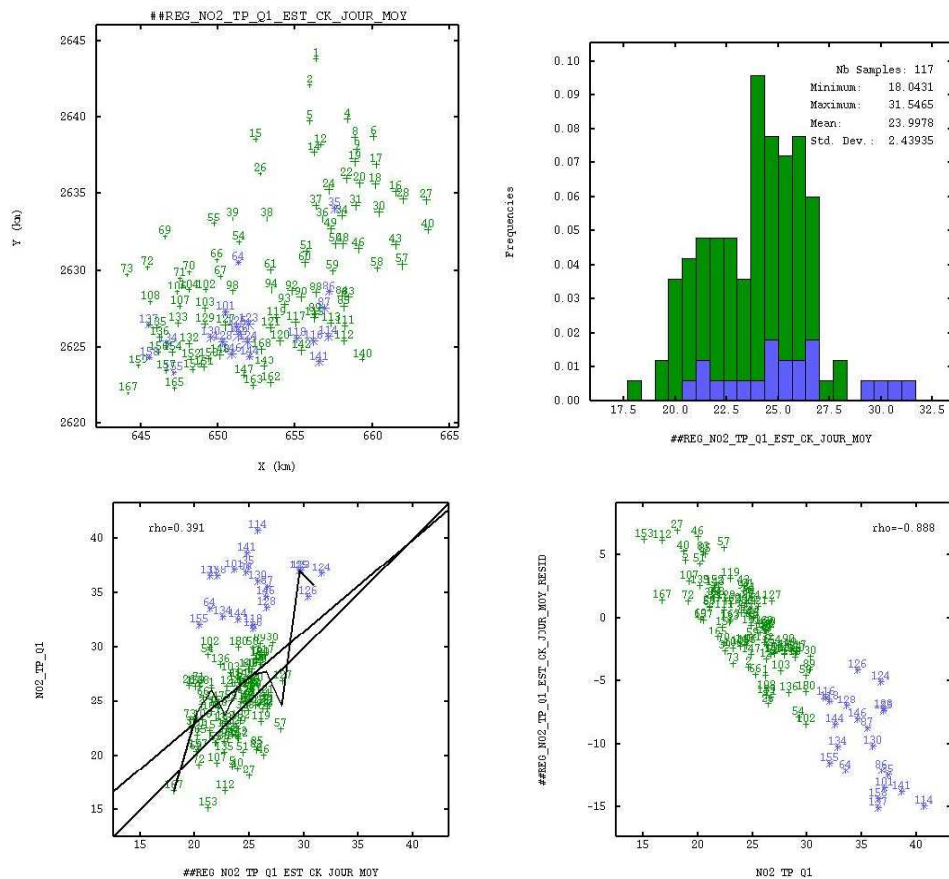


Figure en haut à gauche : localisation des données estimées
 La taille des croix est fonction de la concentration moyenne des 14 estimations
 Figure en haut à droite : Histogrammes de la moyenne de 14 estimations
 Figure en bas à gauche : Nuage de corrélation entre la moyenne de 14 estimations et les concentrations mesurées par échantillonneurs passifs
 Les lignes correspondent à la bissectrice, à la droite de régression et à l'espérance conditionnelle
 Figure en bas à droite : Nuage de corrélation entre les erreurs (Z^*-Z , axe Y) et les concentrations mesurées aux tubes (axe X).
 Les nombres sur les nuages correspondent aux numéros des tubes
 Les symboles en bleu sont les tubes où on a mesuré des concentrations supérieures à $31 \mu\text{g}/\text{m}^3$

Figure 51 : Campagne 1: Comparaison des estimations par cokrigage, à partir de 70 échantillons et les mesures par échantillonneurs passifs

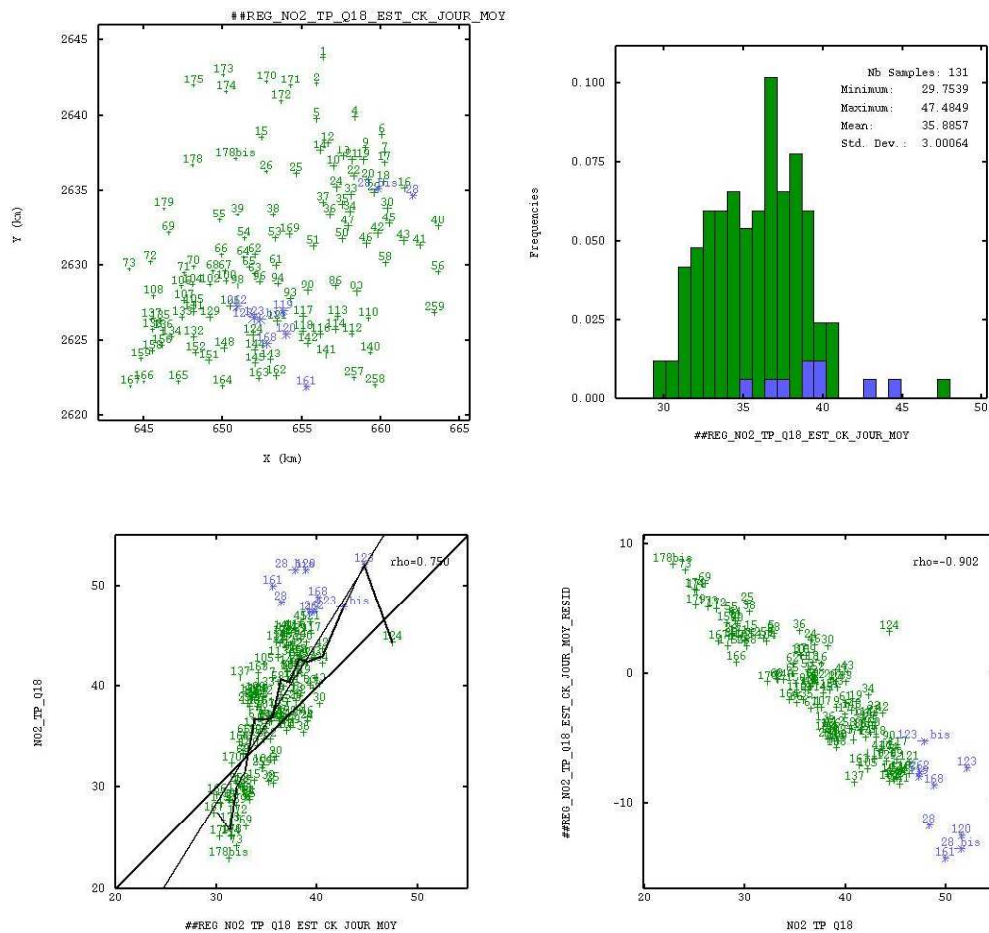
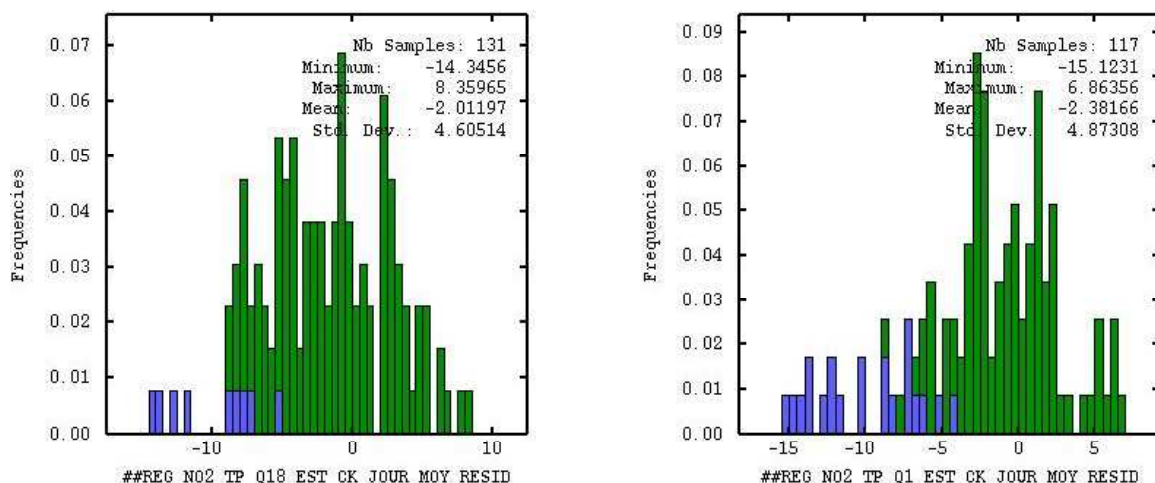


Figure en haut à gauche : localisation des données estimées
 La taille des croix est fonction de la concentration moyenne des 14 estimations
 Figure en haut à droite : Histogrammes de la moyenne de 14 estimations
 Figure en bas à gauche : Nuage de corrélation entre la moyenne de 14 estimations et les concentrations mesurées par échantillonneurs passifs
 Les lignes correspondent à la bissectrice, à la droite de régression et à l'espérance conditionnelle
 Figure en bas à droite : Nuage de corrélation entre les erreurs (Z^*-Z , axe Y) et les concentrations mesurées aux tubes (axe X).
 Les nombres sur les nuages correspondent aux numéros des tubes
 Les symboles en bleu sont les tubes où on a mesuré des concentrations supérieures à $47 \mu\text{g}/\text{m}^3$

Figure 52 : Campagne 18 : Comparaison des estimations par cokrigage, à partir de 62 échantillons et les mesures par échantillonneurs passifs

En Figure 53, nous comparons les statistiques des erreurs des deux campagnes. En été les erreurs peuvent varier entre -15 et $7 \mu\text{g}/\text{m}^3$, et en hiver entre -14 et $8 \mu\text{g}/\text{m}^3$.



Sur l'histogramme des erreurs (Z*-Z) de la campagne 18, à gauche :
 Les symboles en bleu sont les tubes où on a mesuré de concentrations supérieures à 31 µg/m³
 Sur l'histogramme des erreurs (Z*-Z) de la campagne 1, à droite :
 Les symboles en bleu sont les tubes où on a mesuré de concentrations supérieures à 47 µg/m³

Figure 53 : Histogrammes et statistiques des erreurs des deux campagnes

Une façon de comparer les erreurs commises est de calculer l'erreur relative.

Par exemple comme la valeur maximale de la campagne 18 est de 52 µg/m³, une erreur de -14 µg/m³ correspond à une sous-estimation de 27% et comme la valeur minimale est de 23 µg/m³, une erreur de 8 µg/m³ correspond à une sur estimation de 35%.

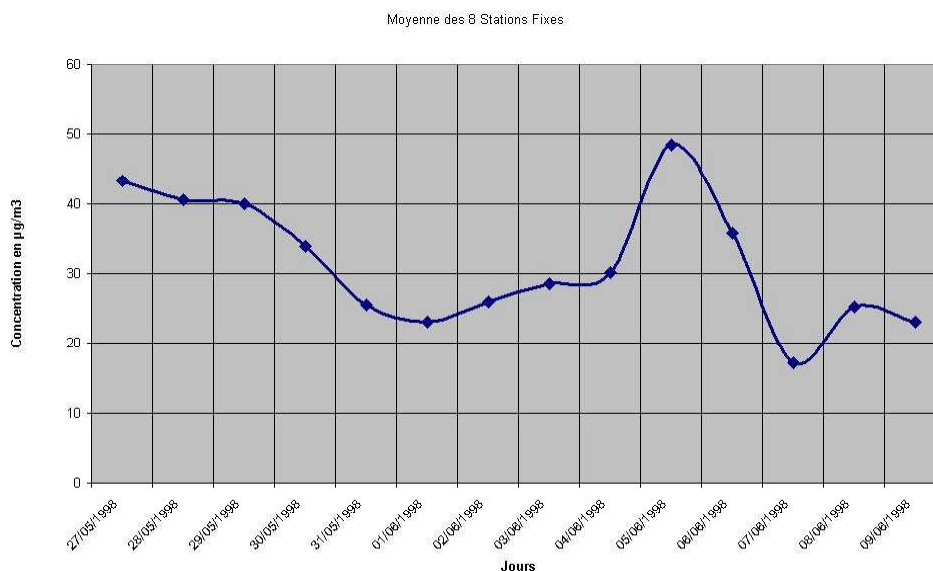
Pour la campagne 1 une erreur de -15 µg/m³ correspond à une sous-estimation de 37% pour la valeur maximale de 41 µg/m³ et une erreur de 7 µg/m³ correspond à une sur estimation de 47% pour la valeur minimale de 15 µg/m³.

Variable	Campagne 18 : Hiver		Campagne 1 : Été	
	Min	Max	Min	Max
Concentrations (µg/m ³)	23	52	15	41
Erreur Brute (µg/m ³)	8	-14	7	-15
Erreur Relative (%)	35	-27	47	-37

Tableau 11 : Erreurs brutes et relatives sur deux campagnes représentatives

4.3.5 PRESENTATION DE QUELQUES CARTES OBTENUES

La Figure 54 montre le comportement journalier de la moyenne des 8 stations fixes pour la période allant du 27 Mai au 9 juin 1998. Les concentrations les plus faibles ont été mesurées lors de trois dernières journées (17.2 $\mu\text{g}/\text{m}^3$ pour le 7 juin) et les plus fortes le 5 juin 1998 (48.4 $\mu\text{g}/\text{m}^3$).



Période	MIN	MAX	MOY
27 Mai au 9 Juin 1988	17.2	48.4	31.45

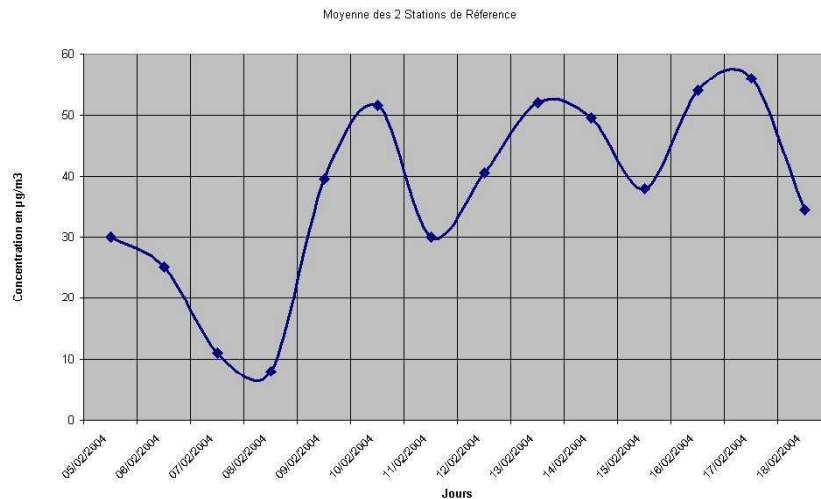
Figure 54 : Statistiques des mesures des 8 stations fixes pour la période allant du 27 Mai au 9 Juin 1998

Parmi les 14 jours de cette période estivale, nous présentons par la suite les estimations obtenues pour la journée du 9 juin 1998 (moy=23.08 $\mu\text{g}/\text{m}^3$).

Pour la période hivernale allant du 5 au 18 février 2004, les mesures des deux stations fixes de référence ont été utilisées, portant le nombre de données sur lequel les estimations ont été réalisées à 64 échantillons.

La Figure 55 montre le comportement de la moyenne journalière des deux stations de référence pour la période allant du 5 au 18 février 2004.

On observe que les valeurs minimales ont été mesurées vers le 7 et le 8 février (8 $\mu\text{g}/\text{m}^3$), et les valeurs maximales alternativement entre le 10, 13, 14, 16 et 17 février 2004 (concentrations supérieures à 50 $\mu\text{g}/\text{m}^3$) ; la moyenne des mesures est de 37.1 $\mu\text{g}/\text{m}^3$.



Période	MIN	MAX	MOY
5 au 18 février 2004	8	56	37.1

Figure 55 : Statistiques des mesures des 2 stations de référence pour la période allant du 5 au 18 février 2004

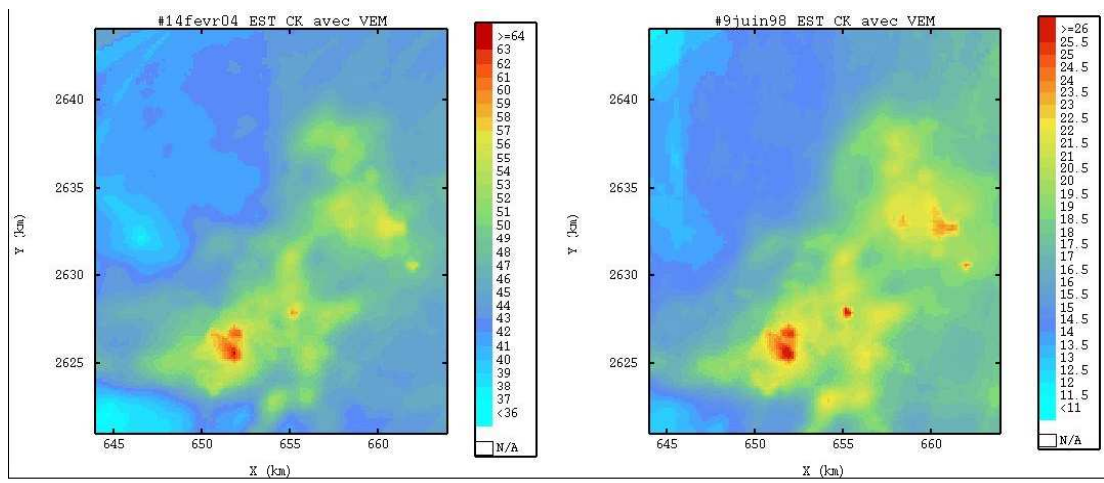
Parmi les 14 jours de cette période hivernale, nous présentons par la suite les estimations obtenues pour la journée du 14 février 2004.

La Figure 56 montre les niveaux de concentrations estimées pour ces deux journées. A noter que l'échelle de couleurs des cartes a été adaptée en fonction des concentrations de chacun des jours, et est donc différente pour chacune des cartes.

Pour la journée estivale, les plus fortes concentrations estimées sont de l'ordre de $25 \mu\text{g}/\text{m}^3$. Ces niveaux de concentrations sont localisés dans le centre ville de Lille, dans l'intersection des autoroutes au nord et dans la zone de Roubaix.

Pour la journée hivernale, les concentrations sont presque trois fois plus grandes, les valeurs les plus fortes peuvent atteindre $63 \mu\text{g}/\text{m}^3$; les zones les plus polluées sont le centre ville de Lille et la partie nord du périurbain.

Le niveau de détail obtenu dans les deux cartes est dû à l'emploi de la variable explicative F1K et à la méthode d'estimation utilisée : cokrigage colocalisé avec VEM. Rappelons que ces deux cartes n'ont été obtenues qu'à partir des données de stations fixes disponibles et de la méthode de corrélation temporelle.



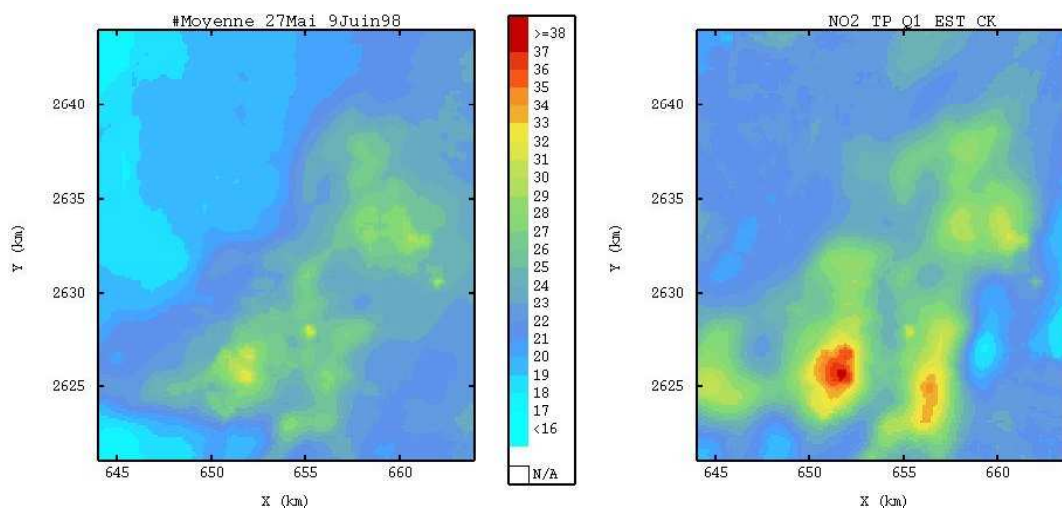
Estimations par cokrigage avec F1K et VEM ($\mu\text{g}/\text{m}^3$)	Minimum	Maximum	Moyenne	Variance ($\mu\text{g}/\text{m}^3$) ²
14 Février 2004 (à gauche)	36.5	63.1	45.4	0.47
9 Juin 1998 (à droite)	11.84	26.3	17	6.01

Figure 56 : Cartes d'estimation des journées 9 Juin 1998 et 14 Février 2004.

Afin d'apprécier la qualité des estimations ainsi réalisées, nous montrons dans les deux figures ci-après la moyenne des estimations journalières réalisées pendant chaque période, comparée avec le krigeage des tubes respectifs.

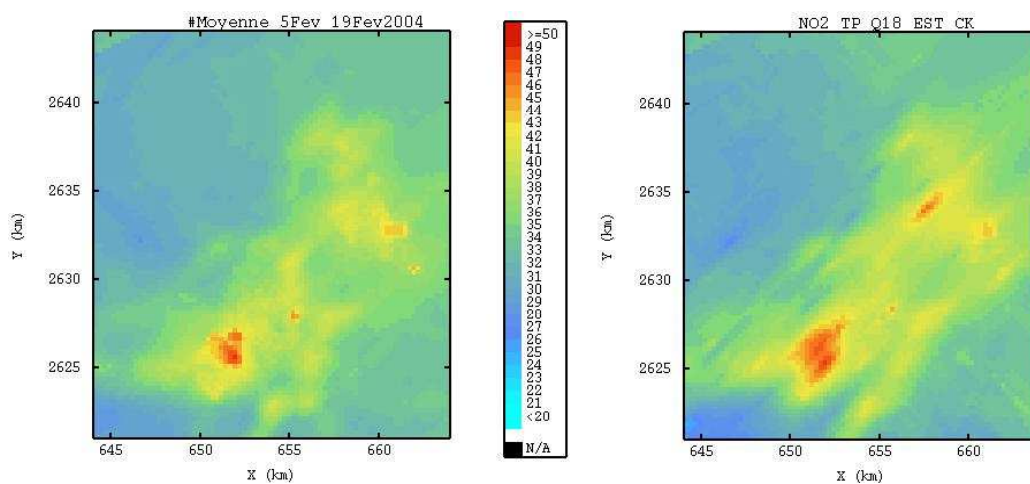
Sur les deux cartes de chaque figure, l'effet de lissage opéré par le cokrigage est clairement représenté. La différence entre les « motifs » de pollution est due principalement aux raisons suivantes:

- Des modèles différents ont été utilisés : un modèle isotrope pour les estimations journalières et un modèle anisotrope pour l'estimation de la campagne des tubes passifs : on observe par exemple sur la campagne hivernale que les estimations sont orientées dans une direction N45°E.
- Des données conditionnelles en nombre différent : pour les estimations journalières nous n'avons qu'une soixantaine d'échantillons, tandis que pour les estimations des quinzaines plus de 100 tubes étaient disponibles.



Campagne 1	Minimum	Maximum	Moyenne	Variance ($\mu\text{g}/\text{m}^3$)²
Moyenne des estimations journalières (à gauche)	17	32.2	22	6.6
Estimation par tubes passifs (à droite)	18	37.4	24	9.3

Figure 57 : Campagne 1 : Comparaison des cartes d'estimation



Campagne 18	Minimum	Maximum	Moyenne	Variance ($\mu\text{g}/\text{m}^3$)²
Moyenne des estimations journalières (à gauche)	27.9	48.4	34.3	9.14
Estimation par tubes passifs (à droite)	24.3	48.2	34.7	13.4

Figure 58 : Campagne 18 : Comparaison des cartes d'estimation

5 CONCLUSIONS

Pour effectuer une cartographie journalière du dioxyde d'azote il est nécessaire de disposer d'un nombre suffisant et relativement important de données de mesures, or cela constitue parfois une limitation, le nombre de stations fixes étant nécessairement limité et parfois insuffisant.

Dans la première partie de cette étude, nous avons présenté une méthode qui peut être utilisée pour compléter, par des données supplémentaires estimées, un nombre restreint de données de mesures, afin de rendre possible une cartographie journalière du dioxyde d'azote sur une agglomération.

Cette méthode est basée sur l'existence d'un historique de campagnes par échantillonneurs passifs dans l'agglomération. Ces campagnes sont exploitées de façon à pouvoir grouper les tubes (appelés désormais stations virtuelles). Pour cela, ils doivent être corrélés temporellement à une station fixe, appelée station de référence.

Dans la seconde partie de cette étude nous avons testé sur un cas d'étude la pertinence de cette méthode et la qualité des estimations effectuées à partir des données ainsi produites.

Les données réestimées peuvent présenter des écarts importants par rapport aux mesures réelles. Cependant la géostatistique permet de prendre en compte ces erreurs sous la forme d'une variance de l'erreur de mesure (VEM) : ce terme est une composante additionnelle de la variabilité des données (comme l'est l'effet de pépite, par exemple) et doit être calculé sur les stations virtuelles.

Un autre apport intéressant des méthodes géostatistiques est le fait de pouvoir exploiter l'information additionnelle apportée par des variables auxiliaires ; dans le cas du dioxyde d'azote, des variables décrivant le milieu tels que la densité de population, le relief, ou les émissions peuvent être utilisées.

La méthode du cokrigeage s'avère être la plus adéquate dans notre cas d'étude, cependant d'autres méthodes multivariées peuvent éventuellement être utilisées.

Pour réaliser une carte journalière du dioxyde d'azote, on ne peut s'affranchir d'une analyse manuelle préalable des campagnes existantes afin de prédéfinir un certain nombre de variables et paramètres (variable auxiliaire la mieux corrélée, structure et portée du variogramme).

Nous avons créé une fonction explicative formée par l'inventaire d'émissions et la densité de population. Cette fonction qui est connue dans tout le domaine d'estimation a permis d'améliorer la précision d'estimation dans les endroits dépourvus de stations virtuelles.

L'automatisation de la procédure de cartographie est réalisée en ajustant un modèle isotrope unique pour chaque saison (hiver – été). Ce choix est fondé par le fait que la corrélation spatiale du dioxyde d'azote dans la zone d'étude, ne varie pas beaucoup d'une année à une autre. Cependant, il serait intéressant de tester cette hypothèse sur des configurations de données différentes, sur d'autres villes, d'autres milieux, etc.

Il est difficile de valider les résultats obtenus sur les estimations journalières. En revanche, il est possible de comparer la moyenne sur une période de 14 jours avec les données mesurées par échantillonneurs passifs.

Deux campagnes ont été considérées pour effectuer ces tests. Malgré l'utilisation des variables auxiliaires, nous observons un lissage des estimations : en hiver, les valeurs fortes sont sous estimées en 27%, et de 37% en été ; de la même façon, en hiver les faibles concentrations sont surestimées de 35% et de 47% en été. (Rappelons que le critère de qualité est fixé à 50% pour ce type de méthodes).

Finalement nous avons effectué des estimations journalières de façon automatique sur l'ensemble du domaine, les cartes sont obtenues par cokrigage avec la fonction explicative F1K et la prise en compte de la VEM. Un modèle prédéfini est utilisé pour chaque saison (été et hiver) : le type de structure et la portée restent fixe, mais le palier est ajusté automatiquement chaque jour.

Les cartes des deux journées étudiées sont assez détaillées, de manière générale elles reconstituent les principales caractéristiques de la pollution par dioxyde d'azote dans le domaine considéré.

Une voie d'amélioration semble être la réduction de l'effet de lissage peut-être par un travail sur les modèles de régression (ceux-ci effectuant déjà un premier lissage des valeurs fortes) et par la création des stations virtuellement supplémentaires.

6 BIBLIOGRAPHIE

[1] Perdrix E., Wroblewski A., Fourché B. et Plaisance H.

Cartographie des concentrations en dioxyde d'azote à partir des mesures en stations fixes.

LCSQA/EMD - Etude n°11 – 2004,

[2] Arrêté du 17 mars 2003 relatif aux modalités de surveillance de la qualité de l'air et à l'information du public, J.O. 167 du 22 juillet 2003.

[3] Directive 1999/30/CE du Conseil du 22 avril 1999 relative à la fixation de valeurs limites pour l'anhydride sulfureux, le dioxyde d'azote et les oxydes d'azote, les particules et le plomb dans l'air ambiant, J.O. des Communautés Européennes 163 du 29 juin 1999.

7 LISTE DES ANNEXES

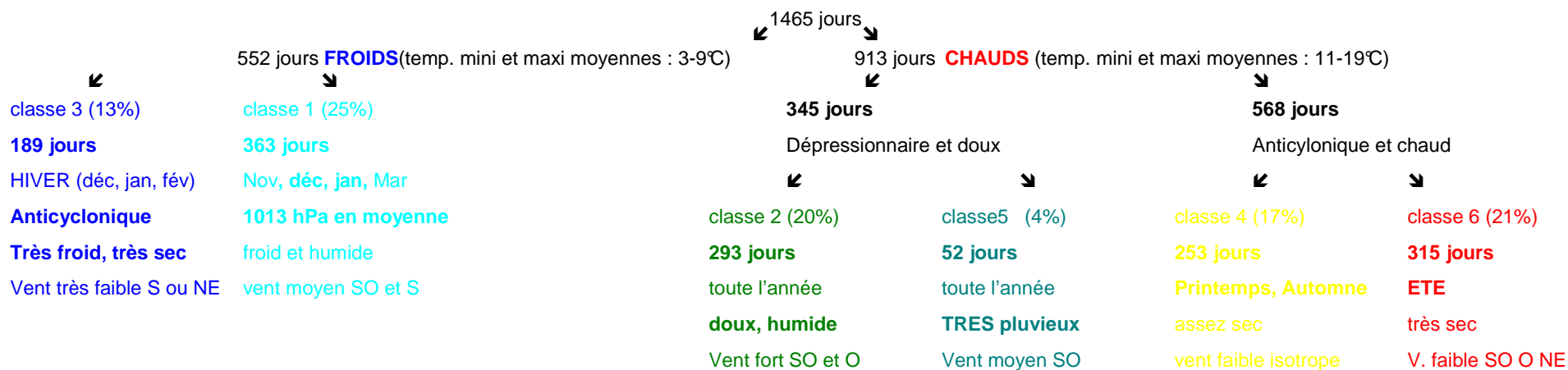
Repère	Désignation	Nombre de pages
Annexe 1	Classification hiérarchique des données météorologiques journalières	1
Annexe 2	Fiche programme	2

ANNEXE 1 : Classification hiérarchique des données météorologiques journalières.

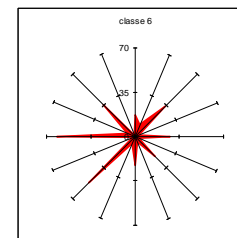
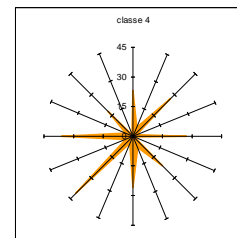
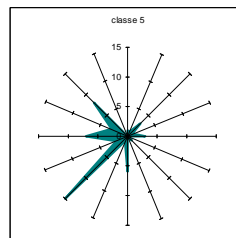
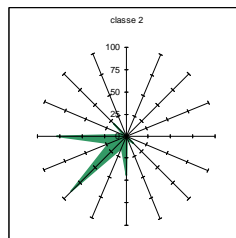
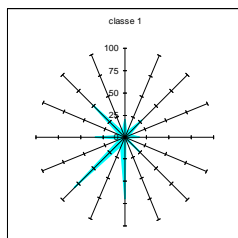
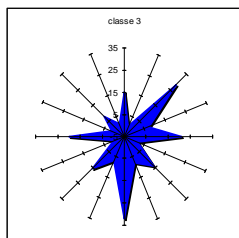
Classification hiérarchique ascendante, méthode de Ward, distance euclidienne au carré, données standardisées (centrées réduites)

Variables : pression, précipitations, température maximale, température minimale, visibilité, vitesse du vent.

2191 données journalières du 01/01/1998 au 31/12/2003, dont 1465 jours totalement renseignés.



	Classe 3	Classe 1	Classe 2	Classe 5	Classe 4	Classe 6
Press. (hPa)	1026	1013	1007	1009	1017	1019
Précip.(mm)	0,2	2,2	3,2	16,1	1,2	0,4
Tmax (°C)	6,3	9,9	15,7	14,8	19,2	22,4
Tmin (°C)	0,6	4,3	10,1	9,4	11,2	12,4
Visib. (km)	9	18	26	13	10	28
VV (km/h)	12	21	30	21	15	16
NO2 H	48	29	17	24	31	24
NO2 R	56	39	25	36	36	30
O3 H	12	32	48	40	37	50



THEME : TRAITEMENTS NUMERIQUES

ADAPTATION METHODES GEOSTATISTIQUES A LA CARTOGRAPHIE AUTOMATIQUE DE NO2

OBJECTIF

Le but de cette fiche est d'exploiter au mieux les données horaires/journalières des stations fixes dont sont équipées les AASQA, c'est-à-dire d'étudier une méthodologie de cartographie des concentrations en NO₂ à partir des mesures horaires/journalières des stations fixes.

CONTEXTE

Dans le cadre de l'application de l'arrêté du 17/03/2003 Art.9-III, les AASQA doivent réaliser et publier régulièrement une information indiquant les niveaux de polluants dans chaque zone. La représentation cartographique de cette information permet une présentation claire et pertinente de la situation.

Jusqu'à présent les méthodes d'interpolation géostatistiques (krigeage) ont été employées pour effectuer des cartographies des niveaux de pollution à partir des données des échantillonneurs passifs, ces données représentent la moyenne du polluant dans une période de plusieurs jours (7 ou 14 jours par exemple) ; on obtient donc des cartes de la pollution de fond représentant la moyenne des niveaux sur des périodes de temps assez longues.

Les données de stations fixes quant à elles sont des mesures quart horaires qui sont utilisées pour surveiller le dépassement du niveau d'alerte de certains polluants, par exemple l'ozone ou le dioxyde d'azote

Dans le cadre de l'étude LCSQA 2004 « Cartographie des concentrations en NO₂ à partir des mesures des stations fixes », nous avons montré que certains sites de mesure par échantillonnage passif présentaient des mesures corrélées à celles des stations fixes automatiques. De tels sites constituent alors des « stations virtuelles » auxquelles on peut affecter, par calcul, une teneur en NO₂ en utilisant la corrélation avec la station automatique réelle. Le nombre de données, réelles et calculées, ainsi obtenues et leur couverture spatiale permet d'envisager la réalisation par géostatistique de cartes de répartition du polluant, avec un pas de temps court (horaire/journalier). Cependant, le raccourcissement du délai de production des cartes nécessite une automatisation de la procédure de cartographie.

TRAVAUX PROPOSES POUR 2005

L'avantage des méthodes d'interpolation géostatistique est la prise en compte de la corrélation spatiale du polluant, cette corrélation est calculée à partir de la covariance spatiale des données disponibles.

L'idée est d'utiliser les mêmes méthodes d'interpolation géostatistiques mais sur des données représentant une période du temps plus petite: une heure, période de huit heures ou un jour par exemple. L'intérêt est d'avoir à disposition et dans un court délai une cartographie (presque) automatique des niveaux de pollution.

Un des principaux objectifs de l'étude sera donc de concevoir l'algorithme de cartographie automatique, capable de réaliser une modélisation automatique de la covariance spatiale, et donc de la représentation spatiale des concentrations de NO₂. La méthode devra également fournir un indice de confiance en fonction de la qualité de la modélisation.

Plusieurs études de sensibilité devront être effectuées de façon à déterminer les conditions d'application de l'algorithme et notamment pour définir le nombre minimal des capteurs nécessaires pour arriver à une cartographie acceptable de la pollution.

DUREE DES TRAVAUX

Les travaux seront menés en 2 parties, conjointement par l'EMD et l'INERIS :

- 1ère partie en 2005 : mise au point de la méthode et définition des conditions d'application dans la région Nord-Pas-de-Calais
- 2ème partie en 2006 : application de la méthode sur une zone géographique différenciée

COLLABORATIONS

AASQA , ADEME