

ECHANTILLONNAGE ET CARTOGRAPHIE DE NO₂

Annexe 2 – Zone urbaine : Cherbourg

Annexe 3 – Zone urbaine : La Rochelle

Annexe 4 – Zone urbaine : Le Robert

Annexe 5 – Région : Champagne-Ardenne

Annexe 6 – Région : Centre

Annexe 2 – Urbain :Cherbourg

I. INTRODUCTION	3
I.1. Présentation du jeu de données initial	3
I.2. Méthodologie de sélection des données	4
II. ETUDE DU JEU DE DONNÉES INITIAL ET ÉLABORATION DE LA CARTE DE RÉFÉRENCE.	7
II.1. Etude du variogramme expérimental	7
II.2. Ajustement du variogramme modélisé et test de validation croisée.....	7
• Modèle 1	8
• Modèle 2	8
II.3.Choix du meilleur modèle et établissement de la carte de référence.....	10
• Toutes les données	10
• Toutes les données	10
III. ETUDE DE LA SÉLECTION DE DONNÉES N°1 (MAILLES DE 1 KM)	13
III.1. Etude du variogramme expérimental	13
III.2. Ajustement du variogramme modélisé et test de validation croisée.....	13
• Modèle 1	14
• Modèle 2	14
• Sélection 1 (1 km)	14
III.3. Test de validation absolue menée sur les données complémentaires de la sélection 1	15
• Données complémentaires de la sélection 1	16
• Données complémentaires de la sélection 1	16
III.4.Choix du meilleur modèle et établissement de la carte de la sélection 1	17
• Sélection 1	18
• Sélection 1	18
IV. ETUDE DE LA SÉLECTION DE DONNÉES N°2 (MAILLES DE 2 KM).....	21
IV.1. Etude du variogramme expérimental	21
IV.2. Ajustement du variogramme modélisé et test de validation croisée	21
IV.3. Test de validation absolue mené sur les données complémentaires de la sélection 2	22
IV.4. Etablissement de la carte pour la sélection 2	23
• Sélection 2	24
V COMPARAISON DES CARTES ET DISCUSSION	26

I. INTRODUCTION

I.1. Présentation du jeu de données initial

Le jeu de données étudié a été élaboré par l'association agréée de surveillance AIRCOM. Il est constitué de deux campagnes de mesure, d'une durée de 14 jours chacune, des concentrations en dioxyde d'azote dans l'air ambiant, par échantillonnage passif, sur la communauté urbaine de Cherbourg-Octeville.

Avec 91 717 habitants, la communauté urbaine de Cherbourg regroupe un peu plus de 18% de la population du département de la Manche (Fig. 40). Au sein de cet ensemble, 48% des habitants résident sur le territoire de Cherbourg-Octeville.

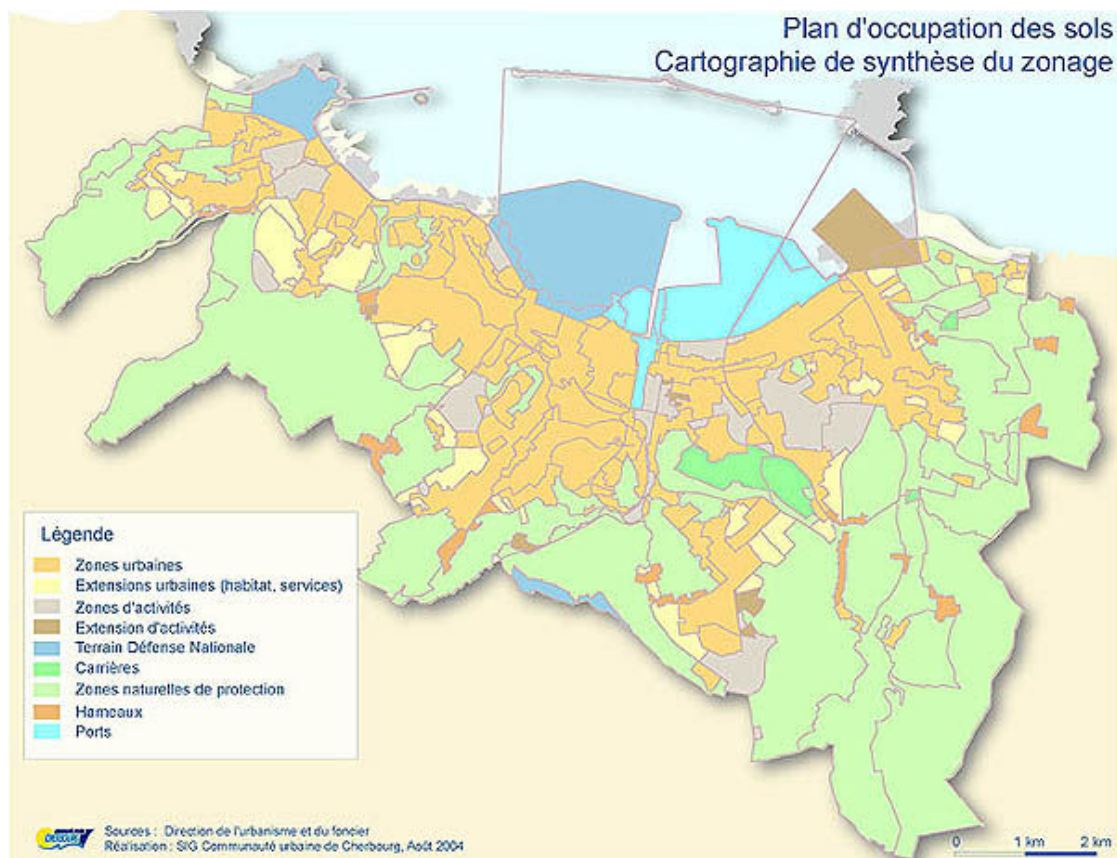


Figure 40 – Communauté Urbaine de Cherbourg-Octeville (plan d'occupation des sols).

88 sites de mesure ont été disposés dans l'agglomération de Cherbourg, sur un domaine d'environ 14 km d'est en ouest et de 10 km du nord au sud (Fig. 41). Les valeurs prises en compte pour cette étude correspondent à la moyenne des deux campagnes ou, lorsque l'une des deux valeurs fait défaut, à la valeur existante.

La stratégie d'échantillonnage suivie pour ces campagnes est celle classiquement adoptée par la plupart des AASQA : un resserrement des points de mesure en centre-ville et une disposition plus espacée en périphérie. Tous les détails sur ces campagnes sont consignés dans le rapport d'AIRCOM [6].

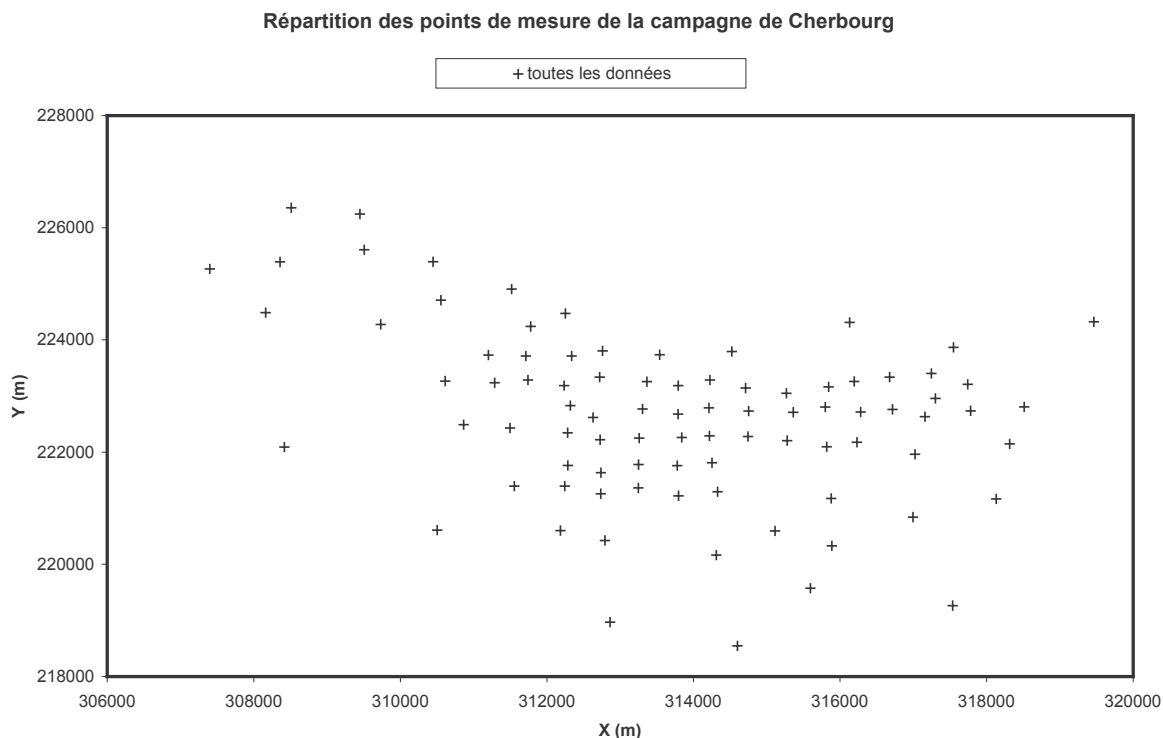


Figure 41 – Carte de répartition des points de mesure de la campagne de Cherbourg.

I.2. Méthodologie de sélection des données

Nous souhaitons échantillonner les données suivant deux maillages réguliers :

- l'un à mailles carrées de 1 km de côté ;
- l'autre à mailles carrées de 2 km de côté.

Pour cela, nous avons calé sur le domaine d'étude un maillage régulier de mailles 1km ou 2km. Puis nous avons sélectionné, pour chaque maille, le point le plus proche du centre de la maille.

Le jeu de données correspondant à une maille régulière de 1 km est appelé « sélection 1 ». Il comprend 43 données (Fig. 42).

Les données non sélectionnées par ce maillage constituent le jeu de données « complémentaires de la sélection 1 », qui comprend 45 données.

Le jeu de données correspondant à une maille régulière de 2 km est appelé « sélection 2 ». Il comprend 21 données (Fig. 43).

Les données non sélectionnées par ce maillage constituent le jeu de données « complémentaires de la sélection 2 », qui comprend 67 données.

L'ensemble de ces sélections est représenté à la Figure 44. Les caractéristiques statistiques des différents jeux de données ainsi constitués sont données au Tableau 9.

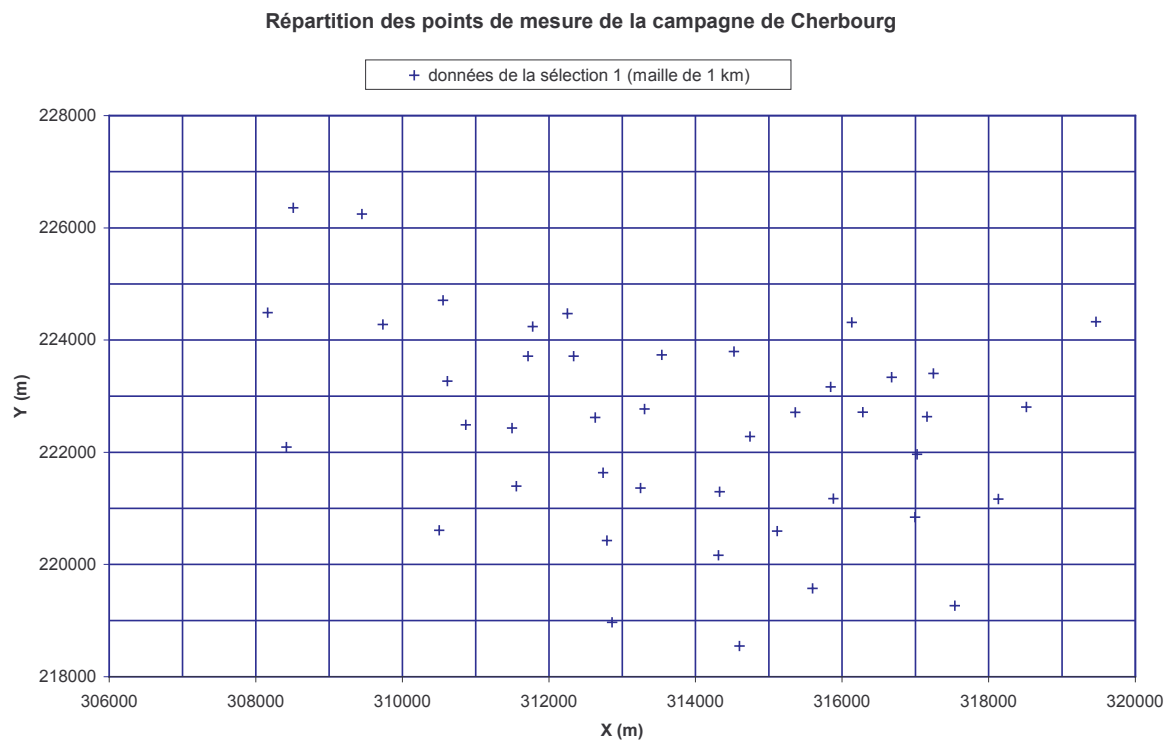


Figure 42 – Carte de répartition des points de la « sélection 1 » (mailles de 1km).

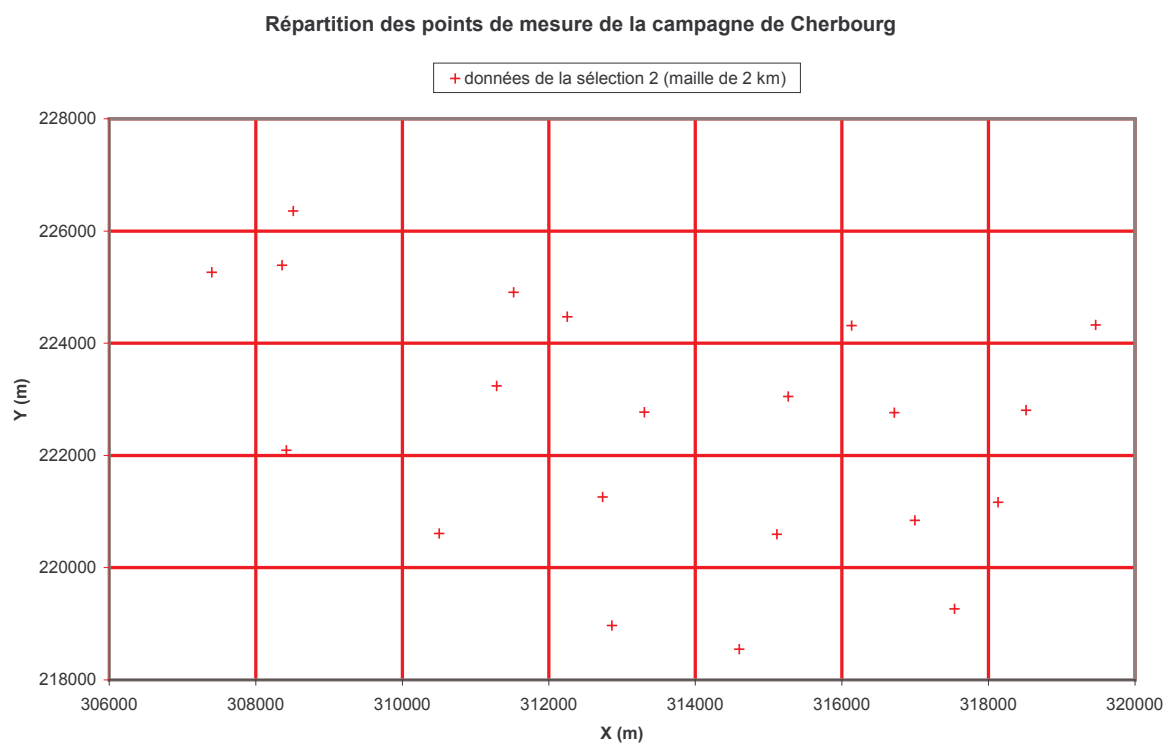


Figure 43 – Carte de répartition des points de mesure de la « sélection 2 » (mailles de 2 km).

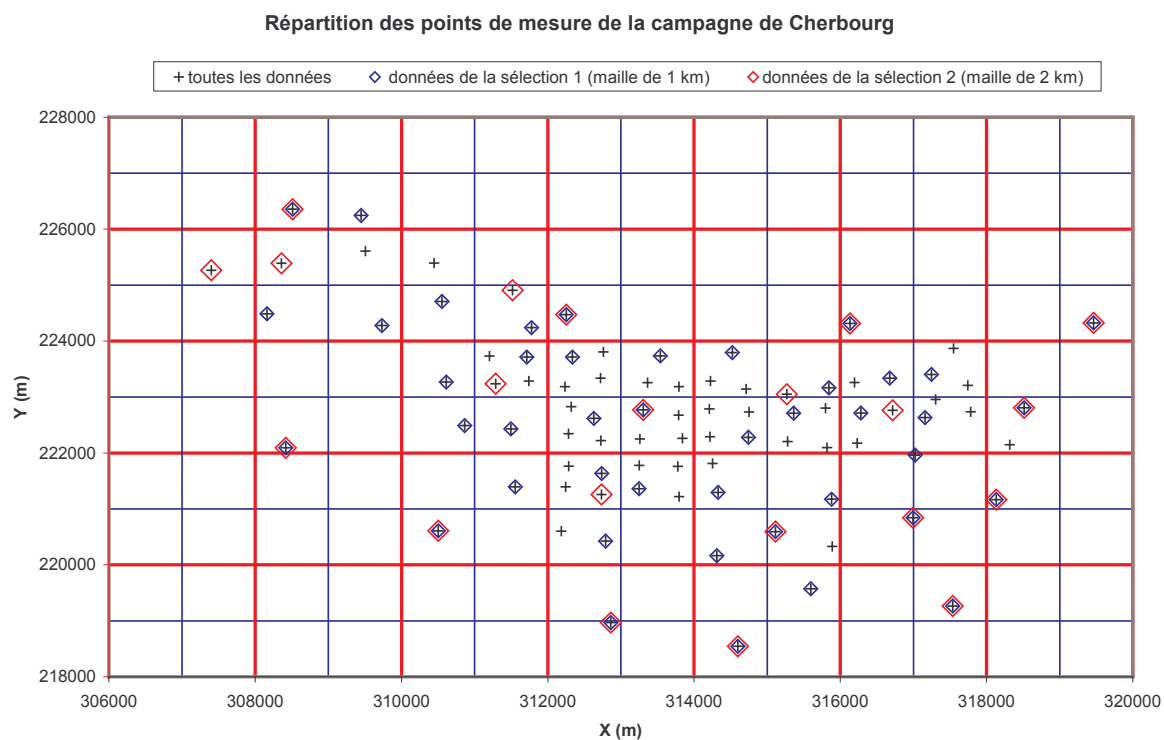


Figure 44 – Carte de répartition des points de mesure échantillonnés selon les différents maillages.

Tableau 9 - Principales caractéristiques des jeux de données étudiés.

	Toutes les données	Sélection 1 (1 donnée par maille de 1km×1km)	Données complémentaires de la sélection 1	Sélection 2 (1 donnée par maille de 2km×2km)	Données complémentaires de la sélection 2
Nombre de données	88	43	45	21	67
Minimum (µg/m3)	6,85	6,85	9,60	7,60	6,85
25 ^{ème} percentile (µg/m3)	12,80	10,50	13,80	11,35	14,05
Médiane (µg/m3)	17,78	15,75	21,05	13,8	20,55
75 ^{ème} percentile (µg/m3)	26,25	24,68	27,40	23,80	27,52
Maximum (µg/m3)	40,80	38,00	40,80	30,25	40,80
Moyenne (µg/m3)	19,81	17,86	21,67	16,51	20,97
Ecart-type (µg/m3)	8,31	7,98	8,20	6,90	8,45
Variance (µg/m3)	69,11	63,70	67,24	47,60	71,40
CV%	0,42	0,45	0,39	0,42	0,40

II. ETUDE DU JEU DE DONNEES INITIAL ET ELABORATION DE LA CARTE DE REFERENCE.

Le but de cette partie est d'établir la carte de référence, c'est-à-dire la carte de répartition des concentrations en dioxyde d'azote établie avec le maximum de points de mesure. Elle servira par la suite à comparer les cartes obtenues avec un nombre de points plus faibles.

II.1. Etude du variogramme expérimental

Le variogramme expérimental retenu est omnidirectionnel, avec 12 pas de 390 m chacun. La nuée variographique ne présentant pas d'anomalie, aucune donnée n'a été écartée (Fig. 45).

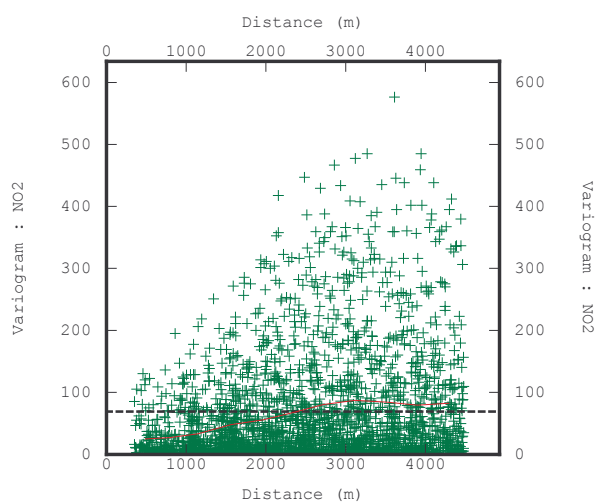


Figure 45 – Variogramme expérimental (pas de 390 m, 12 pas) avec toutes les données.

II.2. Ajustement du variogramme modélisé et test de validation croisée

Deux modèles différents ont été ajustés aux données expérimentales. Ils sont composés tous deux d'une structure pépitique (de même palier=17) additionnée d'une structure gaussienne ou cubique.

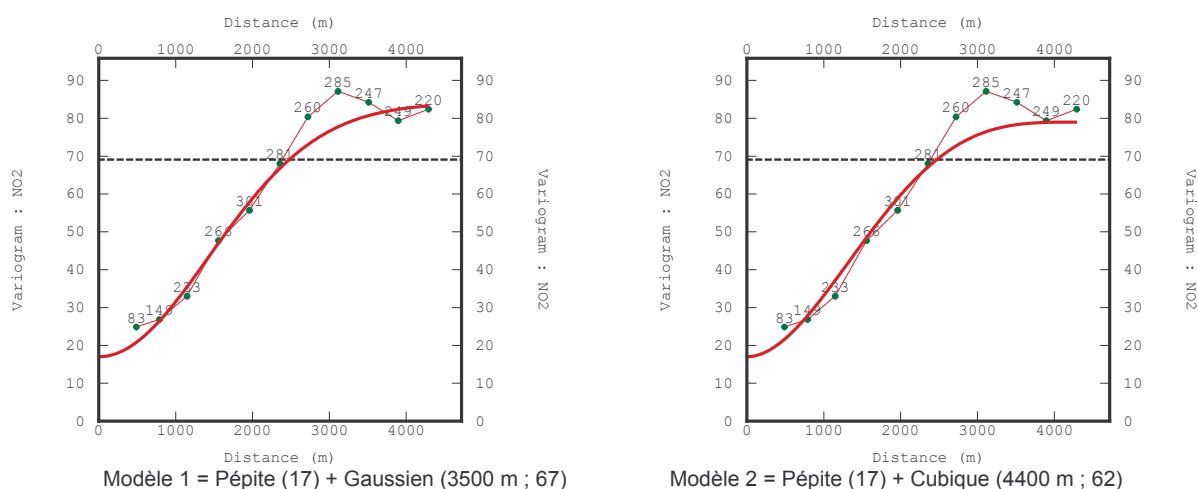


Figure 46 – Variogrammes expérimentaux et modélisés.

Afin d'apprécier leur qualité respective, chacun d'eux est soumis au test de validation croisée. Ce test consiste à calculer, au moyen du modèle, une estimation de la concentration en polluant, pour chaque point de l'espace où celle-ci a été mesurée (pour cela, on ôte momentanément la valeur mesurée du jeu de données). Il faut préalablement définir un voisinage, délimitant de nombre de données mesurées prises en compte pour le calcul d'estimation.

En raison du faible nombre de points de mesure (< 100), c'est le voisinage unique qui a été considéré, c'est-à-dire que, pour chaque estimation, c'est l'ensemble des points de mesure qui est pris en compte. Ce voisinage unique sera celui utilisé dans l'ensemble de l'étude.

Dans le test de validation croisée, les valeurs estimées par le modèle sont comparées aux valeurs mesurées (valeurs vraies), notamment en calculant l'erreur d'estimation. Sous ISATIS, l'erreur d'estimation est définie par la différence entre la valeur estimée moins la valeur vraie.

Tableau 10 – Statistiques de l'erreur d'estimation par validation croisée sur toutes les données

(a) avec le modèle 1 = Pépité (17) + Gaussien (3500 m ; 67)

<i>MODELE 1</i>	moyenne	variance
Erreur	0,13066	19,86196
Erreur standardisée	0,00979	0,91117

1 valeur non robuste sur 88 valeurs.

(b) avec le modèle 2 = Pépité (17) + Cubique (4400 m ; 62).

<i>MODELE 2</i>	moyenne	variance
Erreur	0,14896	19,79324
Erreur standardisée	0,01120	0,84959

0 valeur non robuste sur 88 valeurs.

Les caractéristiques statistiques des erreurs d'estimation (Tab. 10) montrent que les deux modèles conduisent à des résultats semblables et assez satisfaisants : l'erreur moyenne est proche de zéro et la variance de l'erreur standardisée est voisine de 1. Le nombre de valeurs mal estimées (non robustes), pour lesquelles l'écart-type de l'erreur d'estimation standardisée dépasse 2,5 en valeur absolue, est limité (0 ou 1).

Tableau 11 – Résultats de la validation croisée sur toutes les données pour le krigeage ordinaire ponctuel avec le modèle 1 = Pépité (17) + Gaussien (3500 m ; 67).

		Validation croisée par krigeage ordinaire ponctuel selon le modèle 1	
Toutes les données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation standardisée
Nombre de données	88	88	88
Minimum	6,85	8,02	-2,51
25 ^{ème} percentile	12,80	13,92	-0,45
Médiane	17,775	18,51	0,19
75 ^{ème} percentile)	26,25	25,43	0,60
Maximum	40,80	34,44	1,94
Etendue (= Max – Min)	33,95	26,42	4,45
Moyenne)	19,81	19,94	0,01
Ecart-type	8,31	7,03	0,95
Variance	69,11	49,42	0,91
CV %	0,42	0,35	
Corrélation mesures/estimations		r=0,8442	

Tableau 12 – Résultats de la validation croisée sur toutes les données pour le krigeage ordinaire ponctuel avec le modèle 2 = Pépité (17) + Cubique (4400 m ; 62).

		Validation croisée par krigeage ordinaire ponctuel selon le modèle 2	
Toutes les données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation standardisée
Nombre de données	88	88	88
Minimum	6,85	8,20	-2,43
25 ^{ème} percentile	12,80	13,94	-0,41
Médiane	17,775	18,65	0,21
75 ^{ème} percentile	26,25	25,36	0,60
Maximum	40,80	34,76	1,86
Etendue (= Max – Min)	33,95	26,56	4,29
Moyenne	19,81	19,96	0,01
Ecart-type	8,31	6,94	0,92
Variance	69,11	48,13	0,85
CV %	0,42	0,35	
Corrélation mesures/estimations		r=0,8448	

Comme le montrent les tableaux 11 et 12 et la figure 47 suivante, les valeurs estimées sont bien corrélées aux valeurs mesurées, avec un coefficient de corrélation supérieur à 0,84 pour les deux modèles.

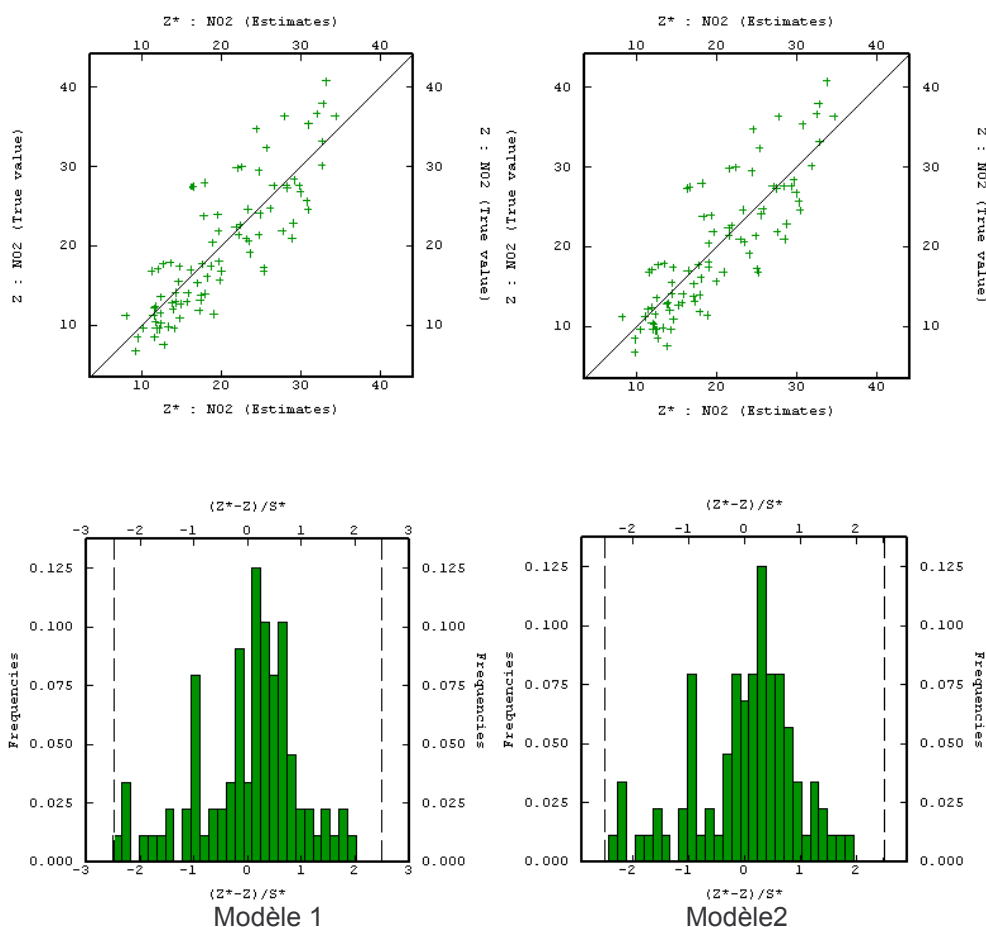


Figure 47 – Résultats de validation croisée pour les modèles 1 et 2.

II.3.Choix du meilleur modèle et établissement de la carte de référence

Bien que les différences entre les deux modèles soient minimales, nous orientons notre choix vers le modèle 2. En effet, celui-ci conduit à des valeurs estimées plus proches des mesures et à une erreur d'estimation plus resserrée (variance de l'écart-type d'erreur plus faible).

Pour dresser la carte de référence de l'étude, c'est-à-dire la carte réalisée à partir du plus grand nombre de mesures disponibles, nous estimons les valeurs de concentration en polluant aux nœuds d'une grille régulière en 2D de 250 m de côté.

Deux techniques de krigeage ordinaire sont utilisées, afin d'être comparées : le krigeage ponctuel et le krigeage de bloc.

Pour le krigeage ponctuel, les caractéristiques de la grille d'estimation sont :

Xmin=307000 m, 57 nœuds, Xmax=321000, $\Delta X=14000$ m,
Ymin=218000 m, 41 nœuds, Ymax=228000, $\Delta X=10000$ m.

Sous ISATIS, lors du krigeage de bloc, les valeurs sont estimées aux nœuds d'une grille correspondant aux centres des blocs. Pour retrouver exactement le même domaine d'estimation que pour le krigeage ponctuel, il faut donc définir une nouvelle grille, décalée d'un demi-bloc (d'une demi-maille).

Pour le krigeage de bloc, les caractéristiques de la grille d'estimation sont ainsi :

Xmin=307125 m, 56 nœuds, Xmax=320875, $\Delta X=13750$ m,
Ymin=218125 m, 40 nœuds, Ymax=227875, $\Delta X=9750$ m.

Tableau 13 – Estimation sur une grille régulière de 250 m de côté, par krigeage ordinaire, ponctuel ou de bloc, selon le modèle 2, pour toutes les données.

<i>TOUTES LES DONNEES</i>	Mesures ($\mu\text{g}/\text{m}^3$)	Estimation par krigeage ordinaire	
		Valeurs estimées modèle 2, krigeage ponctuel ($\mu\text{g}/\text{m}^3$)	Valeurs estimées modèle 2, krigeage de bloc ($\mu\text{g}/\text{m}^3$)
Nombre de données	88	2337	2240
Minimum	6,85	7,26	7,29
25 ^{ème} percentile	12,80	12,85	12,80
Médiane	17,775	15,34	15,32
75 ^{ème} percentile	26,25	17,53	17,76
Maximum	40,80	35,73	35,62
Moyenne	19,81	16,02	16,08
Ecart-type	8,31	5,21	5,30
Variance	69,11	27,14	28,08
CV %	0,42	0,33	0,33

Tableau 14 – Ecart-types de krigeage, sur une grille régulière de 250 m de côté, par krigeage ordinaire, ponctuel ou de bloc, selon le modèle 2, pour toutes les données.

<i>TOUTES LES DONNEES</i>	Estimation par krigeage ordinaire	
	Ecart-types de krigeage modèle 2, krigeage ponctuel ($\mu\text{g}/\text{m}^3$)	Ecart-types de krigeage modèle 2, krigeage de bloc ($\mu\text{g}/\text{m}^3$)
Nombre de données	2337	2240
Minimum	4,42	1,57
Médiane	6,51	4,79
Maximum	9,25	8,25
Moyenne	6,77	5,04
Ecart-type	1,70	2,23
Variance	2,88	4,98
CV %	0,25	0,44

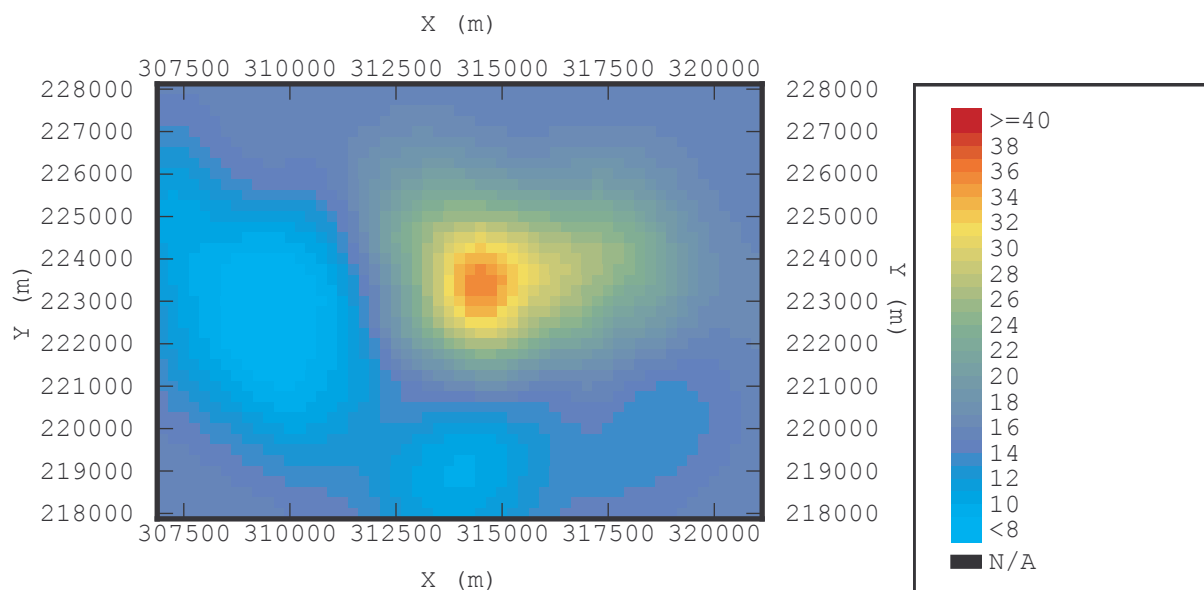


Figure 48 – Estimation par krigeage ponctuel, à partir de toutes les données, avec le modèle 2, sur une grille régulière de 250 m de côté.

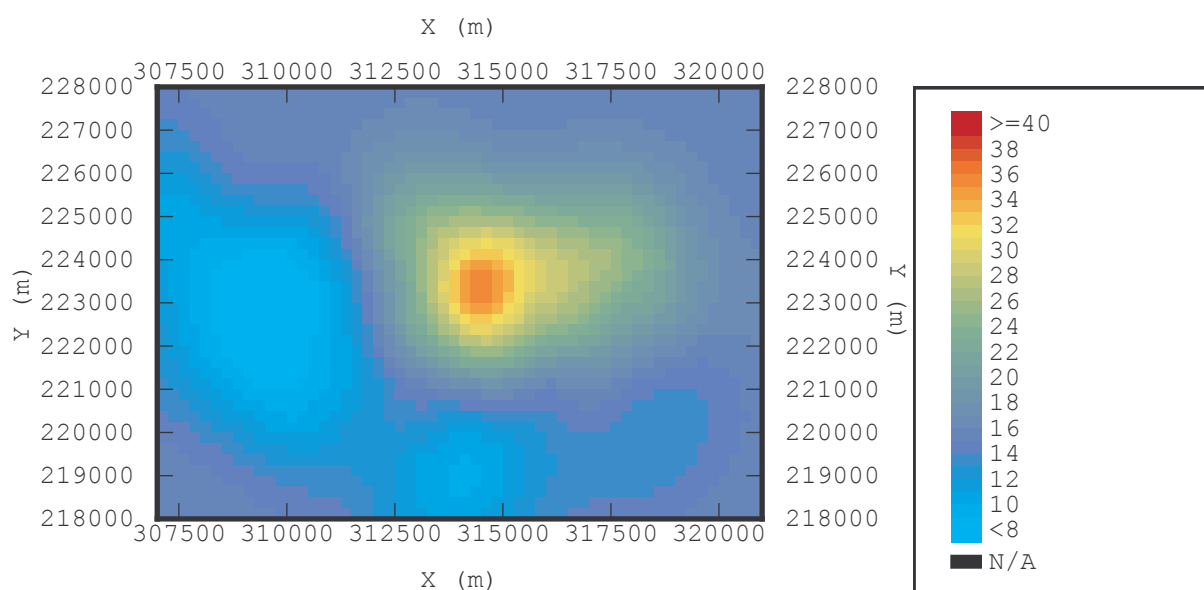


Figure 49 – Estimation par krigeage de bloc, à partir de toutes les données, avec le modèle 2, sur une grille régulière de 250 m de côté.

Les cartes des valeurs estimées (Fig. 48 et 49) montrent que la différence entre krigeage ponctuel et krigeage de bloc est, dans ce cas, peu perceptible : les niveaux de concentrations obtenus sont les mêmes et les contours des zones de pollution sont semblables.

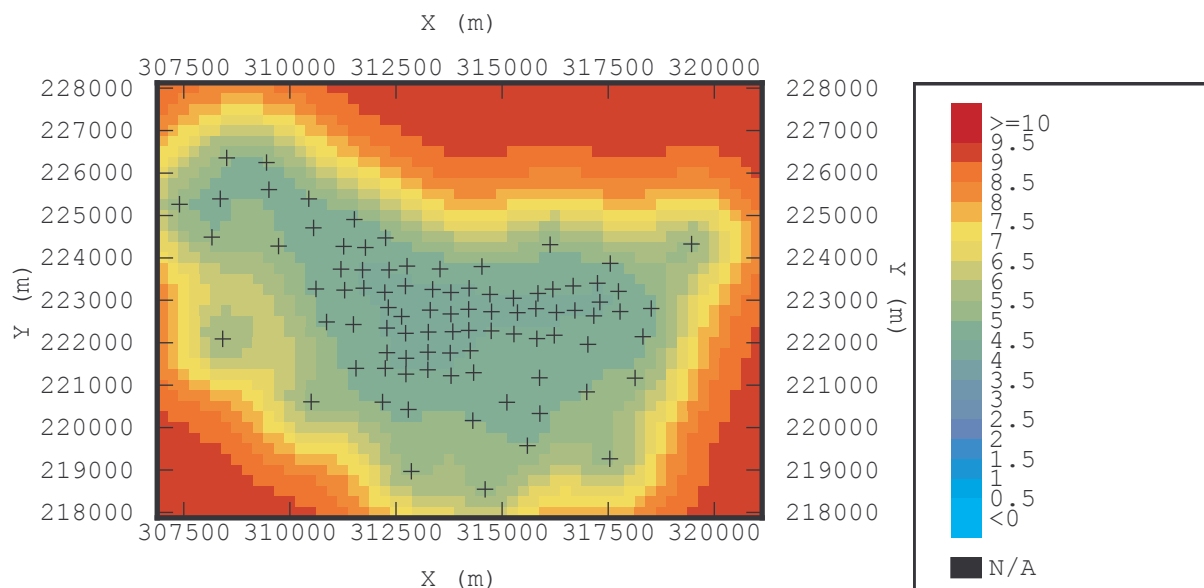


Figure 50 – Ecart-type de krigeage, par krigeage ponctuel, à partir de toutes les données, avec le modèle 2, sur une grille régulière de 250 m de côté.

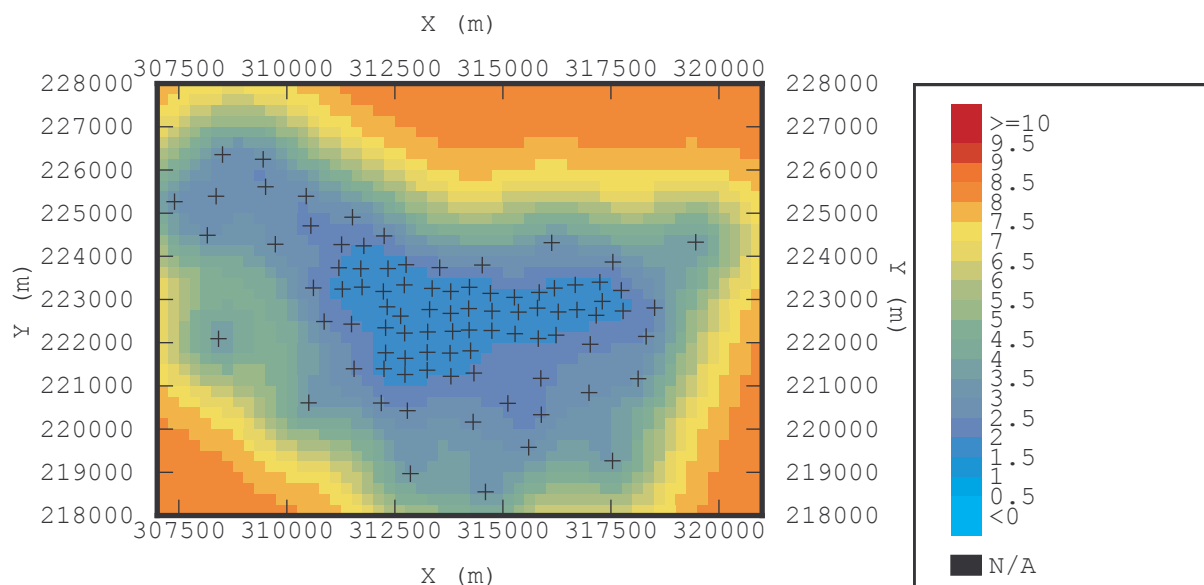


Figure 51 – Ecart-type de krigeage, par krigeage de bloc, à partir de toutes les données, avec le modèle 2, sur une grille régulière de 250 m de côté.

La différence entre krigeage ponctuel et krigeage de bloc est bien plus marquée lorsque l'on examine les cartes d'écarts-type de krigeage (Fig. 50 et 51) : les niveaux des écarts-type de krigeage sont bien plus faibles dans le krigeage de bloc ($5,0 \mu\text{g}/\text{m}^3$ en moyenne) que dans le krigeage ponctuel ($6,8 \mu\text{g}/\text{m}^3$ en moyenne). Ceci reflète le fait que la quantité à estimer est moins variable, donc l'estimation plus précise.

III. ETUDE DE LA SELECTION DE DONNEES N°1 (MAILLES DE 1 KM)

La sélection n°1 comprend 43 données, qui correspondent à 1 point de mesure par maille carrée de 1km².

III.1. Etude du variogramme expérimental

Le variogramme expérimental retenu est omnidirectionnel, avec 13 pas de 368 m chacun.

Aucune donnée n'a été écartée.

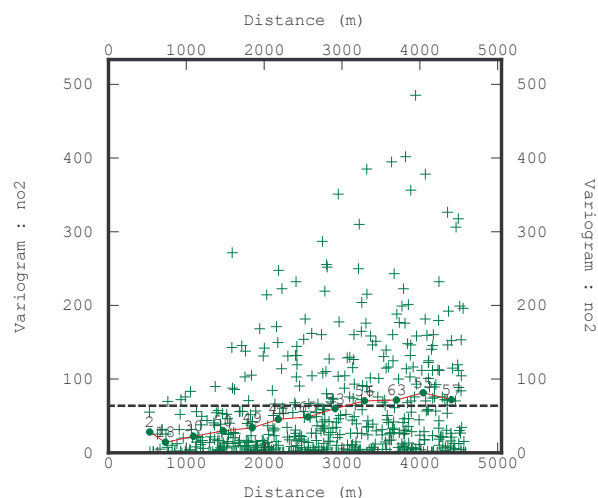
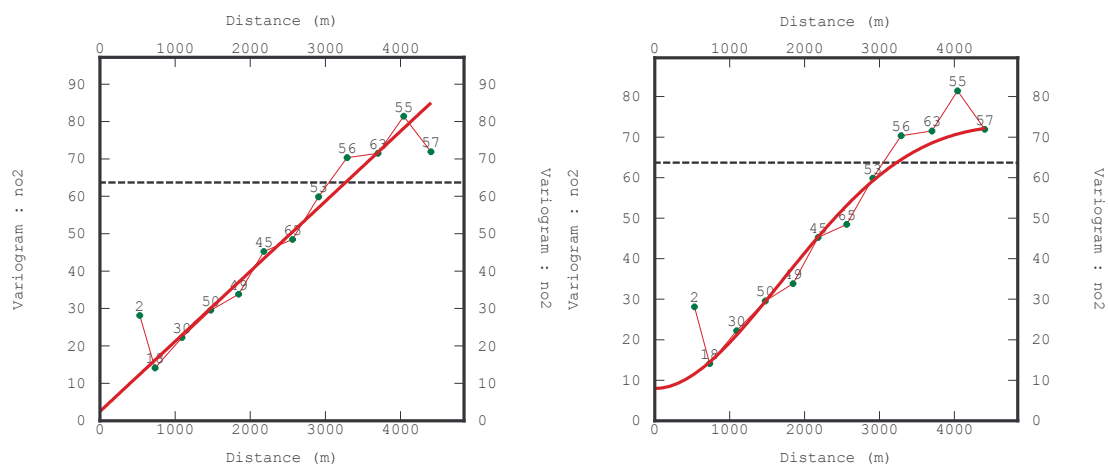


Figure 52 – Variogramme expérimental (pas de 368 m, 13 pas) des données de la sélection 1.

III.2. Ajustement du variogramme modélisé et test de validation croisée

Deux modèles différents ont été ajustés aux données expérimentales. Ils sont composés tous deux d'une structure pépétique additionnée d'une structure linéaire ou cubique.



Modèle 1 = pépète(2,5) + linéaire (2270 m ; 42,5)

Modèle 2 = pépète(8) + cubique (5600 m ; 65)

Figure 53 – Variogrammes modélisés pour la sélection 1.

**Tableau 15 – Statistiques sur l'erreur d'estimation obtenues par validation croisée
(a) avec le modèle 1.**

<i>MODELE 1</i>	moyenne	variance
Erreur	0,11450	17,70800
Erreur standardisée	0,00910	0,88609

0 valeur non robuste sur 43 données.

(b) avec le modèle 2.

<i>MODELE 2</i>	moyenne	variance
Erreur	0,22715	17,73639
Erreur standardisée	0,02043	1,15679

1 valeur non robuste sur 43 données.

Tableau 16 – Résultats de la validation croisée sur toutes les données pour le krigeage ordinaire ponctuel avec le modèle 1= pépité (2,5) + linéaire (2270 m ; 42,5).

		Validation croisée par krigeage ordinaire ponctuel selon le modèle 1	
<i>SELECTION 1 (1 KM)</i>	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation standardisée
Nombre de données	43	43	43
Minimum	6,85	7,13	-2,31
25 ^{ème} percentile	10,50	11,82	-0,47
Médiane	15,75	17,79	0,02
75 ^{ème} percentile	24,68	24,26	0,67
Maximum	38,00	30,81	2,15
Etendue (= Max – Min)	31,15	23,68	4,46
Moyenne	17,86	17,97	0,01
Ecart-type	7,98	7,16	0,94
Variance	63,70	51,29	0,89
CV %	0,45	0,40	
Corrélation mesures/estimations		r=0,8510	

Tableau 17 – Résultats de la validation croisée sur toutes les données pour le krigeage ordinaire ponctuel avec le modèle 2 = pépité (8) + Cubique (5600 m ; 65).

		Validation croisée par krigeage ordinaire ponctuel selon le modèle 2	
<i>Sélection 1 (1 km)</i>	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation standardisée
Nombre de données	43	43	43
Minimum	6,85	7,01	-2,47
25 ^{ème} percentile	10,50	12,28	-0,61
Médiane	15,75	17,66	0,22
75 ^{ème} percentile	24,68	23,51	0,60
Maximum	38,00	31,12	2,50
Etendue (= Max – Min)	31,15	24,11	4,97
Moyenne	17,86	18,09	0,02
Ecart-type	7,98	6,95	1,08
Variance	63,70	48,25	1,16
CV %	0,45	0,38	
Corrélation mesures/estimations		r=0,8497	

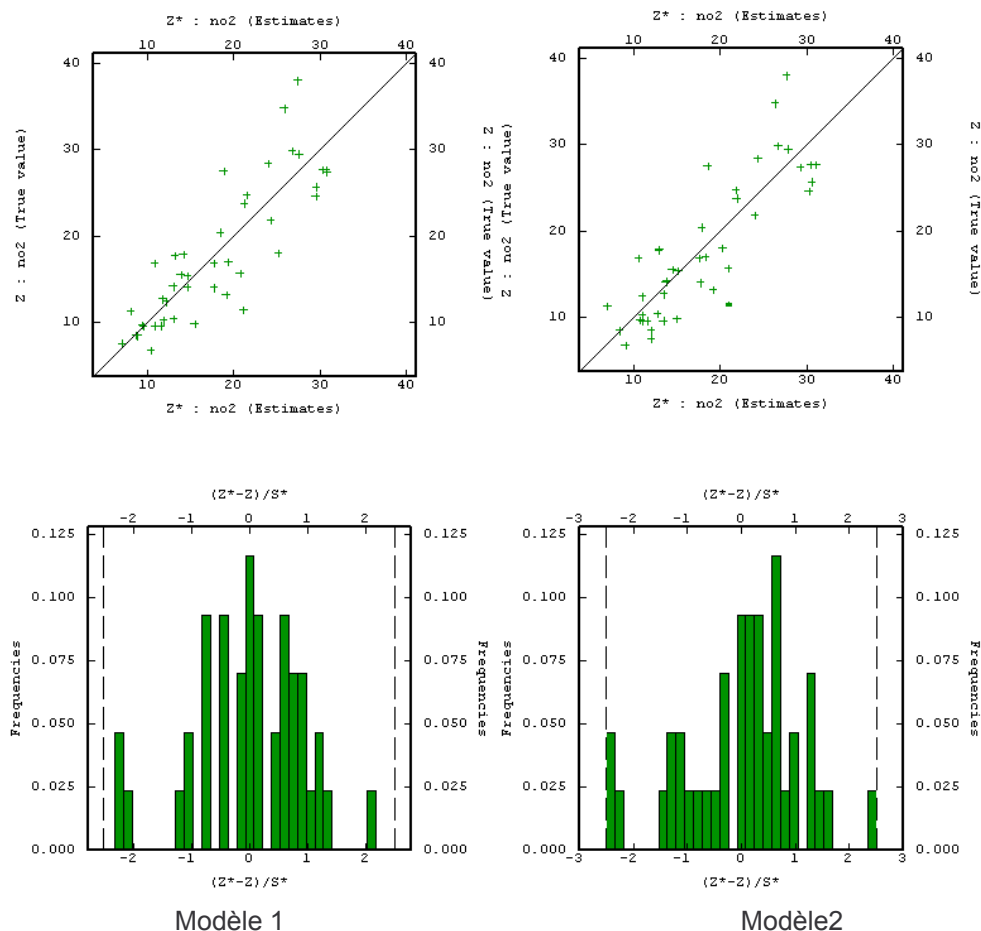


Figure 54 – Résultats de validation croisée pour la sélection 1.

Les résultats de validation croisée montrent que les deux modèles conviennent. Le modèle 1 conduit à des estimations un peu mieux corrélées aux valeurs vraies et à des erreurs d'estimation d'étendue plus petite.

III.3. Test de validation absolue menée sur les données complémentaires de la sélection 1

Le test de validation absolue consiste à effectuer des estimations en des points où les valeurs vraies sont connues, mais sans que ces valeurs aient été préalablement utilisées dans la procédure d'analyse variographique. Il diffère donc du test de validation croisée, où les mêmes valeurs servent à la fois pour l'étude variographique et pour la validation.

Le test de validation absolue est donc ici effectué en estimant les valeurs de concentration, aux points de mesure du jeu de données complémentaires de la sélection n°1 (cf. I.2.). Ces données complémentaires n'ont, en effet, pas été utilisées pour l'étude variographique.

Les informations fournies pour effectuer la validation absolue sont donc :

- le modèle de variogramme,
- les mesures aux points de la sélection 1,
- les coordonnées (X, Y) des points complémentaires de la sélection 1.

Les estimations sont ensuite comparées aux mesures, en calculant notamment l'erreur d'estimation.

Tableau 18 – Résultats de la validation absolue pour le krigeage ordinaire ponctuel avec le modèle 1= pépité(2,5) + linéaire (2270 m ; 42,5).

		Validation absolue par krigeage ordinaire ponctuel selon le modèle 1	
<i>DONNEES COMPLEMENTAIRES DE LA SELECTION 1</i>	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation ($\mu\text{g}/\text{m}^3$)
Nombre de données	45	45	45
Minimum	9,60	11,06	-12,0028
25 ^{ème} percentile	13,80	14,12	-3,9500
Médiane	21,05	20,63	-0,0328
75 ^{ème} percentile	27,40	27,16	2,1734
Maximum	40,80	31,31	12,2417
Moyenne	21,67	21,00	-0,6710
Ecart-type	8,20	6,55	5,2778
Variance	67,24	42,83	27,8547
CV %	0,39	0,31	
Corrélation mesures/estimations		r=0,7658	

Tableau 19 – Résultats de la validation absolue pour le krigeage ordinaire ponctuel avec le modèle 2 = pépité(8) + cubique (5600 m ; 65).

		Validation absolue par krigeage ordinaire ponctuel selon le modèle 2	
<i>DONNEES COMPLEMENTAIRES DE LA SELECTION 1</i>	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation ($\mu\text{g}/\text{m}^3$)
Nombre de données	45	45	45
Minimum	9,60	10,80	-11,5603
25 ^{ème} percentile	13,80	14,09	-3,9272
Médiane	21,05	20,70	-0,1733
75 ^{ème} percentile	27,40	26,22	2,5367
Maximum	40,80	31,09	11,6167
Moyenne	21,67	21,04	-0,6265
Ecart-type	8,20	6,45	5,2406
Variance	67,24	41,59	27,4642
CV %	0,39	0,31	
Corrélation mesures/estimations		r=0,7691	

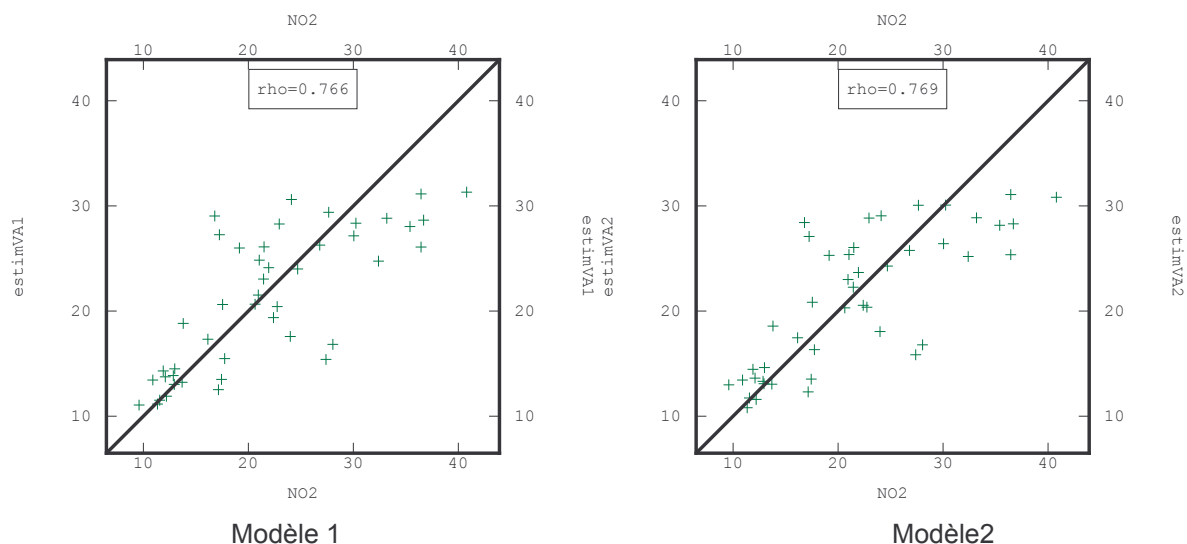


Figure 55 – Test de validation absolue : valeurs estimées en fonction des valeurs vraies.

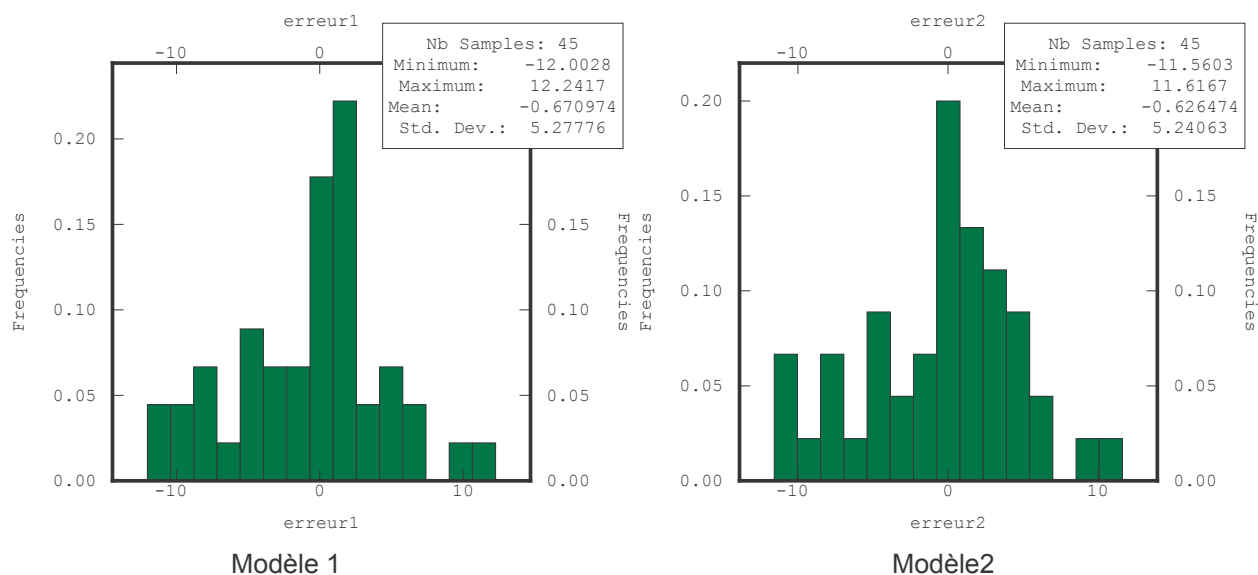


Figure 56 – Test de validation absolue : distribution des erreurs d'estimation.

III.4.Choix du meilleur modèle et établissement de la carte de la sélection 1

Contrairement à ce que pouvait laisser penser le test de validation croisée, la validation absolue montre que c'est plutôt le modèle 2 (à structures pépétique et cubique) qui estime le mieux les données. En effet, les valeurs estimées par le modèle 2 sont mieux corrélées (Fig. 55) et la distribution des erreurs est plus étroite et mieux centrée autour de zéro (Fig. 56).

Pour dresser la carte de la sélection1, nous estimons les valeurs de concentration en polluant par krigeage, soit ponctuel soit de bloc, en utilisant ce modèle et les mêmes grilles que pour les cartes de référence (cf. II.3.).

Tableau 20 – Estimation sur une grille régulière de 250 m de côté, par krigeage ordinaire ponctuel ou de bloc, selon le modèle 2, à partir des données de la sélection 1.

<i>SELECTION 1</i>	Mesures ($\mu\text{g}/\text{m}^3$)	Estimation par krigeage ordinaire	
		Valeurs estimées modèle 2, krigeage ponctuel ($\mu\text{g}/\text{m}^3$)	Valeurs estimées modèle 2, krigeage de bloc ($\mu\text{g}/\text{m}^3$)
Nombre de données	43	2337	2240
Minimum	6,85	6,90	6,92
25 ^{ème} percentile	10,50	12,05	11,97
Médiane	15,75	14,92	14,90
75 ^{ème} percentile	24,68	18,66	19,03
Maximum	38,00	33,40	33,38
Moyenne	17,86	16,10	16,17
Ecart-type	7,98	5,83	5,93
Variance	63,70	34,02	35,17
CV %	0,45	0,36	0,37

Tableau 21 – Ecart-types de krigeage, sur une grille régulière de 250 m de côté, par krigeage ordinaire ponctuel ou de bloc, selon le modèle 2, pour les données de la sélection 1.

<i>SELECTION 1</i>	Estimation par krigeage ordinaire	
	Ecart-types de krigeage modèle 2, krigeage ponctuel ($\mu\text{g}/\text{m}^3$)	Ecart-types de krigeage modèle 2, krigeage de bloc ($\mu\text{g}/\text{m}^3$)
Nombre de données	2337	2240
Minimum	3,19	1,48
25 ^{ème} percentile	3,73	2,36
Médiane	4,95	3,85
75 ^{ème} percentile	7,26	6,45
Maximum	9,04	8,56
Moyenne	5,52	4,44
Ecart-type	1,93	2,26
Variance	3,72	5,10
CV %	0,35	0,51

Comme précédemment, la différence entre krigeage ponctuel et krigeage de bloc est peu perceptible en ce qui concerne les valeurs estimées (Tab. 20 et Fig. 57 et 58), mais bien plus nette en ce qui concerne les écarts-type de krigeage (Tab. 21 et Fig. 59 et 60) : le krigeage de bloc conduit à des écarts-type de krigeage plus faibles et constitue donc la meilleure technique des deux.

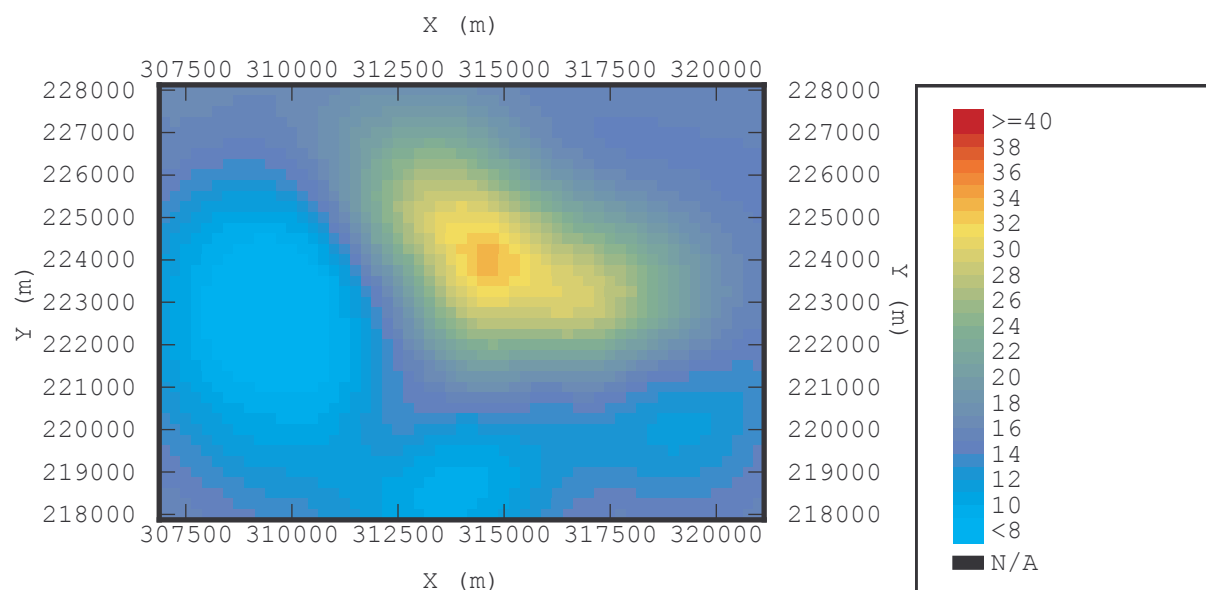


Figure 57 – Estimation par krigeage punctuel, à partir des données de la sélection 1, avec le modèle 2, sur une grille régulière de 250 m de côté

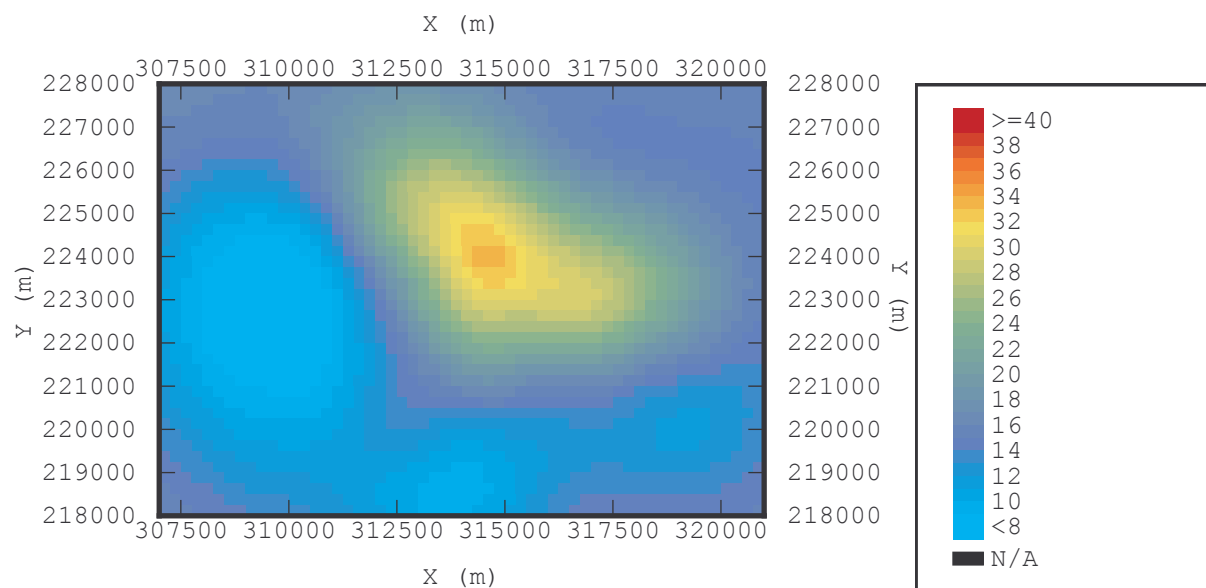


Figure 58 – Estimation par krigeage de bloc, à partir des données de la sélection 1, avec le modèle 2, sur une grille régulière de 250 m de côté

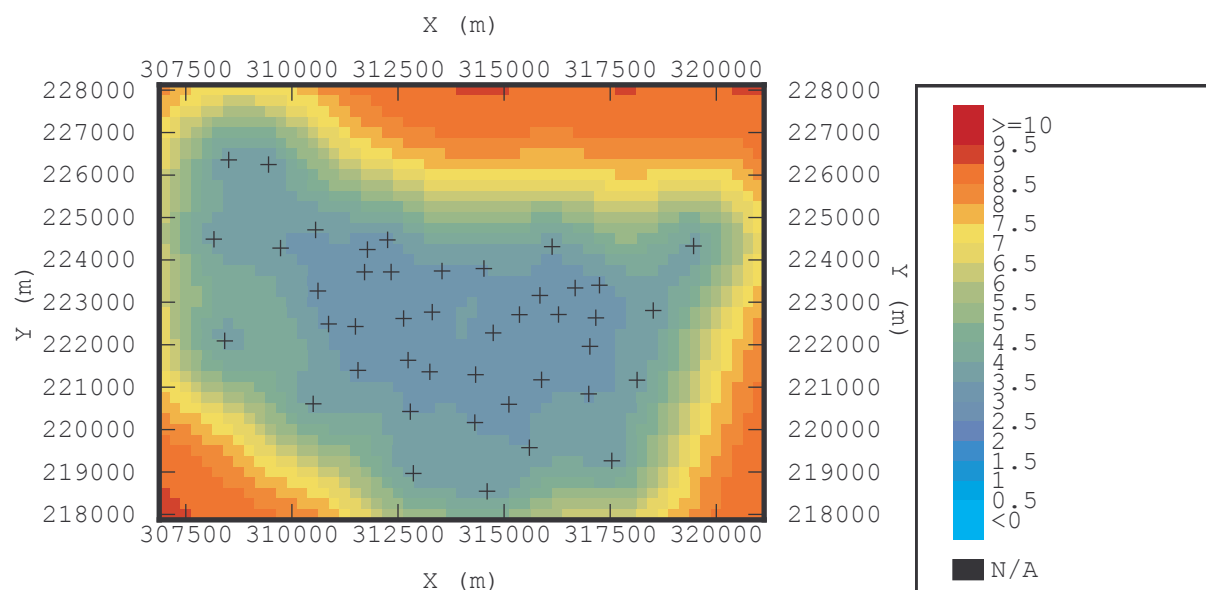


Figure 59 – Ecarts-type de krigeage punctuel, à partir des données de la sélection 1, avec le modèle 2, sur une grille régulière de 250 m de côté

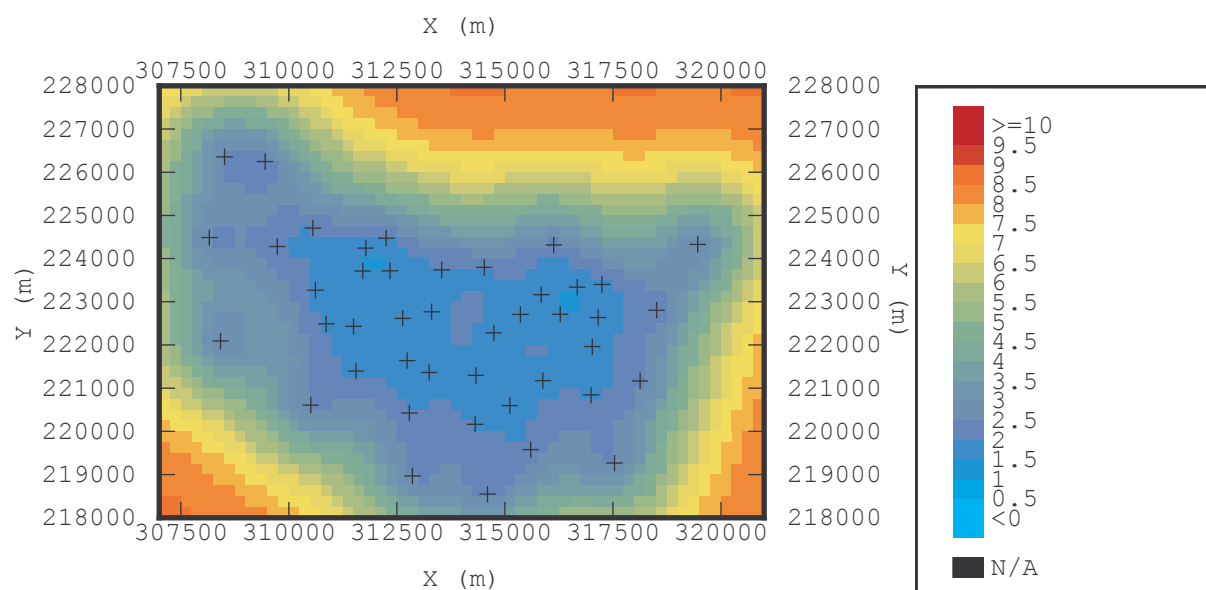


Figure 60 – Ecarts-type de krigeage de bloc, à partir des données de la sélection 1, avec le modèle 2, sur une grille régulière de 250 m de côté

IV. ETUDE DE LA SELECTION DE DONNEES N°2 (MAILLES DE 2 KM)

Sélection des données n°2 : 1 donnée par maille carrée de 2km×2km. Cette sélection comprend 21 données.

IV.1. Etude du variogramme expérimental

Le variogramme expérimental retenu est omnidirectionnel, avec 9 pas de 800 m chacun.

Aucune donnée n'a été écartée.

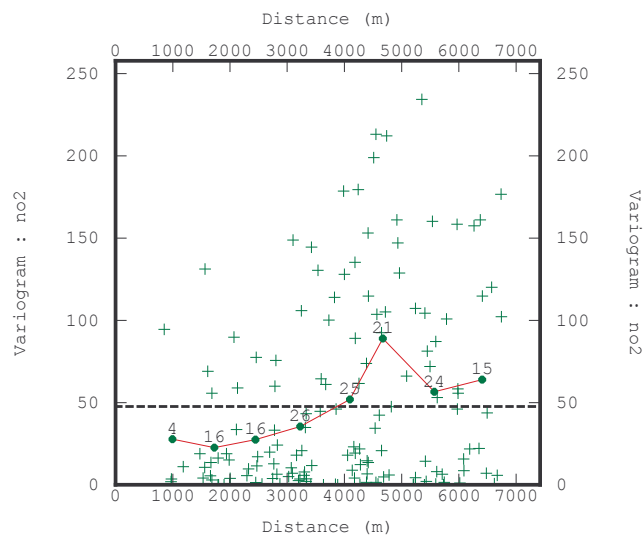


Figure 61 – Variogramme expérimental (pas de 800 m, 9 pas) des données de la sélection 2.

IV.2. Ajustement du variogramme modélisé et test de validation croisée

Le modèle de variogramme qui semble le mieux convenir pour la sélection 2 est :
Modèle = pépite (12) + cubique (10000 m ; 55)

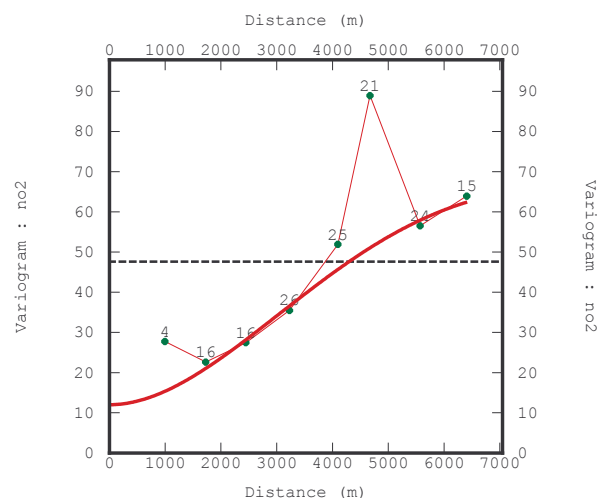


Figure 62 – Variogramme modélisé pour la sélection 2.

Tableau 22 – Statistiques sur l'erreur d'estimation obtenues par validation croisée pour la sélection 2.

	moyenne	variance
Erreur	0,13388	14,08863
Erreur standardisée	0,01407	0,71423

0 valeur non robuste sur 21 données.

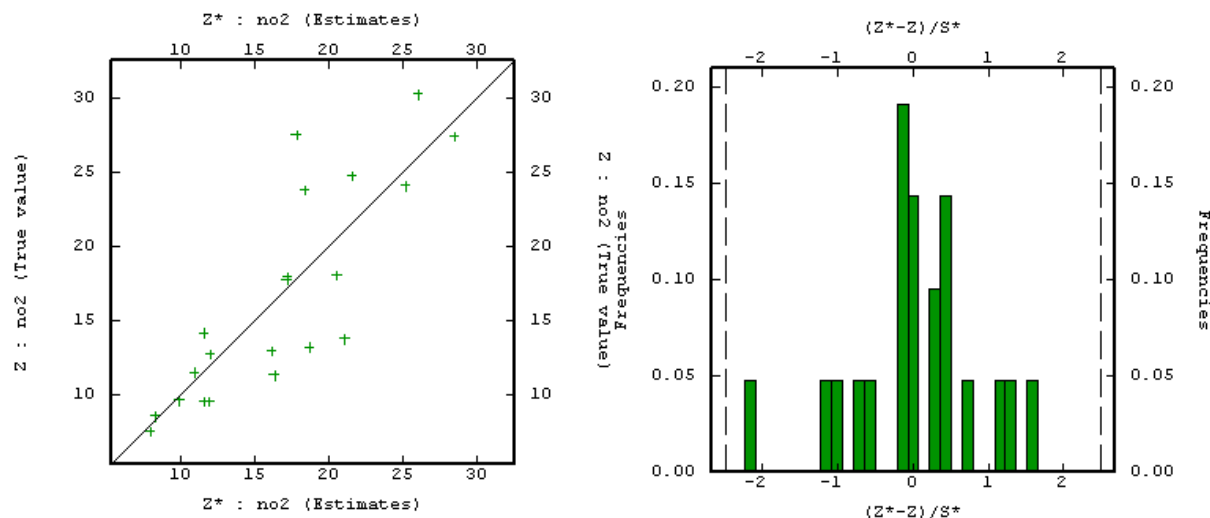


Fig. 63 – Test de validation croisée pour la sélection 2.

IV.3. Test de validation absolue mené sur les données complémentaires de la sélection 2

Comme précédemment, le test de validation absolue est effectué par rapport aux données complémentaires de la sélection n°2, non utilisées dans l'analyse variographique.

Tableau 23 – Résultats de la validation absolue pour le krigeage ordinaire ponctuel avec le modèle 2 = pépite(12) + cubique (10000 m ; 55).

Données complémentaires de la sélection 2	Mesures ($\mu\text{g}/\text{m}^3$)	Validation absolue par krigeage ordinaire ponctuel	
		Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation ($\mu\text{g}/\text{m}^3$)
Nombre de données	66	66	66
Minimum	6,85	9,65	-13,84
25 ^{ème} percentile	14,05	16,22	-4,35
Médiane	20,55	20,33	0,15
75 ^{ème} percentile	27,53	24,42	2,52
Maximum	40,80	28,33	8,54
Moyenne	20,97	20,03	-0,88
Ecart-type	8,45	5,09	5,08
Variance	71,37	25,93	25,83
CV %	0,40	0,25	
Corrélation mesures/estimations		r=0,8300	

Le test de validation absolue montre que le modèle choisi permet de bien estimer les données, puisque le coefficient de corrélation entre estimations et valeurs vraies est de 0,83.

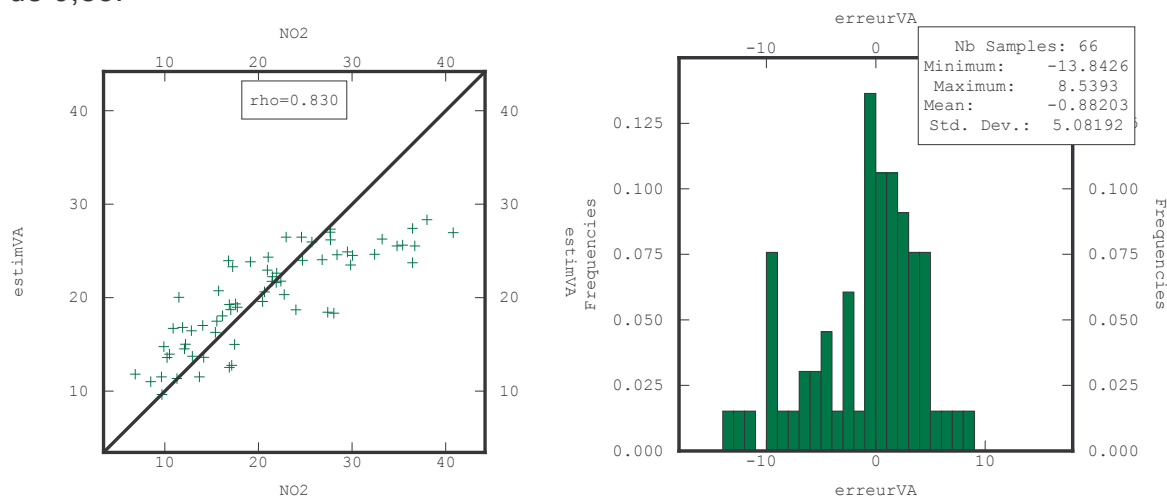


Figure 64 – Résultats de la validation absolue : estimations en fonction des valeurs vraies (à gauche) et distribution des erreurs d'estimation (à droite).

IV.4. Etablissement de la carte pour la sélection 2

La carte de répartition des concentrations en NO₂ est établie à partir du modèle testé précédemment et de la sélection 2 (21 données), sur les mêmes grilles de 250 m que pour les cartes de référence, soit par krigeage ponctuel, soit par krigeage de bloc.

Tableau 24 – Estimation sur une grille régulière de 250 m de côté, par krigeage ordinaire ponctuel ou de bloc, à partir des données de la sélection 2.

Sélection 2	Mesures ($\mu\text{g}/\text{m}^3$)	Estimation par krigeage ordinaire	
		Valeurs estimées krigeage ponctuel ($\mu\text{g}/\text{m}^3$)	Valeurs estimées krigeage de bloc ($\mu\text{g}/\text{m}^3$)
Nombre de données	21	2337	2240
Minimum	7,60	7,65	7,65
25 ^{ème} percentile	11,35	11,22	11,27
Médiane	13,8	15,19	15,37
75 ^{ème} percentile	23,80	20,36	20,64
Maximum	30,25	28,90	28,89
Moyenne	16,51	16,04	16,16
Ecart-type	6,90	5,76	5,82
Variance	47,60	33,18	33,84
CV %	0,42	0,36	0,36

Tableau 25 – Ecart-types de krigeage, sur une grille régulière de 250 m de côté, par krigeage ordinaire ponctuel ou de bloc, pour les données de la sélection 2.

<i>SELECTION 2</i>	Estimation par krigeage ordinaire	
	Ecart-types de krigeage, ponctuel ($\mu\text{g}/\text{m}^3$)	Ecart-types de krigeage de bloc ($\mu\text{g}/\text{m}^3$)
Nombre de données	2337	2240
Minimum	3,95	1,90
Médiane	4,63	2,98
Maximum	8,13	7,21
Moyenne	5,08	3,50
Ecart-type	1,08	1,40
Variance	1,17	1,95
CV %	0,21	0,40

Le krigeage de bloc conduit à des valeurs estimées plus proches des valeurs vraies (Tab. 24 et Fig. 66), et à des écarts-type de krigeage plus faibles (Tab. 25 et Fig. 68) et constitue donc, comme précédemment, la meilleure technique des deux.

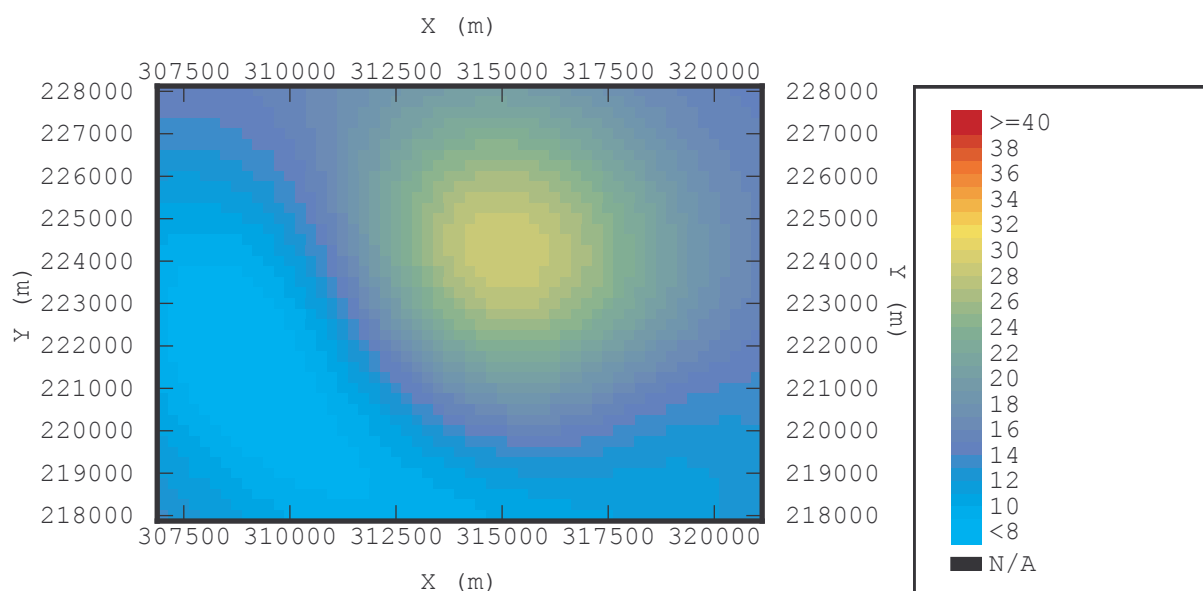


Figure 65 – Carte des estimations par krigeage ponctuel, à partir de la sélection 2.

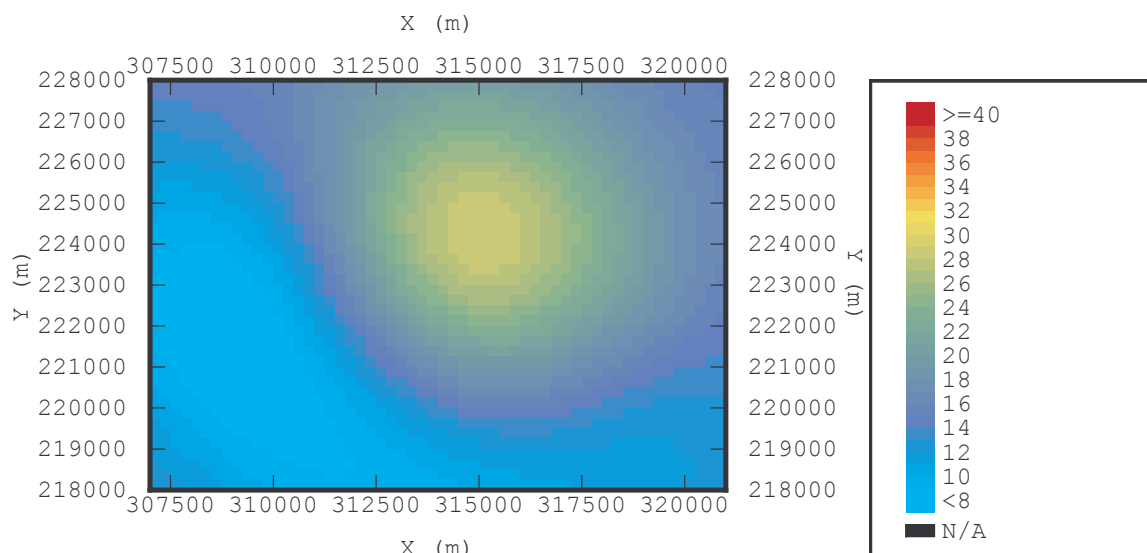


Figure 66 – Carte des estimations par krigeage de bloc, à partir de la sélection 2.

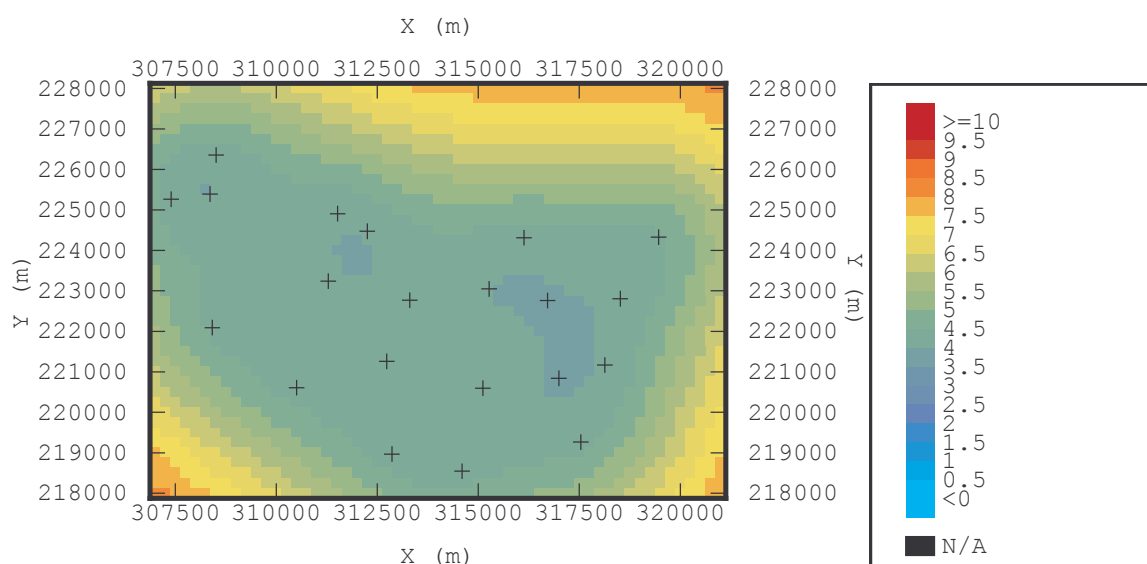


Figure 67 – Carte des écarts-type par krigeage de ponctuel, à partir de la sélection 2.

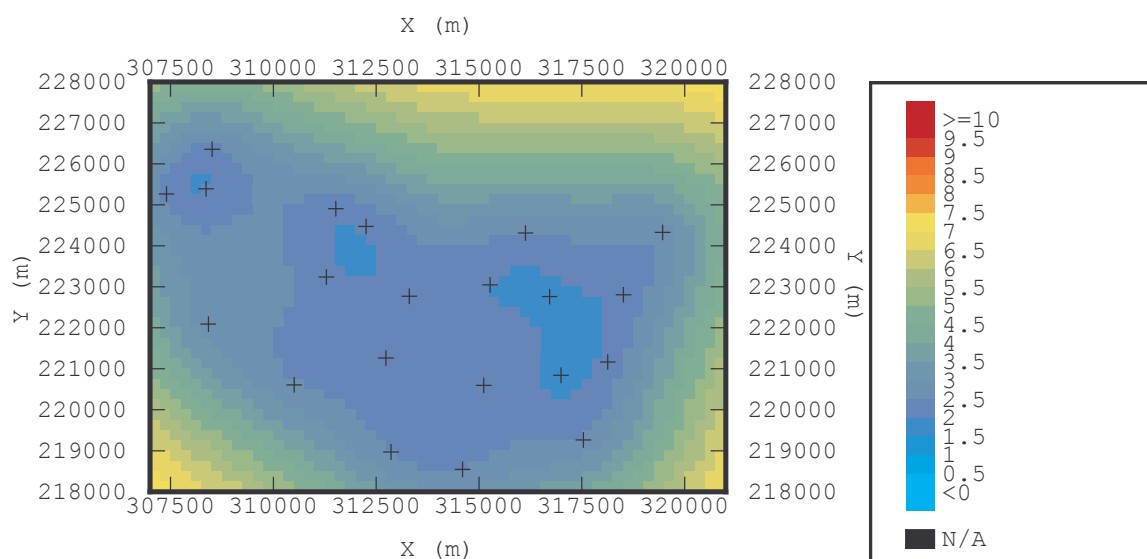


Figure 68 – Carte des écarts-type par krigeage de bloc, à partir de la sélection 2.

V COMPARAISON DES CARTES ET DISCUSSION

A partir des différentes sélections un tableau de synthèse des résultats de validation peut être établi :

Distribution des valeurs		Toutes les données	Sélection 1	Sélection 2
nombre de points		88	43	21
Min ($\mu\text{g}/\text{m}^3$)		6,85	6,85	7,60
Max ($\mu\text{g}/\text{m}^3$)		40,80	38,00	30,25
Moyenne ($\mu\text{g}/\text{m}^3$)		19,81	17,86	16,51
ecart-type ($\mu\text{g}/\text{m}^3$)		8,31	7,98	6,90
Variance ($\mu\text{g}/\text{m}^3$)		69,11	63,70	47,60
Validation croisée				
erreur	mean	0,14896	0,22715	0,13388
	var	19,79324	17,73639	14,08863
std. Error	mean	0,01120	0,02043	0,01407
	var	0,84959	1,15679	0,71423
Validation absolue				
Error	mean	-	- 0,6265	- 0,8820
	var	-	27,4642	25,8259

Tableau 26 : synthèse des résultats de validation des différents sélections.

Mise à part la variance de l'erreur d'estimation obtenue par validation croisée (Fig. 69), les autres paramètres statistiques de validation ne présentent pas de variation monotone avec le nombre de points.

La variance de l'erreur d'estimation obtenue par validation croisée (Fig. 69) augmente lorsque le nombre de points augmente, ce qui pourrait s'expliquer par la difficulté des modèles à estimer une variable dont la distribution des valeurs s'élargit.

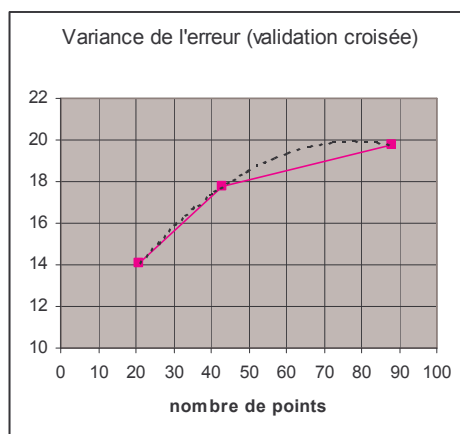


Figure 69 – Evolution de la variance de l'erreur, calculée par validation croisée, en fonction du nombre de points sélectionnés.

Nous avons établi le graphe de synthèse des différents variogrammes expérimentaux (Fig. 70).

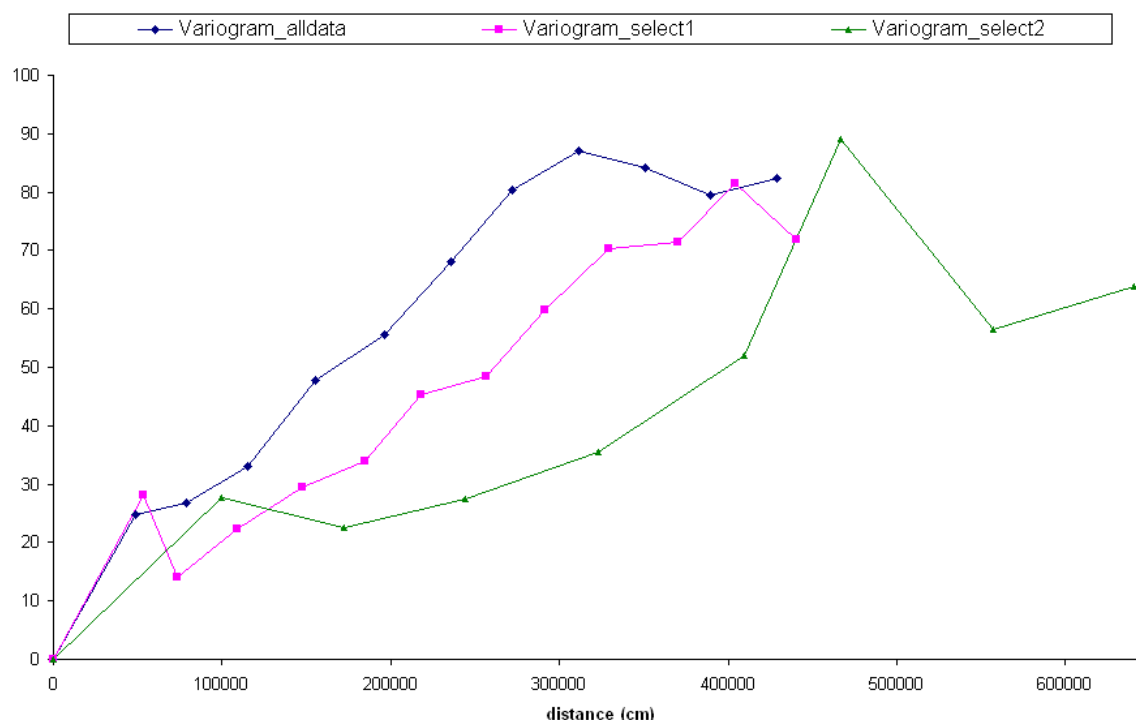


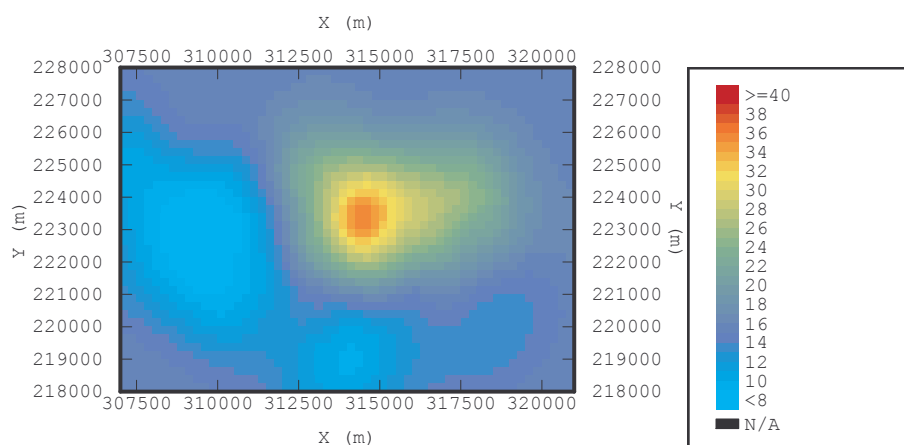
Figure 70 : Comparaison des différents variogrammes

L'analyse de ce graphique montre un net changement de structure du variogramme avec la diminution du nombre de points ; elle apparaît clairement à partir de la sélection 2 (21 points). Cependant, dès la sélection 1 (43 points), on constate que la portée du variogramme devient plus importante.

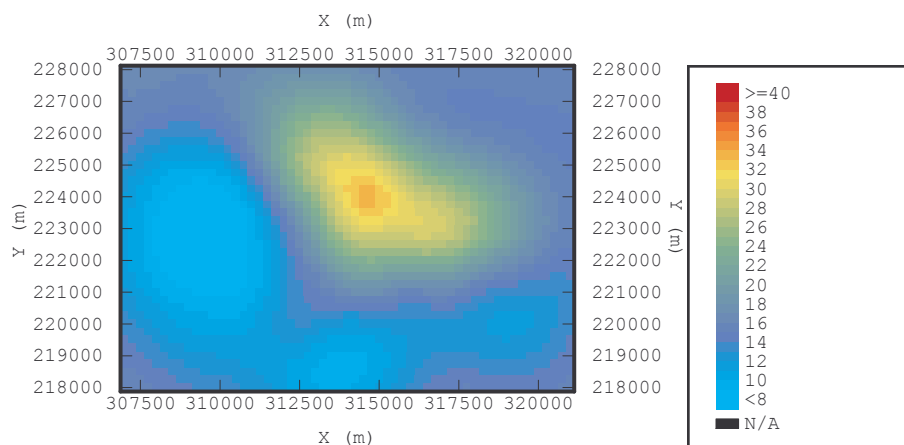
En effet, la diminution homogène (isotrope) du nombre de points, sans privilégier aucune zone particulière, conduit à la diminution des points situés en centre-ville (peu nombreux et localisés) au profit des points situés en périphérie (plus nombreux et bien répartis). Ceci est visible en examinant les paramètres statistiques (moyenne des valeurs ; étendue) qui diminuent avec la diminution du nombre de points.

Or les mesures localisées en périphérie présentent une continuité spatiale (une homogénéité) plus importante que les mesures de centre-ville, ce qui peut se quantifier par la portée spatiale du variogramme qui augmente.

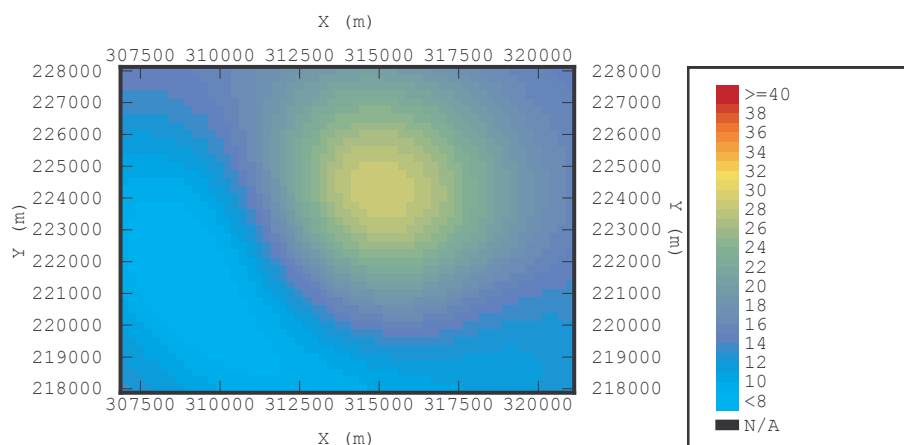
Cela est clairement visible sur les cartes de la figure 71, où l'on voit progressivement disparaître les fortes concentrations (« tâche rouge ») du centre-ville.



(a) Toutes les données (88 points)

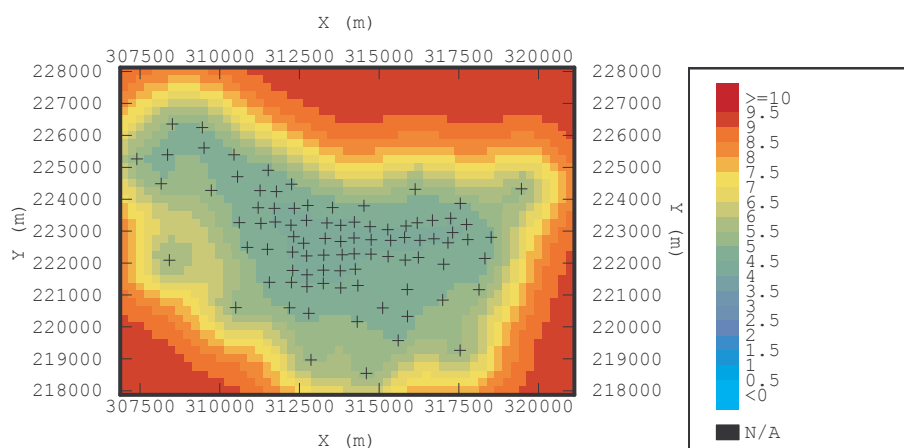


(b) Sélection 1 (43 points)

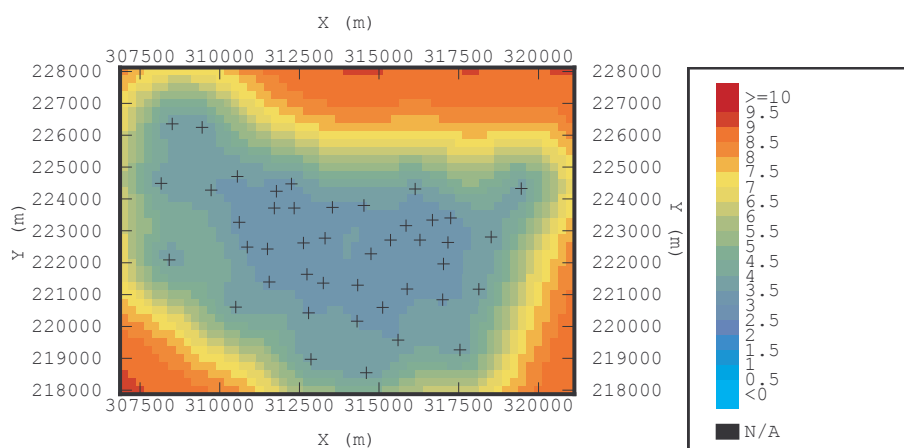


(c) Sélection 2 (21 points)

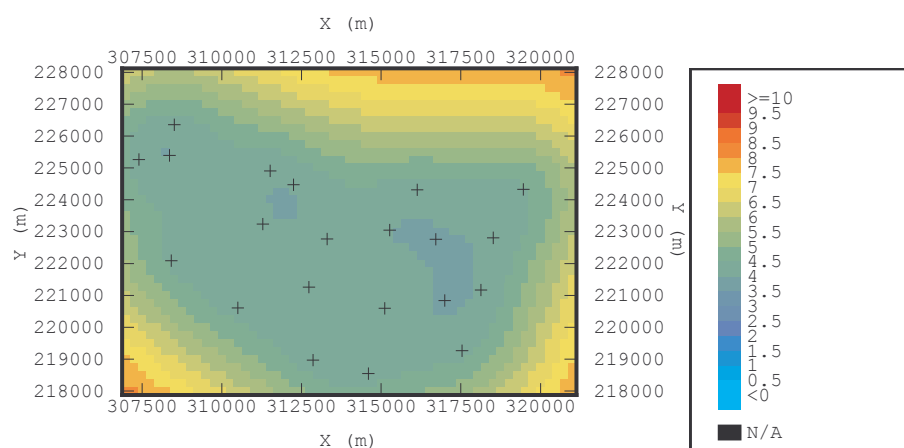
Figure 71 – Comparaison des cartes d'estimation établies par krigeage ponctuel sur la grille de référence.



(a) Toutes les données (88 points)



(b) Sélection 1 (43 points)



(c) Sélection 2 (21 points)

Figure 72 – Comparaison des cartes d'écart-type de krigeage établies par krigeage ponctuel sur la grille de référence.

La comparaison des cartes d'écarts-types de krigeage est, elle aussi, riche d'enseignements. En écartant de notre examen les bords du domaine (peu couverts par les mesures et donc présentant des valeurs d'écarts-types très élevées), on peut noter que la carte correspondant à la sélection 1 présente des écarts-types de krigeage plus faibles que pour les deux autres cartes.

Ceci est d'autant plus remarquable que les trois modèles de variogrammes utilisés comportent tous les mêmes structures :

Toutes les données	: pépité(17) + cubique(4400 m, 62)
Sélection 1	: pépité(8) + cubique(5600 m, 65)
Sélection 2	: pépité(12) + cubique(10000 m, 55)

En effet, on sait qu'un changement de structure peut induire des variations des écarts-type de krigeage.

Ces résultats de krigeage semblent donc accorder une confiance plus importante à la carte établie à partir d'un nombre assez limité de données (43 points) mais bien réparties dans l'espace, plutôt qu'à un nombre plus que double de données (88 points) mais réparties de façon inhomogène dans l'espace (selon un échantillonnage stratifié). Ces résultats, intéressants, méritent d'être confirmés par d'autres études.

Toutefois, lorsque le nombre de données est trop faible (21 points, sélection 2), la confiance accordée aux résultats de krigeage diminue.

Cette étude semble donc montrer qu'un optimum est obtenu pour environ 43 points de mesure pour un domaine de 125 km^2 ($12,5 \text{ km} \times 10 \text{ km}$), soit 34,4 points pour 100 km^2 ou 1 point par maille de $2,90 \text{ km}^2$.

Annexe 3 – Urbain : La Rochelle

B.I. DESCRIPTION DU CAS	33
B.I.1. Présentation du jeu de données initial	33
B.I.2. Méthodologie de sélection des données	33
B.II. ETUDE DU JEU DE DONNÉES INITIAL ET ÉLABORATION DE LA CARTE DE RÉFÉRENCE	35
B.II.1. Etude du variogramme expérimental	36
B.II.2. Modélisation du variogramme et validation croisée	36
B.II.3. Etablissement de la carte de référence	38
B.III. ETUDE DE LA SÉLECTION DE DONNÉES N°1 (MAILLE DE 1 KM)	39
B.III.1. Etude du variogramme expérimental	40
B.III.2. Modélisation du variogramme et test de validation croisée	40
B.III.3. Test de validation absolue sur le complémentaire de la sélection N°1	41
B.III.4. Etablissement de la carte pour la sélection 1	43
B.IV. ETUDE DE LA SÉLECTION DE DONNÉES N°2 (MAILLE DE 1.5 KM)	43
B.IV.1. Etude du variogramme expérimental	43
B.IV.2. Modélisation du variogramme et test de validation croisée	44
B.IV.3. Test de validation absolue sur le complémentaire de la sélection 2	45
B.IV.4. Etablissement de la carte pour la sélection 2	46
B.V. ETUDE DE LA SÉLECTION DE DONNÉES N°3 (MAILLE DE 2 KM)	47
B.V.1. Etude du variogramme expérimental	47
B.V.2. Modélisation du variogramme et test de validation croisée	48
B.V.3. Test de validation absolue sur le complémentaire de la sélection 3	50
B.V.4. Etablissement de la carte pour la sélection 3	51
B.V.5. Test de validation absolue sur l'intersection des complémentaires	51
B.VI. COMPARAISON DES ESTIMATIONS ET DISCUSSION	53
B.VI.1. Influence du nombre de données sur le variogramme expérimental	53
B.VI.2. Influence sur l'estimation du nombre de données de krigeage pour un modèle établi de variogramme	53
• B.VI.2.1. Estimation sur un jeu de données indépendant	53
• B.VI.2.2. Estimation sur la grille	54
B.VI.3. Effet du nombre de données expérimentales sur la qualité des cartes	55

B.VII. CALCULS COMPLEMENTAIRES : UTILISATION DES ÉMISSIONS DE NOX EN VARIABLE AUXILIAIRE – IMPACT SUR LES ESTIMATIONS	57
B.VII.1. Etude du jeu de données complet.....	57
• B.VII.1.1. Corrélation entre NO ₂ et densité d'émissions de NOx.....	57
• B.VII.1.2. Modélisation du variogramme et test de validation croisée	58
• B.VII.1.3. Elaboration de la carte de référence	59
B.VII.2. Etude des sélections n°2 (maille de 1,5km) et n°3 (maille de 2km).....	60
• B.VII.2.1. Corrélation entre NO ₂ et densité d'émissions de NOx.....	60
• B.VII.2.2. Modélisation du variogramme et test de validation croisée	60
• B.VII.2.3. Test de validation absolue.....	62
• B.VII.2.4. Elaboration des cartes pour les deux sélections	63
• B.VII.2.5. BILAN	65

B.I. DESCRIPTION DU CAS

B.I.1. Présentation du jeu de données initial

Le jeu de données étudié a été fourni par l'association agréée de surveillance ATMO Poitou-Charentes. Il concerne la mesure du dioxyde d'azote à La Rochelle (76594 et 147000 habitants pour la ville et la communauté d'agglomération, superficie : 29,2 km² et 206 km² pour la ville et la communauté d'agglomération). Il est constitué de huit campagnes de mesure isolées, d'une durée de 14 jours chacune, conduites par échantillonnage passif au cours de l'année 2004.

65 sites de mesure, dont un site sous influence industrielle, ont été disposés sur un domaine d'environ 17 km d'est en ouest et de 20 km du nord au sud. Leur répartition correspond à la stratégie d'échantillonnage couramment adoptée par les AASQA : un resserrement des points de mesure en centre-ville et une disposition plus espacée en périphérie (Figure 73).

Les valeurs prises en compte pour cette étude correspondent à la moyenne des 8 campagnes. Leurs statistiques sont fournies dans le Tableau 26. Le point de type industriel, qui présente une concentration moyenne très supérieure à celles des autres sites, a été éliminé.

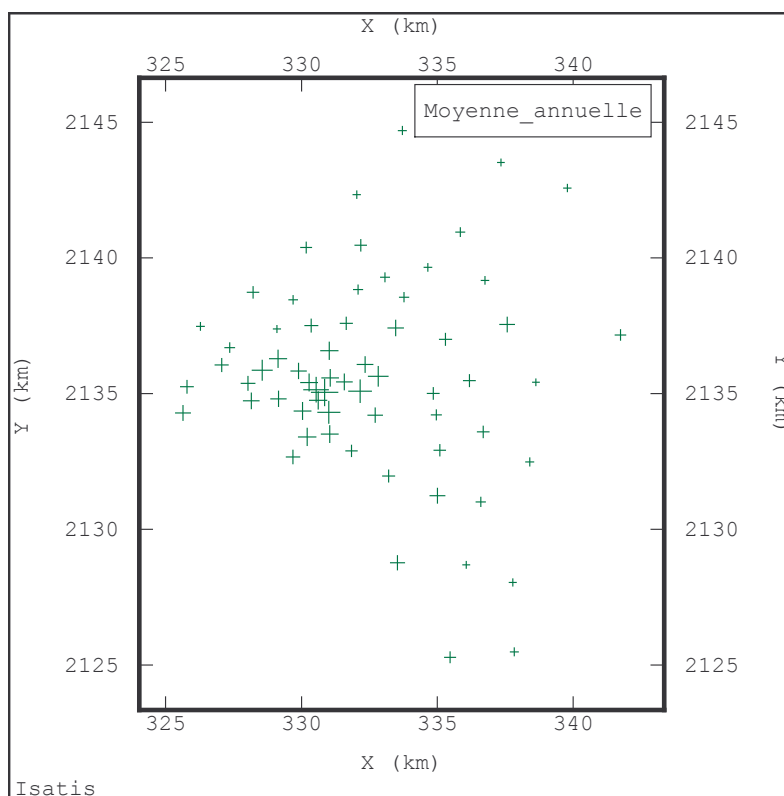


Figure 73 - Carte de répartition des points de mesure

B.I.2. Méthodologie de sélection des données

Les données sont sélectionnées selon trois maillages réguliers :

- mailles carrées de 1 km de côté (sélection 1 notée S1) ;
- mailles carrées de 1,5 km de côté (sélection 2 notée S2)
- mailles carrées de 2 km de côté (sélection 3 notée S3).

La procédure est la suivante : un maillage régulier de résolution 1km, 1.5km ou 2km calé est calé sur le domaine d'étude puis, pour chaque maille, on sélectionne le point de mesure le plus proche du centre de celle-ci.

Les jeux de données complémentaires de ces trois sélections (notés C_{S1} , C_{S2} , C_{S3}) sont également constitués.

Le Tableau 27 présente les statistiques élémentaires de ces différents jeux.

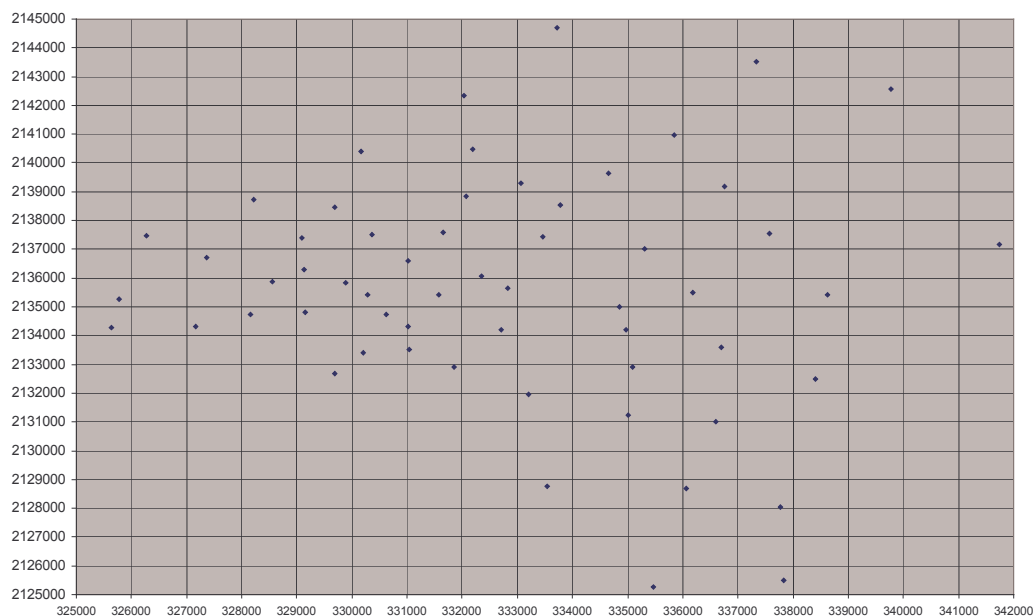


Figure 74 – Carte de répartition des points de la sélection 1 (mailles de 1km).

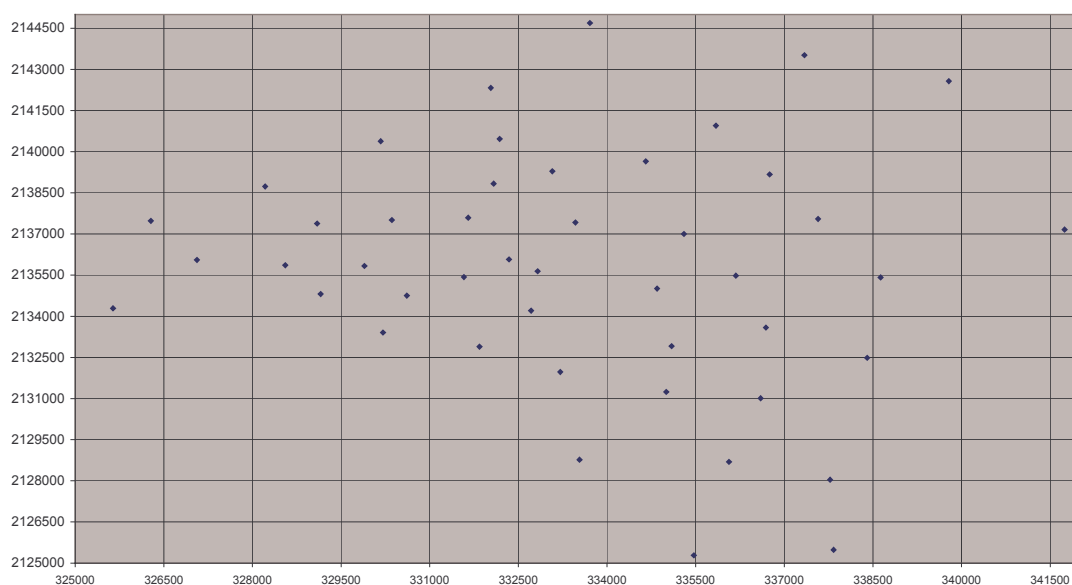


Figure 75 – Carte de répartition des points de la sélection 2 (mailles de 1.5km)

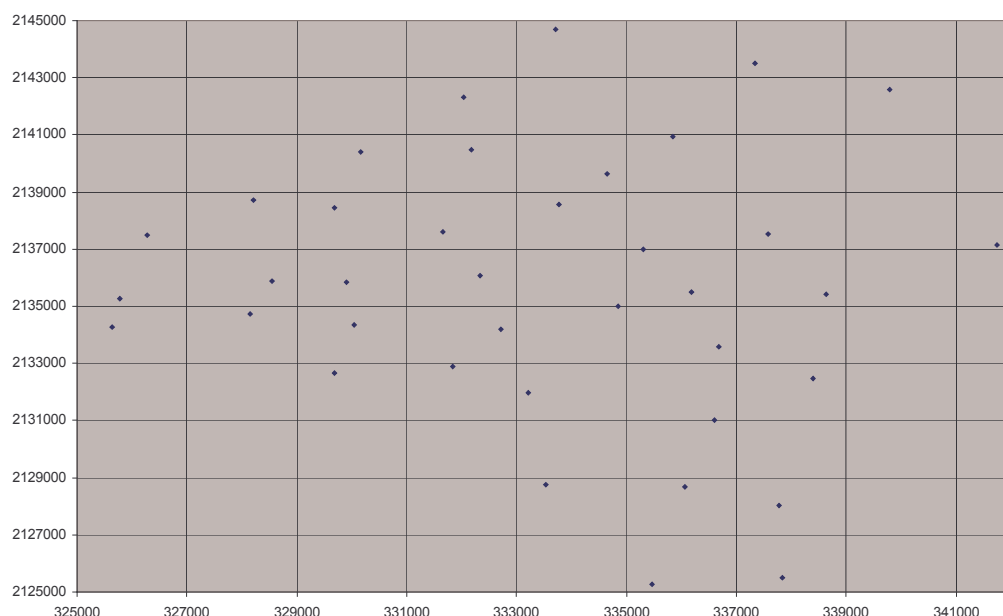


Figure 76 – Carte de répartition des points de la sélection 3 (mailles de 2 km).

Tableau 27 - Principales caractéristiques des jeux de données étudiés.

	Toutes les données	S1	C _{S1}	S2	C _{S2}	S3	C _{S3}
Nombre de données	64	57	7	46	18	38	26
Minimum	7.34	7.34	11.69	7.34	8.58	7.34	7.65
25 ^{ème} percentile	8.89	8.73	12.59	8.23	11.46	8.23	10.52
médiane	10.96	10.59	14.18	10.57	14.14	10.53	13.41
75 ^{ème} percentile	13.39	13.20	19.20	12.74	14.45	11.96	14.44
Maximum	20.66	17.70	20.66	16.28	20.66	16.28	20.66
Moyenne	11.50	10.98	15.77	10.64	13.72	10.34	13.20
Ecart-type	3.07	2.59	3.37	2.41	3.46	2.20	3.37
Variance	9.45	6.71	11.33	5.79	11.95	4.85	11.34
CV%	27	24	21	23	25	21	26

B.II. ETUDE DU JEU DE DONNEES INITIAL ET ELABORATION DE LA CARTE DE REFERENCE

Le but de cette partie est d'élaborer la carte de référence, c'est-à-dire la carte de répartition des concentrations de dioxyde d'azote établie avec le maximum de points de mesure. On comparera par la suite à cette carte les cartes obtenues à partir d'un nombre réduit de points.

Remarque :

Dans une première étape, cette étude est réalisée sans prise en compte de variables auxiliaires. Or des données de population et d'émission, susceptibles d'expliquer les concentrations de NO₂ et par suite, d'améliorer la précision des cartographies, sont disponibles. Dans une seconde étape, il conviendra d'examiner si l'utilisation de telles données peut modifier les conclusions sur l'échantillonnage. Nous reviendrons sur ce point dans la discussion.

B.II.1. Etude du variogramme expérimental

Deux sites de mesure implantés au centre-ville accroissent de façon notable la variabilité spatiale des concentrations. Sans ces deux sites, le variogramme est plus structuré. Quelques points ressortent néanmoins de la nuée variographique (points marqués en bleus dans la figure): ils correspondent à des couples de sites positionnés selon l'axe $N160 \pm 30$. Cette observation est cohérente avec la carte variographique qui met en évidence une anisotropie de directions $N70$ (direction de plus grande continuité) et $N160$.

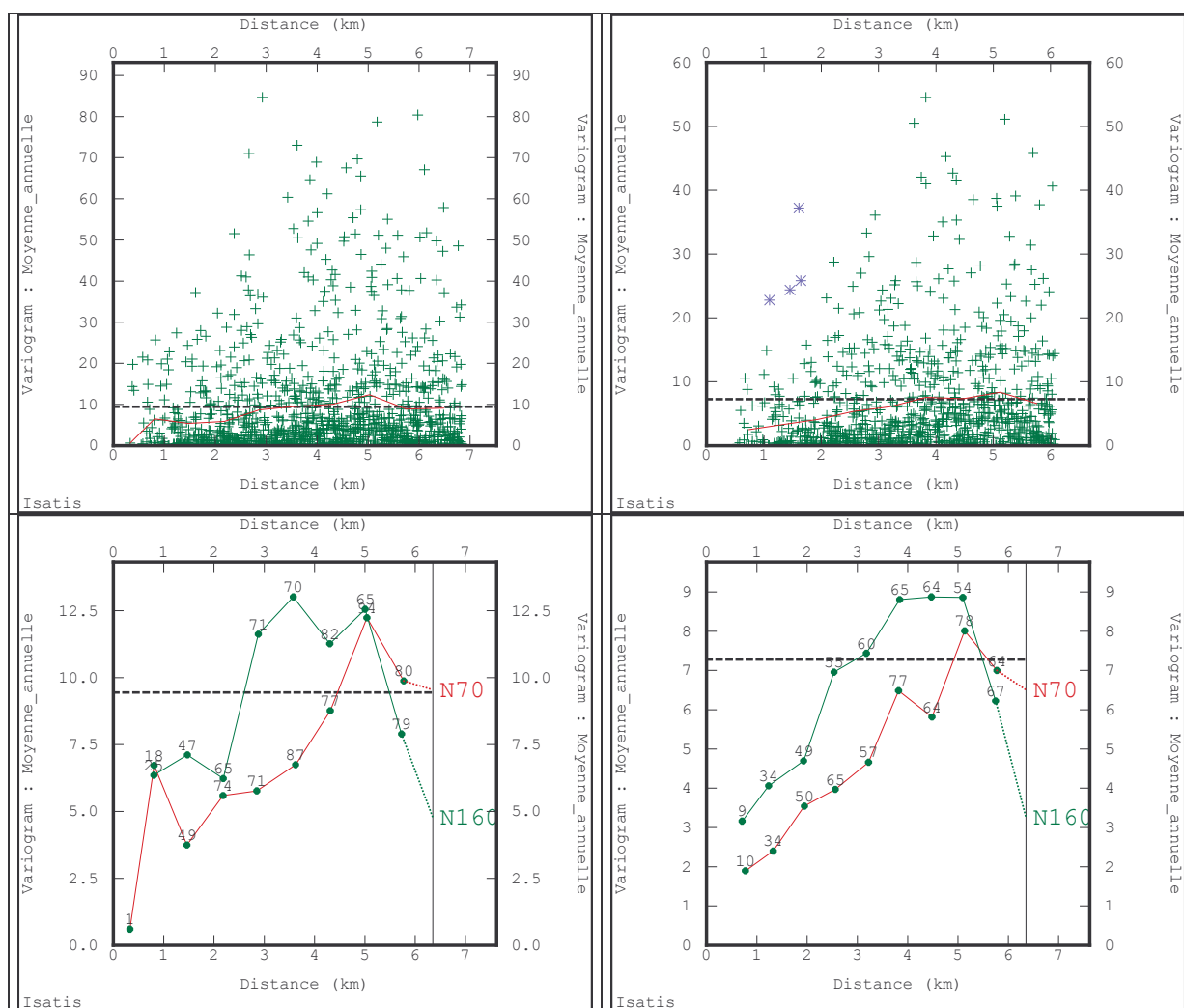


Figure 77 – Ligne du haut : nuée variographique avec (fig. de gauche, pas = 720 m, 10 pas), et sans (fig. de droite, pas = 640 m, 10 pas) les points du centre-ville. Ligne du bas : Variogramme bidirectionnel avec et sans ces deux points.

B.II.2. Modélisation du variogramme et validation croisée

Des modèles ont été ajustés sur les variogrammes expérimentaux omni- et bidirectionnels construits avec et sans les deux points de centre-ville. Ils se composent tous d'une structure pépitique additionnée d'une structure sphérique ou cubique. Avec les deux points du centre, la structure plus erratique du variogramme rend la modélisation moins aisée.

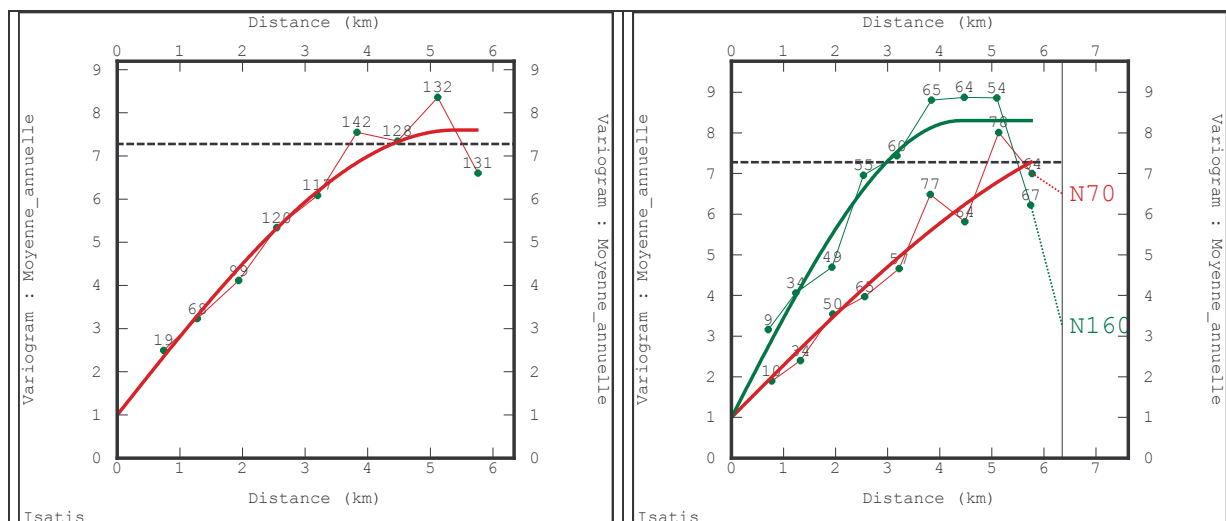


Figure 78– Variogrammes expérimentaux et modélisés sur 62 sites de mesure.

Comme dans l'étude de Cherbourg, la qualité respective des différents modèles est appréciée par validation croisée. Un voisinage unique est considéré pour l'estimation.

Pour tous les modèles, les statistiques d'erreur sont calculées sur les 64 sites d'une part, sur les 62 sites d'autre part.

Le **modèle bidirectionnel sphérique**, ajusté sur le variogramme expérimental calculé avec 62 données, fournit les résultats les plus satisfaisants. Il est pris pour référence.

Tableau 28 – Statistiques sur l'erreur d'estimation obtenues par validation croisée avec le modèle = Pépité (1) + Sphérique (N70- 8500 m, N160-4400m; 7.3)

64 données	moyenne	variance
Erreur	0.07626	3.16898
Erreur standardisée	0.02072	1.09469

2 valeurs non robustes sur 64 valeurs (les deux points du centre).

62 données	moyenne	variance
Erreur	0.06164	2.40735
Erreur standardisée	0.01534	0.67710

0 valeur non robuste sur 62 valeurs.

Tableau 29 – Résultats de la validation croisée sur toutes les données pour le krigeage ordinaire ponctuel, modèle de référence

		Validation croisée par krigeage ordinaire ponctuel sur 64 données			
Toutes les données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	64	64	64	64	64
Minimum	7.34	8.01	-4.59	-3.37	0.23
25 ^{ème} percentile	8.89	9.62	-1.30	-0.62	6.72
Médiane	10.96	10.69	0.43	0.2	11.43
75 ^{ème} percentile	13.39	13.66	1.54	0.74	19.11
Maximum	20.66	17.62	3.24	2.3	39.81
Etendue (= Max – Min)	13.32	9.61	-	-	-
Moyenne	11.50	11.58	0.08	0.02	13.12
Ecart-type	3.07	2.53	1.78	1.05	8.52
Variance	9.45	6.40	3.17	1.09	72.55
CV %	27	22			
Corrélation mesures/estimations		r=0,815			

Les valeurs estimées sont bien corrélées aux valeurs mesurées, avec un coefficient de corrélation supérieur à 0,80 (Tableau 29, Figure 79). Les deux points de plus grande erreur standardisée (données non robustes) sont les deux points du centre-ville, où les concentrations sont sous-estimées. (Pour ces deux sites, il ne s'agit pas d'une validation croisée au sens strict puisqu'ils n'ont pas servi à calculer le variogramme. L'estimation en ces points ne s'améliore cependant pas lorsqu'on utilise le modèle de variogramme établi sur 64 données.)

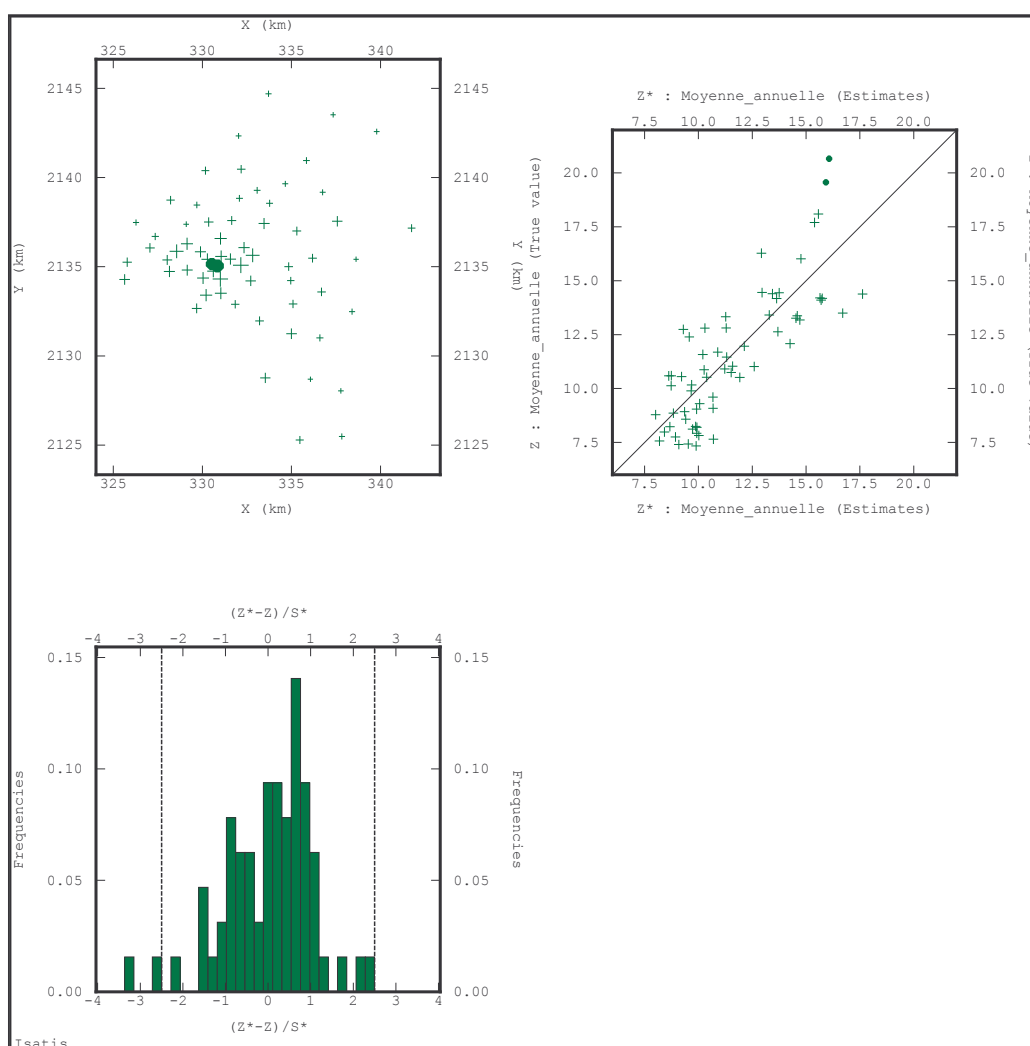


Figure 79 – Résultats de validation croisée. Les deux points qui s'écartent de la bissectrice correspondent aux deux points du centre-ville.

B.II.3. Etablissement de la carte de référence

Afin de dresser la carte de référence de l'étude, c'est-à-dire la carte réalisée à partir du plus grand nombre de mesures disponibles, nous estimons les valeurs de concentration de NO_2 aux nœuds d'une grille 2D régulière. La distance entre sites étant d'au moins 320m dans la zone d'échantillonnage la plus dense, une maille d'estimation de 500m de côté est choisie. Cette maille sera conservée dans toute l'étude.

Les caractéristiques de la grille d'estimation sont :

Xmin=325000 m, 35 nœuds, Xmax=342000, $\Delta X=17000$ m,
Ymin=2125000 m, 41 nœuds, Ymax=2145000, $\Delta Y=20000$ m.
Compte tenu du nombre limité de données (moins d'une centaine), un voisinage d'estimation **unique** est adopté.

Tableau 30 – Estimation sur une grille régulière de 500 m de côté, par krigeage ordinaire ponctuel

Toutes les données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées par krigeage ponctuel ($\mu\text{g}/\text{m}^3$)	écarts-types de krigeage ($\mu\text{g}/\text{m}^3$)
Nombre de données	64	934	934
Minimum	7.34	7.56	1.08
25 ^{ème} percentile	8.89	8.61	1.66
Médiane	10.96	9.72	1.85
75 ^{ème} percentile	13.39	11.12	2.10
Maximum	20.66	17.53	2.77
Moyenne	11.50	10.11	1.89
Ecart-type	3.07	1.86	0.30
Variance	9.45	3.45	0.09
CV %	27	18	16

Statistiques des valeurs estimées à l'intérieur du polygone définissant les contours de l'agglomération

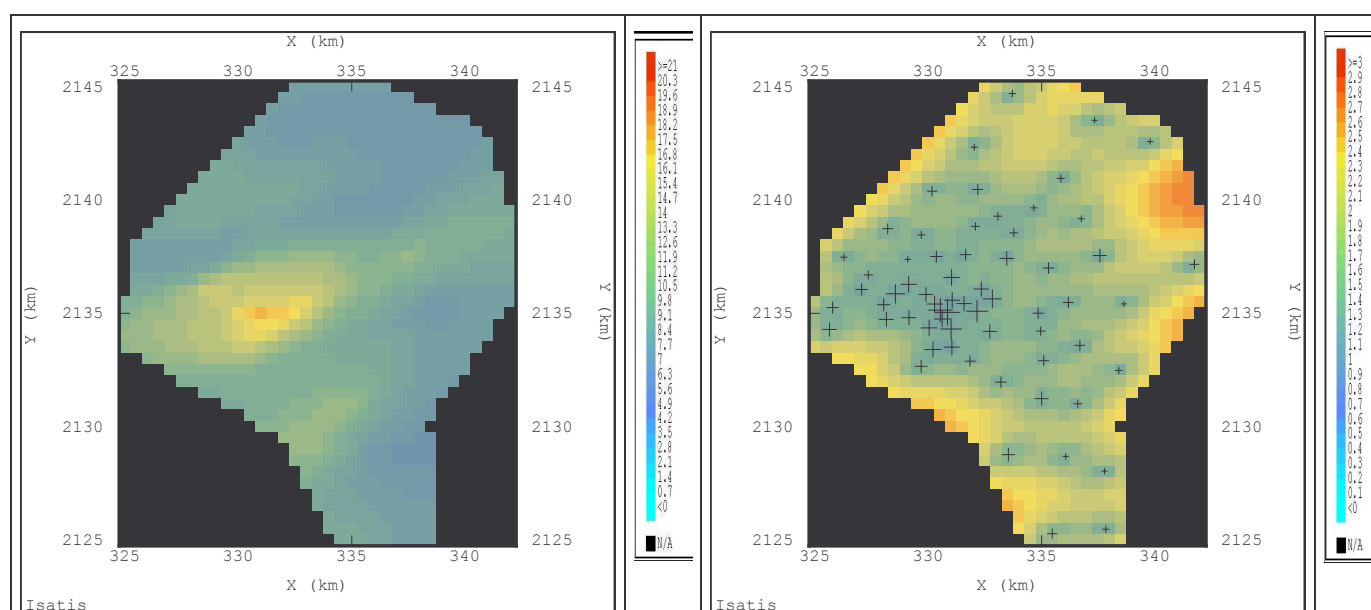


Figure 80 – Estimation par krigeage ponctuel, à partir de toutes les données, avec le modèle de référence (fig. de gauche) et écart-type de krigeage (fig. de droite)

B.III. ETUDE DE LA SELECTION DE DONNEES N°1 (MAILLE DE 1 KM)

La sélection n°1 comprend 57 données, qui correspondent à 1 point de mesure par maille carrée de 1km^2 .

B.III.1. Etude du variogramme expérimental

Le variogramme expérimental comprend 10 pas de 640 m chacun. Il montre là encore une anisotropie (N70, N160). Aucun point n'est éliminé, les sites 43 et 45 précédemment écartés n'étant pas inclus dans cette sélection de points.

Les mêmes couples de points que précédemment ressortent de la nuée variographique.

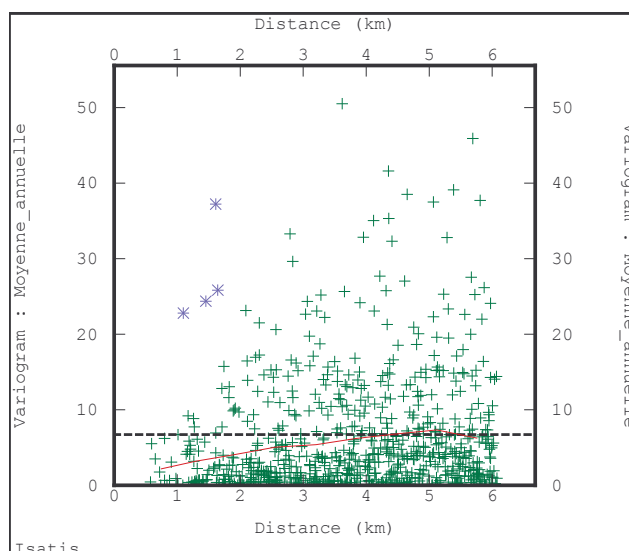


Figure 81– Nuée variographique (pas de 640 m, 10 pas) des données de la sélection 1.

B.III.2. Modélisation du variogramme et test de validation croisée

Comme précédemment, un modèle omnidirectionnel et un modèle bidirectionnel sont ajustés et soumis au test de validation croisée.

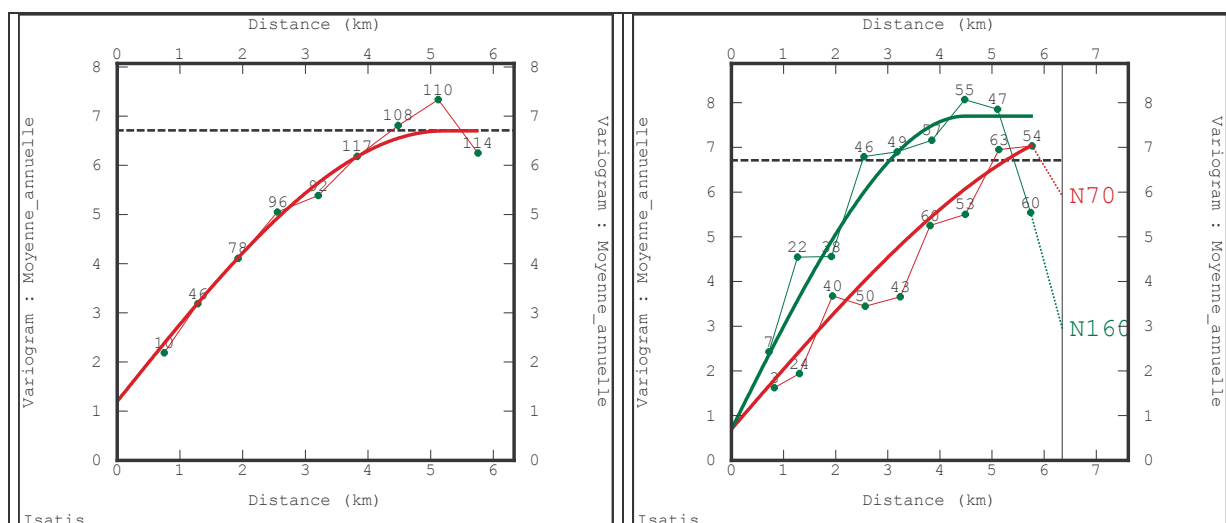


Figure 82 – Variogrammes expérimentaux et modélisés sur 57 sites de mesure

Les statistiques de validation croisée sont favorables au modèle bidirectionnel. Toutes les données sont robustes.

Tableau 31 – Résultats de la validation croisée sur les données de la sélection 1 pour le krigeage ordinaire ponctuel, modèle bidirectionnel de la sélection 1 : modèle = Pépité (0.7) + Sphérique (N70- 7800 m, N160-4500m; 7.3)

Validation croisée par krigeage ordinaire ponctuel sur 57 données					
57 données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	57	57	57	57	57
Minimum	7.34	7.70	-3.43	-2.23	0.09
25 ^{ème} percentile	8.73	9.50	-1.12	-0.56	5.43
Médiane	10.59	10.29	0.53	0.28	10.77
75 ^{ème} percentile	13.20	12.92	1.26	0.70	18.14
Maximum	17.70	15.39	3.09	1.90	40.37
Etendue (= Max – Min)	10.36	7.69	-	-	-
Moyenne	10.98	11.04	0.06	0.02	12.55
Ecart-type	2.59	2.04	1.58	0.86	9.15
Variance	6.71	4.14	2.5	0.74	83.74
CV %	24	18			
Corrélation mesures/estimations		r=0.792			

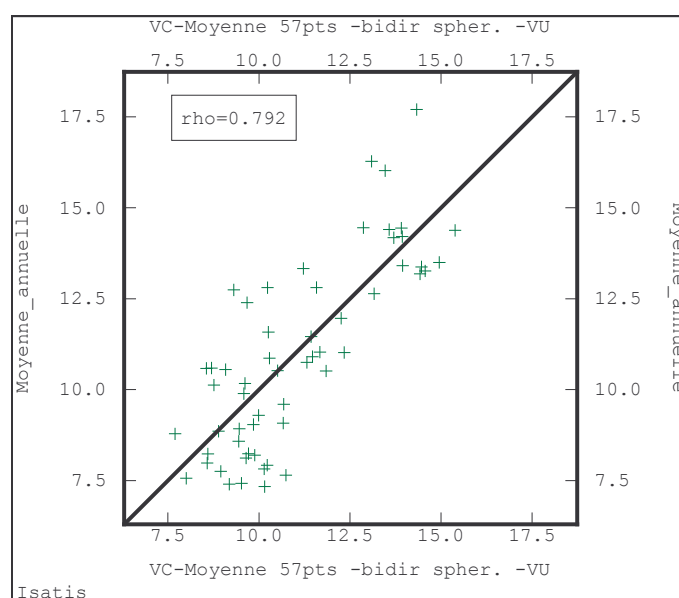


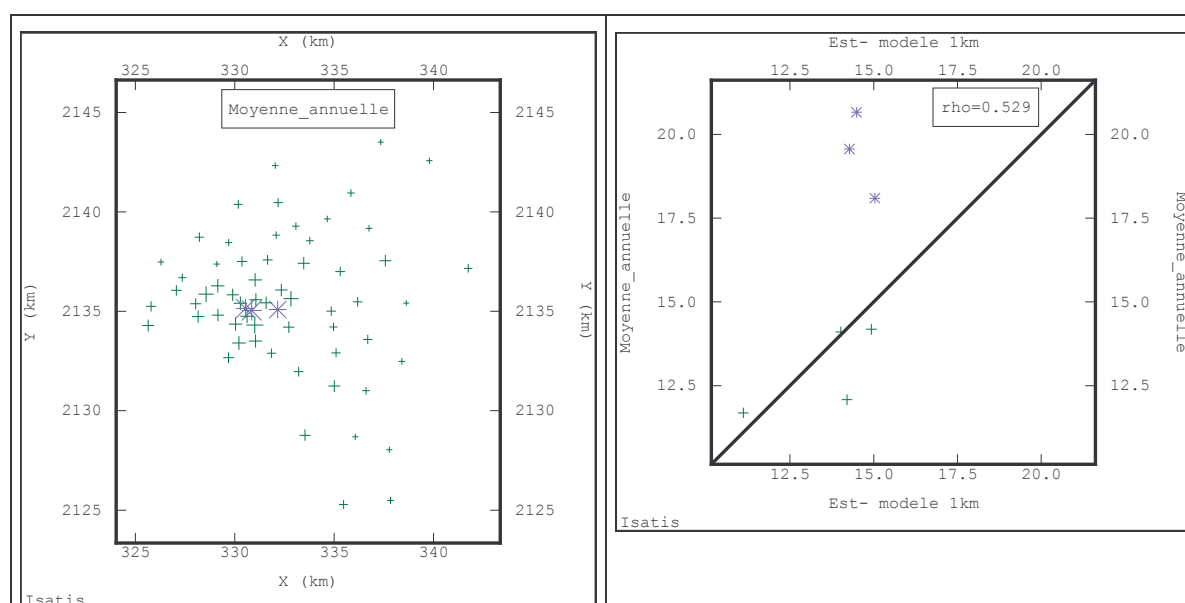
Figure 83 – Validation croisée : nuage de corrélation entre les concentrations estimées et mesurées aux points de la sélection 1

B.III.3. Test de validation absolue sur le complémentaire de la sélection N°1

Les concentrations de NO_2 sont estimées aux points qui n'ont pas servi à la construction du variogramme, puis sont comparées aux valeurs réelles. Les valeurs supérieures à $15 \mu\text{g}/\text{m}^3$ sont sous-estimées, en particulier les mesures aux points 43 et 45 (NB : il s'agit des deux points de centre-ville retirés de l'analyse variographique de référence).

Tableau 32 – Résultats de la validation absolue sur le complémentaire de la sélection 1 pour le krigeage ordinaire ponctuel, modèle de la sélection 1

		Validation absolue par krigeage ordinaire ponctuel sur 7 données			
57 données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	7	7	7	7	7
Minimum	11.69	11.11	-6.18	-4.74	0.64
Maximum	20.66	15.04	2.12	1.43	29.89
Etendue (= Max – Min)	8.97	4.93	-	-	-
Moyenne	15.77	14.01	-1.76	-1.33	14.62
Ecart-type	3.37	1.23	2.91	2.17	10.53
Variance	11.33	1.52	8.46	4.70	110.96
CV %	21	9			
Corrélation mesures/estimations		r=0.529			



B.III.4. Etablissement de la carte pour la sélection 1

La carte d'estimation est proche de la carte initiale ; on note une atténuation des concentrations dans le centre-ville.

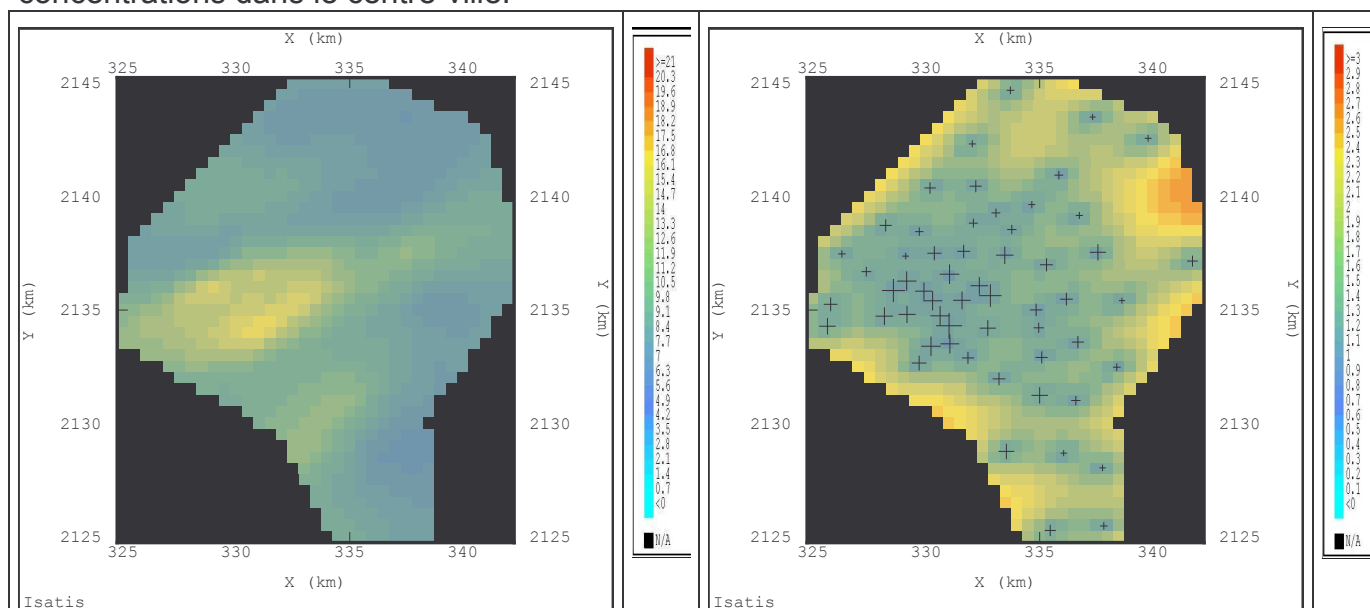


Figure 85 – Estimation par krigeage ponctuel, à partir des données de la sélection 1, avec le modèle de la sélection1 (fig. de gauche) et écart-type de krigeage (fig. de droite)

B.IV. ETUDE DE LA SELECTION DE DONNEES N°2 (MAILLE DE 1.5 KM)

La sélection n°2 comprend 46 données, qui correspondent à 1 point de mesure par maille carrée de 1,5km².

B.IV.1. Etude du variogramme expérimental

Le variogramme expérimental comprend 13 pas de 640 m chacun. Aux distances considérées, il présente une non stationnarité dans la direction N70 tandis que dans la direction perpendiculaire (N160), il atteint rapidement un palier. Trois couples de valeurs contrastées, deux desquels sont alignés dans la direction N160, ressortent de la nuée variographique. Aucun point n'est éliminé (notons que les sites 43 et 45 ne sont pas inclus dans cette sélection de points).

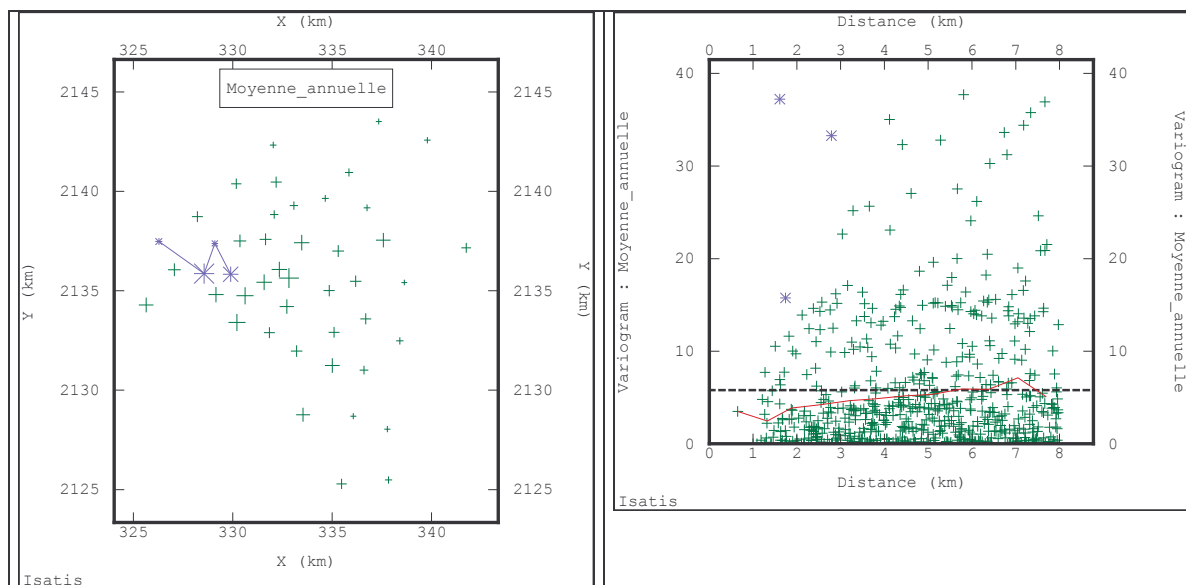


Figure 86– Nuée variographique (pas de 640 m, 13 pas) des données de la sélection 2.

B.IV.2. Modélisation du variogramme et test de validation croisée

Différents modèles omnidirectionnels et bidirectionnels sont ajustés. Le modèle bidirectionnel présenté en Figure 86 fournit les meilleurs résultats de validation croisée. Bien que la qualité d'ajustement se dégrade au-delà de 5km, la variance d'erreur et la corrélation sont respectivement plus faible et plus élevée en voisinage unique.

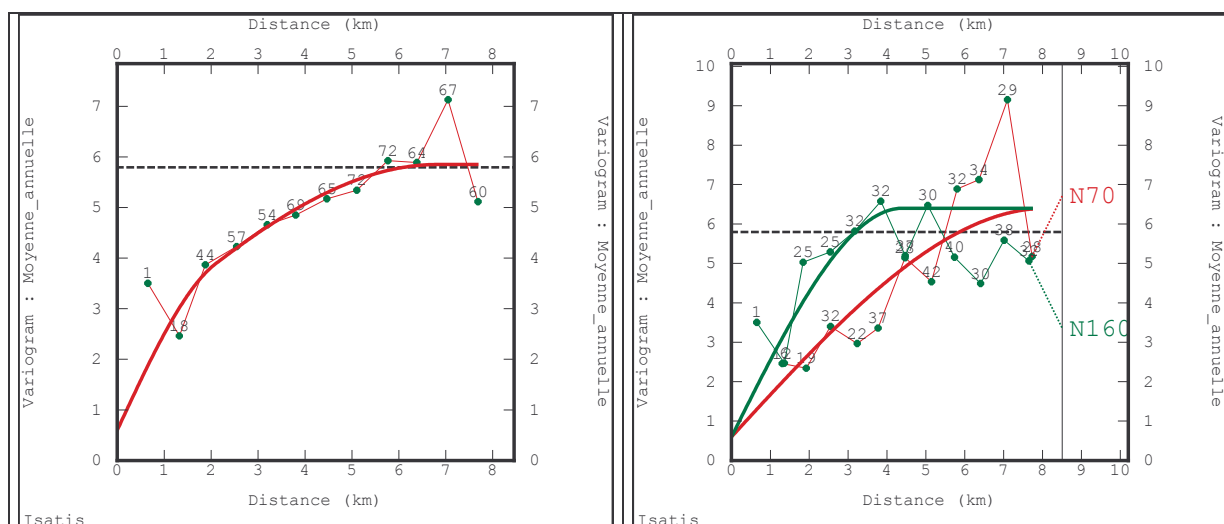


Figure 87 – Variogrammes expérimentaux et modélisés sur 46 sites de mesure

Tableau 33 – Résultats de la validation croisée sur les données de la sélection 2 pour le krigeage ordinaire ponctuel, modèle bidirectionnel de la sélection 2. modèle = Pépité (0.6) + Sphérique (N70- 8100 m, N160-4400m; 7.3)

		Validation croisée par krigeage ordinaire ponctuel sur 46 données			
46 données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	46	46	46	46	46
Minimum	7.34	7.78	-4.16	-2.67	2.16
25 ^{ème} percentile	8.23	9.46	-1.01	-0.63	6.13
Médiane	10.57	10.25	0.54	0.20	13.44
75 ^{ème} percentile	12.74	11.77	1.46	0.76	21.87
Maximum	16.28	14.68	2.81	2.42	38.31
Etendue (= Max – Min)	8.94	6.90	-	-	-
Moyenne	10.64	10.69	0.15	0.01	14.38
Ecart-type	2.41	1.76	1.67	1.00	9.06
Variance	5.79	3.08	2.79	1.00	82.09
CV %	23	16			
Corrélation mesures/estimations		r=0.680			

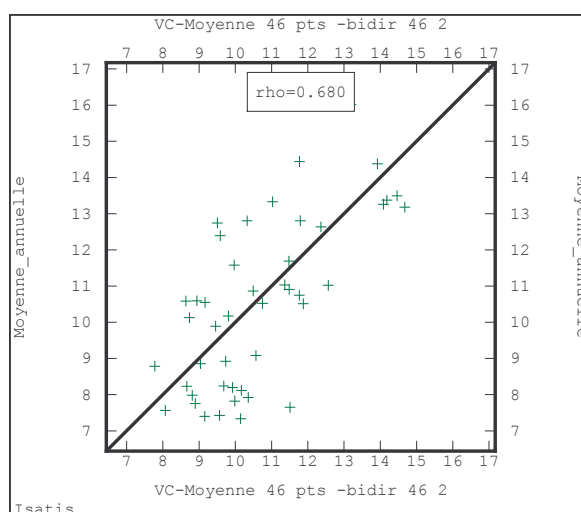


Figure 88 – Validation croisée : nuage de corrélation entre les concentrations estimées et mesurées aux points de la sélection 1

B.IV.3. Test de validation absolue sur le complémentaire de la sélection 2

L'estimation des valeurs inférieures à 15 $\mu\text{g}/\text{m}^3$ est tout à fait satisfaisante. En revanche, comme dans la sélection précédente, les concentrations aux points 43 et 45 sont sous-estimées.

Tableau 34 – Résultats de la validation absolue sur le complémentaire de la sélection 2 pour le krigeage ordinaire ponctuel, modèle de la sélection 2

		Validation absolue par krigeage ordinaire ponctuel sur 18 données			
46 données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	18				
Minimum	8.58	9.50	-6.59	-5.38	0.21
P25	11.46	11.78	-1.72	-1.16	3.45
P50	14.14	13.61	-0.23	-0.17	9.09
P75	14.45	14.08	0.82	0.55	19.59
Maximum	20.66	14.45	2.37	1.66	31.88
Etendue (= Max – Min)	12.08	4.95			
Moyenne	13.72	12.77	-0.95	-0.76	11.24
Ecart-type	3.46	1.63	2.36	1.80	9.07
Variance	11.95	2.66	5.56	3.23	82.21
CV %	25	13			
Corrélation mesures/estimations		r=0.802			

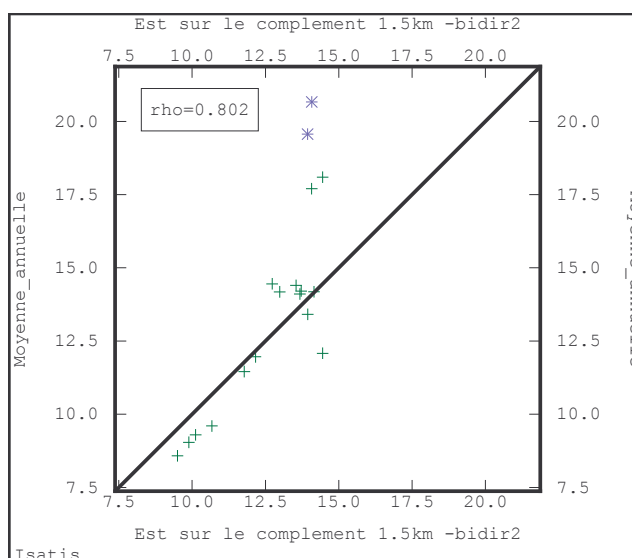


Figure 89 – Nuage de corrélation entre les concentrations estimées et mesurées aux points complémentaires de la sélection 2. En bleu : points où les concentrations sont sous-estimées.

B.IV.4. Etablissement de la carte pour la sélection 2

Visuellement, la carte d'estimation ressemble à la carte obtenue pour la sélection 1 ; elle est légèrement moins contrastée.

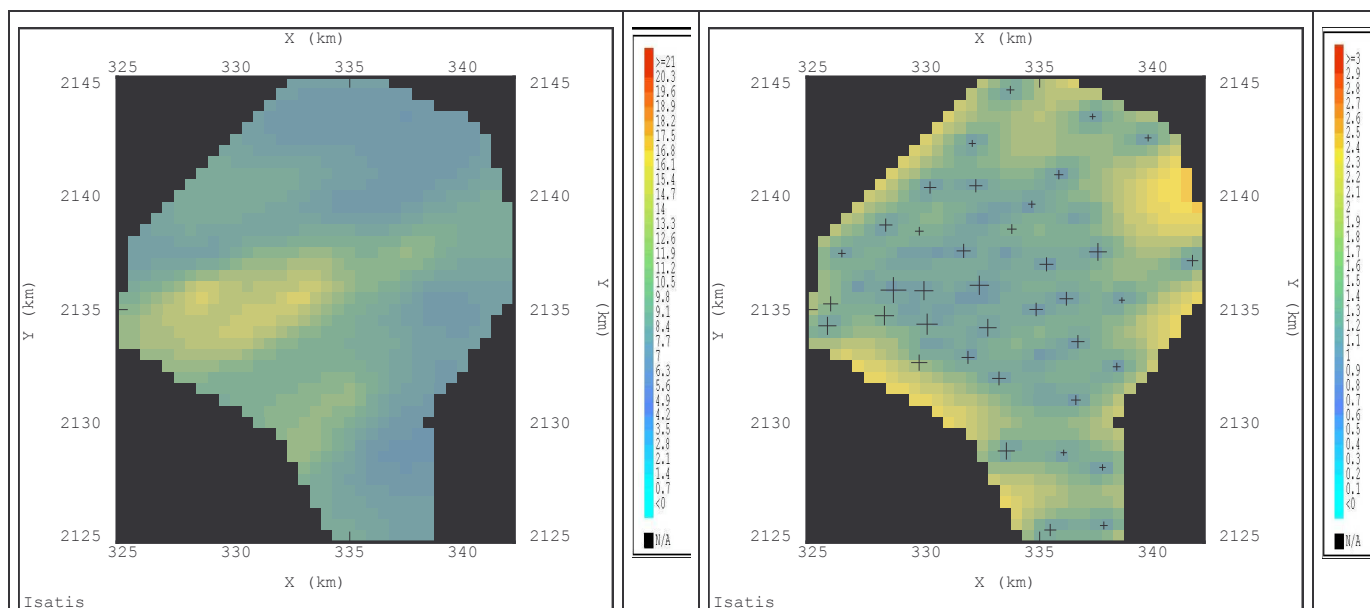


Figure 90 – Estimation par krigeage ponctuel, à partir des données de la sélection 2, avec le modèle bidirectionnel établi sur 46 points (fig. de gauche) et écart-type de krigeage (fig. de droite)

B.V. ETUDE DE LA SELECTION DE DONNEES N°3 (MAILLE DE 2 KM)

B.V.1. Etude du variogramme expérimental

Le variogramme expérimental de la figure 91 comprend 12 pas de 710 m. Il présente un pic à 3km, dû principalement au site 63 : ce point de concentration supérieure à $15 \mu\text{g}/\text{m}^3$ est situé sur la bordure sud d'une zone moins polluée ($<9 \mu\text{g}/\text{m}^3$). Une anisotropie (N70, N160) est encore mise en évidence.

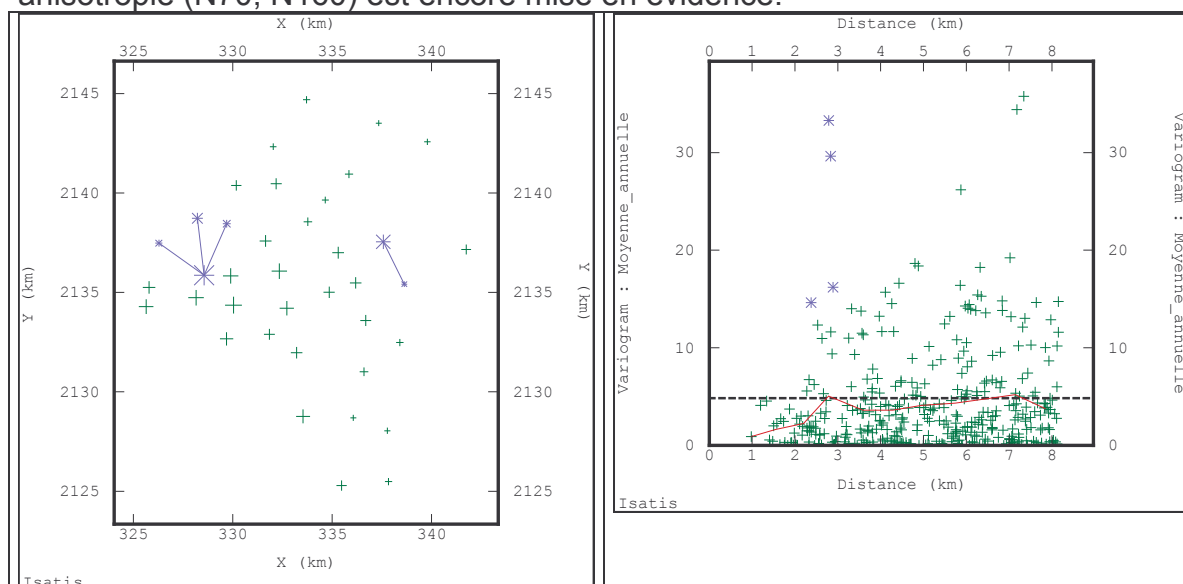


Figure 91– Nuée variographique (pas de 710 m, 12 pas) des données de la sélection 3.

B.V.2. Modélisation du variogramme et test de validation croisée

L'ajustement est effectué avec et sans le point 63, sur les variogrammes omnidirectionnel et bidirectionnel (Figure 92).

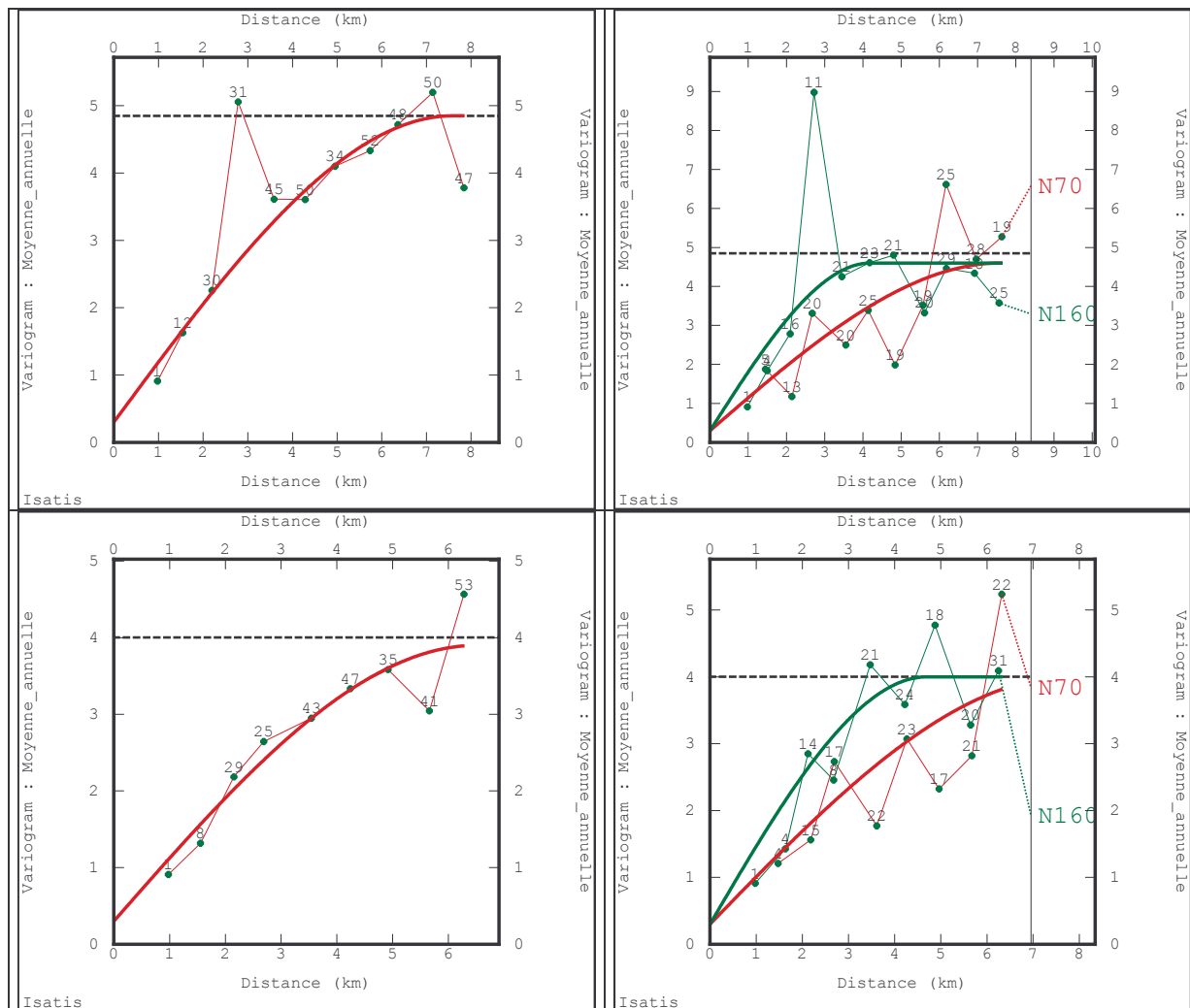


Figure 92 – Variogrammes expérimentaux et modélisés sur 38 sites de mesure (ligne du haut), sur 37 sites de mesure (ligne du bas)

Ces différents modèles sont soumis au test de validation croisée (réestimation sur la totalité des 38 sites à chaque fois). Le modèle bidirectionnel calé sur 38 points est finalement retenu.

Tableau 35 – Résultats de la validation croisée sur les données de la sélection 3 pour le krigeage ordinaire ponctuel, modèle bidirectionnel de la sélection 3. modèle = Pépité (0.3) + Sphérique (N70- 7600 m, N160-4200m; 4.3)

		Validation croisée par krigeage ordinaire ponctuel sur 38 données			
38 données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	38	38	38	38	38
Minimum	7.34	8.18	-4.02	-2.93	3.0
25 ^{ème} percentile	8.23	9.14	-1.21	-0.74	6.53
Médiane	10.53	9.96	0.36	0.22	9.85
75 ^{ème} percentile	11.96	11.51	1.26	0.80	21.20
Maximum	16.28	14.63	3.04	1.79	41.45
Etendue (= Max – Min)	8.94	6.45	-	-	-
Moyenne	10.34	10.38	0.03	0.01	13.99
Ecart-type	2.20	1.59	1.66	1.02	9.40
Variance	4.85	2.53	2.75	1.04	88.43
CV %	21	15			
Corrélation mesures/estimations		$r=0.661$			

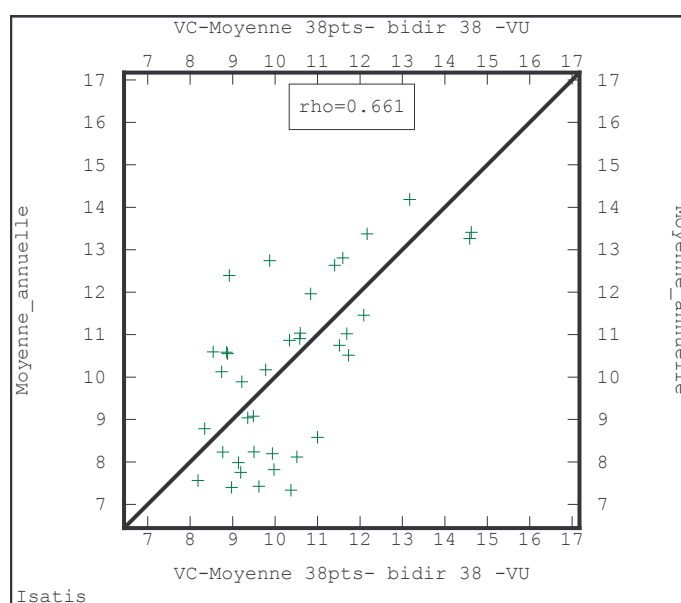


Figure 93 – Validation croisée : nuage de corrélation entre les concentrations estimées et mesurées aux points de la sélection 3

B.V.3. Test de validation absolue sur le complémentaire de la sélection 3

Tableau 36 – Résultats de la validation absolue sur les données de la sélection 3 pour le krigeage ordinaire ponctuel, modèle de la sélection 3

		Validation absolue par krigeage ordinaire ponctuel sur 26 données			
38 données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	26	26	26	26	26
Minimum	7.65	9	-6.85	-5.49	0.31
25 ^{ème} percentile	10.52	10.84	-2.39	-1.63	3.63
Médiane	13.41	12.49	-0.45	-0.35	13.54
75 ^{ème} percentile	14.44	13.81	0.56	0.45	20.55
Maximum	20.66	14.56	3.39	2.41	44.30
Etendue (= Max – Min)	13.01	5.56	-	-	-
Moyenne	13.20	12.24	-0.96	-0.77	14.34
Ecart-type	3.67	1.59	2.50	1.94	11.54
Variance	11.34	2.53	6.24	3.77	133.13
CV %	23	13			
Corrélation mesures/estimations		r=0.712			

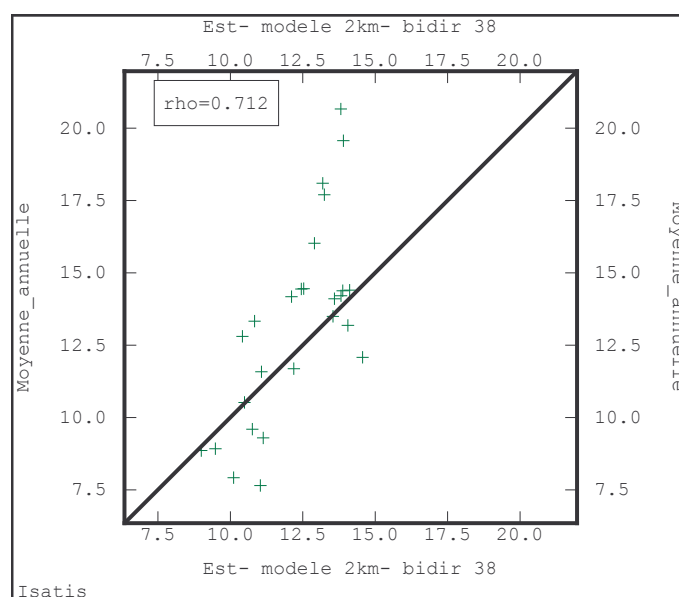


Figure 94 - Nuage de corrélation entre les concentrations estimées et mesurées au points complémentaires de la sélection 32. En bleu : points où les concentrations sont sous-estimées.

B.V.4. Etablissement de la carte pour la sélection 3

La carte d'estimation est plus uniforme que les précédentes.

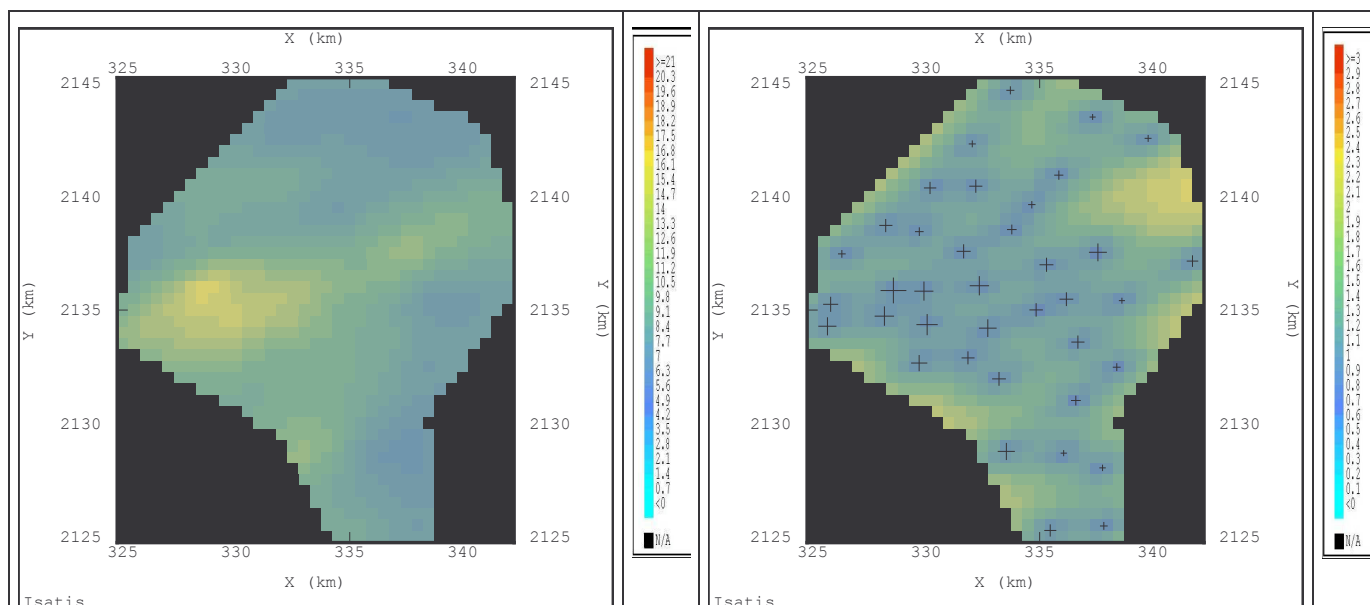


Figure 95 – Estimation par krigeage ponctuel, à partir des données de la sélection 3, avec le modèle bidirectionnel établi sur 38 points (fig. de gauche) et écart-type de krigeage (fig. de droite)

B.V.5. Test de validation absolue sur l'intersection des complémentaires

Trois jeux de données indépendants sont définis :

- l'intersection des complémentaires des sélections 1 et 2 (6 données) ;
- l'intersection des complémentaires des sélections 2 et 3 (12 données) ;
- l'intersection des complémentaires des sélections 1, 2 et 3 (5 données).

Les concentrations sont estimées sur ces intersections à l'aide des données des sélections S1, S2 ou S3 puis elles sont confrontées aux valeurs expérimentales. Dans la quasi totalité des cas, les valeurs estimées s'écartent de plus ou plus de la bissectrice au fur et à mesure que le nombre de points d'échantillonnage diminue. La variance de l'erreur augmente.

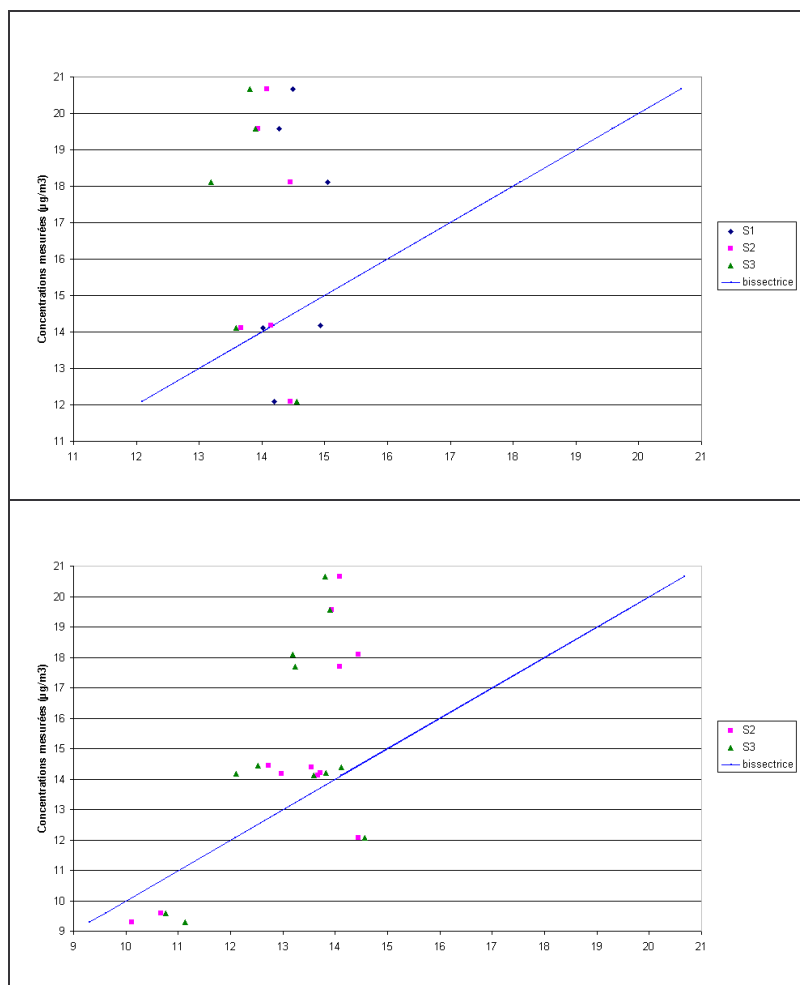


Figure 96- Estimations sur les intersections $C_{S1} \cap C_{S2}$ et $C_{S1} \cap C_{S2} \cap C_{S3}$ (figure du haut) et sur l'intersection $C_{S2} \cap C_{S3}$ (figure du bas)

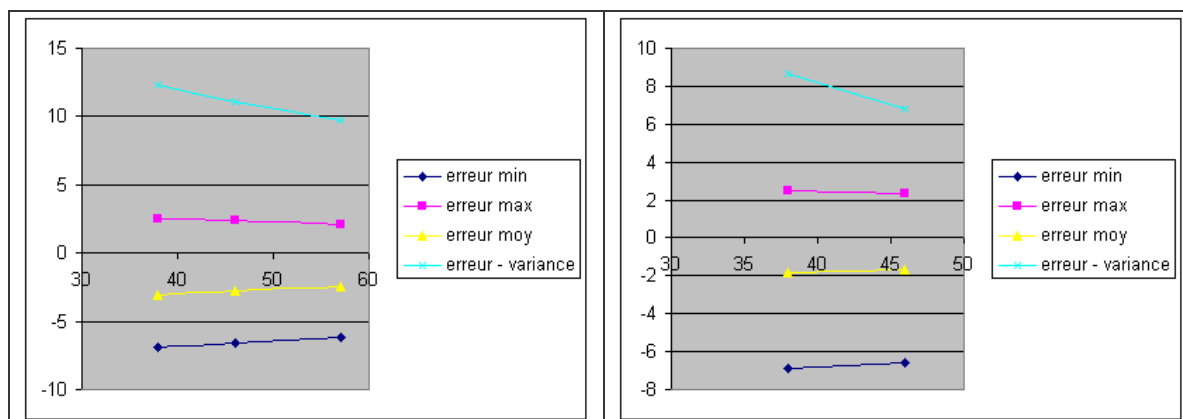


Figure 97- Statistiques d'erreur en fonction du nombre de données – Estimations effectuées sur les intersections $C_{S1} \cap C_{S2} \cap C_{S3}$ (figure de gauche) et sur l'intersection $C_{S2} \cap C_{S3}$ (figure de droite)

B.VI. COMPARAISON DES ESTIMATIONS ET DISCUSSION

B.IV.1. Influence du nombre de données sur le variogramme expérimental

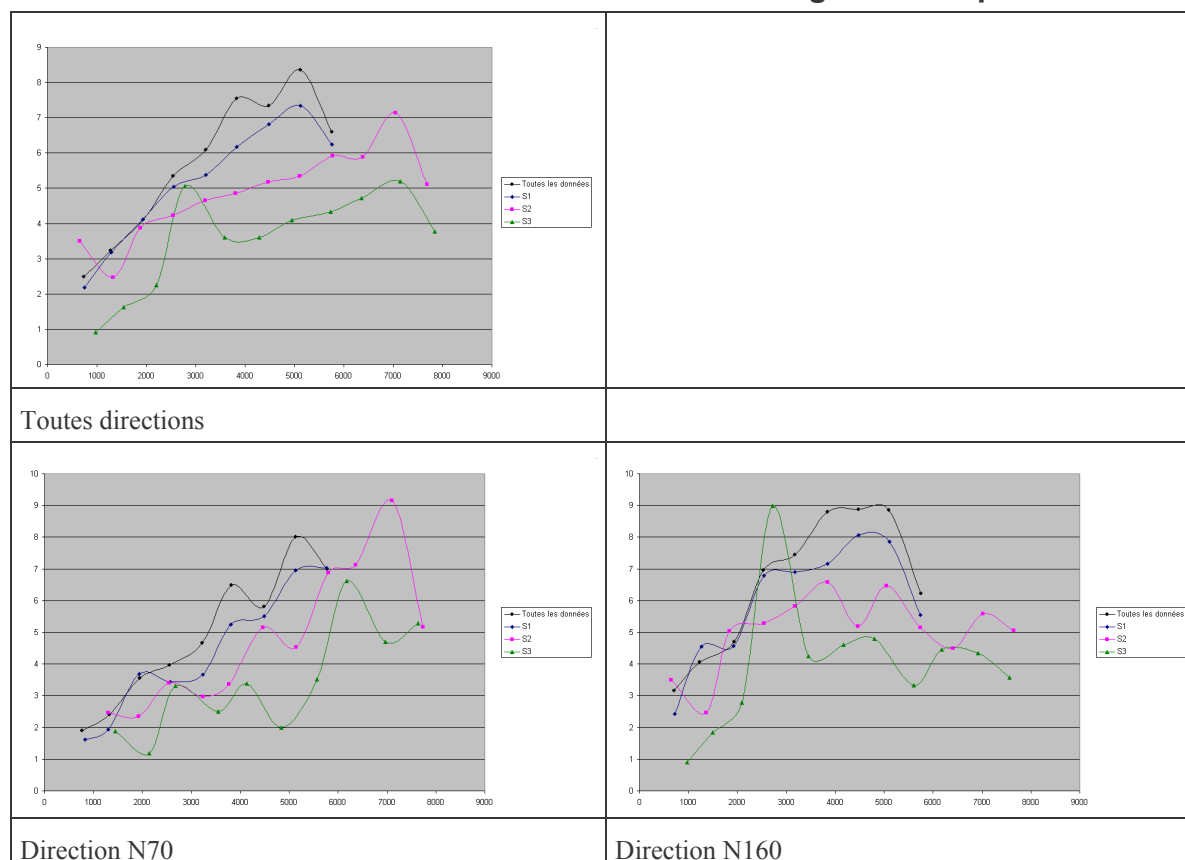


Figure 98 – Variogrammes expérimentaux omnidirectionnels (figure du haut) et bidirectionnels (figures du bas) obtenus pour les trois sélections

Quand la maille d'échantillonnage se relâche, le variogramme expérimental s'écarte du variogramme de référence. Ce changement de structure est plus particulièrement marqué entre la sélection 2 et la sélection 3.

B.VI.2. Influence sur l'estimation du nombre de données de krigeage pour un modèle établi de variogramme

B.VI.2.1. ESTIMATION SUR UN JEU DE DONNEES INDEPENDANT

Le but de ces calculs est d'évaluer isolément l'effet du nombre de données expérimentales sur l'estimation, en supposant la structure spatiale connue.

Le modèle de variogramme utilisé est le modèle de référence.

Les données d'entrée du krigeage sont successivement les sélections S1, S2 et S3.

Les concentrations sont estimées sur les intersections des complémentaires.

Tableau 37 – Estimation sur les intersections des complémentaires avec le variogramme de référence – effet du nombre de données

Jeu de validation	$C_{S1} \cap C_{S2} \cap C_{S3}$		$C_{S1} \cap C_{S2}$		$C_{S2} \cap C_{S3}$	
Nb de données	5		6		12	
Statistique	ρ	$\sigma_{K\text{moy}}$	ρ	$\sigma_{K\text{moy}}$	ρ	$\sigma_{K\text{moy}}$
S1 (57 données)	0.490	1.52	0.255	1.53		
S2 (46 données)	-0.057	1.58	-0.049	1.59	0.752	1.65
S3 (38 données)	-0.445	1.75			0.615	1.78

ρ : corrélation mesures-estimations

$\sigma_{K\text{moy}}$: écart-type moyen de l'erreur de krigeage

A modèle constant, la diminution du nombre de données utilisées dans le krigeage détériore systématiquement la corrélation entre concentrations estimées et mesurées et accroît l'écart-type de l'erreur de krigeage.

B.VI.2.2. ESTIMATION SUR LA GRILLE

Si aux modèles propres à chaque sélection, on substitue le modèle de référence, les estimations sont peu modifiées. Ce résultat traduit la robustesse du krigeage vis-à-vis du modèle variographique. Seule l'estimation de la sélection 3, i.e. de la sélection la plus pauvre en mesures, montre une légère sensibilité au variogramme (Figure 89).

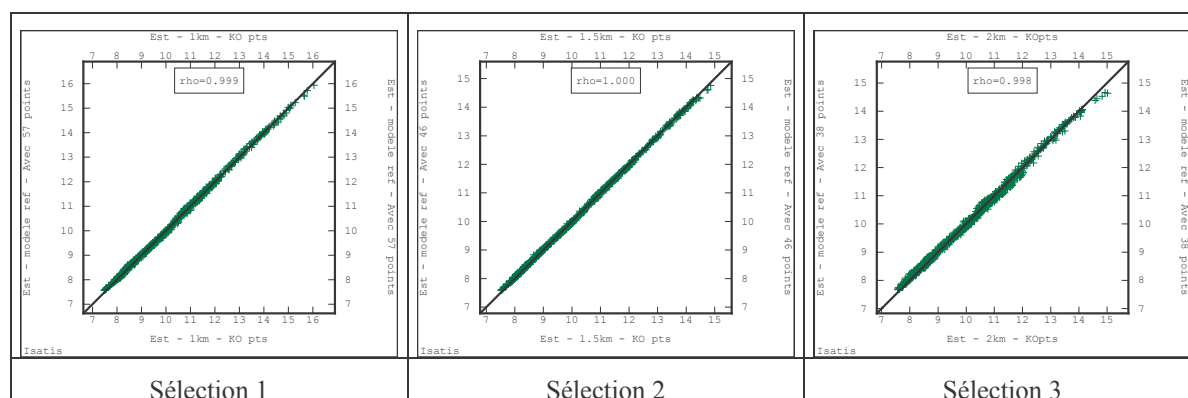


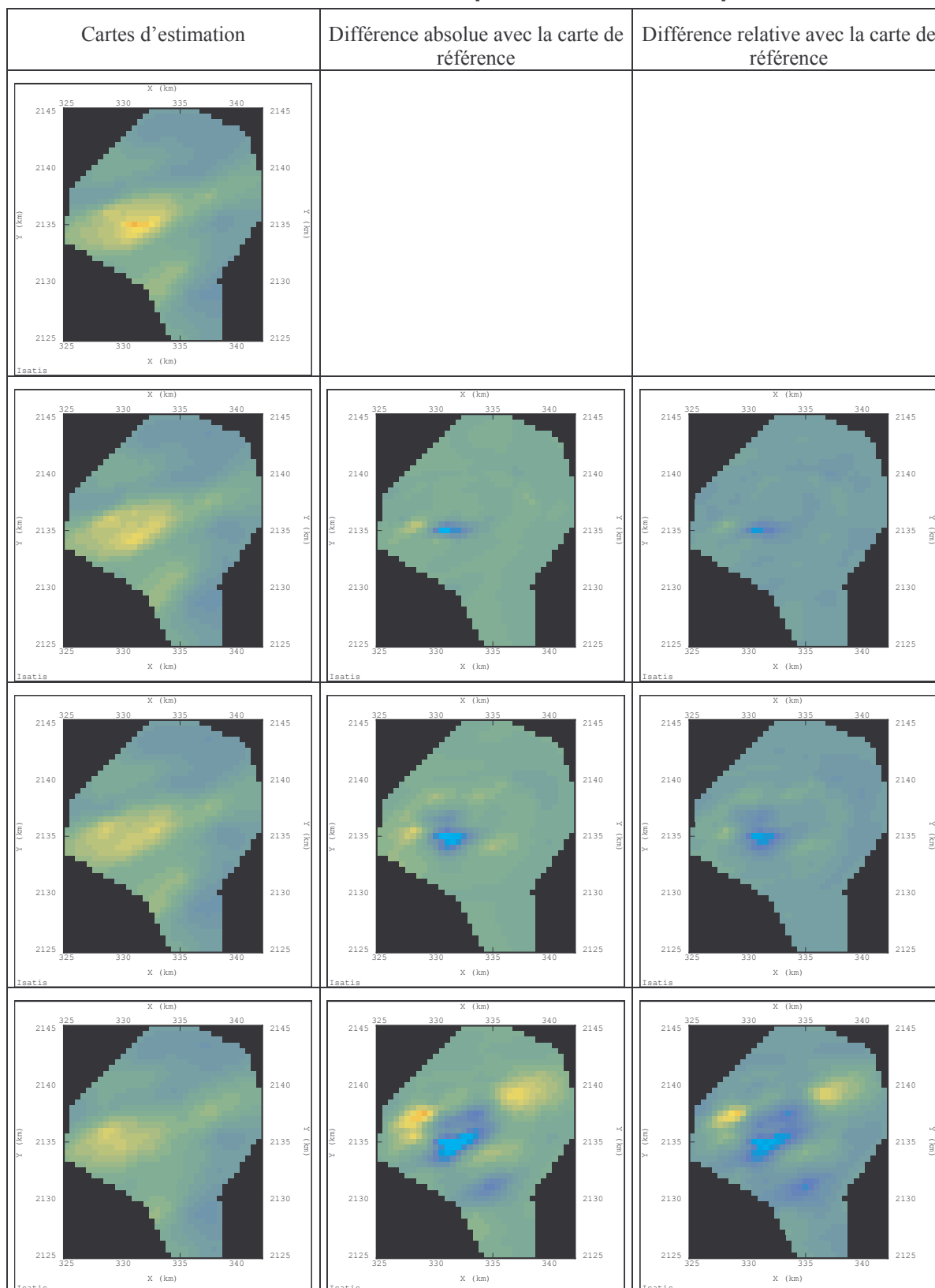
Figure 99 – Comparaison entre les estimations obtenues avec le modèle de référence (Y) et le modèle propre à la sélection (X).

Ainsi la dégradation des cartes est moins liée à celle du variogramme (perte d'information sur la structure spatiale du phénomène) qu'à la diminution du nombre de données. L'évolution à modèle constant de l'écart-type moyen de krigeage l'atteste (on rappelle que cet écart-type ne dépend que du modèle de variogramme et de la configuration des points.)

Tableau 38 – Statistiques de l'écart-type de l'erreur de krigeage. Le modèle de variogramme est le modèle de référence.

	Référence	S1	S2	S3
Min	1.28	1.29	1.38	1.41
Max	2.77	2.77	2.78	2.82
Moy	1.89	1.90	1.93	1.97
Variance	0.089	0.086	0.079	0.073

B.VI.3. Effet du nombre de données expérimentales sur la qualité des cartes



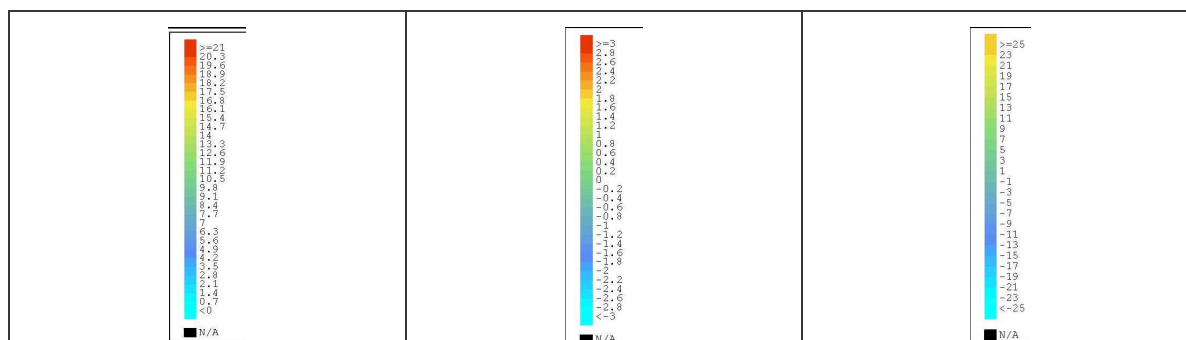


Figure 100 – Comparaison entre les cartes d'estimation obtenues avec un jeu réduit de points (de haut en bas : S1, S2, S3) et la carte de référence

En valeur absolue, l'écart avec la carte de référence reste limité à 3 $\mu\text{g}/\text{m}^3$. En valeur relative, il peut atteindre 25 % sur une surface d'autant plus étendue que le nombre de points est plus faible. Pour les sélections 1 et 2, la dégradation affecte la zone la plus polluée ; pour la sélection 3, elle touche également les zones périphériques, avec un net effet de lissage: diminution des estimations dans le centre-ville, augmentation dans la partie nord du domaine.

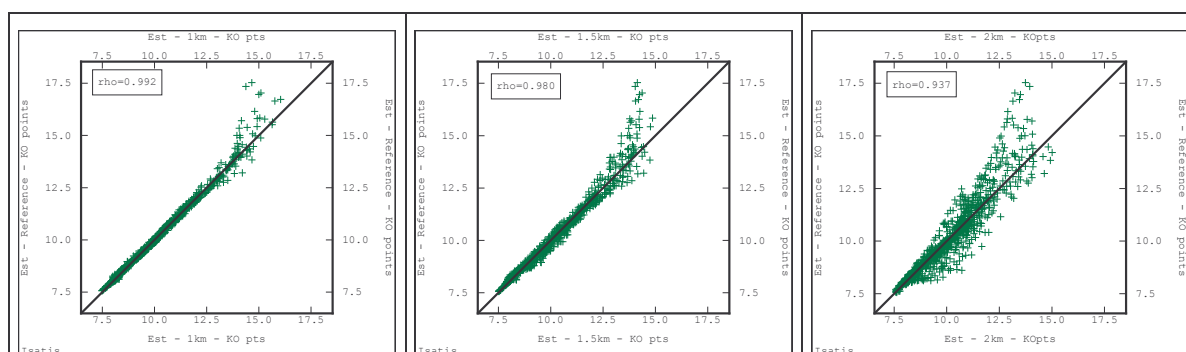


Figure 101 – Comparaison entre l'estimation de référence et les estimations obtenues à partir des trois sélections étudiées (intérieur du polygone de La Rochelle)

Tableau 39 – Statistiques de l'estimation de référence et des estimations obtenues à partir des trois sélections étudiées (intérieur du polygone de La Rochelle)

	Mesures ($\mu\text{g}/\text{m}^3$)	Ref	Sélection 1km	Sélection 1.5km	Sélection 2km
Nombre de données	64	934	934	934	934
Minimum	7.34	7.56	7.48	7.52	7.57
25 ^{ème} percentile	8.89	8.61			
Médiane	10.96	9.72	9.75	9.81	9.91
75 ^{ème} percentile	13.39	11.12	11.13	11.19	11.02
Maximum	20.66	17.53	16.04	14.86	15.01
Moyenne	11.50	10.11	10.07	10.08	10.05
Ecart-type	3.07	1.86	1.81	1.75	1.59
Variance	9.45	3.45	3.28	3.06	2.54
CV %	27	18	18	17	16

B.VII. CALCULS COMPLEMENTAIRES : UTILISATION DES EMISSIONS DE NOX EN VARIABLE AUXILIAIRE – IMPACT SUR LES ESTIMATIONS

Cette partie complémentaire vise à répondre aux questions : l’inventaire des émissions de NOx permet-il d’améliorer la précision de l’estimation et de pallier la dégradation de l’échantillonnage ? Les recommandations sur la maille sont-elles les mêmes qu’en l’absence de telles données ?

B.VII.1. Etude du jeu de données complet

B.VII.1.1. CORRELATION ENTRE NO₂ ET DENSITE D’EMISSIONS DE NOX

L’inventaire fourni par ATMO Poitou-Charentes couvre la totalité du domaine d’étude. Il représente la quantité de NOx (en tonnes) émis par maille de 500m de côté. A partir de ces données brutes, quatre grilles de valeurs ont été créées¹ :

- densité d’émissions de NOx dans une maille de 500m de côté ;
- densité d’émissions de NOx dans une fenêtre carrée glissante de 1500m de côté ;
- densité d’émissions de NOx dans une fenêtre carrée glissante de 2500m de côté ;
- densité d’émissions de NOx dans une fenêtre carrée glissante de 3500m de côté.

Les densités d’émissions aux 64 points de mesure ont été estimées par migration à partir de la grille². Après transformation logarithmique, la densité d’émissions dans une fenêtre de 2500m est la plus corrélée à la concentration de NO₂. Le nuage de corrélation apparaît néanmoins dispersé. Un cokrigage colocalisé est ici mis en œuvre.

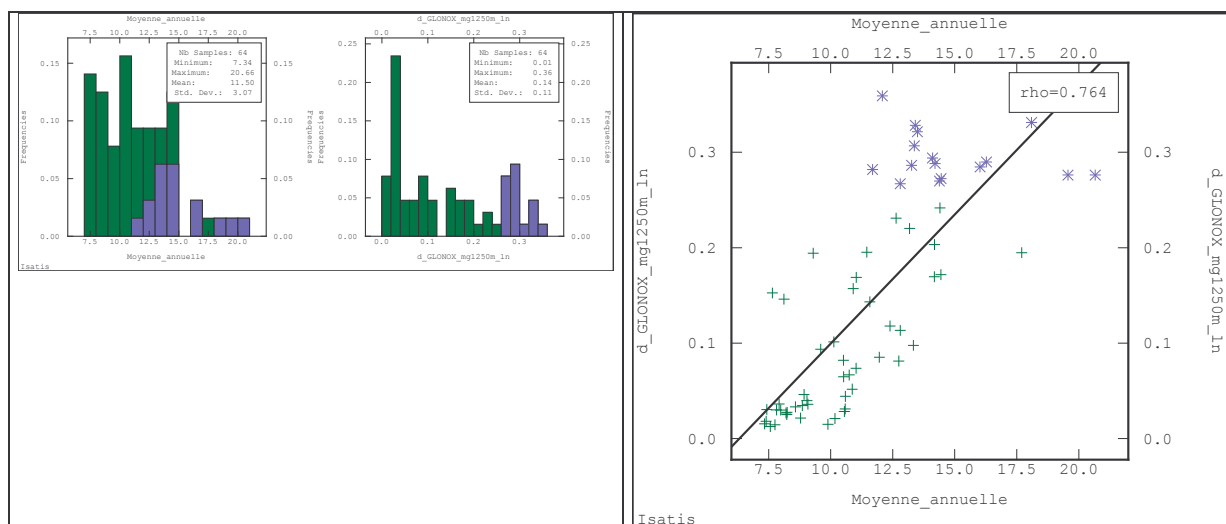


Figure 102 – Relation entre concentration de NO₂ et logarithme de la densité d’émissions dans une fenêtre de 2500m de côté

Remarque : A partir des données de population mises à disposition par ATMO Poitou-Charentes, des densités de population dans un rayon de 250 m, 1km et 2km ont été également

¹ Pour des raisons de temps de calcul, les densités d’émissions sont calculées dans des fenêtres carrées glissantes (opération aisément réalisable dans Isatis) au lieu de fenêtres circulaires (SIG requis).

² Pour les raisons évoquées ci-dessus, les densités d’émissions aux points expérimentaux sont obtenues par migration à partir de la grille plutôt que par un rigoureux calcul d’intersection de mailles. L’approximation est cependant limitée, un site d’échantillonnage étant distant d’au plus 354m d’un centre de maille.

calculées. Comme pour la densité d'émissions, la corrélation entre la concentration de NO₂ et ces trois variables (après transformation logarithmique) est d'environ 0,7-0,8 avec un nuage de corrélation relativement dispersé. La densité de population, bien corrélée à la densité d'émissions ($\rho=0,87$), montre une certaine redondance par rapport à cette dernière et n'est pas ici exploitée.

B.VII.1.2. MODELISATION DU VARIOGRAMME ET TEST DE VALIDATION CROISEE

Comme en krigeage ordinaire monovarié, le modèle bidirectionnel est le plus satisfaisant. Son ajustement ne tient pas compte des deux points de plus forte concentration qui perturbent la structure. Il se compose d'un effet de pépite, d'une structure cubique et d'une structure sphérique. Les résultats de validation croisée sont légèrement meilleurs que dans le cas monovarié. Seule l'erreur standardisée est supérieure, ce qui s'explique par de plus faibles écarts-types de krigeage (la variable d'émissions étant connue au point cible).

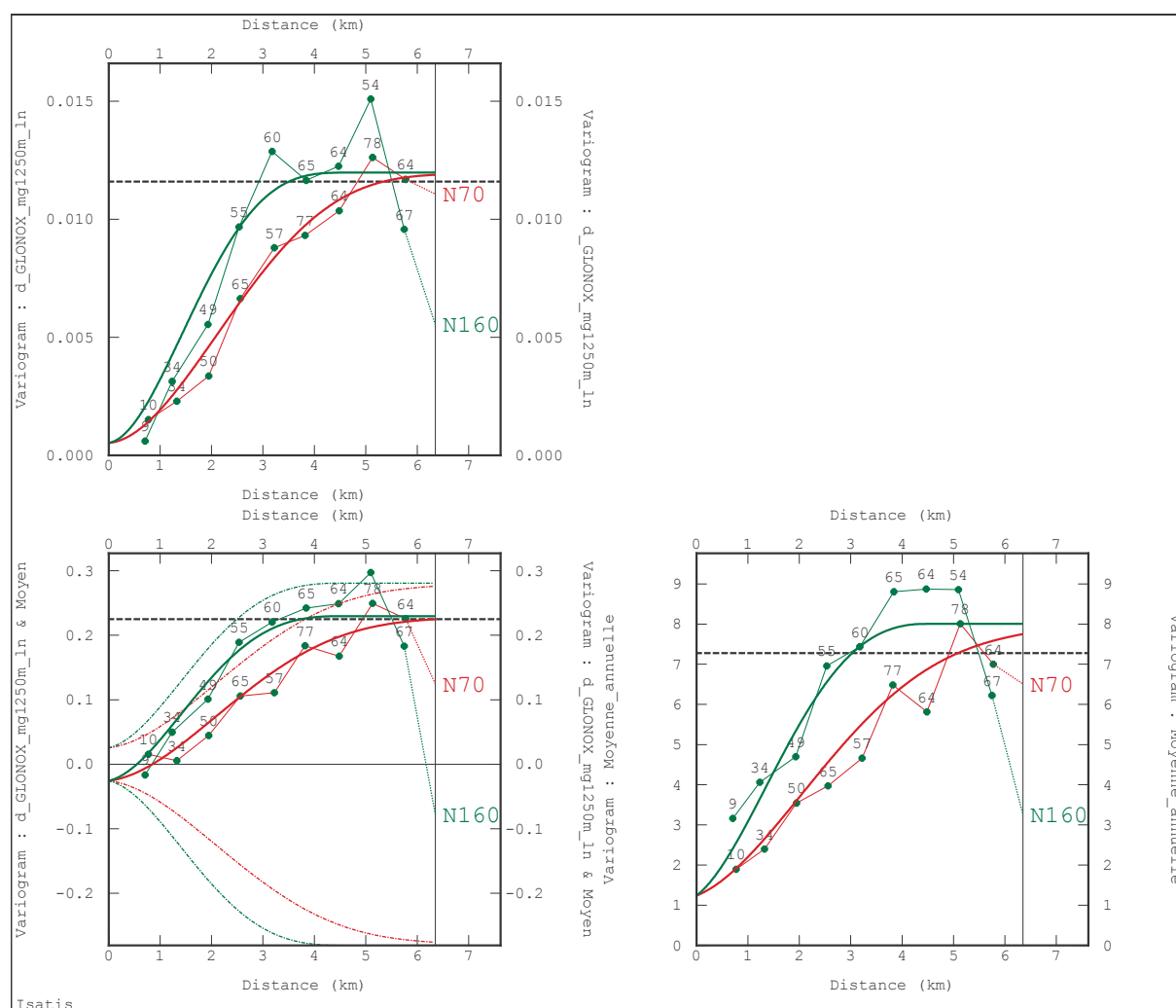


Figure 103 – Modèle bidirectionnel de corégionalisation

Tableau 40 – Résultats de la validation croisée sur toutes les données pour le cokrigage colocalisé, modèle de référence bidirectionnel

Toutes les données	Mesures ($\mu\text{g}/\text{m}^3$)	Validation croisée par krigage ordinaire ponctuel sur 64 données			
		Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	64	64	64	64	64
Minimum	7.34	8.26	-3.75	-4.09	0.01
25 ^{ème} percentile	8.89	9.23	-1.24	-0.72	4.43
Médiane	10.96	10.84	0.22	0.14	9.34
75 ^{ème} percentile	13.39	13.89	1.09	0.69	15.78
Maximum	20.66	17.27	3.35	2.95	34.74
Etendue (= Max – Min)	13.32	9.01			
Moyenne	11.50	11.57	0.07	0.03	11.10
Ecart-type	3.07	2.53	1.58	1.22	7.96
Variance	9.45	6.40	2.48	1.50	63.29
CV %	27	23			
Corrélation mesures/estimations		r=0,859			

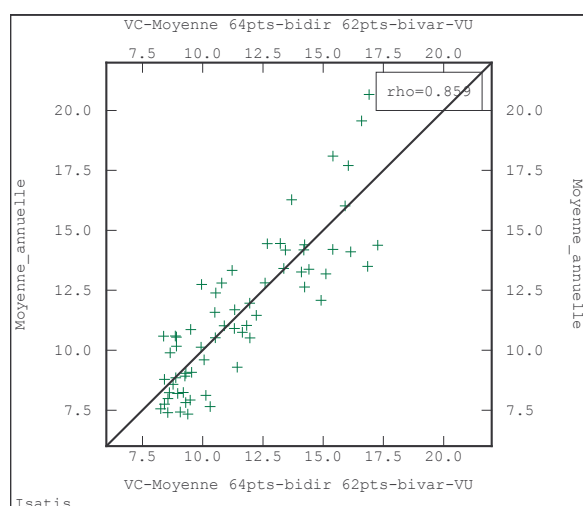


Figure 104 – Validation croisée : nuage de corrélation entre les concentrations mesurées et estimées sur les 64 points urbains

B.VII.1.3. ELABORATION DE LA CARTE DE REFERENCE

La grille d'estimation, de résolution 500m, coïncide avec la grille des émissions. La carte d'estimation obtenue ressemble à la cartographie de référence établie par krigage ordinaire. Les niveaux de concentration sont légèrement rehaussés dans le centre et en bordure ouest de la ville (zone sans mesures).

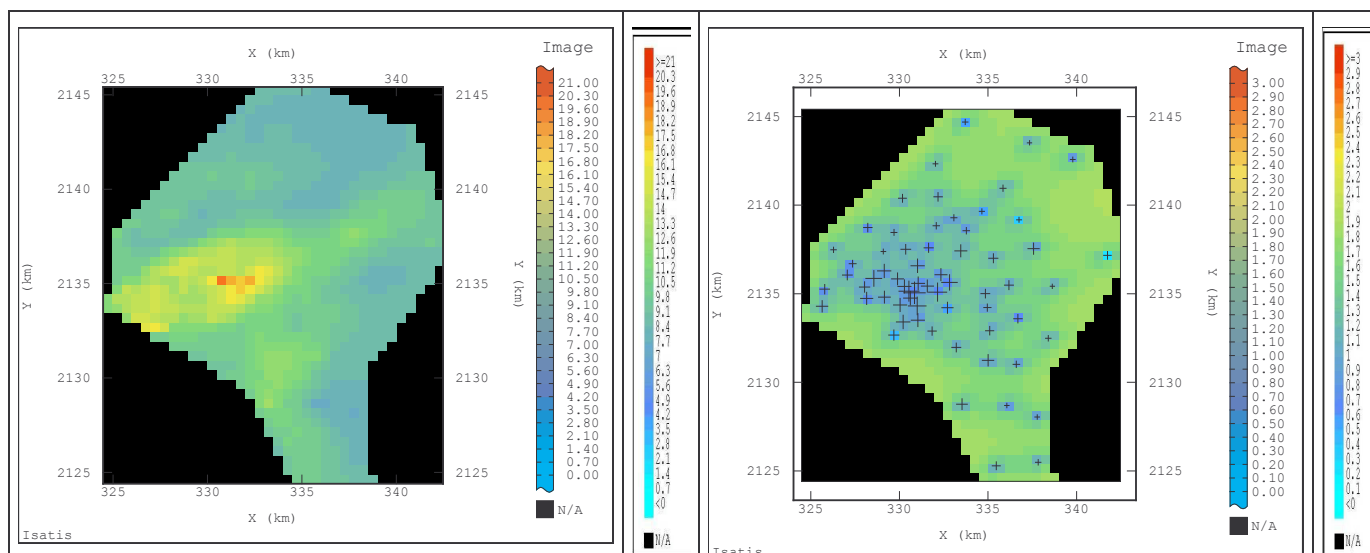


Figure 105 – Estimation par cokrigage colocalisé, à partir des 64 données urbaines, avec le modèle bidirectionnel établi sur 62 points de mesure (fig. de gauche) et écart-type de krigage (fig. de droite)

B.VII.2. Etude des sélections n°2 (maille de 1,5km) et n°3 (maille de 2km)

B.VII.2.1. CORRELATION ENTRE NO_2 ET DENSITE D'EMISSIONS DE NOX

Pour les deux sélections, la racine carrée de la densité d'émissions dans une fenêtre de 2500m est la plus corrélée à la concentration de NO_2 (Figure 106).

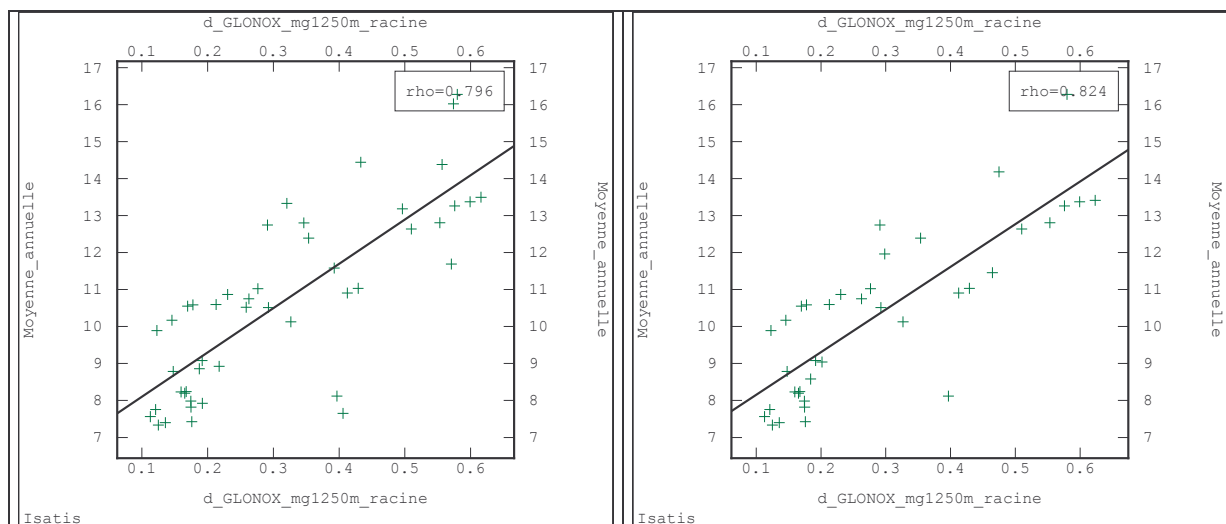


Figure 106 – Relation entre concentration de NO_2 et logarithme de la densité d'émissions dans une fenêtre de 2500m de côté. A gauche : sélection 1,5km ; à droite : sélection 2km

B.VII.2.2. MODELISATION DU VARIOGRAMME ET TEST DE VALIDATION CROISEE

Pour la sélection 2, des modèles mono- et bidirectionnels sont définis (Figure 107). Pour la sélection 3, les variogrammes bidirectionnels simples et croisé se laissent moins aisément modéliser ; seul un modèle monodirectionnel est ajusté (Figure 108).

Dans tous les cas, les modèles se composent d'un effet de pépite et d'une structure sphérique. Ils sont testés par validation croisée (Tableaux 40 et 41, Figure 109). Pour la sélection 2, on se

limite à présenter le modèle bidirectionnel auquel les statistiques d'erreurs sont légèrement plus favorables.

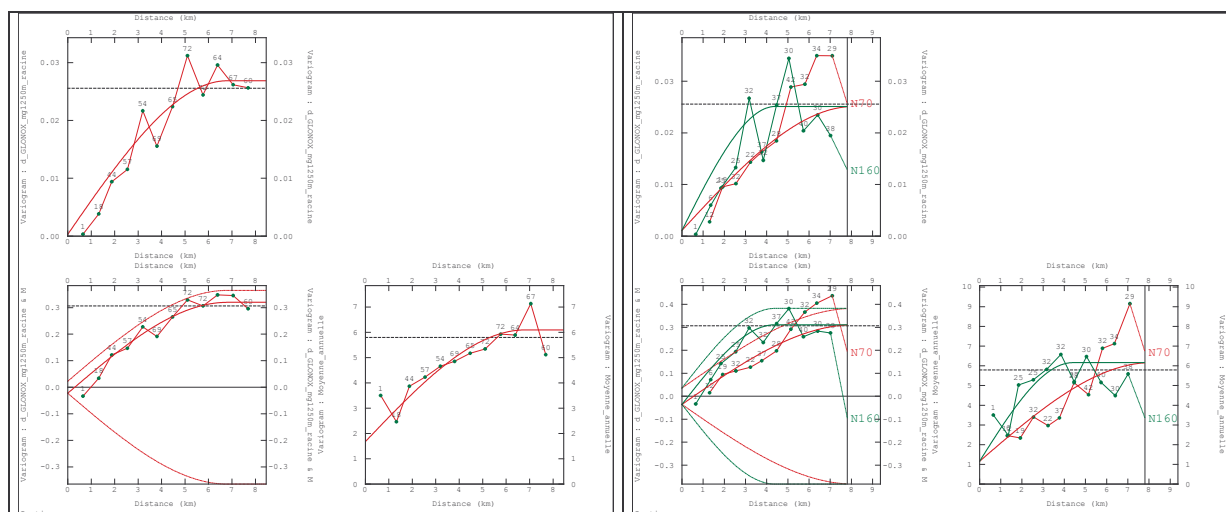


Figure 107 – Sélection 2. Modèles mono et bidirectionnels de corégionalisation

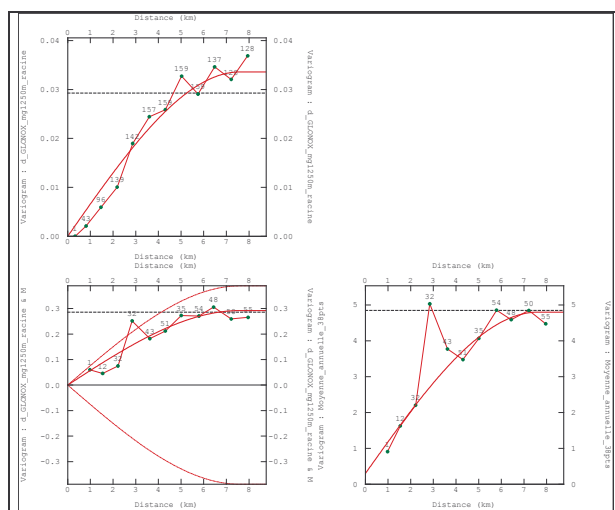


Figure 108 – Sélection 3. Modèle monodirectionnel de corégionalisation

Tableau 40 – Résultats de la validation croisée sur les 46 données de la sélection 2 pour le cokrigage colocalisé, modèle bidirectionnel

46 données	Mesures ($\mu\text{g}/\text{m}^3$)	Validation croisée par krigage ordinaire ponctuel sur 64 données			
		Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	46	46	46	46	46
Minimum	7.34	7.95	-2.83	-1.88	1.02
25 ^{ème} percentile	8.23	8.80	-1.45	-0.95	4.51
Médiane	10.57	10.30	0.30	0.21	8.35
75 ^{ème} percentile	12.74	12.00	0.89	0.59	15.71
Maximum	16.28	14.83	4.33	2.98	56.55
Etendue (= Max – Min)	8.94	6.78			
Moyenne	10.64	10.63	-0.03	0.00	11.42
Ecart-type	2.41	2.13	1.50	1.01	10.11
Variance	5.79	4.56	2.25	1.01	102.26
CV %	23	20			
Corrélation mesures/estimations		r=0,791			

Tableau 41 – Résultats de la validation croisée sur les 38 données de la sélection 3 pour le cokrigage colocalisé, modèle bidirectionnel

Validation croisée par krigage ordinaire ponctuel sur 64 données					
38 données	Mesures (µg/m³)	Valeurs estimées (µg/m³)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	38	38	38	38	38
Minimum	7.34	7.56	-3.29	-3.51	0.00
25 ^{ème} percentile	8.23	8.68	-0.59	-0.52	2.83
Médiane	10.53	10.21	0.00	0.00	8.45
75 ^{ème} percentile	11.96	11.82	0.00	0.78	17.97
Maximum	16.28	15.04	3.70	3.19	45.55
Etendue (= Max – Min)	8.94	7.52			
Moyenne	10.34	10.35	0.00	0.00	10.95
Ecart-type	2.20	1.86	1.45	1.35	10.32
Variance	4.85	3.45	2.10	1.81	106.52
CV %	21	18			
Corrélation mesures/estimations		r=0,757			

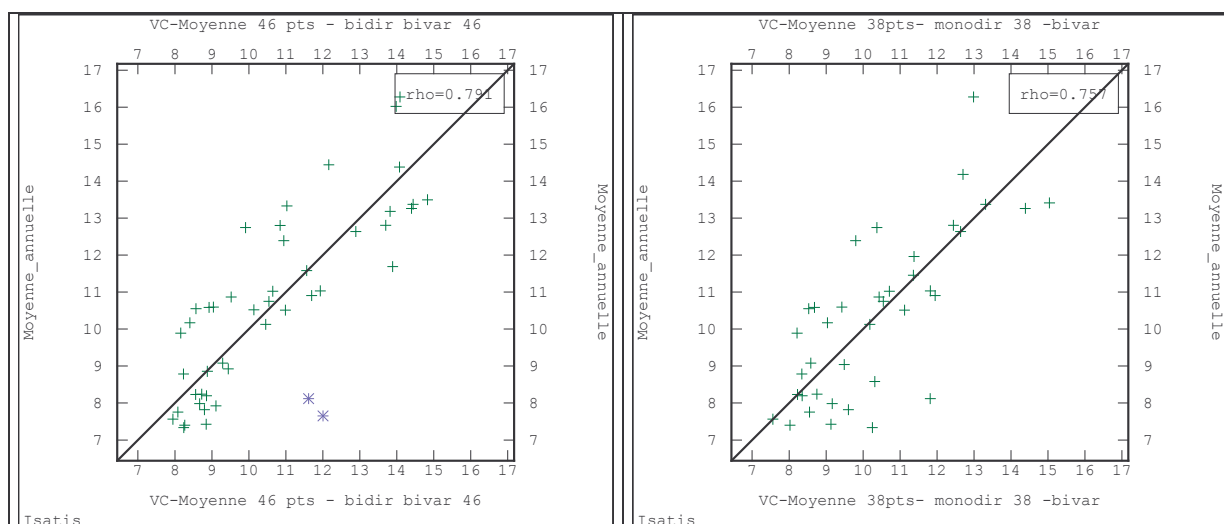


Figure 109 – Validation croisée : nuage de corrélation entre les concentrations mesurées et estimées sur les 46 points de la sélection 2 (figure de gauche), sur les 38 points de la sélection 3 (figure de droite).

La prise en compte des émissions par cokrigage réduit les erreurs de validation croisée ; les statistiques des valeurs estimées sont plus proches de celles des mesures.

En deux points de la sélection 2 cependant, les niveaux de concentration sont surestimés d'environ 40% à 50%. Ces deux points situés près du littoral se caractérisent par des densités d'émissions assez élevées mais de faibles concentrations. Il s'agit des deux points qui, à la figure 106, se détachent du nuage de corrélation.

B.VII.2.3. TEST DE VALIDATION ABSOLUE

Les concentrations sont estimées :

- sur les 18 points complémentaires de la sélection 2, en utilisant les 46 données et le modèle bidirectionnel relatifs à cette sélection ;
- sur les 26 points complémentaires de la sélection 3, en utilisant les 38 données et le modèle monodirectionnel relatifs à cette sélection.

A la différence de la validation croisée, L'examen des résultats ne montre pas d'amélioration par rapport au krigeage monovarié (Figures 89, 94, 110) :

- Des concentrations élevées de centre-ville (points V16, V18, V25, V28) restent sous-estimées.
- Des concentrations bien estimées en krigeage ordinaire sont surestimées en cokrigeage (les deux points en bleu dans le complémentaire de la sélection 2, situés dans des zones de plus forte densité d'émissions).

Si le modèle bivariable construit avec 62 points est substitué aux modèles propre à chaque sélection, on note un petit resserrement des nuages.

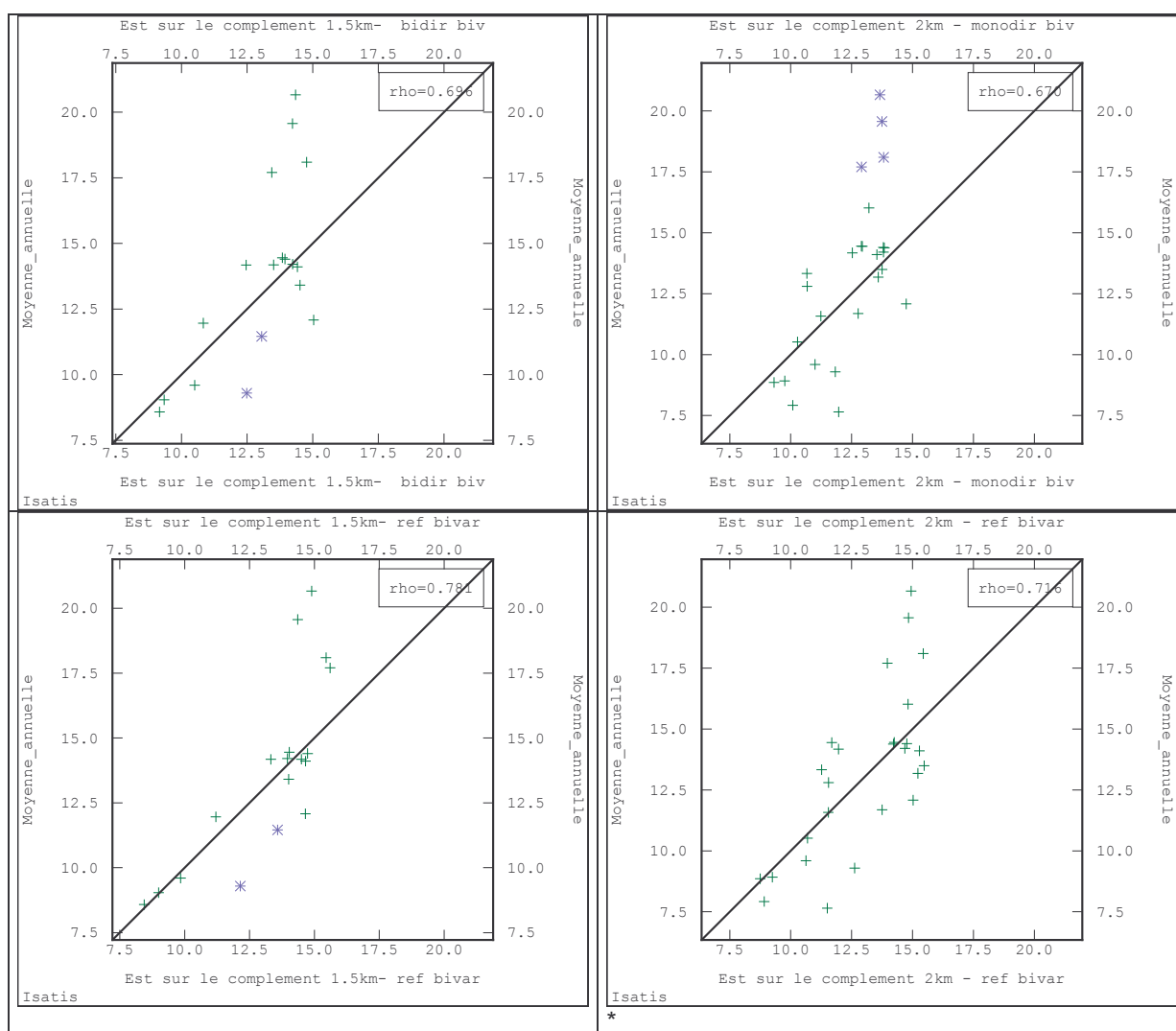


Figure 110 – Nuage de corrélation entre les concentrations mesurées et estimées aux points complémentaires de la sélection 2 (à gauche) et de la sélection 3 (à droite).
Ligne du haut : avec les modèles propres à chaque sélection (modèle bivariable bidirectionnel pour la sélection 2 et bivariable monodirectionnel pour la sélection 3).
Ligne du bas : avec le modèle bivariable de référence (62 points)

B.VII.2.4. ELABORATION DES CARTES POUR LES DEUX SELECTIONS

Comme dans le cas de référence, les estimations par cokrigeage ressemblent à celles du krigeage ordinaire mais la prise en compte des émissions diminue l'effet de lissage (pour la

sélection 2, la variance des estimations est de 3.06, respectivement 3.58, en krigeage et cokrigeage, pour une moyenne de 10.08, respectivement 9.98 ; pour la sélection 3, cette variance est de 2.54, respectivement 2.90, en krigeage et cokrigeage, pour une moyenne de 10.05, respectivement 10.07.)

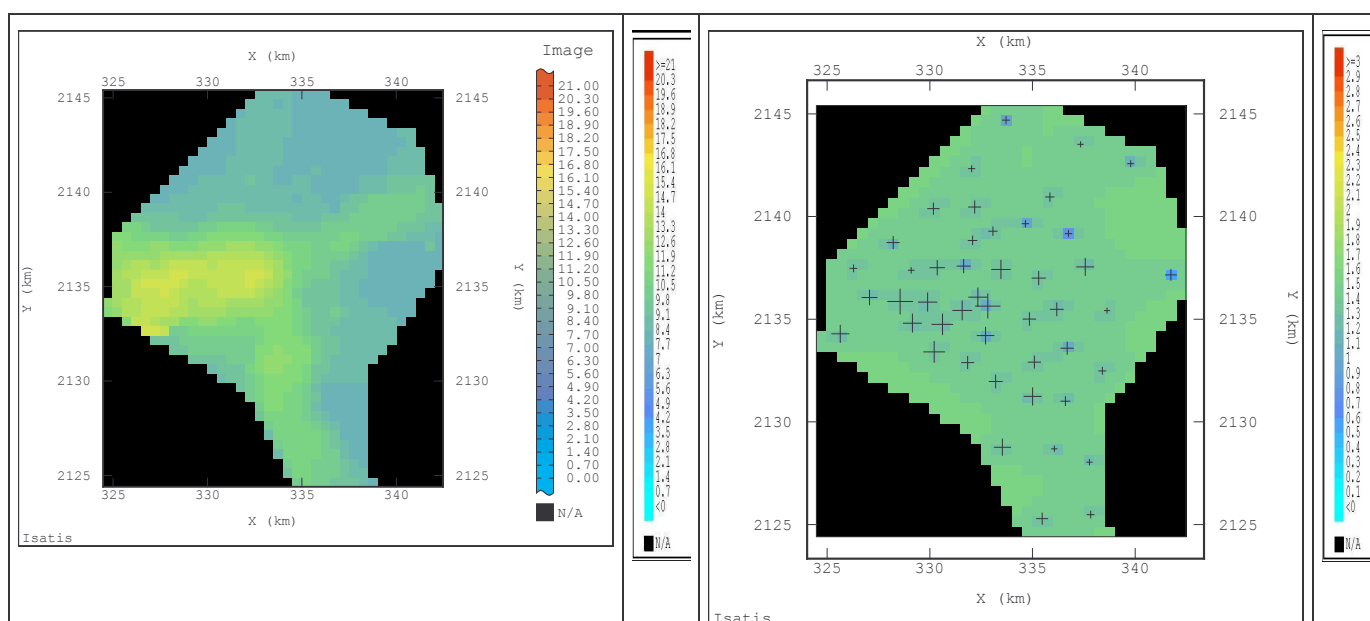


Figure 111 – Sélection 2. Estimation par cokrigeage colocalisé avec le modèle bidirectionnel établi sur 46 points de mesure (fig. de gauche) et écart-type de krigeage (fig. de droite)

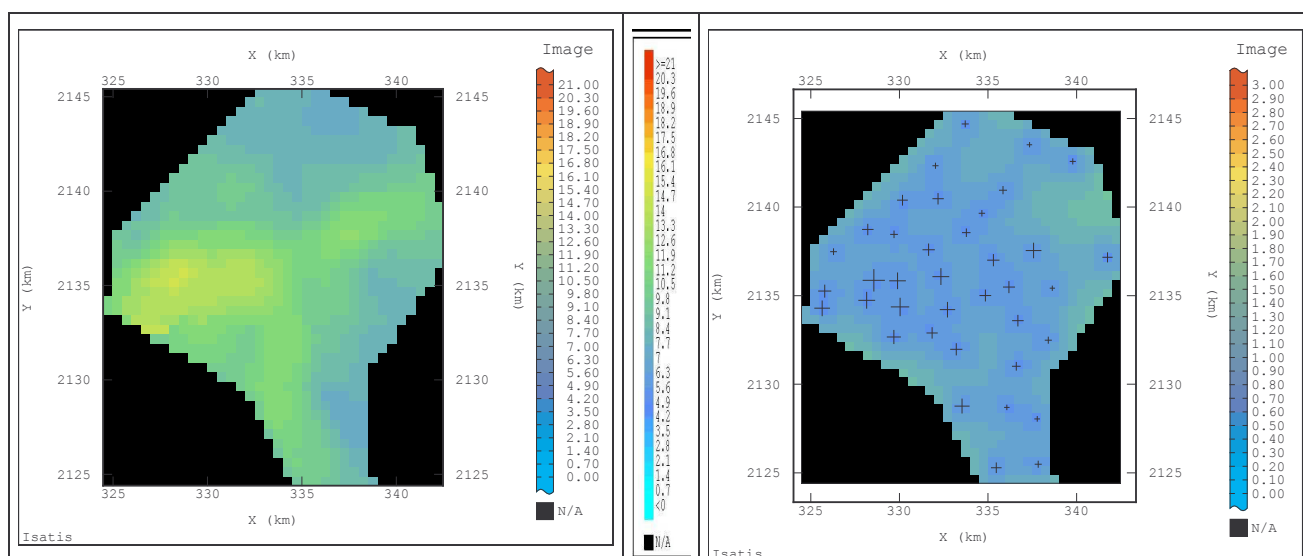


Figure 112 – Sélection 3. Estimation par cokrigeage colocalisé avec le modèle monodirectionnel établi sur 38 points de mesure (fig. de gauche) et écart-type de krigeage (fig. de droite)

Remarque : bien que les données soient moins nombreuses, l'écart-type de krigeage de la sélection 3 est inférieur en moyenne à celui de la sélection 2. Ce résultat s'explique par le variogramme de la concentration : du fait de la suppression de quelques points de centre-ville, l'effet de pépite et le palier total sont abaissés.

B.VII.2.5. BILAN

Les résultats obtenus nous conduisent à réitérer les préconisations émises en krigeage ordinaire. Une cinquantaine de points correspondant à une maille d'échantillonnage de 1 à 1,5 km de côté s'impose pour construire un modèle bivarié satisfaisant. Si le cokrigeage révèle son efficacité, notamment en validation croisée, les tests effectués montrent l'influence possible de la covariable et l'importance de soigneusement définir cette variable auxiliaire.

Annexe 4 – Urbain : Le Robert

NO₂ / Etude d'une ville de moins de 100 000 habitants : cas du Robert (Martinique)
--

1. DESCRIPTION DU CAS	68
2. ETUDE DU JEU DE DONNÉES INITIAL	69
2.1 ETABLISSEMENT DU JEU DE DONNÉES DE RÉFÉRENCE J ₁	69
2.2 ETABLISSEMENT DU JEU DE DONNÉES DE RÉFÉRENCE J ₂	71
3. CONCLUSION.....	74
4. RÉFÉRENCES	74

DESCRIPTION DU CAS

Les données ont été recueillies auprès de l'Association agréée de surveillance de la qualité de l'air en Martinique (Madininair), suite à trois campagnes de quatorze jours de mesure des concentrations de NO₂ dans l'air ambiant par échantillonnage passif sur la commune du Robert entre la mi-octobre et la fin mai 2002.

Cette commune située sur le littoral Est de l'île est la 3^{ème} ville de Martinique avec 21 240 habitants (INSEE, recensement 1999) sur une superficie de 47,3 km². C'est un territoire de plaine coupé de petits hameaux et avec un littoral très découpé. Le bourg du Robert se niche au centre d'un profond havre enserré par des mornes : plages, mangroves, forêts tropicales.

103 sites de mesure ont été répartis sur le territoire de la commune dans un domaine d'environ 10 km d'Est en Ouest, et 8 km du Nord au Sud, avec un resserrement des points de mesure au niveau du centre-ville (Fig. 1).



Figure 1 – Commune du Robert et implantation des sites de mesure (carte fournie par l'Office de tourisme)

Les valeurs prises en compte pour cette étude correspondent à la moyenne des trois campagnes ou, lorsqu'une seule valeur est inexistante, à la moyenne des deux valeurs restantes. On obtient alors un jeu de données de 101 points. Les détails de cette campagne sont disponibles dans le rapport de Madininair [1].

ETUDE DU JEU DE DONNEES INITIAL

L'étude géostatistique est réalisée grâce au logiciel Isatis (version 5.0.0).

Etablissement du jeu de données de référence J_1

Dans un premier temps, l'étude variographique a montré que 34 sites perturbaient notablement le variogramme expérimental. Ces 34 sites ont été écartés du jeu de données initial J_0 (Fig. 2), correspondant :

- d'une part, aux mesures effectuées à proximité de la circulation automobile (principalement le long des axes routiers et dans les bourgs et hameaux), qui présentent une hétérogénéité plus importante et donc accroissent la variabilité spatiale des concentrations ;
- et d'autre part, aux valeurs moyennes de concentrations de NO_2 inférieures à la limite de quantification définie pour la méthode des tubes passifs, soit $7 \mu\text{g.m}^{-3}$ [2].

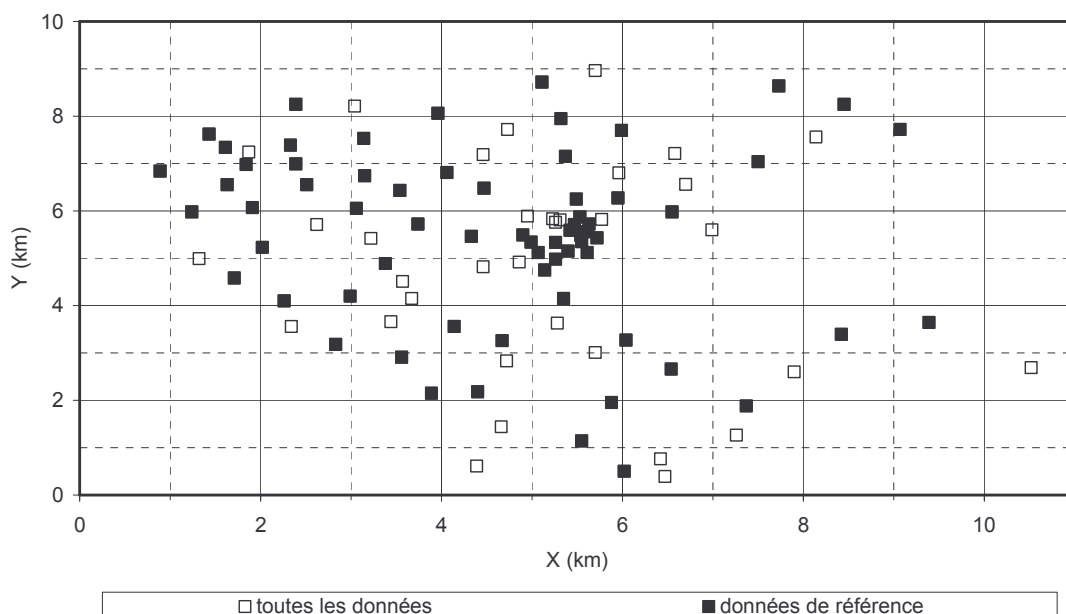


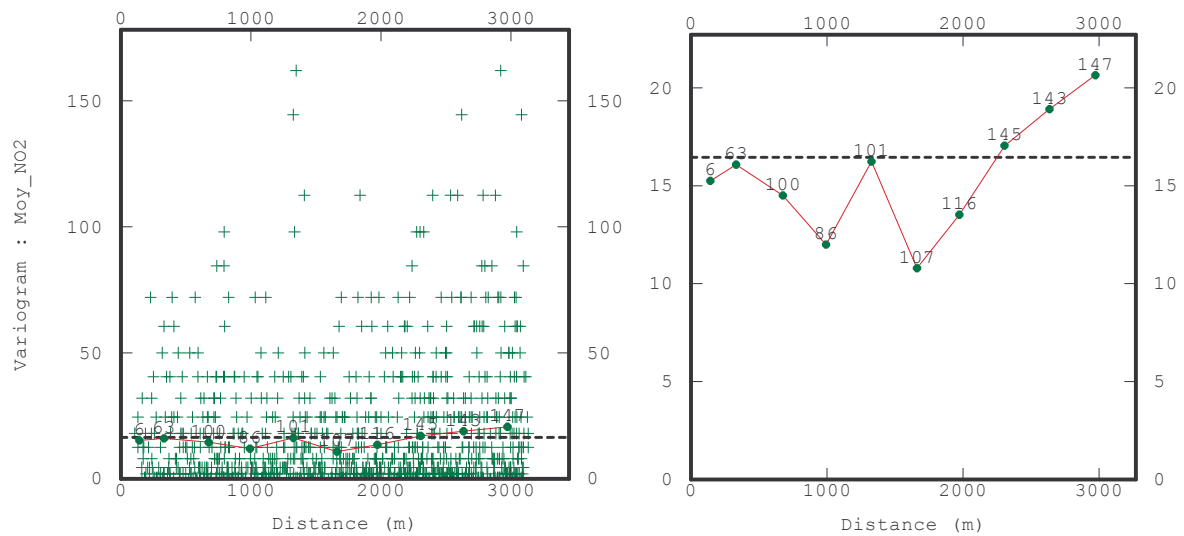
Figure 2 – Carte de répartition des 67 données de référence et des 34 données non sélectionnées

Le jeu de données de référence J_1 se compose donc de 67 points de mesure, dont les caractéristiques statistiques sont présentées dans le Tableau 1.

Tableau 1 – Caractéristiques statistiques du jeu de données initial J_0 et de celui de référence J_1

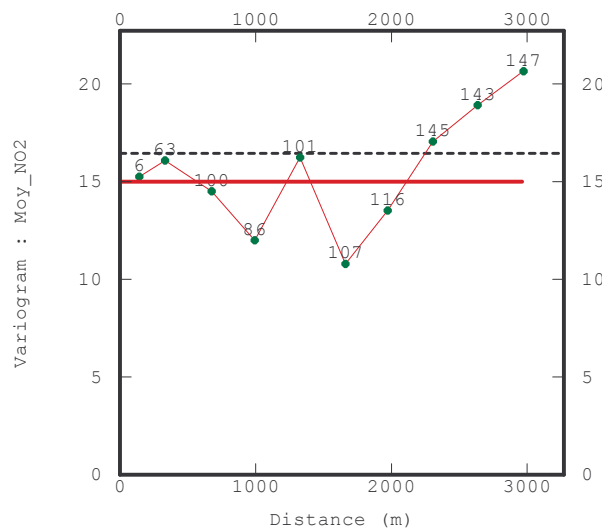
		J_0	J_1
$\mu\text{g/m}^3$	Nombre de données	101	67
	Minimum	1	7
	Percentile 25	6	9
	Médiane	10	11
	Percentile 75	14	14
	Maximum	41	25
	Moyenne	11,32	12
	Ecart-type	7,38	4,06
	Variance	54,44	16,45
	Coefficient de variation	0,65	0,34

Le variogramme retenu est omnidirectionnel avec 10 pas de 330 m chacun (Fig. 3)



**Figure 3 – (gauche) Nuée variographique des 67 données de référence
(droite) Variogramme omnidirectionnel (pas : 330 m, 10 pas)**

Quel que soit le pas retenu, le variogramme expérimental n'est correctement modélisé que par une structure pépétique pure, dont l'effet est proche de la valeur de la variance du jeu de données de référence (16,45), représentée en traits pointillés sur la Figure 4.



**Figure 4 – Variogramme modélisé sur 67 sites de mesure
Modèle = pépité (15)**

Ce modèle traduit une absence de corrélation spatiale entre les différentes valeurs de NO_2 , même pour des sites d'échantillonnage proches.

A noter que les variogrammes expérimentaux ont été également calculés indépendamment pour chaque campagne, menant de nouveau à des structures pépétiques pures.

Etablissement du jeu de données de référence J₂

Dans une deuxième approche, les données ont été examinées au vu des conditions météorologiques spécifiées par Madinainair. Lors des campagnes 1 et 3, des épisodes pluvieux ont en effet entraîné une diminution des concentrations en NO₂ sur la majorité des sites. Les sites ne présentant pas cette tendance ont été invalidés. De plus, les coefficients de variation pour chaque site ont été calculés (bien que le nombre de mesures par site, 3, soit légèrement insuffisant), et le critère préconisé par le Laboratoire européen de référence de la pollution de l'air (ERLAP) a été utilisé, à savoir : les sites dont le CV est compris entre 10 et 30% sont directement validés ; ceux présentant un CV supérieur à 30% ou inférieur à 10% doivent être examinés selon les cas proposés dans le tableau 2.

Tableau 2 – Cas des sites dont les valeurs de CV sont extérieures au domaine 10-30% [2]

Cas	CV	[NO ₂]	Type de site
1	> 30%	faible	en périphérie, loin des sources
2	< 10%	forte	centre-ville, influencé de manière homogène
3	< 10%	faible	éloigné de toute source
4	> 30%	forte	en périphérie, soumis à une source fixe selon le vent

Dans notre cas, la valeur moyenne des concentrations de NO₂ observées pour les 3 campagnes est de 11 µg m⁻³. Nous considérons donc que la valeur de [NO₂] est « faible » lorsqu'elle est inférieure ou égale à cette valeur, et « forte » au dessus.

Le tableau 3 présente le nombre de sites correspondant à chaque critère, ainsi que le nombre de sites finalement validés pour les coefficients de variation hors du domaine [10-30%]. Les sites invalidés correspondent pour les cas 1 et 3 aux sites de centre-ville et de proximité automobile, et pour le cas 4 aux sites de centre-ville.

Tableau 3 – Nombre de sites selon les critères liés au coefficient de variation et à la concentration en NO₂

Critères		Nombre de sites	
		initial	validés
<i>météorologique</i>		101	79
10% ≤ CV ≤ 30%		19	19
CV < 10%	[NO ₂] ≤ 11 µg m ⁻³	5	2
	[NO ₂] > 11 µg m ⁻³	1	1
CV > 30%	[NO ₂] ≤ 11 µg m ⁻³	37	28
	[NO ₂] > 11 µg m ⁻³	17	8
<i>Total</i>		79	58

Le jeu de données est réduit à 58 valeurs moyennes de concentrations de NO₂. Les données écartées correspondent principalement aux sites du Bourg, à ceux du hameau du Vert-Pré au nord-ouest de la commune, et aux axes de circulation (voir figure ci-dessous). Enfin

PROJET

pour l'analyse du variogramme expérimental, 3 sites de proximité automobile n'ont pas été pris en compte.

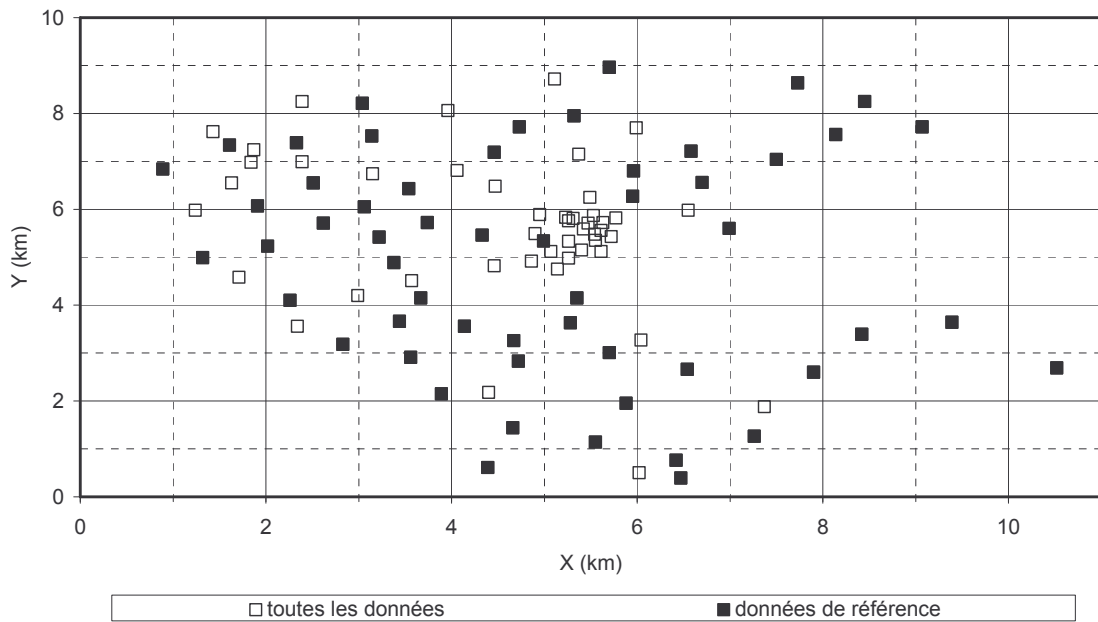


Figure 5 – Carte de répartition des 55 données de référence et des 46 données non sélectionnées

Le tableau 4 présente les statistiques relatives à ce nouveau jeu de données de référence. Dans ce cas, le critère correspondant à la limite de quantification de $7 \mu\text{g m}^{-3}$ n'a pas été retenu, pour garder un nombre de données suffisant pour l'étude géostatistique (Fig. 6).

Tableau 4 – Caractéristiques statistiques du jeu de données initial J_0 et de ceux de référence J_1 et J_2

		J_0	J_1	J_2
$\mu\text{g/m}^3$	Nombre de données	101	67	55
	Minimum	1	7	3
	Percentile 25	6	9	5
	Médiane	10	11	7
	Percentile 75	14	14	10
	Maximum	41	25	16
	Moyenne	11,34	12	7,96
	Ecart-type	7,42	4,06	3,34
	Variance	55,11	16,45	11,18
	Coefficient de variation	0,65	0,34	0,42

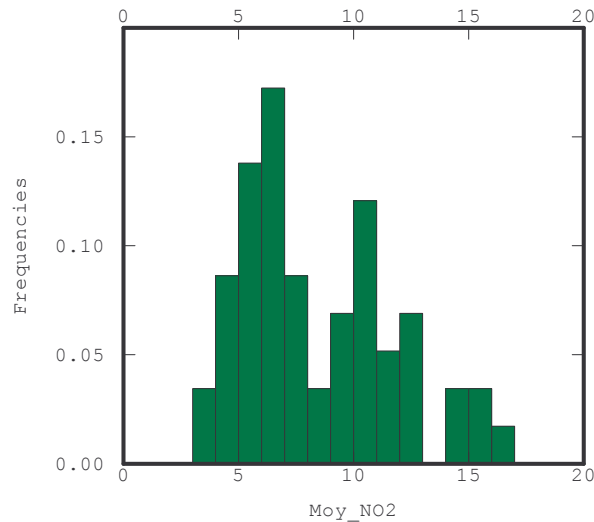


Figure 6 – Répartition des données J_2 en classes d'étendue 1 ; 25 sites présentent une concentration moyenne de NO_2 inférieure à $7 \mu\text{g m}^{-3}$

Le variogramme présenté est omnidirectionnel avec 15 pas de 200 m chacun (Fig. 7)

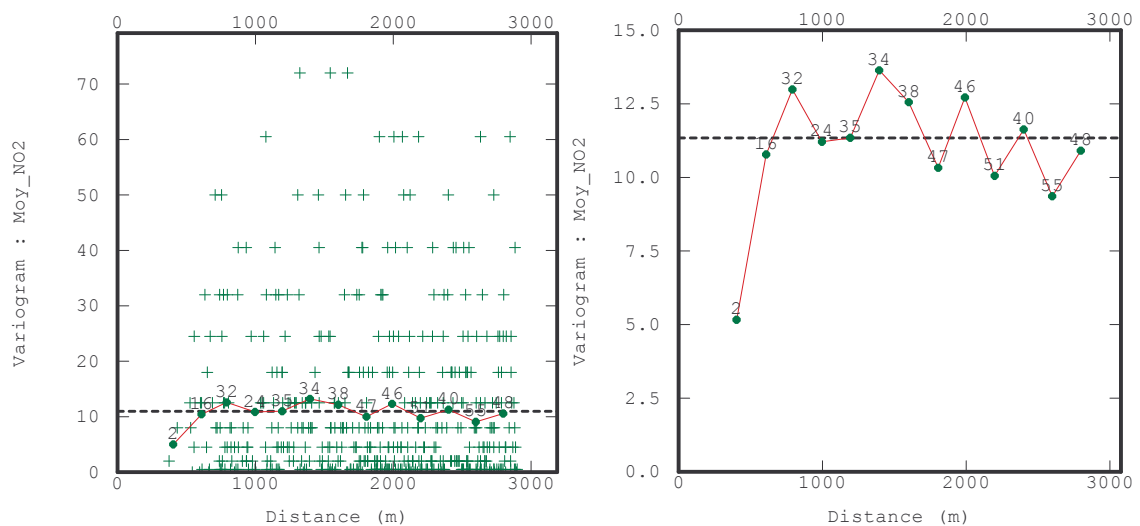


Figure 7 – (gauche) Nuée variographique des 55 données de référence (droite) Variogramme omnidirectionnel (pas : 200 m, 15 pas)

Quel que soit le pas retenu, on obtient à nouveau un variogramme expérimental pour lequel une modélisation correcte s'effectue à l'aide d'une structure pépitique pure, dont l'effet est égal à la variance du jeu de données.

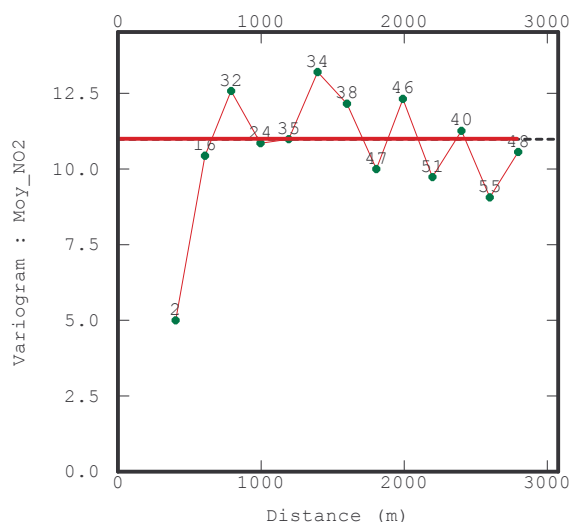


Figure 8 – Variogramme modélisé sur 55 sites de mesure
Modèle = pépité (11)

CONCLUSION

Les deux approches destinées à constituer un jeu de données de référence pour l'étude des concentrations moyennes de NO₂ sur la commune du Robert, à partir des données expérimentales obtenues lors de trois campagnes par échantillonnage passif, ont conduit à un résultat équivalent.

Le variogramme expérimental est correctement modélisé par une structure pépétique pure, indiquant une absence de corrélation spatiale même pour des points situés à faible distance. Une explication possible peut être avancée considérant le jeu de données qui contient un grand nombre de valeurs inférieures ou proches de la limite de quantification de 7 µg m⁻³.

Dans le cas d'un effet pépétique pur, l'estimation des concentrations peut se faire par les méthodes statistiques traditionnelles, à savoir la moyenne et l'écart-type. On peut donc estimer qu'avec une cinquantaine de tubes (pour obtenir un jeu de données statistiquement représentatif) répartis régulièrement sur le territoire de la commune du Robert, on obtiendrait une valeur moyenne et un écart-type qui serait une estimation satisfaisante des concentrations de NO₂.

REFERENCES

- [1] Madinainair, juin 2003 – Etude de la qualité de l'air au Robert : étude par tubes passifs ; étude par camion laboratoire – *S. Gandar*.
- [2] LCSQA-EMD, mai 2000 – Synthèse de l'expérience acquise par les réseaux sur l'échantillonnage passif du NO₂ – *S. Garcia-Fouqué, J.-L. Houdret, H. Plaisance*.

Annexe 5 – Région : Champagne-Ardenne

TABLE DES MATIERES

1. DESCRIPTION DU CAS.....	77
1.1 Présentation du jeu de données initial.....	77
1.2 Présentation des données auxiliaires.....	78
1.3 Méthodologie de sélection des données	78
• 1.3.1 Echantillonnage régulier	78
• 1.3.2 Echantillonnage stratifié en fonction des émissions de NOx.....	80
1.4 Périodes d'échantillonnage traitées	82
2. ETUDE DU JEU DE DONNÉES INITIAL ET ÉTABLISSEMENT DE LA CARTE DE RÉFÉRENCE.....	83
2.1 Etude des quatre périodes de mesure et de la moyenne (hiver).....	83
• 2.1.1 Histogrammes.....	83
• 2.1.2 Corrélations.....	85
2.2 Etude du variogramme expérimental (hiver)	86
2.3 Corrélation avec les variables de coordonnée	87
2.4 Corrélation avec les variables auxiliaires	88
2.5 Modélisation géostatistique et validation croisée	89
• 2.5.1 Cas monovarié.....	89
2.5.1.1 Etude variographique	89
2.5.1.2 Estimation sur toute la région.....	90
• 2.5.2 Cas multivarié	93
2.5.2.1 Régression multilinéaire	93
2.5.2.2 Etude variographique	94
2.5.2.3 Estimation sur toute la région.....	96
3. DÉGRADATION DE L'ÉCHANTILLONNAGE : EVOLUTION DU VARIOGRAMME EXPÉRIMENTAL ET DE LA RELATION NO₂/VARIABLES AUXILIAIRES - HIVER	97
3.1 Evolution du variogramme expérimental	97
3.2 Evolution de la relation avec les variables auxiliaires.....	98
4. DÉGRADATION DE L'ÉCHANTILLONNAGE : IMPACT SUR LA MODÉLISATION DU VARIOGRAMME ET SUR L'ESTIMATION - HIVER ...	99
4.1 Krigeage ordinaire.....	99
• 4.1.1 Modélisation du variogramme	99
• 4.1.2 Validation croisée.....	101

PROJET

• 4.1.3 Vérification sur les jeux de données complémentaires	101
• 4.1.4 Krigeage sur la grille d'estimation	104
4.2 Krigeage des résidus.....	105
• 4.2.1 Modélisation du variogramme.....	106
• 4.2.2 Validation croisée.....	106
• 4.2.3 Vérification sur les jeux de données complémentaires	107
• 4.2.4 Krigeage sur la grille d'estimation	108
4.3 Complément sur l'incertitude d'estimation.....	112
5. STRATIFICATION DE L'ÉCHANTILLONNAGE EN FONCTION DES ÉMISSIONS	113
6. ETUDE SANS LA DENSITÉ DE POPULATION	117
7. ETUDE DE L'ÉTÉ.....	120
7.1 Etude du jeu de données initial	120
• 7.1.1 Histogrammes.....	120
• 7.1.2 Cas monovarié : modélisation du variogramme – validation croisée.....	121
• 7.1.3 Cas bivarié : utilisation des données hivernales.....	122
• 7.1.4 Cas multivarié : modélisation du variogramme – validation croisée	124
• 7.1.5 Estimation sur la grille	125
7.2 Etude d'une sélection	127
• 7.2.1 Estimation par krigeage des résidus	127
• 7.2.2 Remarque subsidiaire sur le cokrigeage avec la moyenne hivernale ..	129
7.3 Commentaires sur l'estimation des concentrations estivales	130
8. RÉFÉRENCES	131

Etude d'une région : surveillance du dioxyde d'azote en Champagne-Ardenne

DESCRIPTION DU CAS

Présentation du jeu de données initial

Le jeu de données étudié a été fourni par l'association agréée de surveillance ATMO Champagne-Ardenne. Il est issu d'une campagne de mesure par échantillonnage passif conduite durant l'année 2005 sur l'ensemble de la région.

Il se compose de huit campagnes de mesure, quatre campagnes consécutives en été et quatre campagnes consécutives en hiver, d'une durée de 14 jours chacune.

La région Champagne-Ardenne s'étend sur une superficie totale de 25 606 km² dont 61% de surface agricole utilisée et 27% de bois et forêts. Elle compte huit agglomérations de plus de 20000 habitants dont deux de plus de 120000 habitants (Reims et Troyes), deux de plus de 60000 habitants (Charleville-Mézières et Châlons-en-Champagne), deux de plus de 30000 habitants (Saint-Dizier et Epernay) et deux de plus de 25000 habitants (Sedan et Chaumont). Sa population totale est d'environ 1,34 million d'habitants, avec une densité moyenne de 52 habitants/km² [source : CCI Champagne-Ardenne].

235 sites de mesure ont été installés dans la région, sur un domaine d'environ 185 km d'est en ouest et de 280 km du nord au sud (Figure). Les points de mesure se distribuent de façon homogène dans l'espace, selon un pas d'échantillonnage de 15km dans les deux directions. Ils sont resserrés dans les principales zones urbaines et dans le Nord-Est de la région (la Vallée de la Meuse, où l'on rencontre des villes de plus de 1400 habitants et des sites industriels, notamment des fonderies).

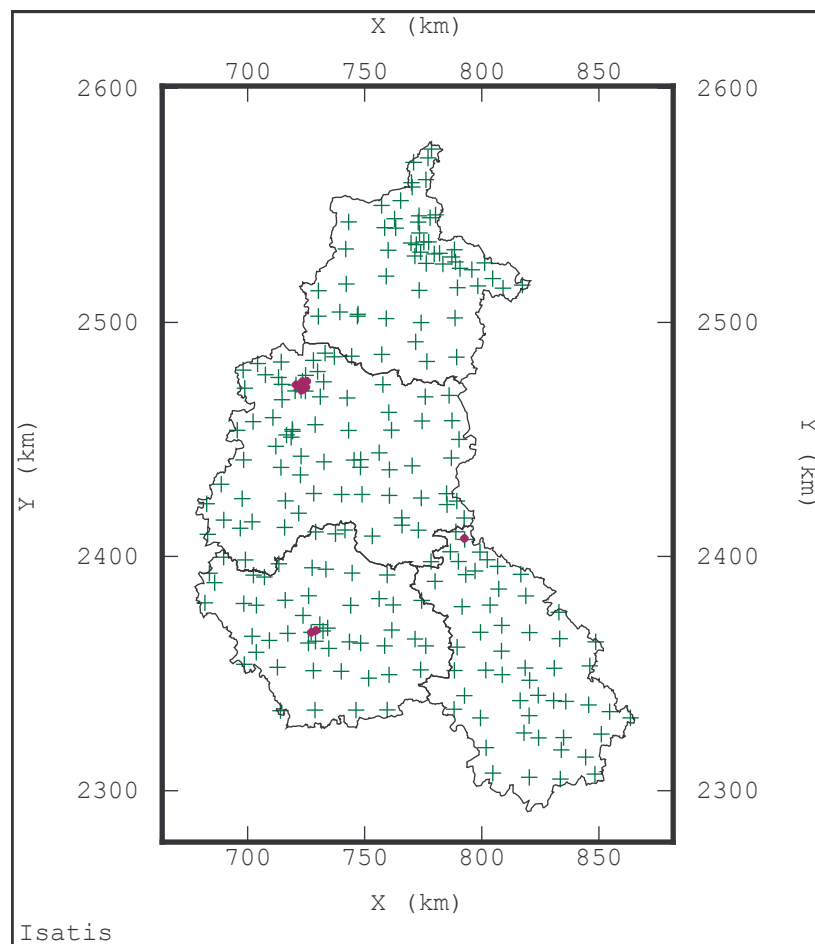


Figure 1 - Carte de répartition des points de mesure. En vert : tubes à échantillonnage passif. En violet : stations fixes.

Les données considérées dans l'étude sont les concentrations moyennes par saison, dont le Tableau fournit les principales statistiques. Un examen des huit campagnes individuelles est effectué préalablement.

Présentation des données auxiliaires

Deux types d'informations ont été fournis par ATMO Champagne Ardenne :

- l'inventaire régional kilométrique des émissions de NO_x (données ATMO Champagne Ardenne) ;
- les données de population par secteur IRIS et par îlot sur la région (données INSEE).

A partir de ces données brutes, un traitement par SIG (Arcview, version 2.0) a permis de calculer différentes variables de densité, ainsi qu'il est décrit dans l'analyse exploratoire :

- densité d'émissions dans un rayon de 1km, 2km et 5km autour du point considéré (notée dE1km, dE2km, dE5km) ;
- densité de population dans un rayon de 1km, 2km et 5km autour du point considéré (notée dpop1km, dpop2km, dpop5km).

Méthodologie de sélection des données

ECHANTILLONNAGE REGULIER

Les données sont sélectionnées selon différents maillages réguliers (Figure 2) :

- mailles carrées de 15 km de côté (deux sélections, décalées de 7500m en y et notées S15km1 et S15km2 dans le texte) ;

- mailles carrées de 20 km de côté (deux sélections, décalées de 2500m en x et en y et notées S20km1 et S20km2 dans le texte) ;
- mailles carrées de 25 km de côté (deux sélections, décalées de 12500m en x et en y et notées S25km1 et S25km2 dans le texte) ;
- mailles carrées de 30 km de côté (deux sélections, décalées de 7500m en x et en y et notées S30km1 et S30km2 dans le texte).

La procédure est la suivante : un maillage régulier de résolution 15 km, 20 km, 25 km ou 30 km est calé sur le domaine d'étude puis, pour chaque maille, on sélectionne le point de mesure le plus proche du centre de celle-ci.

Les jeux de données complémentaires de ces sélections (notés C_{SX}) sont également constitués. A la différence du maillage d'origine, aucun resserrement n'est effectué dans les zones urbaines. Toutefois, afin de n'omettre aucune des principales villes, les sites suivants sont systématiquement ajoutés aux sélections (si celles-ci ne les prennent pas déjà en compte) : Châlons (101), Charleville (23), Chaumont (204), Epernay (85), Langres (226), Reims (71), Rethel (47), Saint-Dizier (134), Sedan (31), Troyes (175), Vitry (129).

Le Tableau présente les statistiques élémentaires de ces différents jeux.

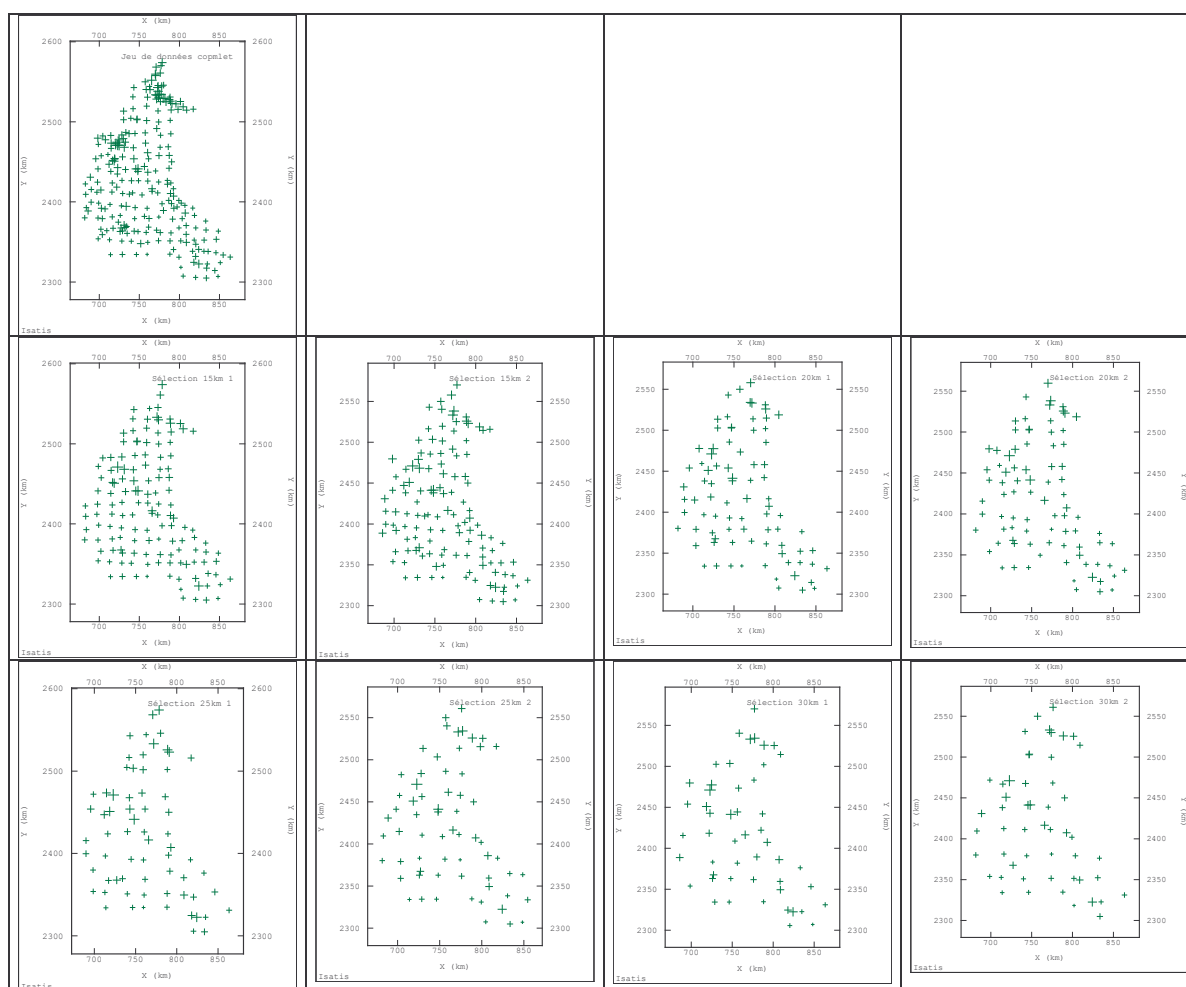


Figure 2 – Plans d'échantillonnage considérés. De gauche à droite : = jeu de données initial (en haut) et sélection selon une maille de 15km (en bas) puis sélections 20km (1 et 2), 25km (1 et 2), 30km (1 et 2)

Tableau 1 - Principales caractéristiques des jeux de données étudiés (hiver).

Données	Toutes *	S15km1	S15km2	S20km1	S20km2	S25km1	S25km2	S30km1	S30km2
Nombre de données valides	233 (234)	137	125	84	85	67	64	50	52
Minimum	13	12.7	12.7	12.7	14.0	12.7	15.3	15.3	14.0
25 ^{ème} percentile	21	20.0	20.7	20.7	20.3	20.4	21.0	21.4	20.5
médiane	24	22.3	23.3	23.3	22.3	23.0	23.7	25.9	22.4
75 ^{ème} percentile	28	25.5	27.3	27.0	26.5	27.6	28.3	28.8	28.1
Maximum	40	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0
Moyenne	24.84 (24.92)	23.15	24.25	24.29	23.82	24.25	24.86	26.00	24.31
Ecart-type	5.04 (5.18)	4.59	4.87	5.13	5.30	5.27	5.18	5.56	5.65
Variance	25.39 (26.85)	21.04	23.73	26.31	28.07	27.75	26.88	30.86	31.97
CV%	20.29 (20.79)	19.81	20.09	21.11	22.24	21.72	20.86	21.37	23.26

* Entre parenthèses, avec un point supplémentaire – 235- ajouté sur une station urbaine du centre de Reims

Tableau 2 - Principales caractéristiques des jeux de données étudiés (été).

Données	Toutes	S15km1	S15km2	S20km1	S20km2	S25km1	S25km2	S30km1	S30km2
Nombre de données valides	225	132	121	80	81	64	60	47	47
Minimum	2	2.0	2.0	3.0	2.3	3.0	2.0	2.8	3.0
25 ^{ème} percentile	5	4.5	4.7	4.6	4.5	4.7	4.5	4.9	4.4
médiane	7	5.5	6.3	6.3	6.0	5.8	6.8	8.0	6.0
75 ^{ème} percentile	9	7.3	8.7	8.7	8.3	8.7	8.8	10.9	7.6
Maximum	21	21.0	18.0	21.0	21.0	21.0	18.0	18.0	18.0
Moyenne	7.39	6.39	6.98	7.23	7.00	7.21	7.24	8.23	6.99
Ecart-type	3.20	3.13	3.07	3.54	3.55	3.75	3.38	3.77	3.70
Variance	10.24	9.81	9.40	12.55	12.60	14.07	11.40	14.18	13.69
CV%	43.29	49.03	43.92	49.00	50.69	52.04	46.65	45.73	52.96

ECHANTILLONNAGE STRATIFIE EN FONCTION DES EMISSIONS DE NOX

La procédure d'échantillonnage testée (voir corps du guide) consiste à quadriller la zone par une cinquantaine de points implantés à intervalles réguliers – ce qui, d'après l'expérience, correspond au minimum requis pour modéliser la structure spatiale de façon satisfaisante - puis à compléter ce maillage de façon que les sites d'échantillonnage couvrent l'ensemble du spectre des émissions tout en se répartissant au mieux dans l'espace.

Les sélections 30km1 (50 points) et 30km2 (52 points) ont été choisies comme maillages de départ. Elles ont été enrichies de respectivement 13 et 7 points (Figure 3). Chaque classe d'émissions, sauf lorsqu'elle n'est pas représentée dans le jeu de données complet, comprend au moins deux sites de mesure.

A titre d'essai, la dégradation du maillage initial a été poussée plus loin et deux nouvelles sélections décalées dans l'espace ont été créées : S40km1 et S40km2. Les points représentatifs des principales communes leur sont ajoutés. Ils appartiennent logiquement aux classes d'émissions les plus élevées. Les autres strates sont complétées selon la procédure évoquée précédemment. Les jeux de données ainsi obtenus se composent (pour l'hiver) de respectivement 52 et 48 sites (Figure 4). La répartition des points ressemble à celle des sélections 30km1 et 30km2 complétées.

Les statistiques associées à ces différents jeux sont présentées dans les tableaux suivants (Tableau 3 à Tableau 6).

Remarque :

La définition d'un plan d'échantillonnage stratifié étant plus longue que celle d'un maillage régulier, nous n'avons pu, faute de temps, évaluer la sensibilité des résultats :

- au nombre de strates (20 classes d'émissions délimitées par les quantiles 5, 10, 15,... , et une classe extrême ont été considérées);
- au support spatial de la variable auxiliaire qui détermine la stratification (les émissions par maille kilométrique ont été utilisées).

Echantillonnage élaboré à partir d'une maille de 30km

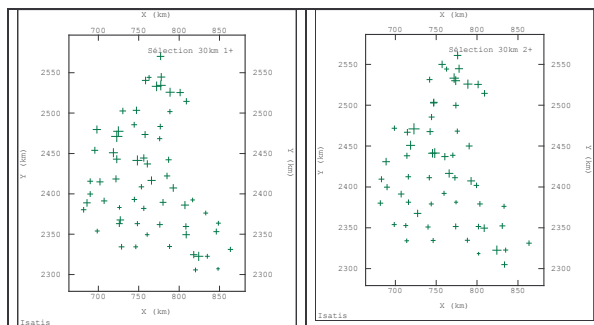


Figure 3 – Echantillonnage stratifié : sélections S30km1 et S30km2 complétées.

Tableau 3 - Principales caractéristiques des jeux de données étudiés (hiver).

Données	S30km1	S30km1 complété	S30km2	S30km2 complété
Nombre de données valides	50	63	52	59
Minimum	15.3	15.3	14.0	14.0
25 ^{ème} percentile	21.4	20.7	20.5	20.3
médiane	25.9	24.5	22.4	22.5
75 ^{ème} percentile	28.8	28.3	28.1	28.0
Maximum	40.0	40.0	40.0	40.0
Moyenne	26.00	25.25	24.31	24.21
Ecart-type	5.56	5.37	5.65	5.41
Variance	30.86	28.79	31.97	29.25
CV%	21.4	21.	23.3	22.

Tableau 4 - Principales caractéristiques des jeux de données étudiés (été).

Données	S30km1	S30km1 complété	S30km2	S30km2 complété
Nombre de données valides	47	60	47	54
Minimum	2.8	2.75	3.0	3.0
25 ^{ème} percentile	4.9	4.5	4.4	4.5
médiane	8.0	7.0	6.0	6.0
75 ^{ème} percentile	10.9	9.5	7.6	7.5
Maximum	18.0	18.0	18.0	18.0
Moyenne	8.23	7.75	6.99	6.90
Ecart-type	3.77	3.55	3.70	3.50
Variance	14.18	12.62	13.69	12.27
CV%	45.73	46.	52.96	51.

Echantillonnage élaboré à partir d'une maille de 40km

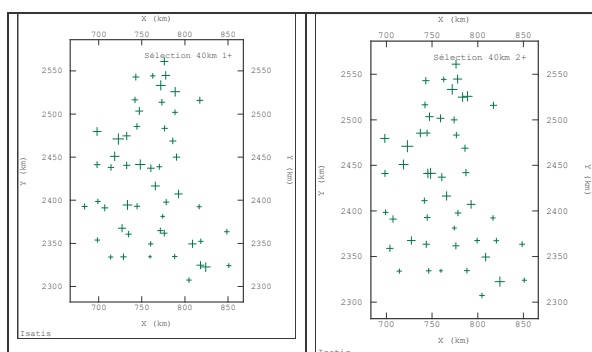


Figure 4 – Echantillonnage stratifié : sélections S30km1 et S30km2 complétées

Tableau 5 - Principales caractéristiques des jeux de données étudiés (hiver).

Données	S40km1 complétée	S40km2 complétée
Nombre de données valides	52	48
Minimum	12.7	14.0
25 ^{ème} percentile	19.8	19.3
médiane	23.2	23.7
75 ^{ème} percentile	28.3	28.6
Maximum	40.0	40.0
Moyenne	24.4	24.7
Ecart-type	5.75	5.41
Variance	33.1	29.3
CV%	24	24

Tableau 6 - Principales caractéristiques des jeux de données étudiés (été).

Données	S40km1 complétée	S40km2 complétée
Nombre de données valides	49	46
Minimum	2.75	2.75
25 ^{ème} percentile	4.5	4.5
médiane	5.5	6.0
75 ^{ème} percentile	9.5	8.7
Maximum	20.0	18.0
Moyenne	7.2	7.2
Ecart-type	4.14	3.70
Variance	17.2	13.7
CV%	58	52

Périodes d'échantillonnage traitées

L'étude est conduite en détail pour la campagne hivernale, durant laquelle des concentrations de NO₂ plus élevées ont été mesurées. Des résultats pour l'été sont présentés plus succinctement à la fin du rapport.

ETUDE DU JEU DE DONNEES INITIAL ET ETABLISSEMENT DE LA CARTE DE REFERENCE

Le but de cette partie est d'élaborer pour chaque saison la carte de référence, c'est-à-dire la carte de répartition des concentrations de dioxyde d'azote établie avec le maximum de points de mesure. Par la suite, on comparera à cette carte les représentations obtenues à partir d'un nombre réduit de points.

Etude des quatre périodes de mesure et de la moyenne (hiver)

HISTOGRAMMES

Les histogrammes des concentrations sont unimodaux (Figure 5, Figure 7), hormis peut-être celui de la troisième période. Le second mode, de moindre ampleur, correspond principalement à la Vallée de la Meuse, qui se caractérise par la présence de petites villes de plus de 1400 habitants et de sites industriels (Figure 6). Sauf au point 235 qui, systématiquement, présente la valeur maximale (site de la station permanente Mairie, au centre de Reims), les concentrations les plus élevées n'apparaissent pas nécessairement aux mêmes points d'une période à l'autre, ni dans les mêmes zones. Entre la deuxième et la troisième période, les niveaux de concentration augmentent sensiblement sur l'ensemble de la région.

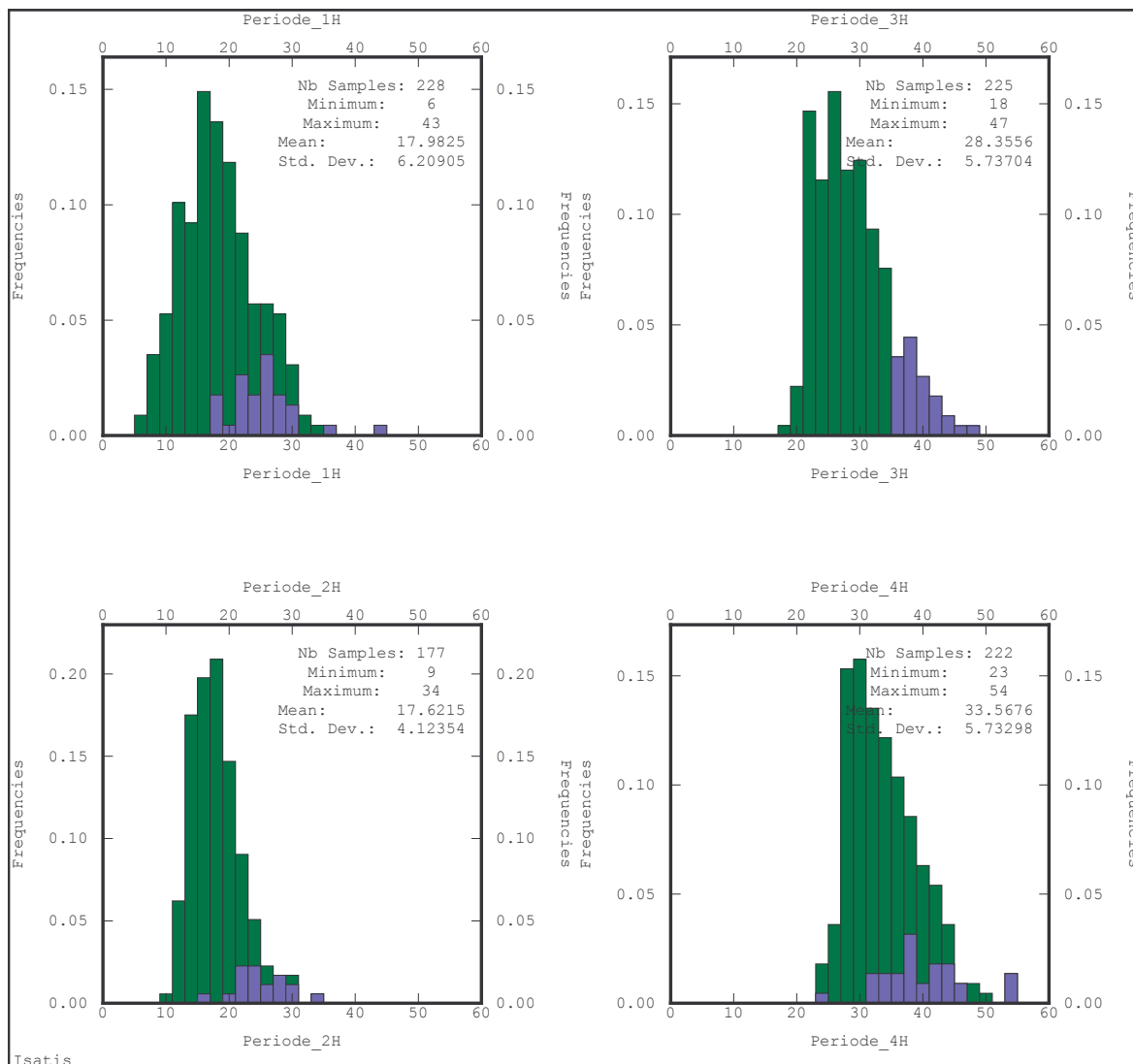


Figure 5 - Histogrammes des quatre périodes de mesure. En bleu, on examine la si les fortes valeurs d'une période correspondent aux fortes valeurs des autres campagnes. Les points de mesure correspondants sont représentés sur la carte-ci dessous.

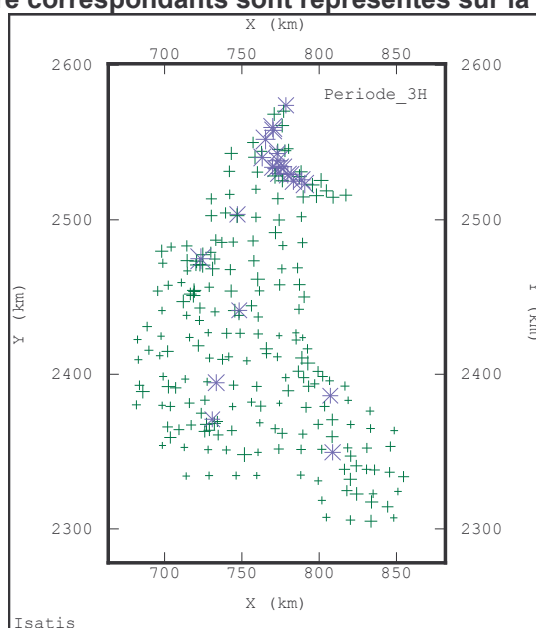


Figure 6 – Localisation des points associés aux plus fortes concentrations de la 3^e période.

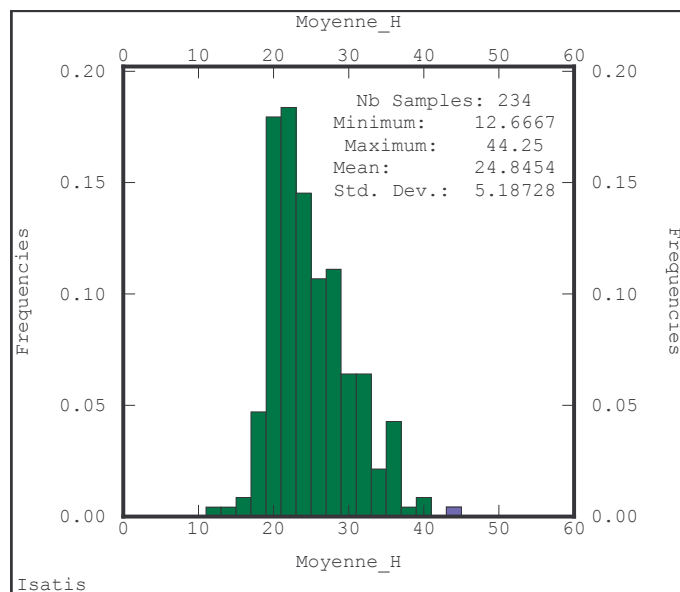


Figure 7 – Histogramme des concentrations moyennes hivernales

CORRELATIONS

Pour juger du degré de ressemblance entre les concentrations mesurées pendant chaque période d'échantillonnage et les niveaux moyens saisonniers, on effectue une analyse en composantes principales (Figure 8).

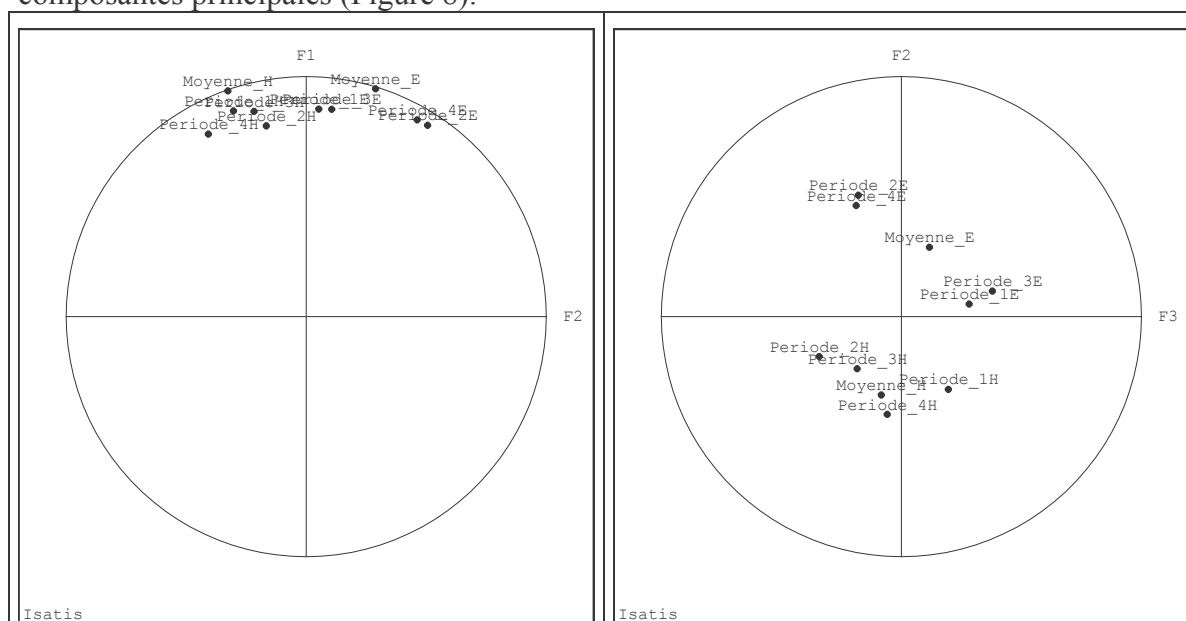


Figure 8 – Analyse en composantes principales sur les concentrations. Projection sur le plan 1-2 (à gauche) et 2-3 (à droite)

La projection des variables sur le premier plan factoriel met en évidence une corrélation positive entre toutes les périodes de mesure. On observe néanmoins :

- sur l'axe 2 : une séparation entre les campagnes hivernales et estivales ;
- sur les axes 2 et 3 : une séparation entre les campagnes d'une même saison. Ce résultat s'accorde avec les remarques relatives aux histogrammes. La deuxième période, qui se démarque légèrement pour l'hiver, comporte un assez grand nombre de données manquantes.

Les corrélations entre quinze d'hiver nous semblent cependant suffisantes pour limiter notre travail à l'analyse de la moyenne hivernale.

Dans toute la suite, on désigne par concentration de NO_2 la concentration moyenne hivernale de NO_2 .

Etude du variogramme expérimental (hiver)

A grande échelle, le variogramme expérimental croît avec la distance, révélant une non stationnarité (Figure 9).

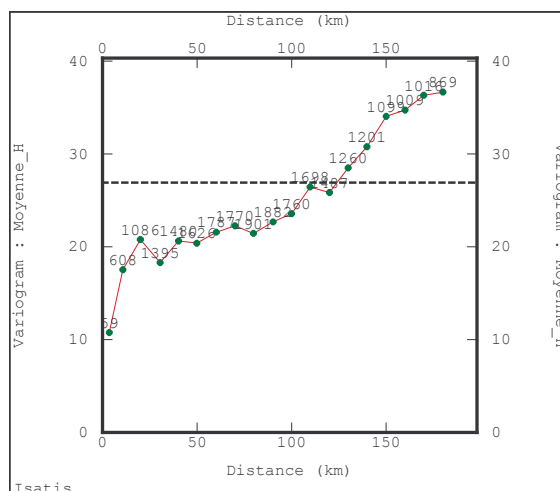
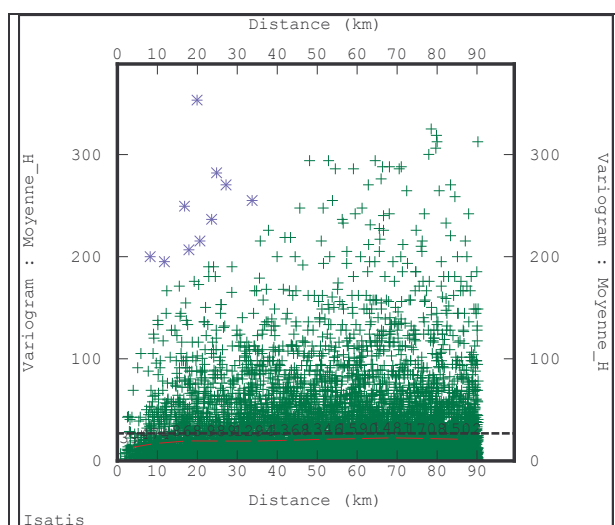


Figure 9 – Comportement du variogramme aux grandes distances

L'échelle d'étude est réduite aux 100 premiers kilomètres, ce qui, selon la direction, correspond approximativement à la moitié ou au tiers du champ, conformément à la pratique. Le rayon du voisinage d'estimation utilisé en krigeage ordinaire n'excèdera pas cette distance au-delà de laquelle l'hypothèse de stationnarité n'est plus vérifiée.

Dans la Figure 10, le variogramme expérimental a été calculé avec et sans le point du centre de Reims (station Mairie). Ce point, qui s'écarte systématiquement des histogrammes (cf. 0), augmente la variabilité spatiale des concentrations, en particulier aux courtes distances. Il ne sera pas pris en compte dans l'étape de modélisation.

Dans tous les cas, l'effet de pépite demeure élevé. Un examen du variogramme sur les premiers kilomètres montre qu'il est en grande partie attribuable à la variabilité de petite échelle (Figure 11).



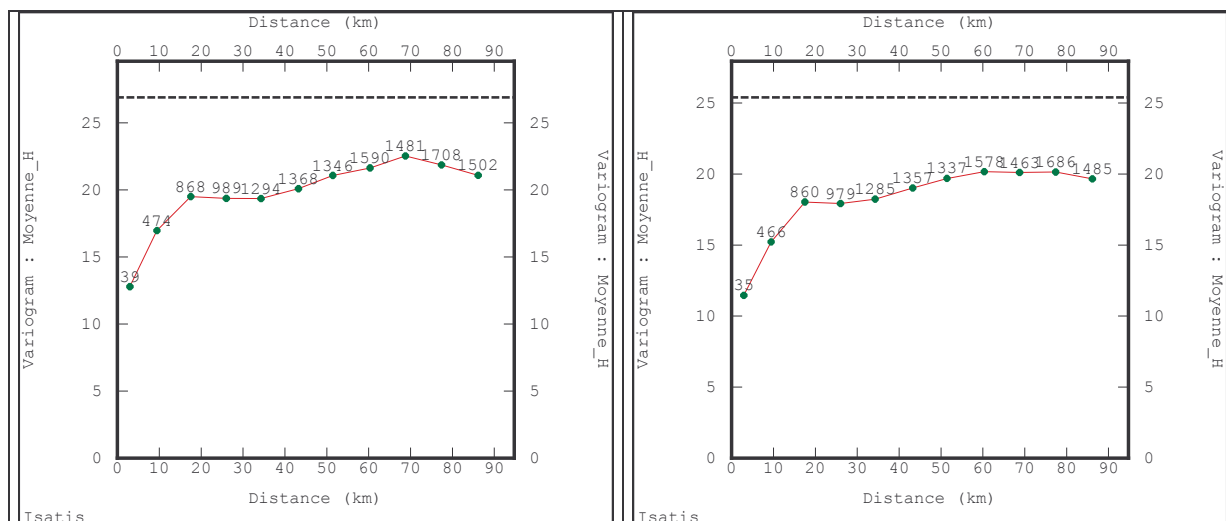


Figure 10 – Ligne du haut : nuée variographique ; les points en bleu correspondent à des paires de points formées avec le site 235. Ligne du bas : Variogramme omnidirectionnel : avec (fig. de gauche, pas = 8600 m, 11 pas), et sans (fig. de droite, pas = 8600 m, 11 pas) le point 235.

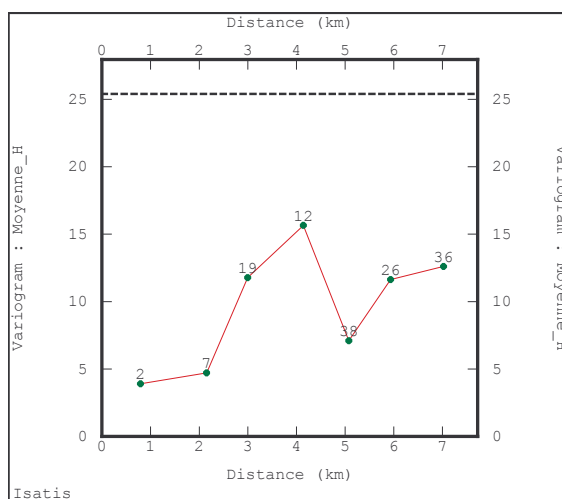


Figure 11 – Variogramme expérimental à petite échelle.

La variabilité des concentrations est plus grande dans la direction N40, comme le montrent la carte variographique et le calcul par direction du variogramme. Entre les autres directions, la différence est moindre. Un variogramme omnidirectionnel est conservé.

Corrélation avec les variables de coordonnées

Afin d'identifier de possibles dérives, les concentrations sont représentées en fonction des coordonnées spatiales. Elles augmentent en moyenne avec la coordonnée Y, comme l'illustre la Figure 12. Aucune relation n'est mise en évidence avec la coordonnée X.

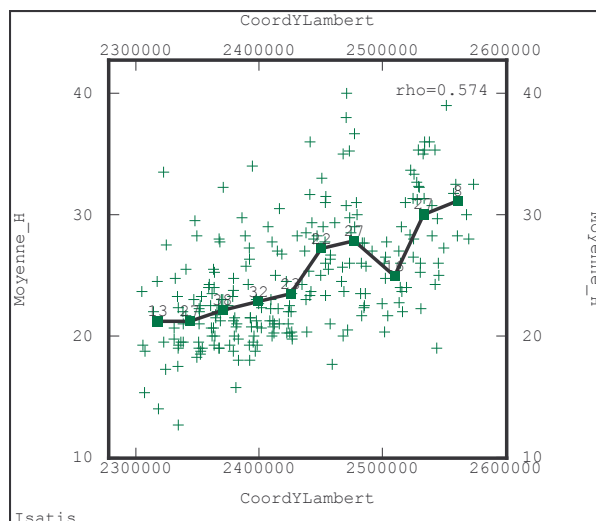


Figure 12 – Représentation des concentrations en fonction de Y. En noir : courbe d'espérance conditionnelle.

Correlation avec les variables auxiliaires

Les concentrations de NO₂ sont corrélées positivement avec les différentes variables auxiliaires considérées (Tableau 7, Figure 13). Celles-ci ont subi préalablement une transformation logarithmique, ce qui a pour effet de linéariser les nuages de corrélation.

Tableau 7 – Coefficient de corrélation entre la variable de concentration et les variables auxiliaires.

	Corrélation avec :		Corrélation avec :
ln(1+dE1km)	0.65	ln(1+dpop1km)	0.75
ln(1+dE2km)	0.66	ln(1+dpop2km)	0.74
ln(1+dE5km)	0.68	ln(1+dpop5km)	0.73

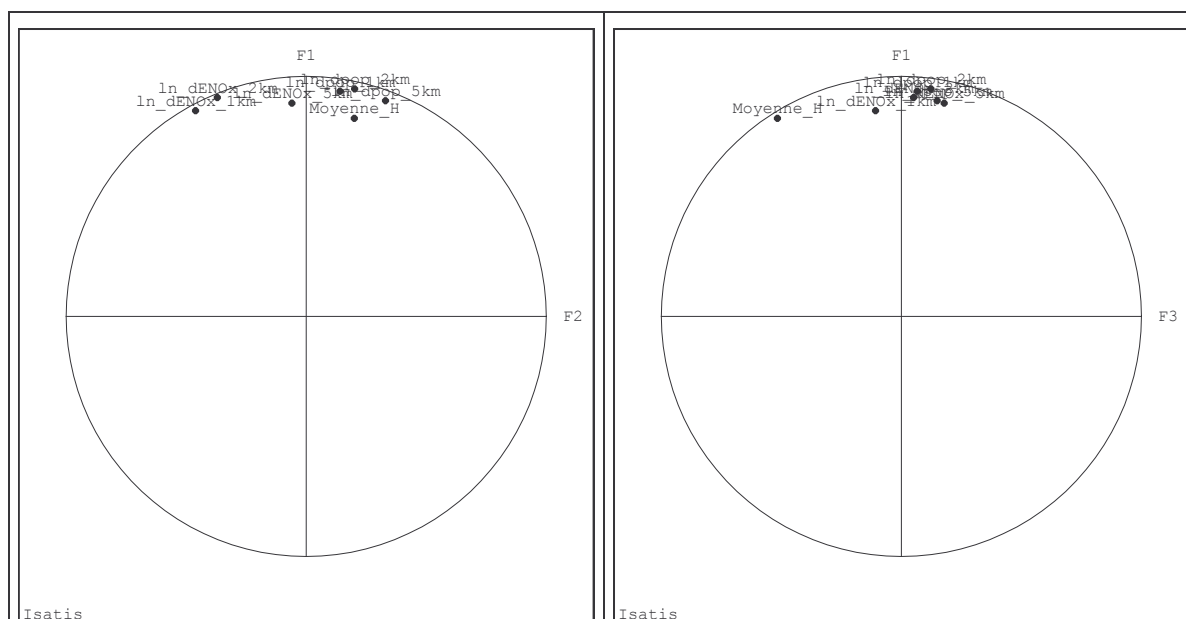


Figure 13 – Analyse en composantes principales sur la variable de concentration (moyenne hivernale) et les variables auxiliaires. Projection sur les plans 1-2 et 1-3.

Modélisation géostatistique et validation croisée

CAS MONOVARIABLE

Etude variographique

Deux modèles isotropes sont ajustés sur le variogramme expérimental omidirectionnel : l'un, continu à l'origine, ne tient pas compte de l'incertitude de mesure; l'autre, discontinu à l'origine, la modélise par un effet de pépité (évalué visuellement sur la Figure 11 et d'après des mesures de triplets). D'autre part, ils comprennent tous deux :

- une structure exponentielle de courte portée, qui représente la variabilité des concentrations à petite distance ;
- une structure sphérique de longue portée, qui représente les phénomènes de transport à plus grande échelle.

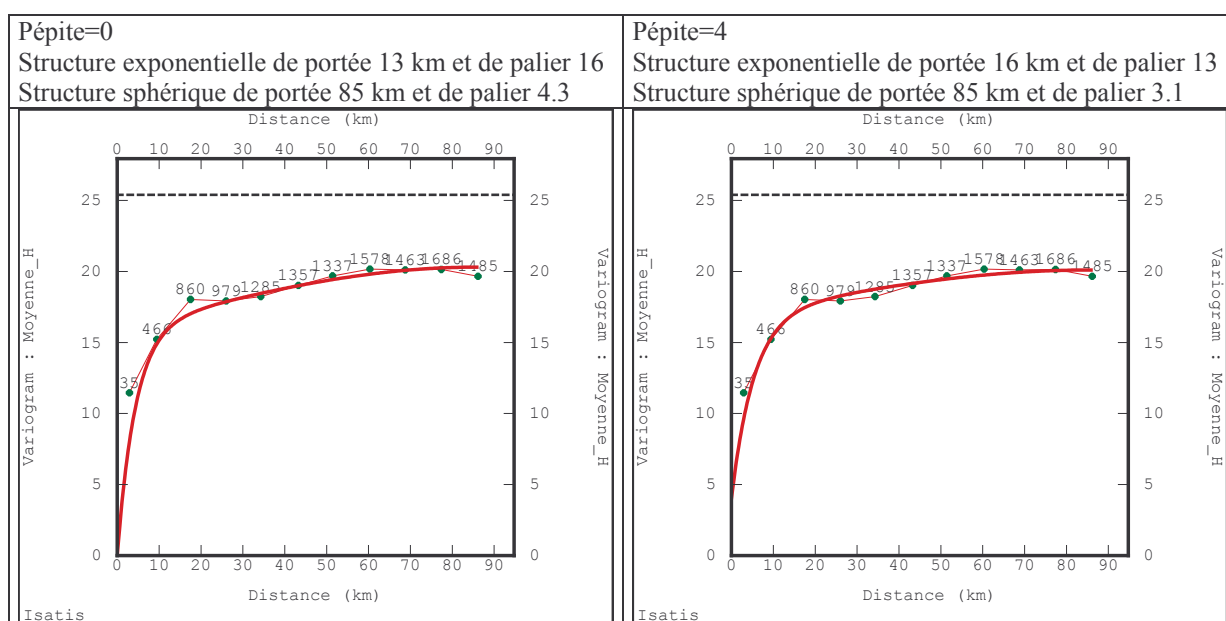


Figure 14– Variogrammes expérimentaux et modélisés sans (figure de gauche) et avec (figure de droite) effet de pépité

Le choix des structures et des paramètres variographiques découle d'ajustements successifs testés par validation croisée.

Les validations croisées associées aux deux modèles sélectionnés fournissent les résultats suivants :

Tableau 8 – Résultats de la validation croisée sur toutes les données pour le krigeage ordinaire ponctuel, modèle de référence sans effet de pépité

Voisinage d'estimation : glissant de rayon 90 km		Validation croisée par krigeage ordinaire ponctuel			
Toutes les données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	233	233	233	233	233
Minimum	12.67	19.68	-12.97	-3.250	0.04
25 ^{ème} percentile	21.00	21.95	-2.04	-0.522	5.64
Médiane	23.67	24.07	0.72	0.218	9.95
75 ^{ème} percentile	28.00	26.98	2.58	0.625	16.37
Maximum	40.00	36.72	11.47	3.164	66.46
Etendue (= Max – Min)	27.33	17.04	24.44	6.415	66.41
Moyenne	24.76	24.94	0.18	0.025	11.97

Ecart-type	5.05	3.58	3.58	0.938	9.78
Variance	25.51	12.85	12.80	0.879	95.74
CV %	90.60	146.38			
Corrélation mesures/estimations		r=0,706			
Nombre de données non robustes		4			

Tableau 9 – Résultats de la validation croisée sur toutes les données pour le krigeage ordinaire ponctuel, modèle de référence avec effet de pépité

Voisinage d'estimation : glissant de rayon 90 km		Validation croisée par krigeage ordinaire ponctuel			
Toutes les données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	233	233	233	233	233
Minimum	12.67	19.78	-12.90	-3.172	0.03
25 ^{ème} percentile	21.00	22.05	-2.04	-0.541	5.43
Médiane	23.67	24.09	0.53	0.142	10.10
75 ^{ème} percentile	28.00	26.96	2.63	0.628	15.71
Maximum	40.00	35.21	11.39	3.026	67.64
Etendue (= Max – Min)	27.33	15.43	24.29	6.198	67.61
Moyenne	24.76	24.91	0.15	0.020	11.99
Ecart-type	5.05	3.46	3.58	0.903	9.78
Variance	25.51	11.99	12.83	0.815	95.65
CV %	90.60	161.47			
Corrélation mesures/estimations		r=0,705			
Nombre de données non robustes		4			

Les modèles avec et sans effet de pépité conduisent à des résultats et à des statistiques d'erreur comparables. Les valeurs estimées sont assez bien corrélées aux valeurs mesurées, avec un coefficient de corrélation égal à 0,71 (Tableau , Tableau 9, Figure 15). Avec effet de pépité, le lissage des valeurs est un peu plus marqué.

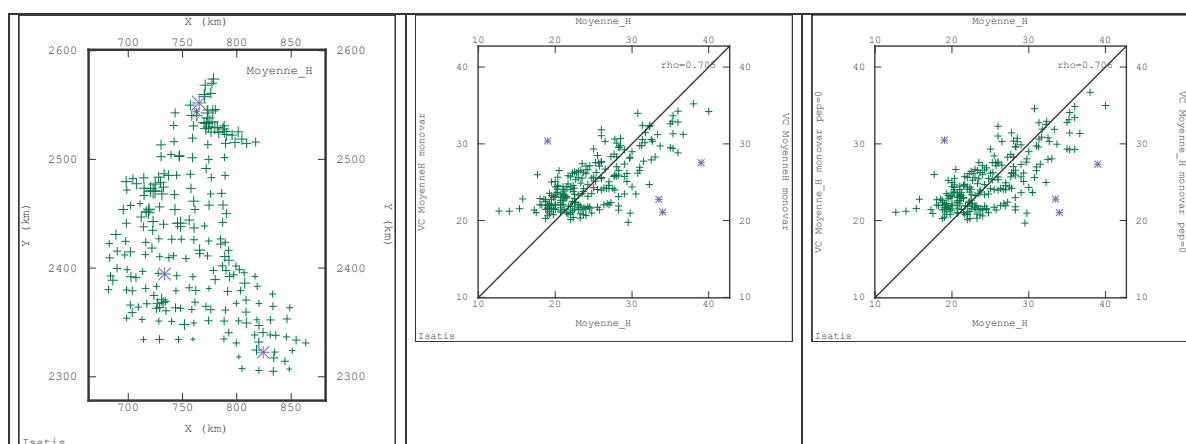


Figure 15 – Comparaison entre les valeurs estimées par validation croisée (modèle avec (à gauche) et sans (à droite) effet de pépité) et les valeurs mesurées. En bleu : localisation (en bleu) des points de plus grande erreur standardisée.

Estimation sur toute la région

Après discussion avec ATMO Champagne-Ardenne, deux grilles d'estimation sont considérées.

- La première coïncide avec la grille du cadastre des émissions. La taille de la maille est de 1km.
- La seconde est plus lâche ; la taille de la maille est de 4km.

Les caractéristiques de ces grilles sont respectivement :

$X_{min}=678500$ m, 186 mailles $X_{max}=864500$, $\Delta X=1000$ m,
 $Y_{min}=2291500$ m, 285 mailles, $Y_{max}=2576500$, $\Delta X=1000$ m.
 et
 $X_{min}=680000$ m, 46 mailles $X_{max}=864000$, $\Delta X=4000$ m,
 $Y_{min}=2293000$ m, 71 mailles, $Y_{max}=2573000$, $\Delta X=4000$ m.

Le voisinage d'estimation le plus satisfaisant au vu des résultats de la validation croisée est un voisinage **glissant de 90km de rayon**. Notons que ce rayon est compatible avec le domaine de validité du variogramme.

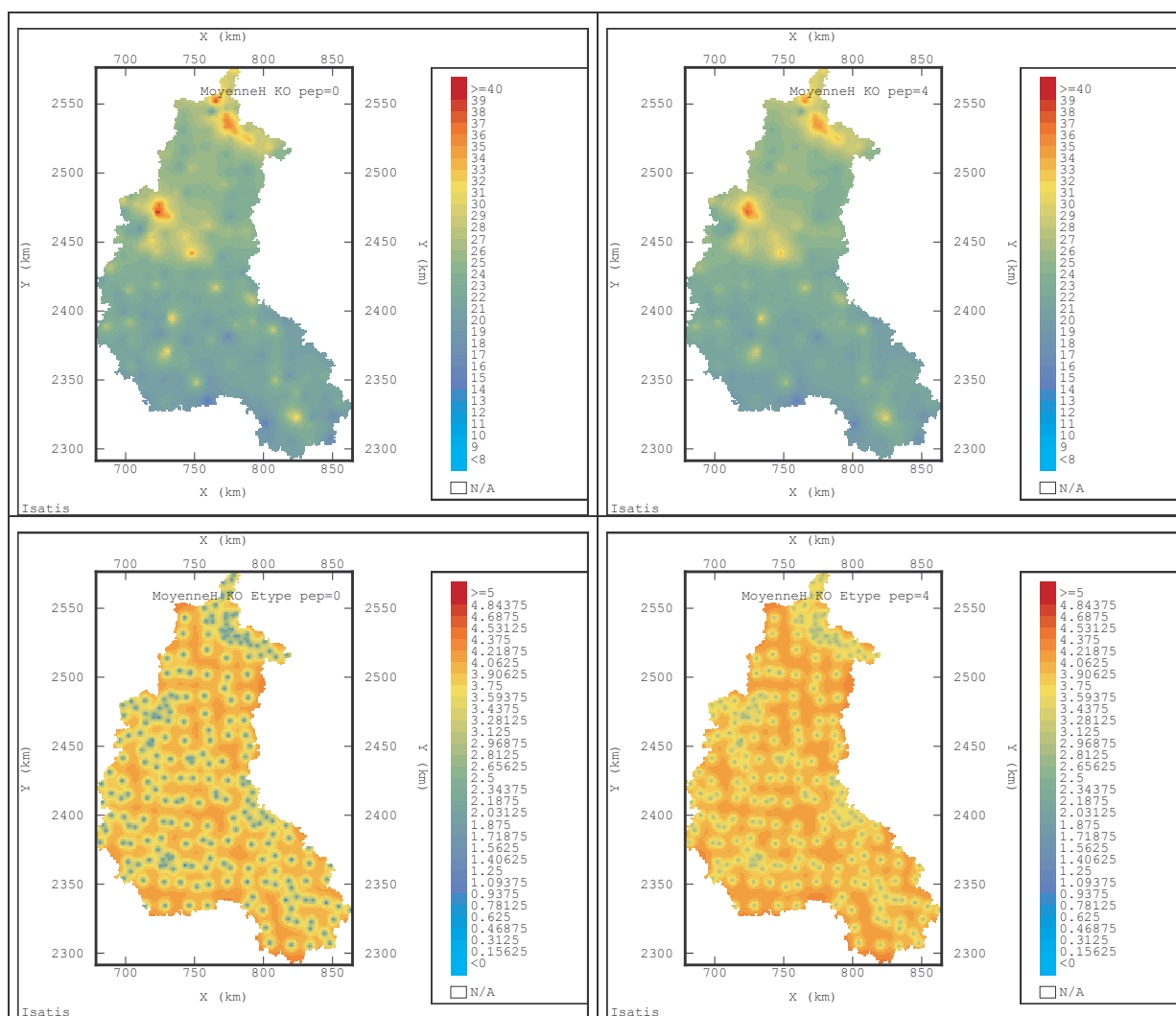


Figure 16 – Estimation par krigeage ponctuel à partir des données, sur une grille de pas égal à 1km. A gauche, modèle sans effet de pépité ; à droite : modèle avec effet de pépité. Ligne du haut : concentrations de NO₂ estimées [µg/m³] ; ligne du bas : écarts-types de krigeage.

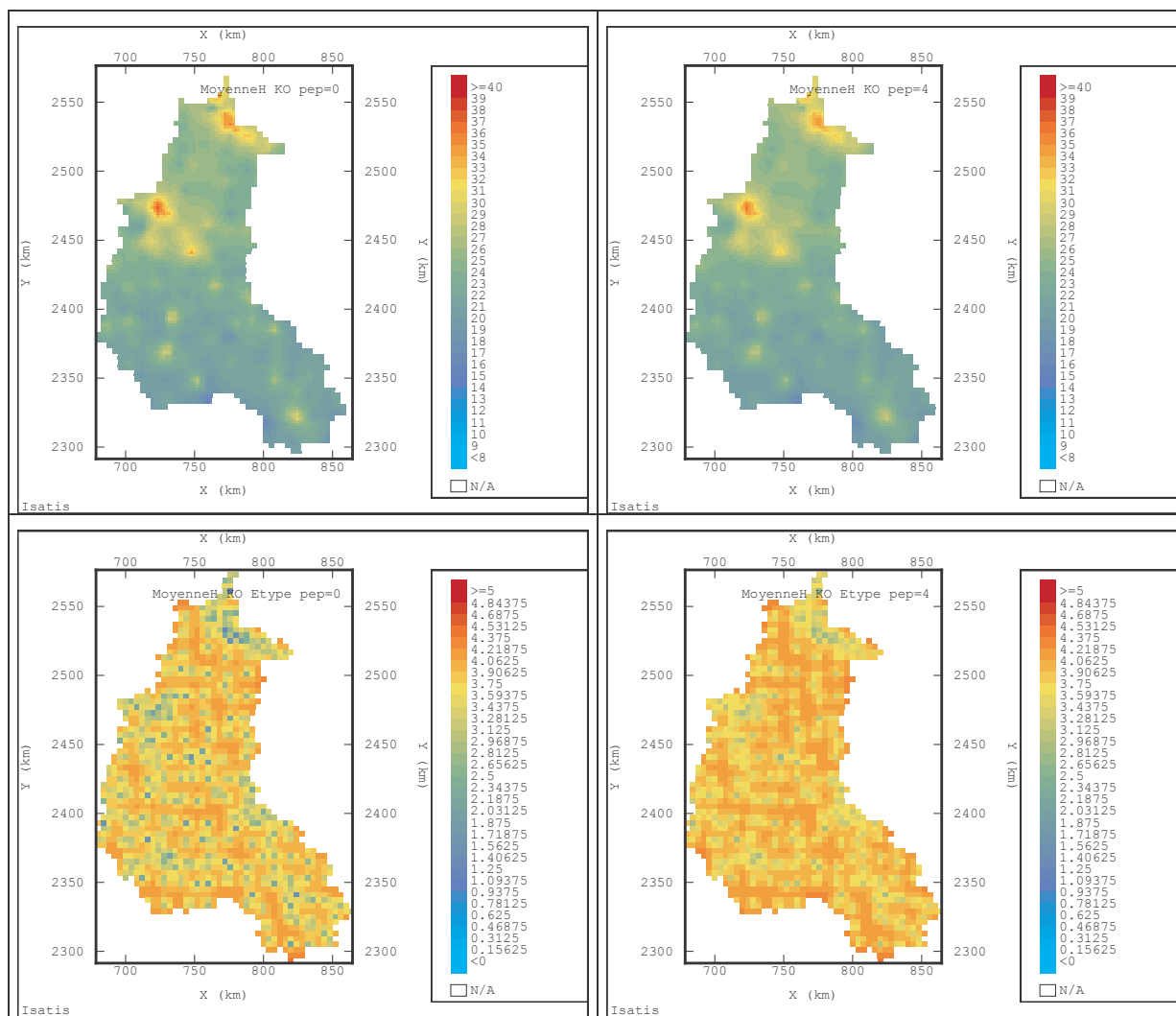


Figure 17 – Estimation par krigeage ordinaire ponctuel à partir de toutes les données, sur une grille de pas égal à 4km. A gauche, modèle sans effet de pépité ; à droite : modèle avec effet de pépité. Ligne du haut : concentrations de NO₂ estimées [µg/m³] ; ligne du bas : écarts-types de krigeage.

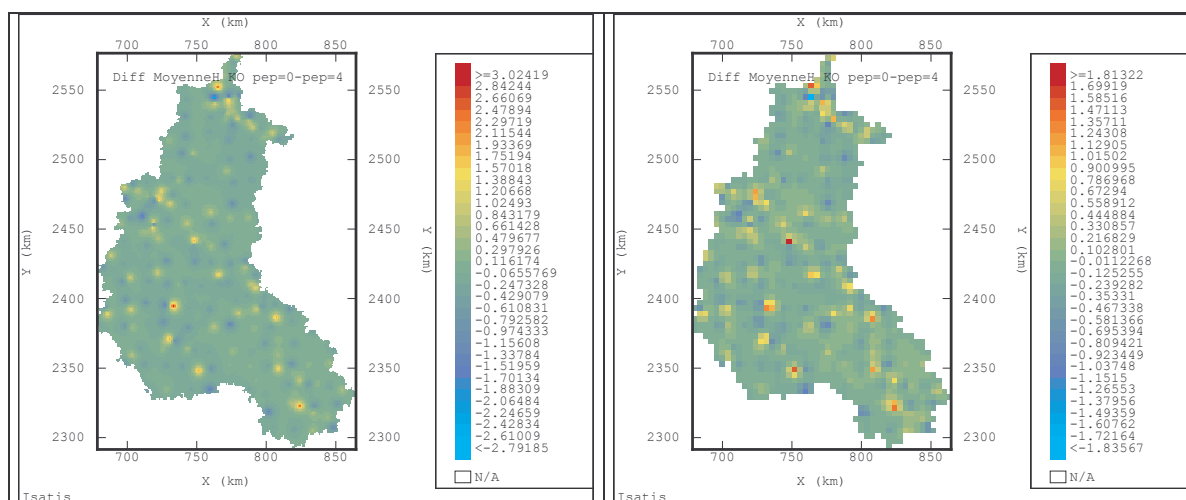


Figure 18 – Différence (µg/m³) entre les estimations sans et avec effet de pépité. A gauche : maille de 1km, à droite : maille de 4km.

Les estimations avec et sans effet de pépité se ressemblent (Figure 16). Les différences les plus importantes s'observent au voisinage des points de mesure (avec un effet de pépité, les

estimations sont plus lisses et moins influencées par les mesures les plus proches) mais elles n'excèdent jamais $4 \mu\text{g}/\text{m}^3$. D'autre part, avec un effet de pépité, les écarts-types de krigeage sont légèrement plus élevés, reflétant la perte de structure modélisée par cet effet.

Visuellement, l'élargissement de la maille d'estimation altère assez peu le détail de la carte (Figure 17). Pour la comparaison des plans d'échantillonnage, et puisque les variables auxiliaires sont disponibles à fine résolution (cf. 0), on utilisera la maille la plus fine.

CAS MULTIVARIABLE

Plusieurs techniques de modélisation et d'estimation ont été considérées afin de prendre en compte les variables auxiliaires :

- régression (multi)linéaire, modélisation du variogramme des résidus et krigeage des résidus
- modélisation non stationnaire (dérives proposées : coordonnées d'espace et variables de population et d'émissions ; structures proposées : celles du variogramme des résidus, avec différentes portées) et krigeage en dérive externe
- régression (multi)linéaire, modélisation automatique par le module *bundled external drift* (dérives possibles : coordonnées d'espace et fonction de régression) et krigeage en dérive externe
- régression (multi)linéaire, modélisation bivariée (seconde variable : fonction de régression) et cokrigeage colocalisé

Ces différentes méthodes ont été comparées par validation croisée. Les statistiques d'erreur et d'erreur relative sont favorables au krigeage des résidus suivi du cokrigeage. Cette observation s'accorde avec des études antérieures qui ont montré l'efficacité du krigeage des résidus pour cartographier le NO_2 à l'échelle d'une agglomération (Cori, 2005) ou d'une région (Malherbe et Cardenas, 2005).

Ce rapport étant centré sur l'impact de l'échantillonnage plus que sur le choix d'une méthode multivariable, seuls les résultats du krigeage des résidus sont présentés. Pour tous les schémas d'échantillonnage étudiés par la suite, on se limitera à l'utilisation du krigeage ordinaire et du krigeage des résidus.

Régression multilinéaire

Dans le krigeage des résidus, on suppose que la concentration de NO_2 en un point x se décompose en une dérive $Y(x)$ et un résidu aléatoire $R(x)$:

$$Z_{\text{NO}_2}(x) = Y(x) + R(x)$$

L'expression de $Y(x)$ est obtenue ici par régression multilinéaire sur les variables auxiliaires disponibles.

L'étude des corrélations (§0) et l'application d'une procédure de régression stepwise³ conduit à sélectionner deux variables explicatives :

- la densité d'émission à 5km ($\ln(1+dE5km)$) ;
- la densité de population à 1km ($\ln(1+dpop1km)$) .

³ Procédure mise en œuvre avec le logiciel R, version 2.2.0.

Bien qu'elles soient corrélées entre elles et que par suite, ce choix puisse être discuté, ces deux variables ne sont pas considérées comme entièrement redondantes. En particulier, elles ne correspondent pas tout à fait à la même échelle d'information spatiale. (Dans une ACP sur les variables auxiliaires, elles se trouvent séparées par l'axe 3.) Elles sont donc conservées toutes deux.

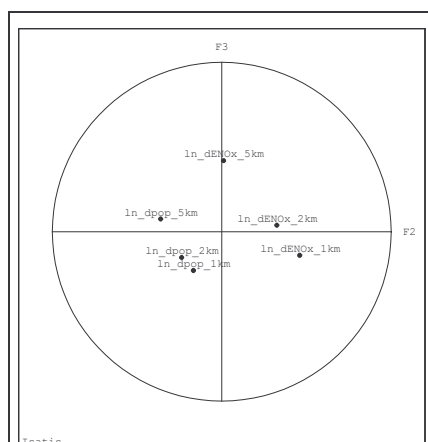
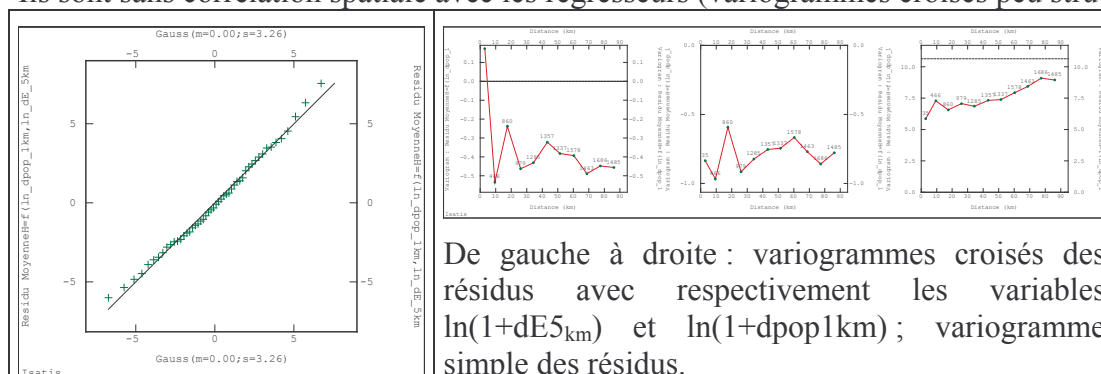


Figure 19 – Analyse en composantes principales sur les variables auxiliaires. Projection sur le plans 2-3

La fonction de régression s'écrit :

$$Y(x) = 1,193 \cdot \ln(1 + dE5km(x)) + 1,626 \cdot \ln(1 + dpop1km(x)) + 9,273$$

Les résidus de la régression n'affichent pas de tendance et se distribuent selon la loi normale. Ils sont sans corrélation spatiale avec les régresseurs (variogrammes croisés peu structurés).



De gauche à droite : variogrammes croisés des résidus avec respectivement les variables $\ln(1 + dE5km)$ et $\ln(1 + dpop1km)$; variogramme simple des résidus.

L'estimation de la concentration de NO_2 en un point de l'espace s'écrit :

$$\hat{Z}_{NO_2}(x) = Y(x) + \hat{R}_{KO}(x)$$

où $\hat{R}_{KO}(x)$ est l'estimation par krigeage ordinaire du résidu.

Etude variographique

La structure spatiale des résidus de la régression multilinéaire est modélisée par (Figure 20) :

- une structure exponentielle de courte portée, qui représente la variabilité des concentrations à petite distance ;
- une structure sphérique de longue portée, qui représente les phénomènes de transport à plus grande échelle.

Pépite=3

Structure exponentielle de portée 7,5 km et de palier

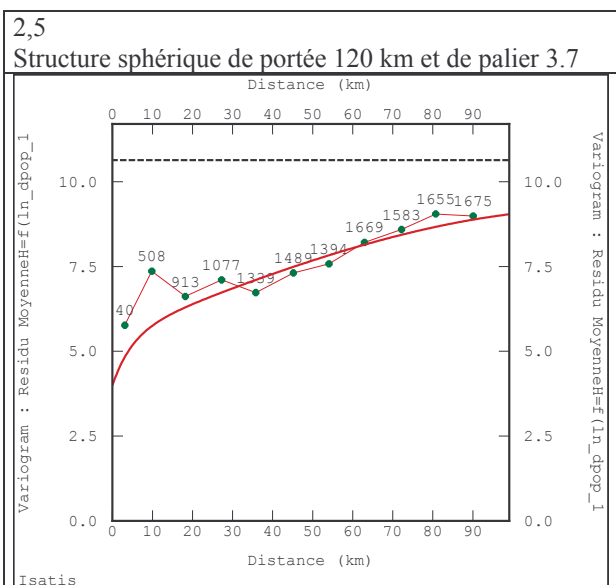


Figure 20– Variogramme expérimental des résidus et variogramme modélisé

Comme dans l'étude précédente, le choix des structures et des paramètres variographiques découle d'ajustements successifs testés par validation croisée. En comparaison du krigeage ordinaire monovarié, le modèle multivarié reproduit plus fidèlement la distribution statistique des données de concentration et améliore nettement les statistiques d'erreur (Tableau 10, Figure 21).

Tableau 10 – Résultats de la validation croisée sur toutes les données pour le krigeage des résidus

Voisinage d'estimation : glissant de rayon 90 km		Validation croisée par krigeage des résidus			
Toutes les données	Mesures ($\mu\text{g}/\text{m}^3$)	Valeurs estimées ($\mu\text{g}/\text{m}^3$)	Erreur d'estimation	Erreur d'estimation standardisée	Erreur d'estimation relative (%)
Nombre de données	233	233.00	233.00	233.00	233.00
Minimum	12.67	16.46	-7.22	-2.87	0.10
25 ^{ème} percentile	21.00	21.42	-1.60	-0.63	3.43
Médiane	23.67	23.66	-0.06	-0.03	6.29
75 ^{ème} percentile	28.00	27.72	1.64	0.65	11.62
Maximum	40.00	40.52	8.80	3.40	69.46
Etendue (= Max – Min)	27.33	24.06	16.02	6.27	69.36
Moyenne	24.76	24.76	0.00	0.00	8.34
Ecart-type	5.05	4.69	2.53	1.01	7.52
Variance	25.51	22.03	6.40	1.03	56.48
CV %	90.60	102.91			
Corrélation mesures/estimations		r=0.87			
Nombre de données non robustes		4			

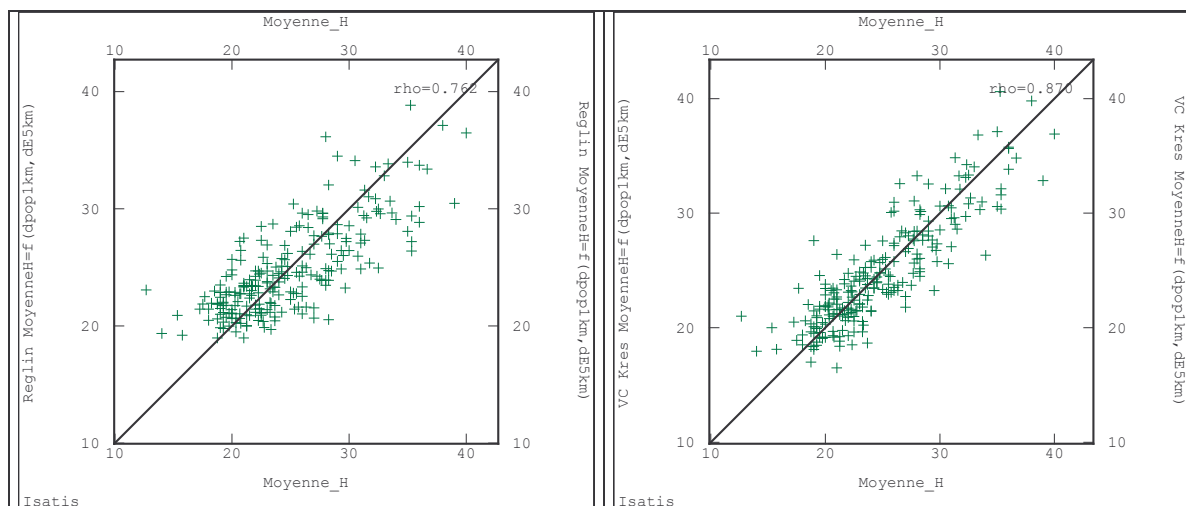


Figure 21 – Gain de précision apporté par le krigeage des résidus par rapport à la régression multilinéaire seule. Comparaison des valeurs estimées par régression (figure de gauche) et par krigeage des résidus en validation croisée (figure de droite) avec les valeurs mesurées.

N.B. : Comme la concentration de NO_2 tend à augmenter avec la coordonnée y (cf. § 0), on a fait également l'essai d'introduire y dans la régression multilinéaire. Les résultats finaux sont moins satisfaisants.

Estimation sur toute la région

Pour des raisons de temps de calcul, la densité de population n'a été calculée qu'aux nœuds de la grille la plus fine (maille de 1km de côté). En conséquence, l'estimation est faite uniquement sur cette grille. Ce choix est justifié puisque la résolution spatiale des données auxiliaires brutes, qu'il s'agisse des émissions (inventaire kilométrique) ou de la population (données IRIS+îlots), est relativement élevée. Il est donc préférable de conserver la richesse de cette information. Comme dans l'estimation monovariée, l'interpolation des résidus est réalisée avec un voisinage glissant de rayon 90 km.

L'influence des routes apparaît plus nettement et elle est plus circonscrite qu'en krigeage ordinaire (Figure 22).

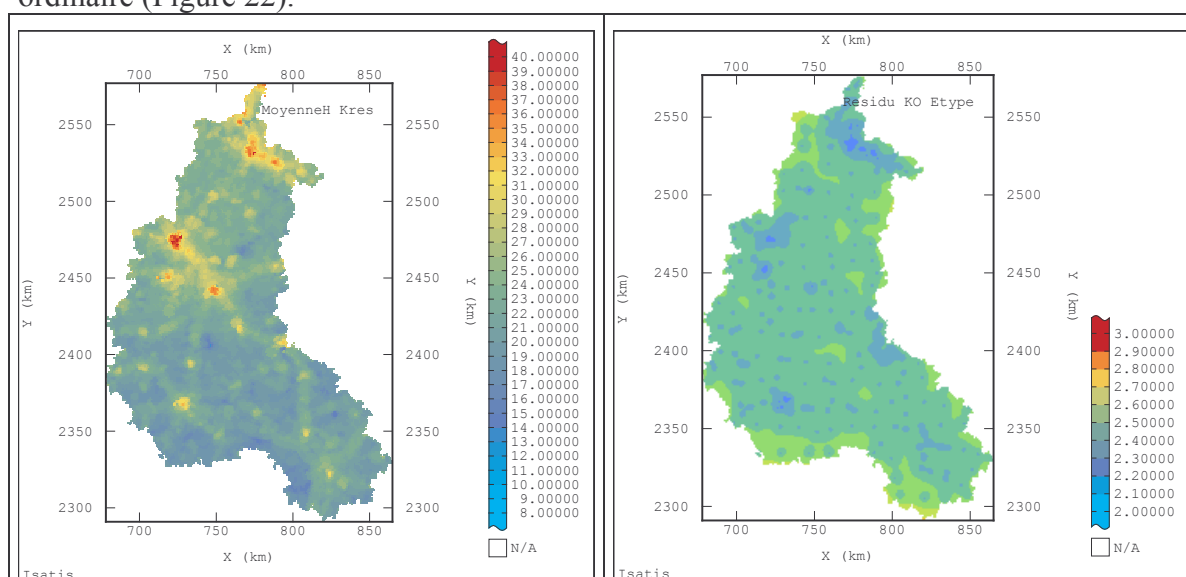


Figure 22 – Estimation par régression linéaire et krigeage des résidus. A gauche : concentration estimée [$\mu\text{g}/\text{m}^3$]. A droite : écart-type de l'erreur de krigeage [$\mu\text{g}/\text{m}^3$]

L'écart-type de l'erreur de krigeage correspond à l'erreur d'estimation non pas de la concentration mais du résidu, qui est plus faible. Il devrait être inférieur à l'écart-type obtenu par krigeage ordinaire monovarié.

Le krigeage est un interpolateur exact : effectué aux points de mesure, il restitue exactement la donnée. Dans la cartographie cependant, le krigeage est réalisé aux centres de maille. Afin de contrôler la cohérence entre les valeurs mesurées et estimées aux points de grille les plus proches (distants d'au plus 710m pour un maillage kilométrique), on a migré ces dernières vers les sites d'échantillonnage. Les deux série de valeurs sont en accord (Figure 23).

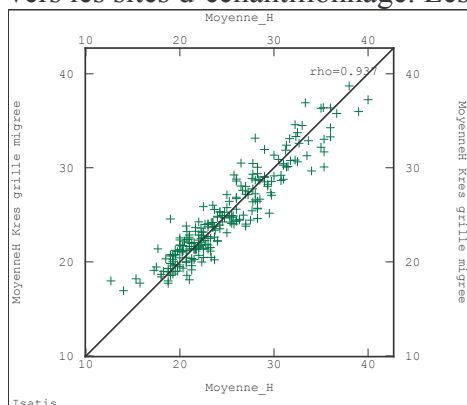


Figure 23 – Nuage de corrélation entre les moyennes hivernales mesurées et les valeurs estimées aux points de grille les plus proches.

DEGRADATION DE L'ECHANTILLONNAGE : EVOLUTION DU VARIOGRAMME EXPERIMENTAL ET DE LA RELATION NO₂/VARIABLES AUXILIAIRES – HIVER

Evolution du variogramme expérimental

Plus la maille d'échantillonnage est lâche, plus le variogramme expérimental s'écarte du variogramme de référence et perd en structure. Ce changement est plus particulièrement marqué à partir des sélections S25km (Figure 24).

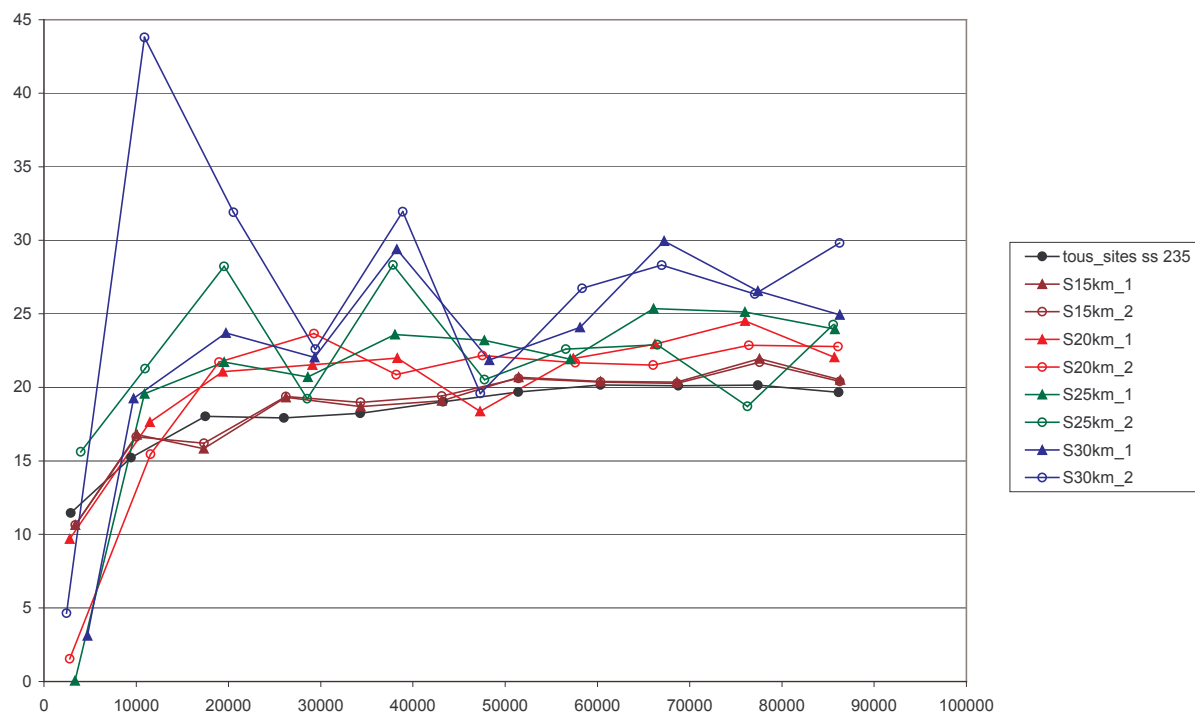


Figure 24 – Variogrammes expérimentaux associés au jeu de données complet (en noir) et aux différentes sélections

Evolution de la relation avec les variables auxiliaires

Les tableaux ci-dessous présentent les coefficients de corrélation (ρ) et les paramètres a et b de la régression linéaire entre les concentrations de NO_2 et les variables auxiliaires ($Z=a.\Phi+b$).

**Tableau 11 – Relation avec la densité d'émissions de NO_x (transformée logarithmique).
Hiver**

	1km			2km			5km		
	ρ	a	b	ρ	a	b	ρ	a	b
Toutes (sauf 235)	0.65	2.56	4.87	0.66	2.66	4.31	0.68	3.11	1.20*
S15km1	0.72	2.90	1.77	0.71	3.15	2.73	0.70	3.08	0.63*
S15km2	0.66	2.46	5.21	0.69	2.64	4.01	0.69	3.17	0.62*
S20km1	0.73	3.16	-0.14*	0.75	3.00	1.35*	0.77	3.31	-0.37*
S20km2	0.81	3.24	-1.26*	0.79	3.06	0.45*	0.77	3.59	-2.79*
S25km1	0.78	2.89	2.13*	0.78	2.88	2.18*	0.75	3.24	-0.02*
S25km2	0.72	2.78	3.30*	0.68	2.51	5.61	0.72	3.12	1.51*
S30km1	0.70	2.97	1.78*	0.71	2.90	2.60*	0.80	3.55	-1.76*
S30km2	0.82	3.37	-1.40*	0.75	2.65	3.98*	0.75	3.13	0.75*

* : constante non significative au seuil 5%

Tableau 12 – Relation avec la densité de population (transformée logarithmique). Hiver

	1km			2km			5km		
	ρ	a	b	ρ	a	b	ρ	a	b
Toutes (sauf 235)	0.75	2.21	16.10	0.74	2.27	15.67	0.73	2.52	14.66
S15km1	0.73	1.96	16.46	0.72	1.96	16.21	0.71	2.44	14.48
S15km2	0.76	2.05	16.48	0.76	2.21	15.75	0.76	2.71	14.00
S20km1	0.76	1.97	16.65	0.78	2.13	15.91	0.75	2.48	14.78
S20km2	0.77	2.02	16.14	0.77	2.15	15.50	0.77	2.68	13.70
S25km1	0.76	1.95	16.51	0.75	2.10	15.88	0.73	2.58	14.19
S25km2	0.77	1.90	17.22	0.77	2.06	16.54	0.77	2.47	14.98
S30km1	0.79	2.01	16.76	0.81	2.21	15.92	0.81	2.66	14.46
S30km2	0.85	2.08	16.01	0.84	2.13	15.60	0.83	2.78	13.14

N.B. : ces coefficients ont été calculés dans R (v. 2.2.0) ; les coefficients a et b, estimés par la méthode des moindres carrés, peuvent légèrement différer des valeurs calculées par Isatis.

La relation entre concentration de NO₂ et densité d'émissions à petite échelle (densité calculée dans un rayon de 1 ou 2km) est assez sensible à la dégradation de l'échantillonnage. Avec la densité d'émissions à 5km, l'impact de la maille d'échantillonnage est moins marqué.

La relation entre concentration de NO₂ et densité de population est relativement stable.

DEGRADATION DE L'ECHANTILLONNAGE : IMPACT SUR LA MODELISATION DU VARIOGRAMME ET SUR L'ESTIMATION - HIVER

Krigeage ordinaire

MODELISATION DU VARIOGRAMME

Pour chaque sélection, en s'aidant de la validation croisée, on a recherché un modèle le plus en adéquation possible avec les données expérimentales. Les fonctions utilisées, les portées et les paliers peuvent ainsi différer d'une sélection à une autre.

De façon générale, les modèles se composent d'une première structure de petite portée (entre 15km et 25km) à laquelle, éventuellement, s'ajoute une seconde structure (de portée supérieure ou égale à 85km). D'autre part, pour quelques sélections, des modèles avec et sans effet de pépité ont été comparés. Les résultats sont proches, qu'il s'agisse des estimations par validation croisée ou des cartographies obtenues (le lissage dû à l'effet de pépité est peu perceptible). L'ajout ou non d'un petit effet de pépité n'affecte pas ici les conclusions sur l'échantillonnage.

Dans la suite, afin de ne pas multiplier les figures, la présentation des résultats se limitera aux modèles sans effet de pépité.

Exemples d'ajustement avec effet de pépité :

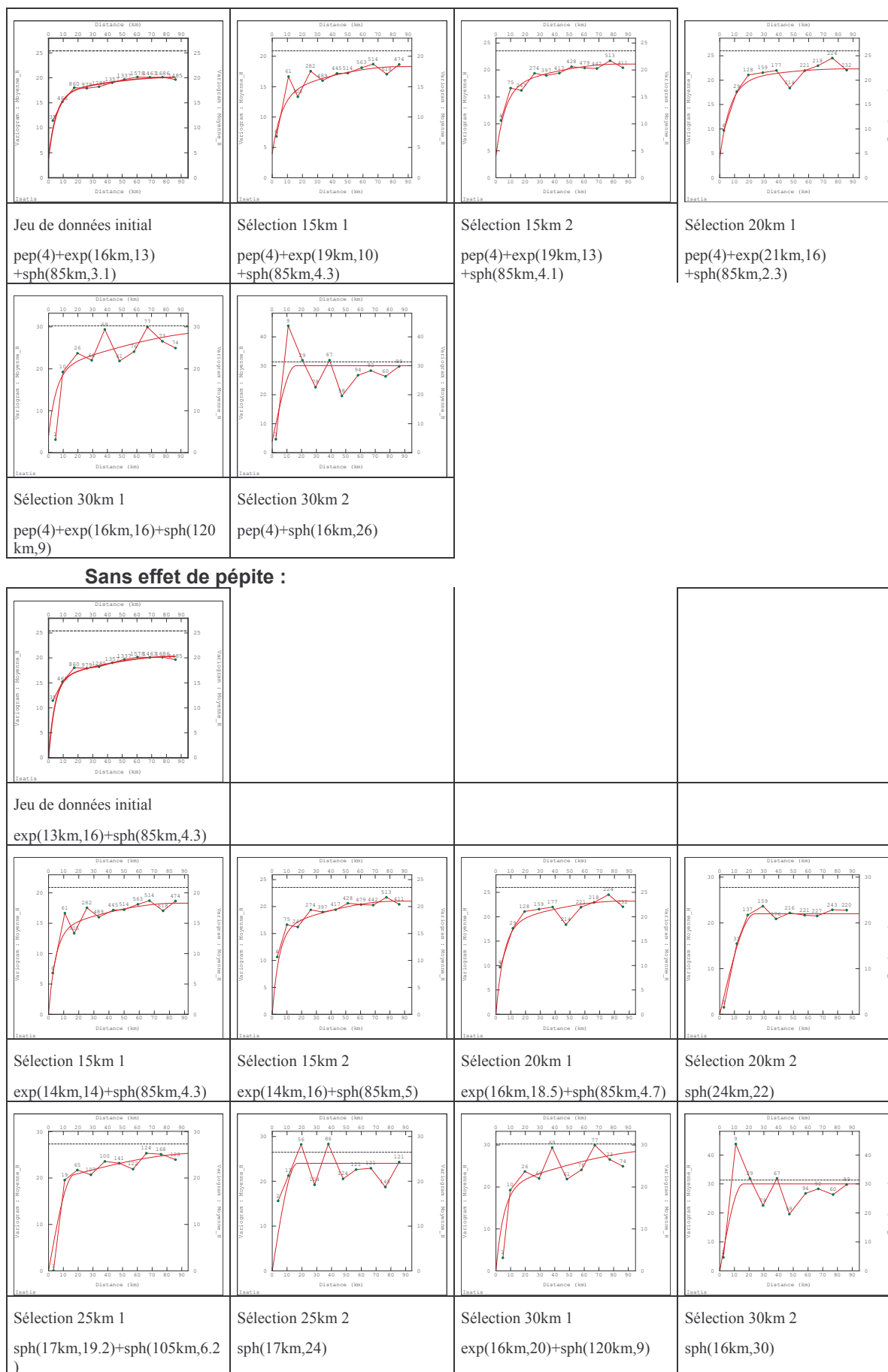


Figure 25 – Modélisation isotrope du variogramme de la concentration de NO₂

Jusqu'à un pas d'échantillonnage de 20km, le variogramme conserve une structure et hormis l'incertitude sur l'effet de pépite, il se prête relativement bien à un ajustement. A partir de la maille de 25km, il devient plus sensible à l'implantation des points et plus erratique (Figure 25).

VALIDATION CROISEE

Tous les modèles ajustés ont fait l'objet d'une validation croisée. Nous nous limitons à la présentation des nuages de corrélation entre valeurs mesurées et estimées (Figure 26). Ces nuages restent approximativement centrés sur la bissectrice ; cependant, ils sont assez dispersés, indiquant une précision d'estimation plutôt médiocre, notamment dans la gamme des concentrations supérieures à 30 $\mu\text{g}/\text{m}^3$.

Rem. : bien qu'ils soient présentés côte à côte, ces nuages ne sont pas à effectif constant. Le nombre de points diminue avec l'élargissement de la maille d'échantillonnage.

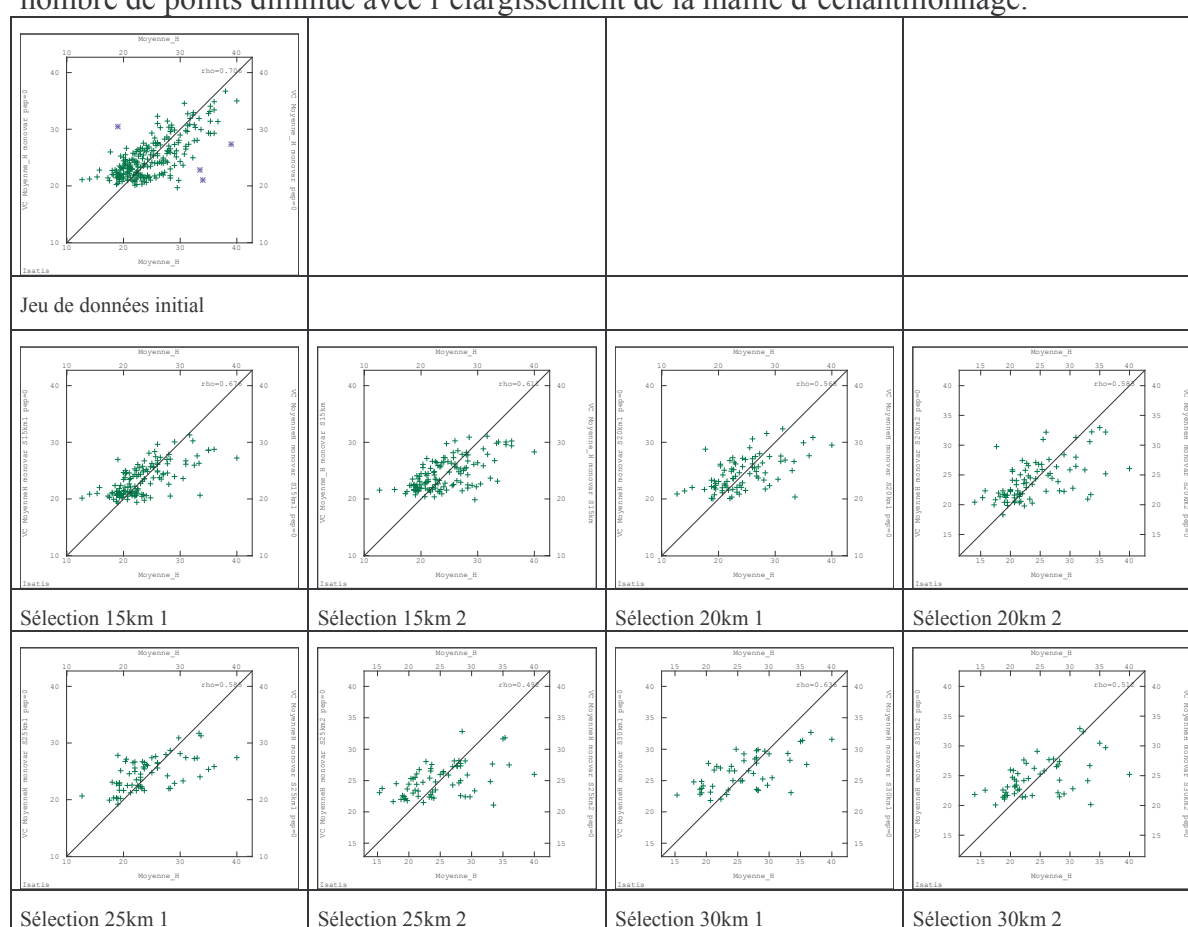


Figure 26 – Nuages de corrélation entre valeurs estimées par validation croisée et valeurs mesurées.

VERIFICATION SUR LES JEUX DE DONNEES COMPLEMENTAIRES

A chaque sélection de données est associé un jeu de données complémentaire, composé des points qui ont été exclus de l'analyse (Figure 27). Les concentrations de NO_2 en ces points sont estimées par krigeage ordinaire, à l'aide des données de la sélection et du variogramme associé (Figure 28, Figure 29). Elles sont comparées aux valeurs expérimentales. Les nuages de corrélation sont orientés selon la bissectrice, révélant une assez bonne concordance entre l'estimation et la mesure. Ils montrent néanmoins une certaine dispersion, avec en particulier la sous-estimation de quelques valeurs fortes.

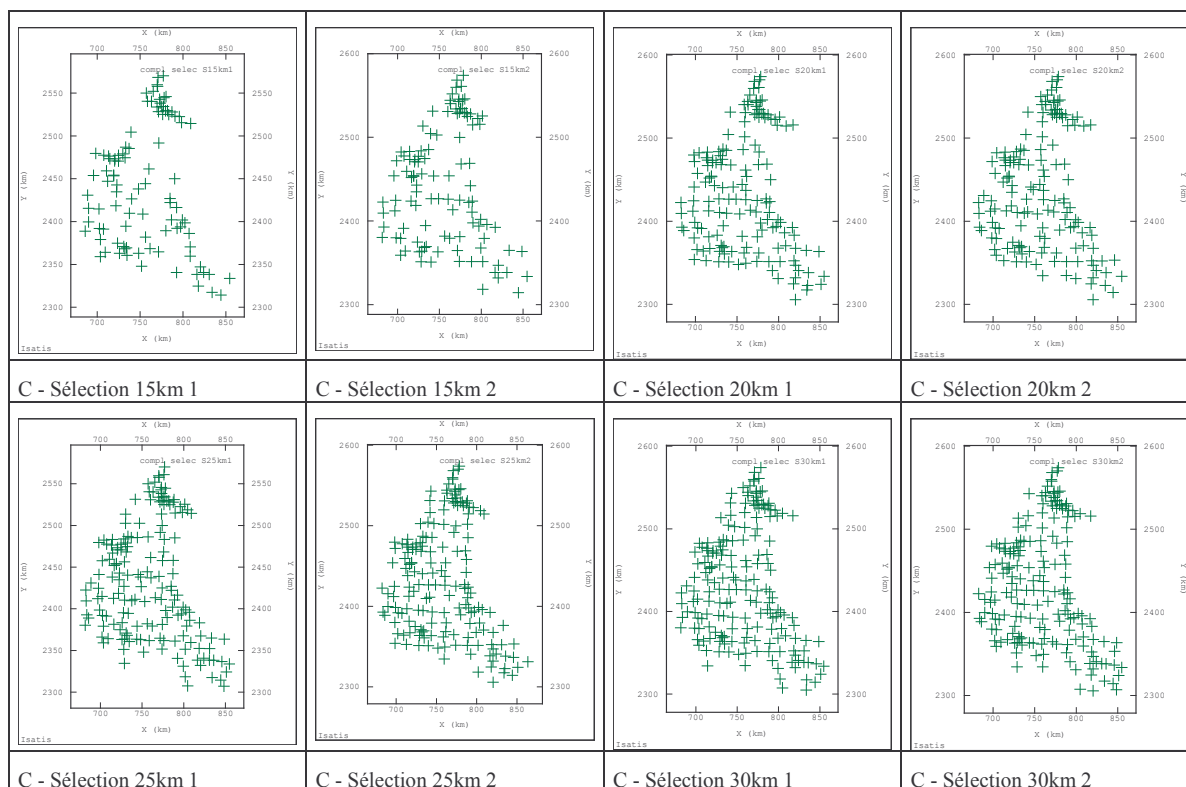


Figure 27 – Jeux de données complémentaires des huit sélections.

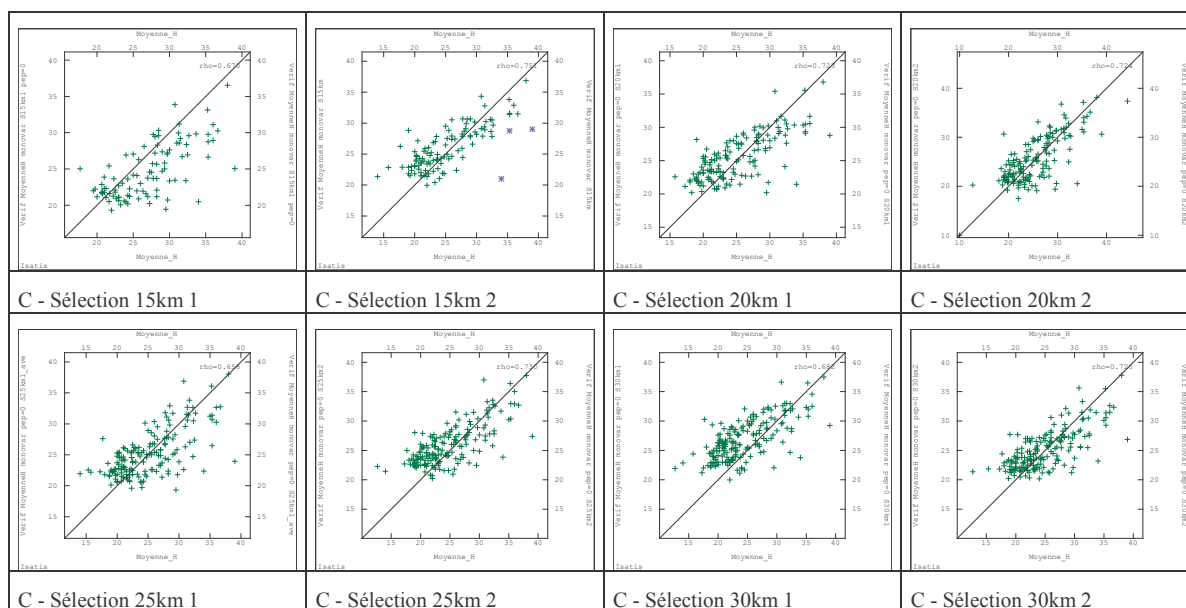


Figure 28 – Nuages de corrélation entre valeurs estimées par validation absolue sur les complémentaires de chaque sélection et valeurs mesurées

Les sélections 15km1 et 25km1 contiennent le point 10 : de faible concentration, ce point est situé dans un environnement plus pollué (Vallée de la Meuse). Cette valeur de concentration est cohérente avec la densité d'émissions, elle aussi peu élevée dans la maille kilométrique entourant le point. Ainsi elle ne semble pas due à un éventuel problème de mesure mais à une particularité locale dans les émissions de NOx. Un nouveau krigeage est effectué en retirant le point 10 des données d'entrée. L'effet sur les résultats est peu perceptible. Pour la sélection 25km1, la sous-estimation d'un petit nombre de valeurs est atténuée (Figure 29).

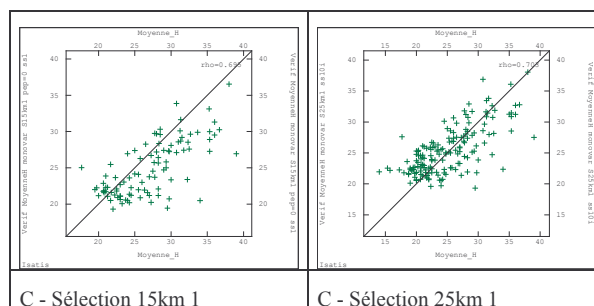


Figure 29 – Nuages de corrélation entre valeurs estimées par validation absolue sur les complémentaires de chaque sélection et valeurs mesurées.

La représentation des nuages de corrélation est limitée à l'intersection des complémentaires de façon que les points de vérification soient identiques pour toutes les sélections. La Figure 30 ne met pas en évidence d'influence particulière de la maille d'échantillonnage sur la qualité de l'estimation. Ce résultat pourrait s'expliquer par l'implantation préférentielle des points, qui ne permet pas de comparer les estimations sur l'ensemble de la région. Toutefois, si l'on compare une sélection à une autre sur l'intersection de leurs complémentaires respectifs (les points de mesure sont alors plus nombreux et mieux répartis dans l'espace), la détérioration liée à la dégradation du maillage est légère voire nulle. Ainsi, statistiquement, les plans d'échantillonnage considérés sont d'efficacité comparable.

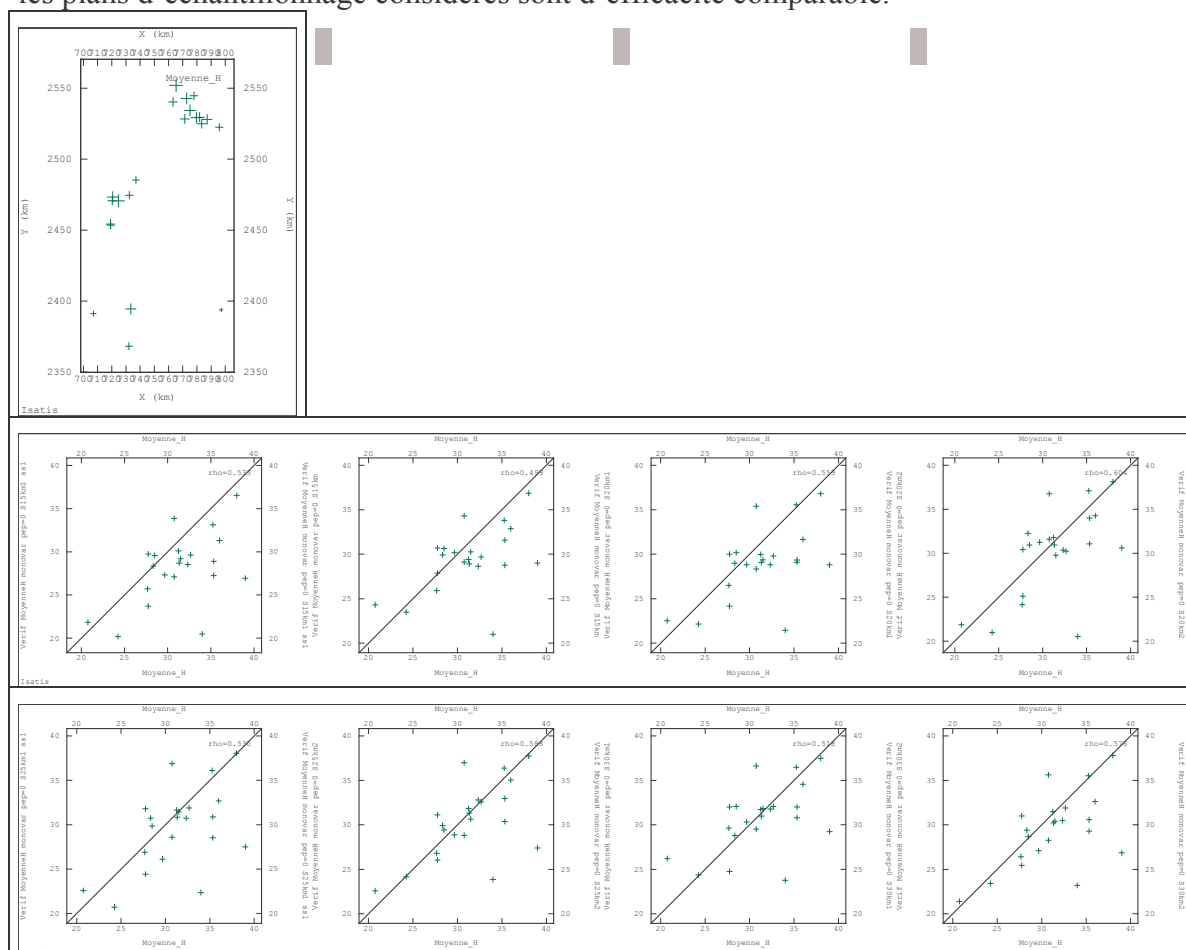
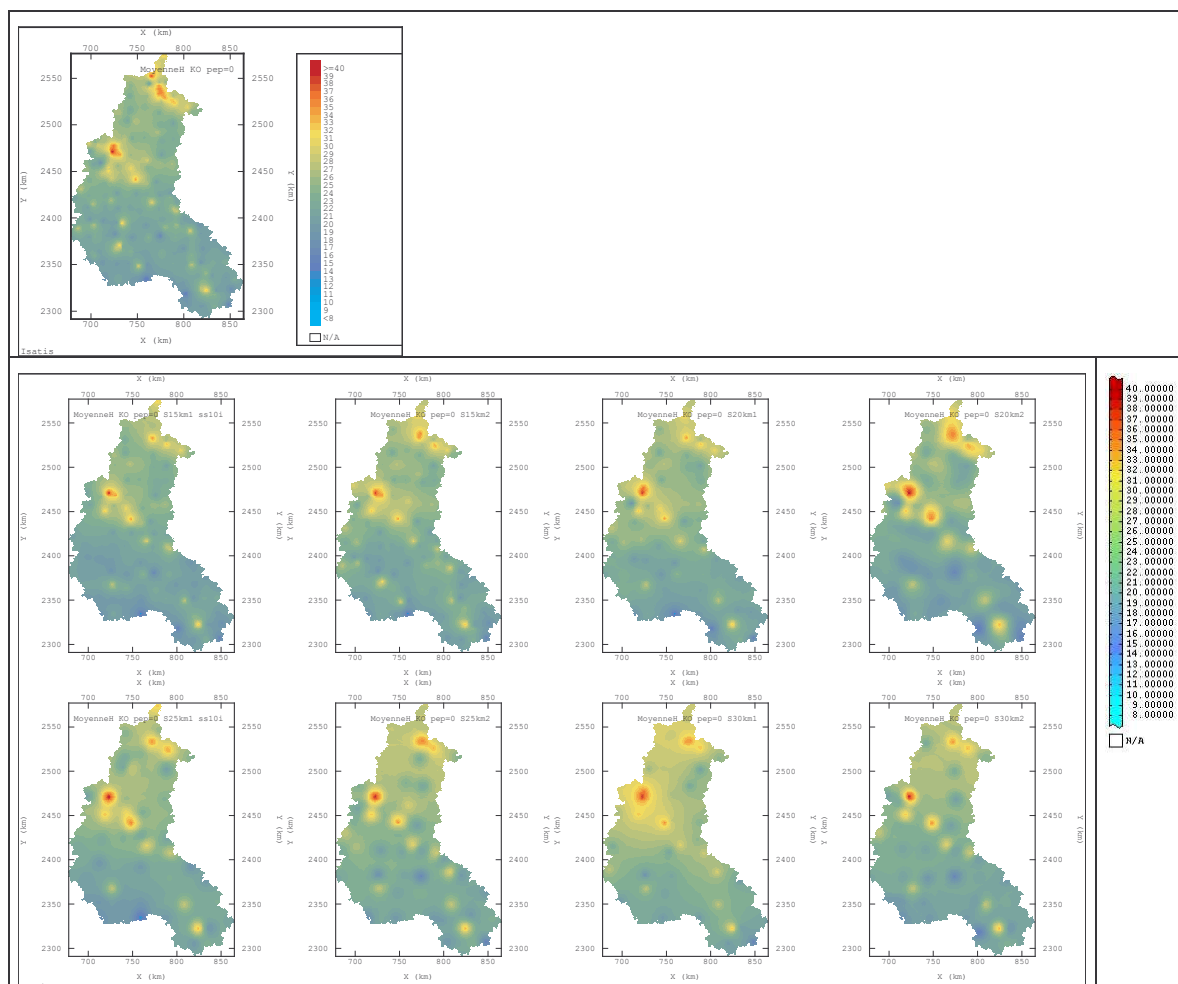


Figure 30 – Nuages de corrélation entre valeurs estimées par validation absolue et valeurs mesurées sur l'intersection des complémentaires. Ligne du haut : localisation des points.

KRIGEAGE SUR LA GRILLE D'ESTIMATION

Les concentrations de NO₂ sont estimées par krigeage ordinaire sur la grille kilométrique. La carte d'estimation est plus sensible au plan d'échantillonnage que les résultats de la vérification. Elle se différencie rapidement de l'estimation de référence au fur et à mesure que le nombre de données diminue (Figure 31, Figure 32). La suppression de points de mesure fait logiquement augmenter l'écart-type de krigeage.



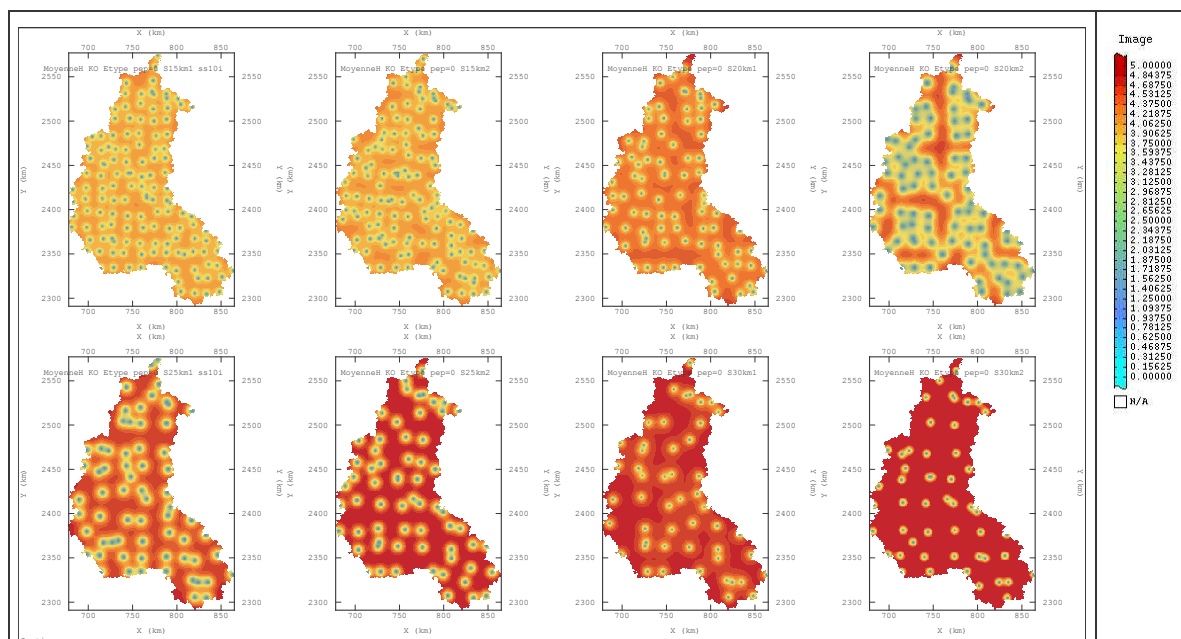


Figure 31 – Estimation par krigeage ordinaire ponctuel dans le cas de référence et pour une maille d'échantillonnage de plus en plus lâche. Partie supérieure : concentrations de NO₂ estimées [µg/m³] ; partie inférieure : écarts-types de l'erreur krigeage [µg/m³]. Les figures présentées correspondent à un modèle sans effet de pépite.

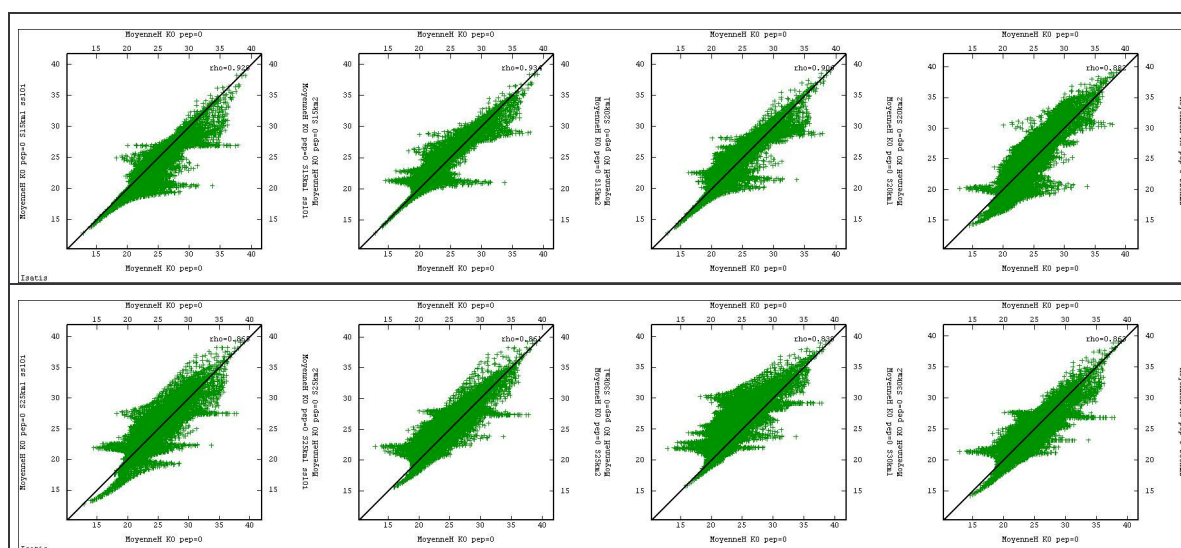


Figure 32 – De gauche à droite et de haut en bas : comparaison entre les estimations associées à des mailles d'échantillonnage de plus en plus lâches et l'estimation de référence

Krigeage des résidus

Pour chaque sélection, on reproduit la démarche décrite en 0, complétée d'une vérification sur les jeux de données complémentaires :

- Régression multilinéaire : la concentration de NO₂ est exprimée en fonction de la densité d'émissions à 5km et de la densité de population à 1km

NB : 1.) Afin de faciliter la démarche et la comparaison entre plans d'échantillonnage, les mêmes variables auxiliaires sont utilisées pour toutes les sélections, même si pour quelques unes d'entre elles, la concentration de NO₂ est plus corrélée à la densité d'émissions dans un rayon d'1km ou 2km 2.) Pour la sélection 30km² la densité d'émissions n'est pas retenue En effet, le coefficient de

régression associé à cette variable n'est pas significatif dans la régression bilinéaire.

- Modélisation du variogramme des résidus et validation croisée ;
- Validation absolue ;
- Krigeage des résidus.

MODELISATION DU VARIOGRAMME

Comme en krigeage ordinaire, des modèles sont ajustés au mieux sur les variogrammes expérimentaux en s'aidant de la validation croisée. Ils sont composés d'un effet de pépite, pour la plupart d'une structure exponentielle, et pour tous, d'une structure sphérique (Figure 33).

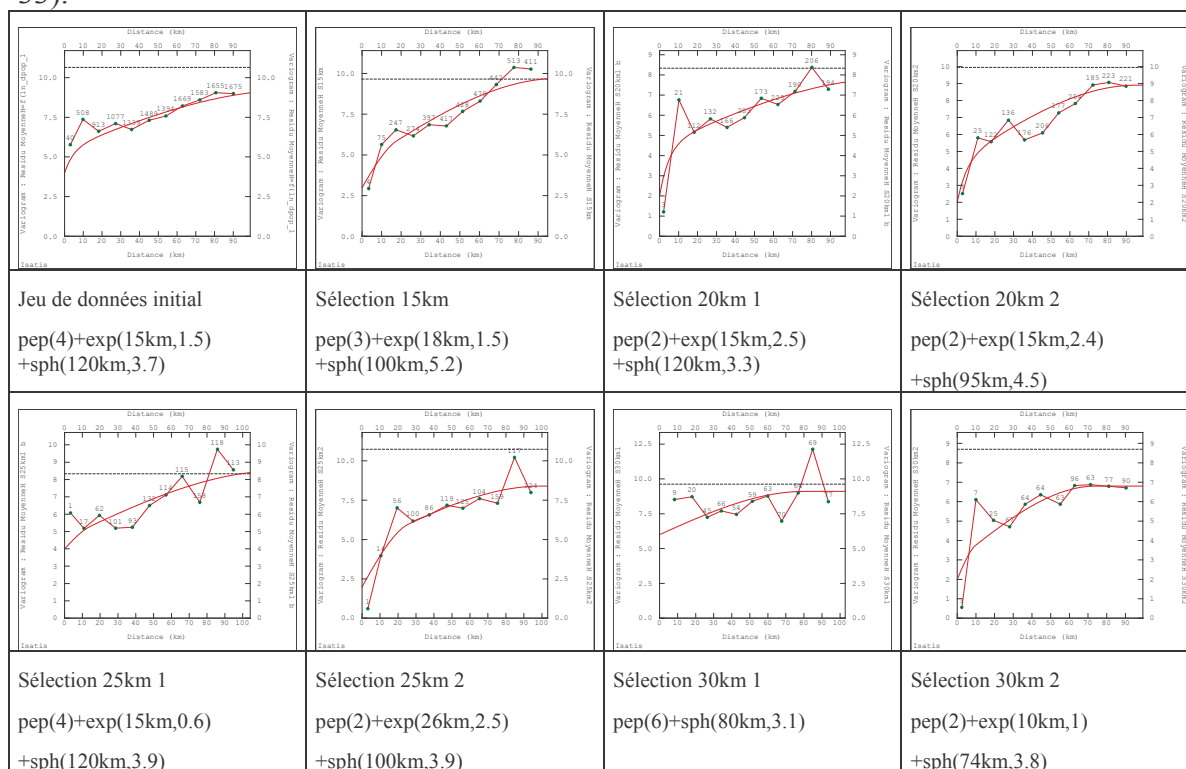


Figure 33 – Modélisation isotrope du variogramme des résidus de la régression $\text{NO}_2=f[\ln(1+\text{dpop1km}),\ln(1+\text{dE5km})]$

Pour les sélections 15km1 et 25km1, la régression et le variogramme des résidus sont construits avec le point 10. En effet, si la concentration mesurée y est plus faible qu'aux points environnants, les émissions de NO_x en ce point sont elles aussi plus faibles.

VALIDATION CROISEE

Quelle que soit la sélection, la validation croisée estime les valeurs mesurées avec une précision supérieure à celle du krigeage ordinaire (les nuages de corrélation de la Figure 34 sont plus resserrés sur la bissectrice).

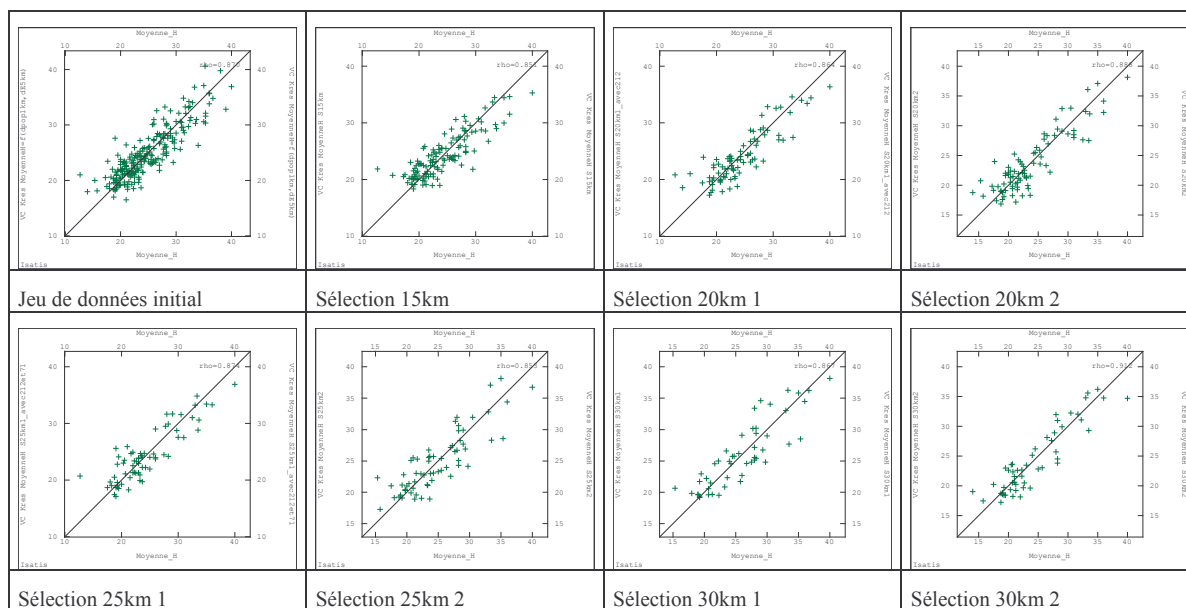


Figure 34 – Nuages de corrélation entre valeurs estimées par validation croisée et valeurs mesurées

VERIFICATION SUR LES JEUX DE DONNEES COMPLEMENTAIRES

Le gain de précision par rapport au krigeage ordinaire, que l'on a observé en validation croisée, se confirme en validation absolue (Figure 35).

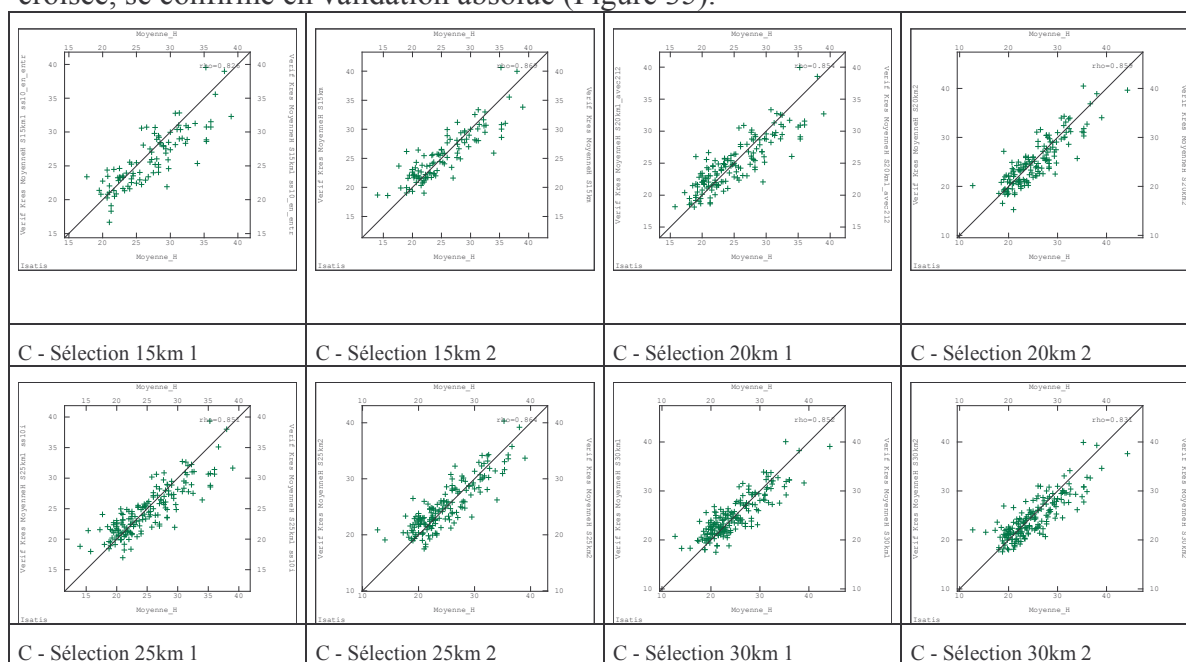


Figure 35 – Nuages de corrélation entre valeurs estimées par validation absolue sur les complémentaires de chaque sélection et valeurs mesurées

Rem. : le point 10 a été retiré des sélections S15km1 et S25km1 avant de réaliser l'estimation sur leurs complémentaires.

Comme en krigeage ordinaire, l'estimation sur l'intersection des complémentaires n'est pas sensible à l'échantillonnage (Figure 36).

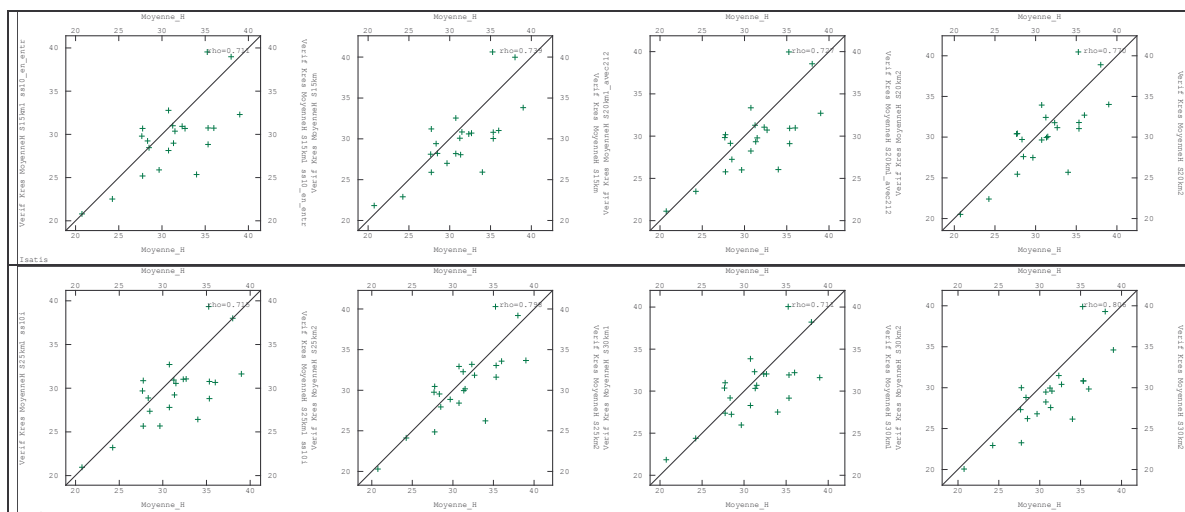


Figure 36 – Nuages de corrélation entre valeurs estimées par validation absolue sur l'intersection des complémentaires et valeurs mesurées

KRIGEAGE SUR LA GRILLE D'ESTIMATION

Les concentrations de NO₂ sont estimées sur la grille kilométrique en faisant la somme de la régression et des résidus interpolés par krigeage. A la différence du krigeage ordinaire, les cartographies obtenues restent proches de la cartographie de référence (Figure 37, Figure 38). Des écarts un peu plus marqués apparaissent localement avec la maille de 30km, et notamment avec la sélection S30km2 pour laquelle la densité d'émissions n'a pas été utilisée (Figure 39).

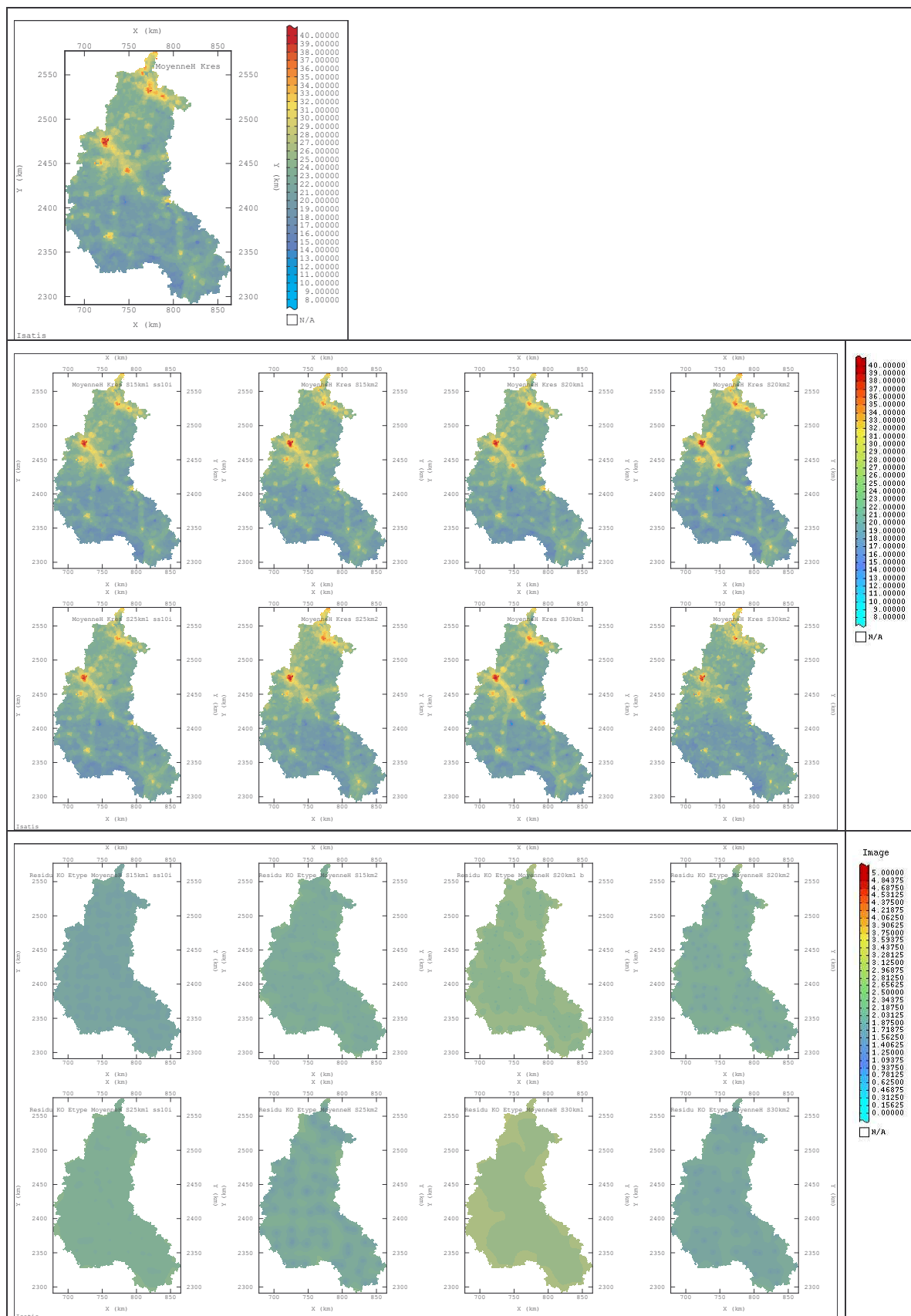


Figure 37 – Estimation par krigeage des résidus dans le cas de référence et pour une maille d'échantillonnage de plus en plus lâche. Partie supérieure : concentrations de NO₂ estimées [µg/m³] ; partie inférieure : écarts-types de l'erreur krigeage [µg/m³].

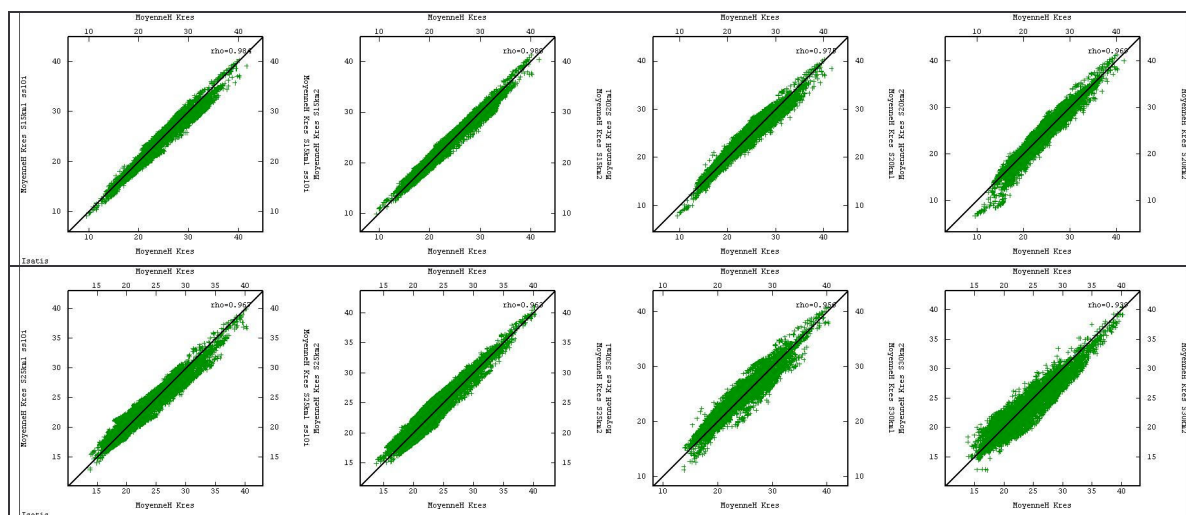


Figure 38 – De gauche à droite et de haut en bas : comparaison entre les estimations associées à des mailles d'échantillonnage de plus en plus lâches et l'estimation de référence

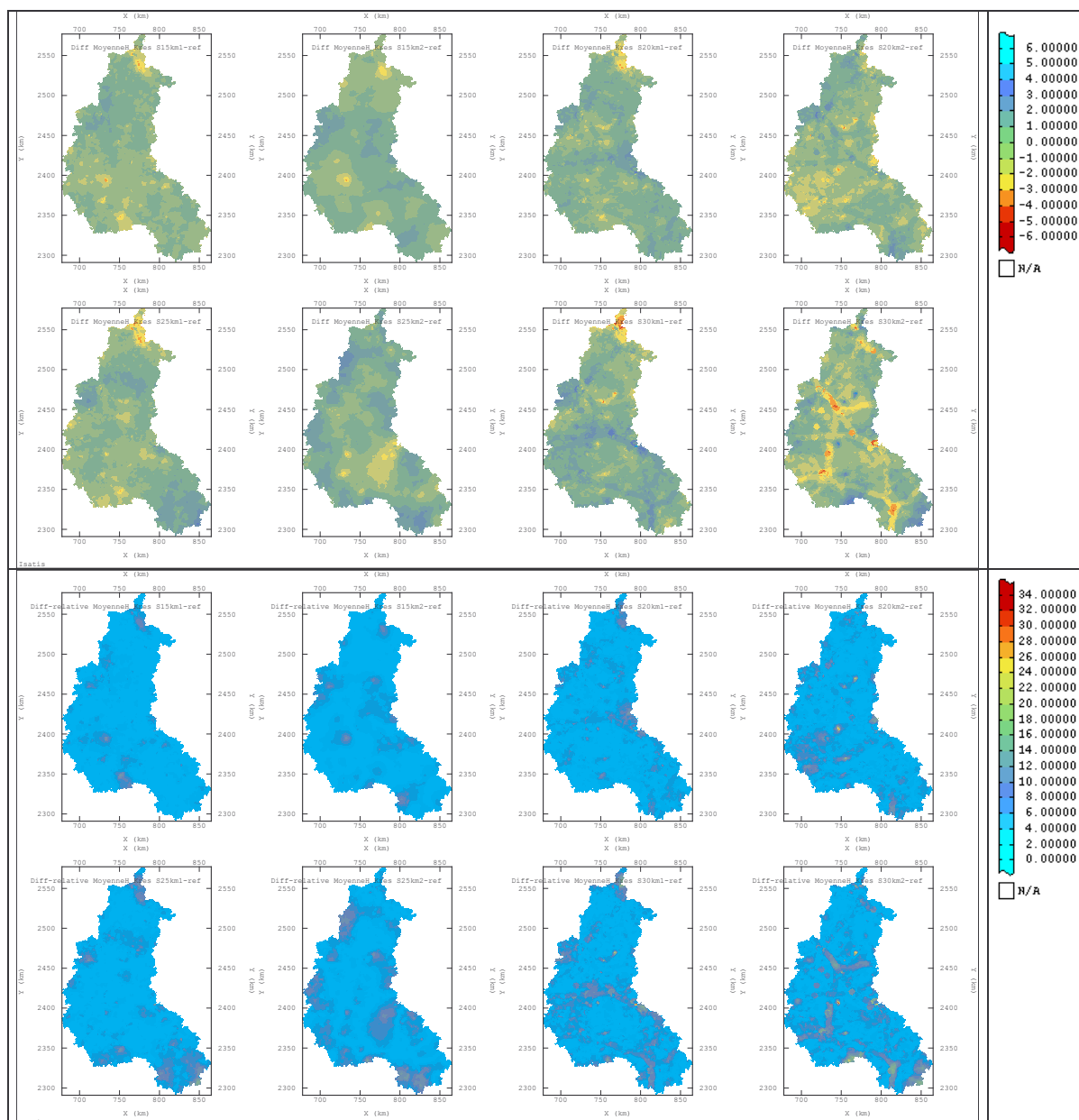


Figure 39 – Cartes des différences (les huit figures du haut) $[\mu\text{g}/\text{m}^3]$ et des différences relatives (les huit figures du bas)%, en valeur absolue] entre les estimations obtenues pour chaque sélection et l'estimation de référence.

Complément sur l'incertitude d'estimation

En complément des nuages de corrélation, quelques statistiques des erreurs relatives, c'est-à-dire des écarts relatifs (en valeur absolue) entre estimations et mesures, sont fournies sous forme graphique. On rappelle que l'erreur relative correspond à la définition réglementaire de l'incertitude de modélisation.

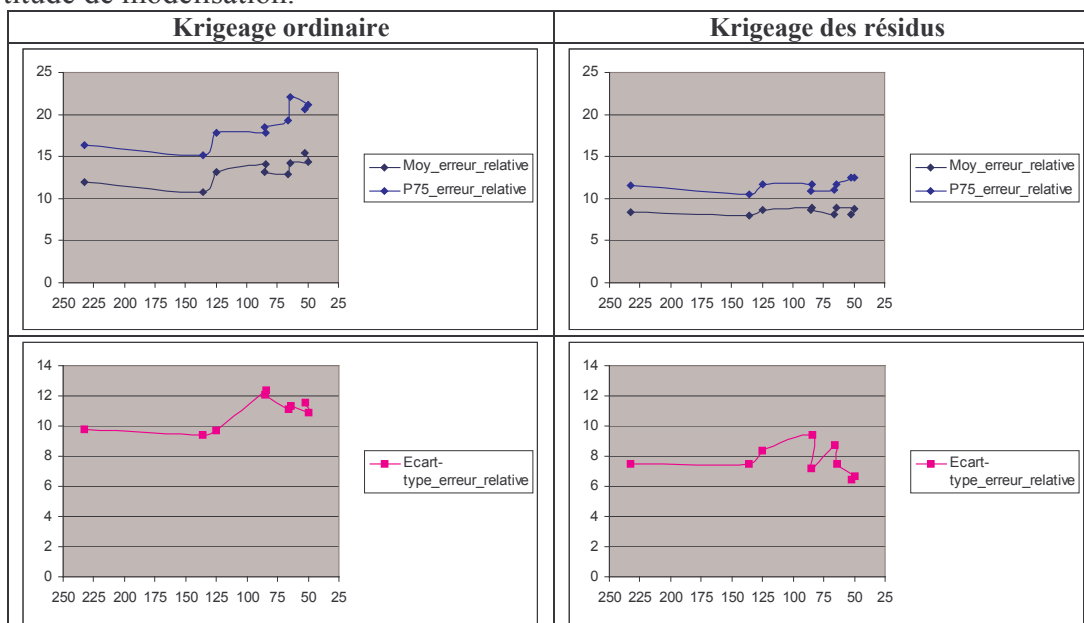


Figure 40 – Moyenne, centile 75 et écart-type des erreurs relatives de validation croisée. La valeur en abscisse indique l'effectif du jeu de données d'entrée ; celui-ci est égal à l'effectif des erreurs sur lesquelles sont calculées les statistiques.

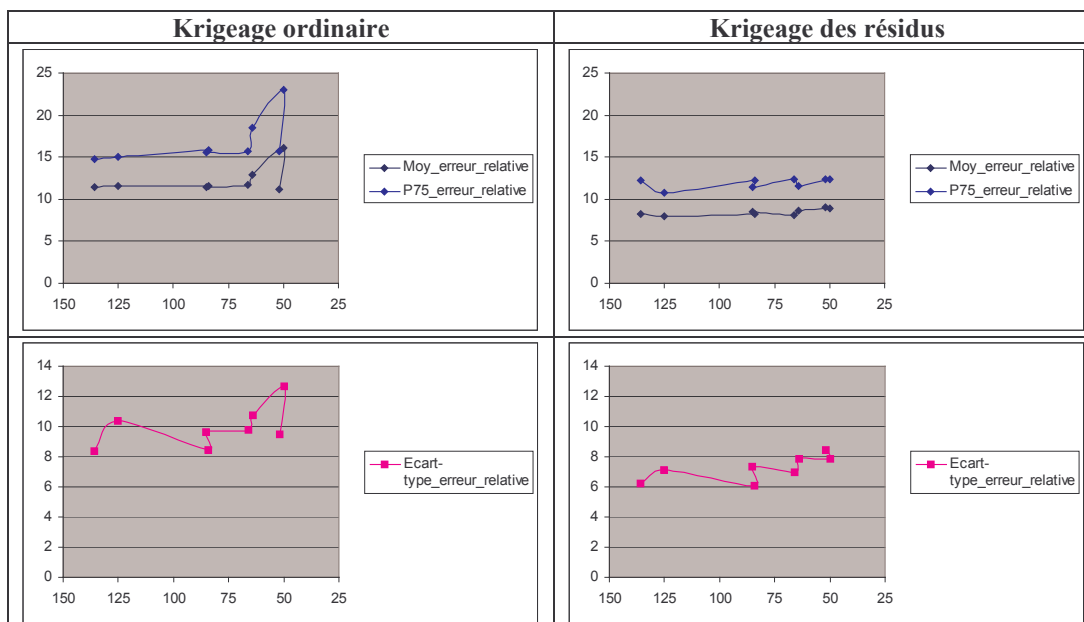


Figure 41 – Moyenne, centile 75 et écart-type des erreurs relatives de validation absolue. La valeur en abscisse indique l'effectif du jeu de données d'entrée ; l'effectif des erreurs sur lesquelles sont calculées les statistiques est de : 96, 108, 149, 149, 166, 170, 184, 182

Si l'erreur relative maximale peut localement dépasser 30%, l'objectif de qualité fixé pour la moyenne annuelle de NO₂ (30% en modélisation) est généralement respecté. Les courbes de la Figure 41 font également voir l'effet favorable des variables auxiliaires.

STRATIFICATION DE L'ECHANTILLONNAGE EN FONCTION DES EMISSIONS

A partir des plans d'échantillonnage stratifiés définis au paragraphe 0, un krigeage des résidus a été mis en œuvre selon la démarche décrite en 0 :

- étude des corrélations avec les variables auxiliaires (Tableau 13, Tableau 14) ;
- modélisation du variogramme et validation croisée ;
- validation absolue (Figure 42, Figure 43, Figure 44) ;
- estimation sur la grille (Figure 45) ;

Les estimations obtenues ressemblent à celles des sélections 30km1 et 30km2. Pour la sélection S30km2+, la densité d'émissions à 5km a été conservée parce qu'elle améliore les résultats de validation croisée et de vérification. Toutefois, son poids relativement faible dans la régression lui donne peu d'influence sur la cartographie. Pour la sélection 40km1+, on distingue quelques zones assez étendues où l'estimation diffère de plus de 20% de la cartographie de référence. (A titre d'information, la Figure 47 met en correspondance les différences relatives *estimation de la sélection-estimation de référence* avec la densité de population.)

Tableau 13 – Relation avec la densité d'émissions de NOx (transformée logarithmique). Hiver

	1km			2km			5km		
	ρ	a	b	ρ	a	b	ρ	a	b
Toutes (sauf 235)	0.65	2.56	4.87	0.66	2.66	4.31	0.68	3.11	1.20*
S30km1	0.70	2.97	1.78*	0.71	2.90	2.60*	0.80	3.55	-1.76*
S30km1plus	0.65	2.53	5.22*	0.68	2.68	4.27*	0.79	3.61	-2.40*
S30km2	0.82	3.37	-1.40*	0.75	2.65	3.98*	0.75	3.13	0.75*
S30km2plus	0.73	2.87	2.58	0.70	2.44	5.71	0.73	3.12	0.79*
S40km1plus	0.70	2.53	5.06*	0.70	2.71	6.05	0.76	3.35	-0.58*
S40km2plus	0.72	2.77	3.40*	0.74	2.65	4.28*	0.79	3.43	-1.18*

* : constante non significative au seuil 5%

Tableau 14 – Relation avec la densité de population (transformée logarithmique). Hiver

	1km			2km			5km		
	ρ	a	b	ρ	a	b	ρ	a	b
Toutes (sauf 235)	0.75	2.21	16.10	0.74	2.27	15.67	0.73	2.52	14.66
S30km1	0.79	2.01	16.76	0.81	2.21	15.92	0.81	2.66	14.46
S30km1plus	0.77	2.00	16.49	0.78	2.19	15.76	0.81	2.70	13.98
S30km2	0.85	2.08	16.01	0.84	2.13	15.60	0.83	2.78	13.14
S30km2plus	0.82	2.02	16.22	0.80	2.07	15.85	0.81	2.76	13.10
S40km1plus	0.75	1.89	16.79	0.73	2.01	16.19	0.74	2.70	13.7
S40km2plus	0.78	1.98	16.43	0.78	2.13	15.72	0.81	2.95	12.61

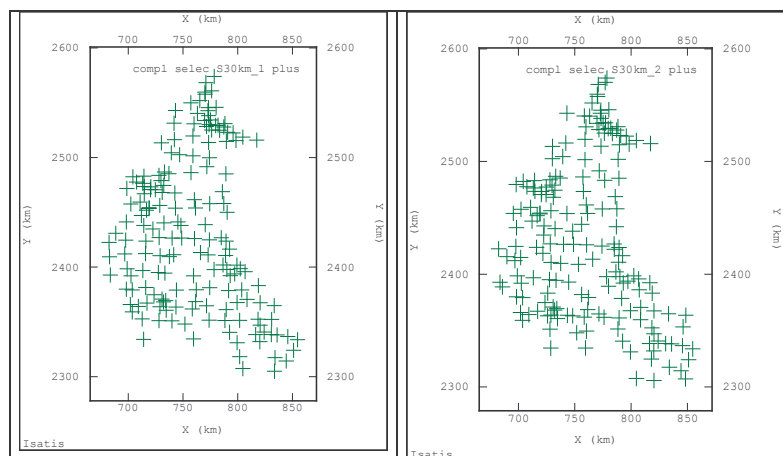


Figure 42 – Jeux de données complémentaires des sélections S30km1+ et S30km2+.

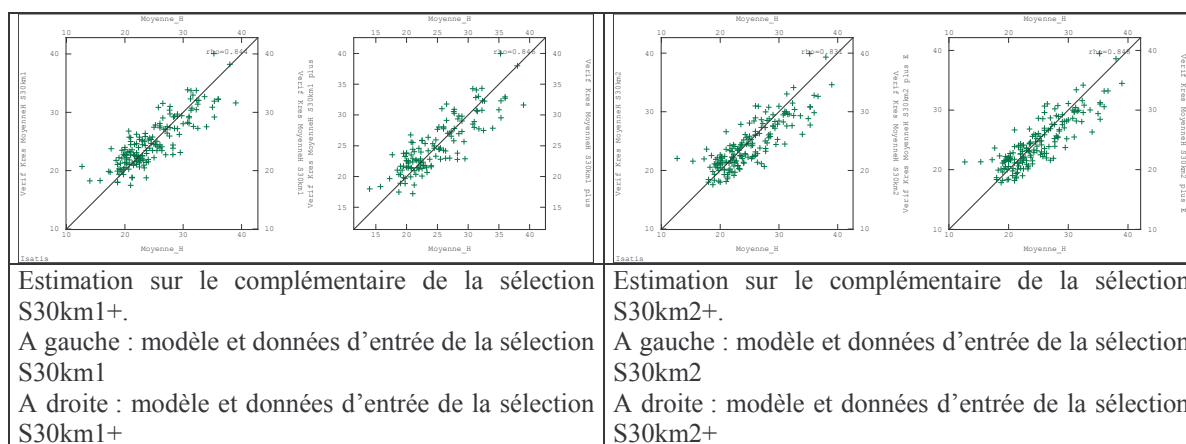


Figure 43 - Nuages de corrélation entre valeurs estimées par validation absolue et valeurs mesurées. Krigeage des résidus.

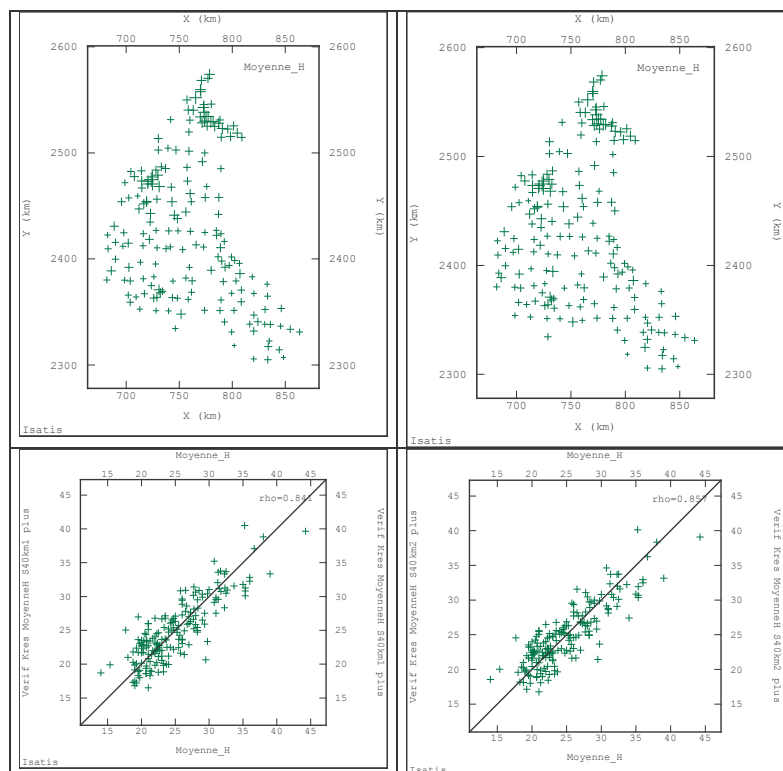


Figure 44 – Jeux de données complémentaires des sélections S40km1+ et S40km2+.
Nuages de corrélation entre valeurs estimées par validation absolue et valeurs mesurées.
Krigeage des résidus.

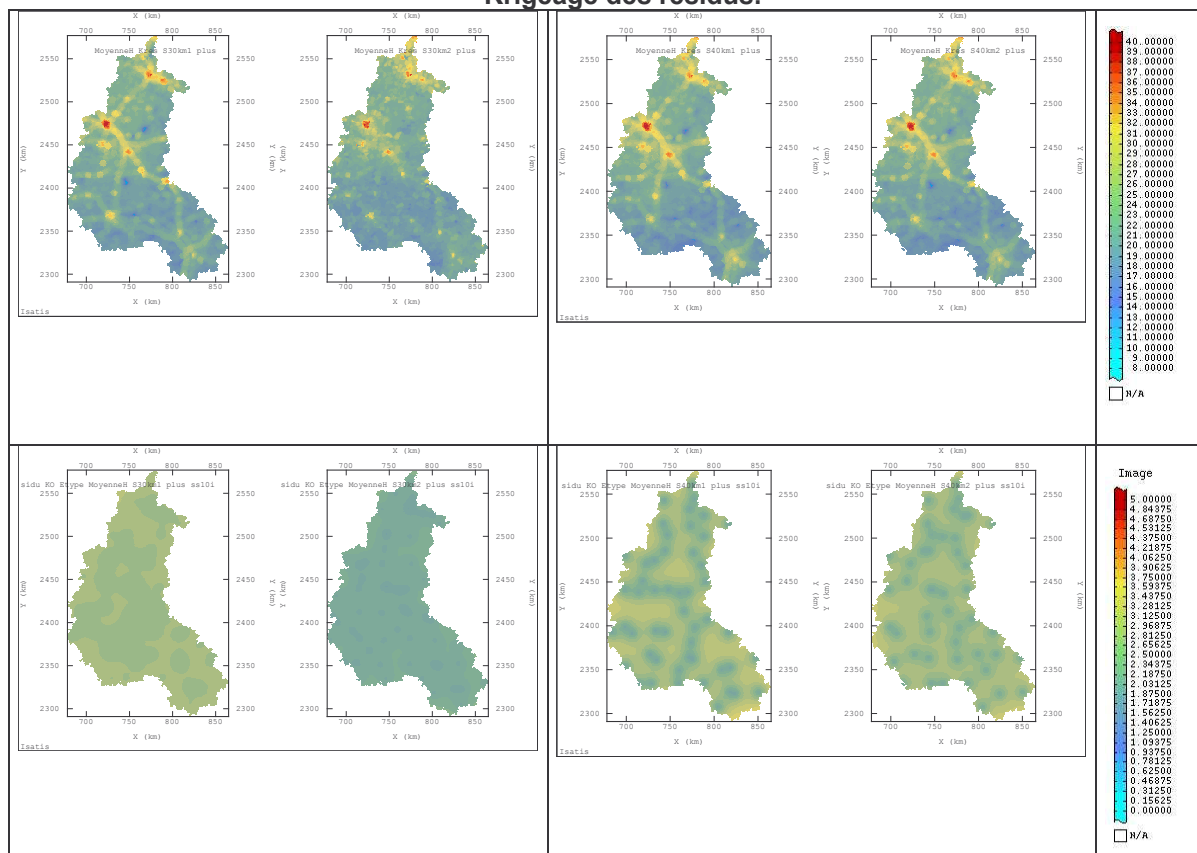


Figure 45 – Concentrations de NO₂ estimées par krigeage des résidus (ligne du haut) et écarts-types de l'erreur de krigeage (ligne du bas) [µg/m³]. A gauche : sélections 30km1+ et 30km2+. A droite : sélections 40km1+ et 40km2+.

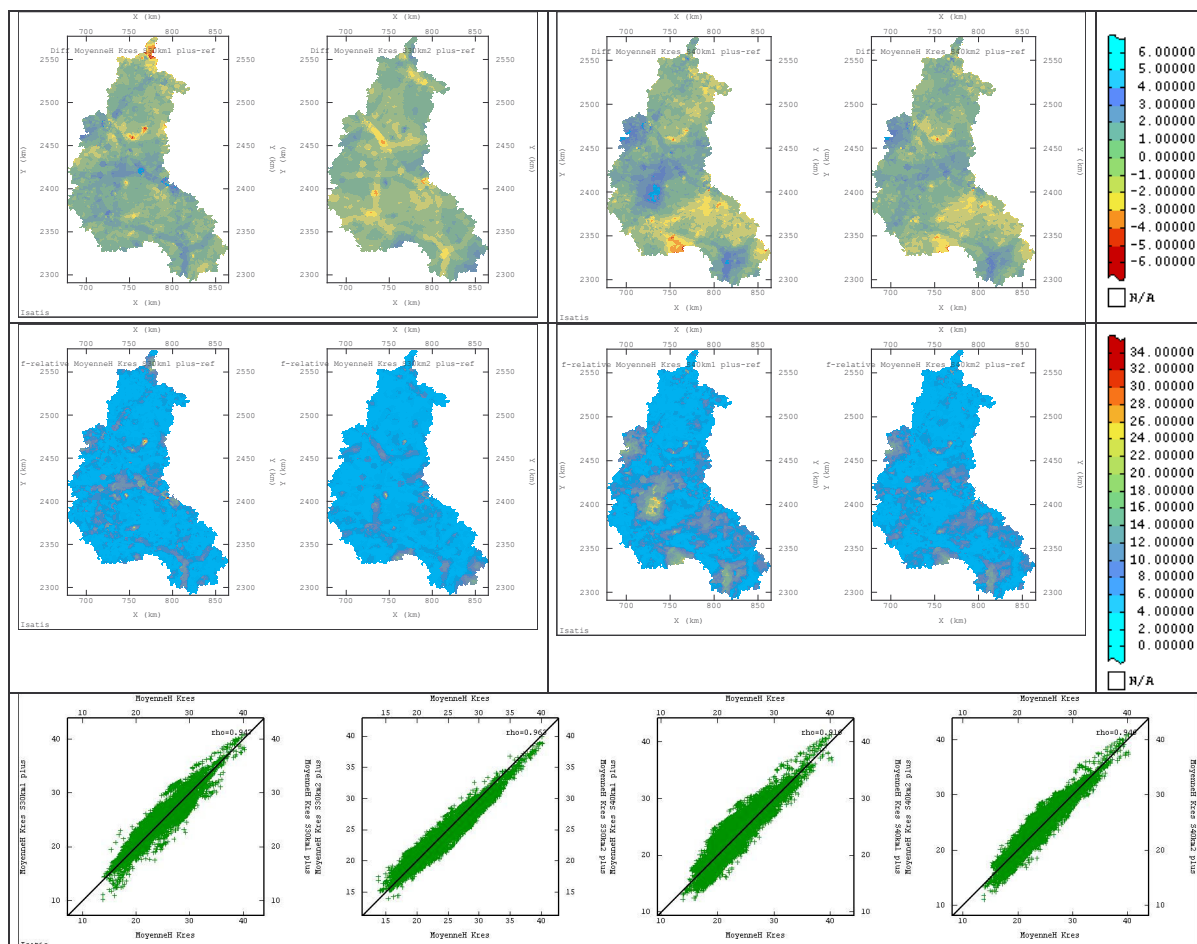


Figure 46 – Cartes des différences (ligne du haut) $[\mu\text{g}/\text{m}^3]$, cartes des différences relatives (ligne du milieu) [%], en valeur absolue] et nuages de corrélation (ligne du bas) entre les estimations obtenues pour chaque sélection et l'estimation de référence. A gauche : sélections 30km1+ et 30km2+. A droite : sélections 40km1+ et 40km2+.

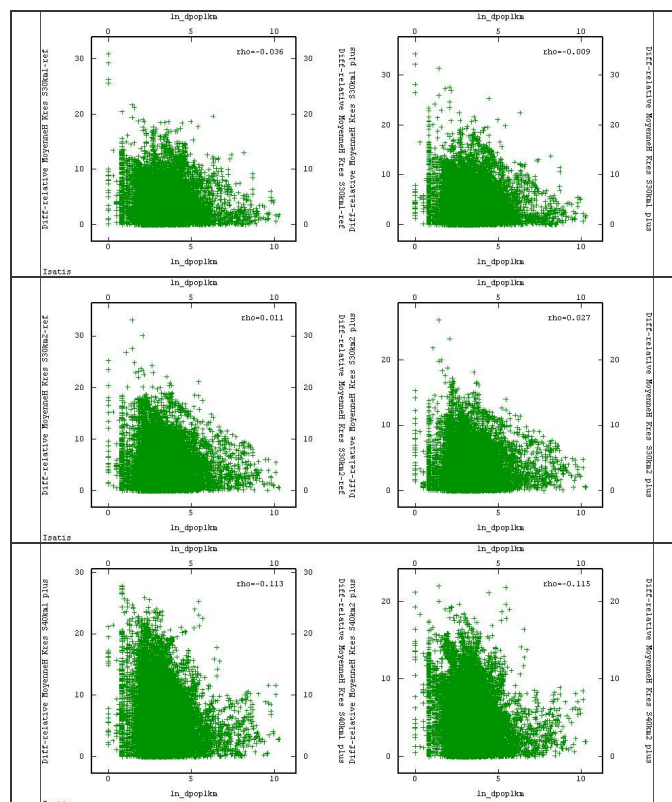


Figure 47 – Différences relatives par rapport à l'estimation de référence (ordonnée) et densité de population (abscisse). De haut en bas : sélections 30km1 et 30km1+ ; 30km2 et 30km2+ ; 40km1+ et 40km2+ Dans les zones de plus forte densité de population, ces écarts sont compris entre 0 et 15% environ.

ETUDE SANS LA DENSITE DE POPULATION

Comme cela a été mis en évidence, l'exploitation de données fines d'émissions et de population permet d'accroître la précision de l'estimation. L'objet de ce chapitre est d'apprécier la sensibilité de l'estimation aux variables auxiliaires disponibles. Avec la sélection 30km2, on a déjà vu qu'ignorer la variable d'émissions influençait la cartographie finale dans laquelle les axes routiers ressortent moins nettement. Dans cette partie, on se limite à l'utilisation des données d'émissions.

La sélection 30km2+ est considérée, ainsi que la densité d'émissions à 1km, légèrement plus corrélée à la concentration de NO₂ que la densité d'émissions à 5km (Figure 48).

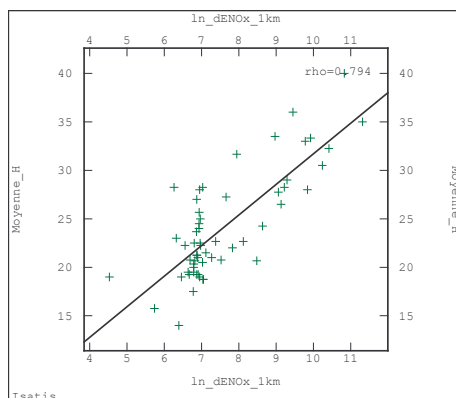


Figure 48 – Nuage de corrélation entre concentration moyenne hivernale de NO₂ et densité d'émissions dans un rayon de 1km (transformée logarithmique)

Une relation linéaire est établie par régression entre la concentration de NO₂ et la densité d'émissions à 1km.

Un modèle composé d'un effet de pépite et de deux structures sphériques est ajusté sur le variogramme expérimental des résidus de cette régression. La Figure 49 présente les résultats de validation croisée et de vérification.

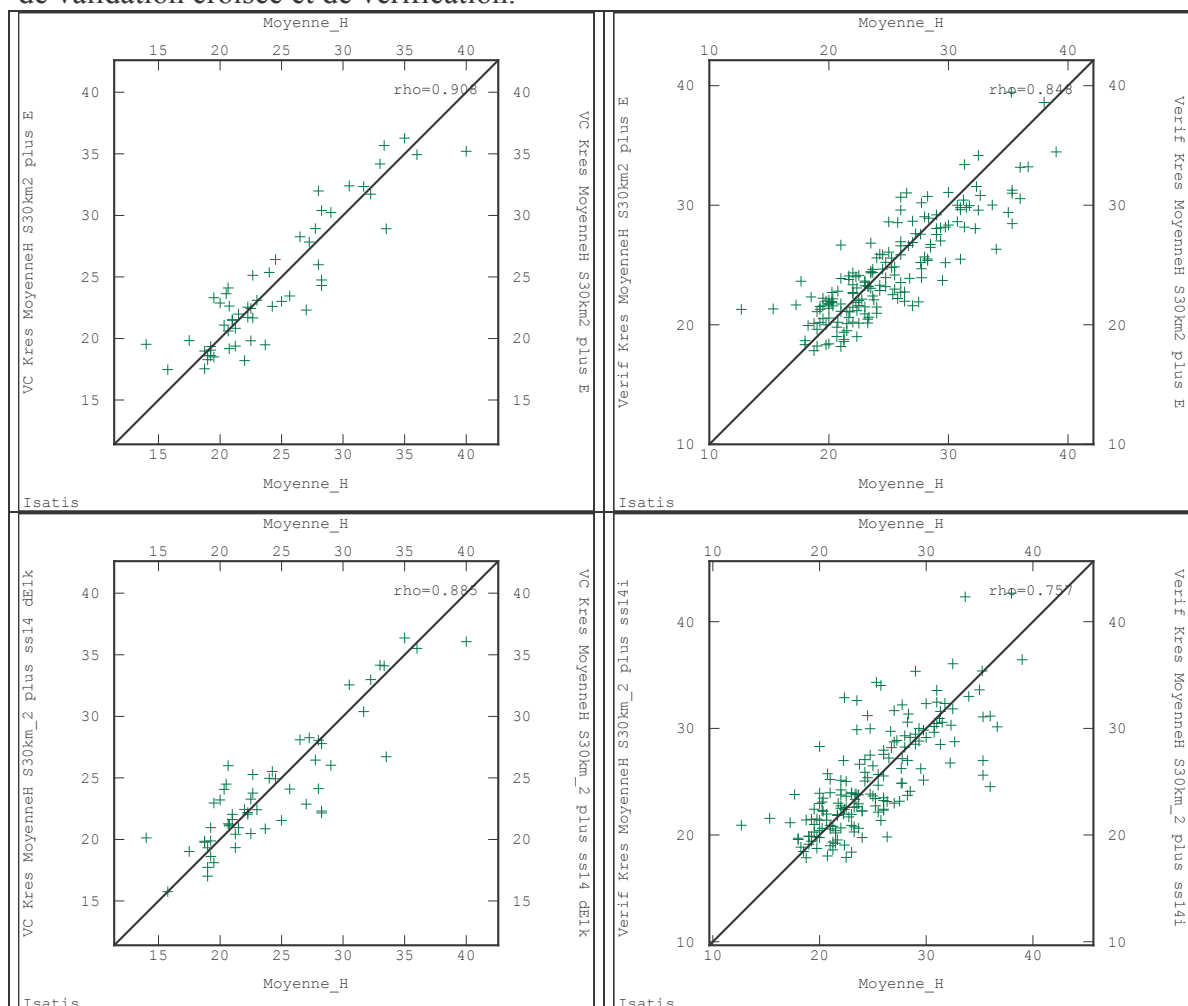


Figure 49 – Nuages de corrélation entre les valeurs estimées par validation croisée (fig. de gauche)/validation absolue sur le complémentaire (fig. de droite) et les valeurs mesurées.

Ligne du haut : estimation avec population à 1km et émissions à 5km

Ligne du bas : estimation avec émissions à 1km

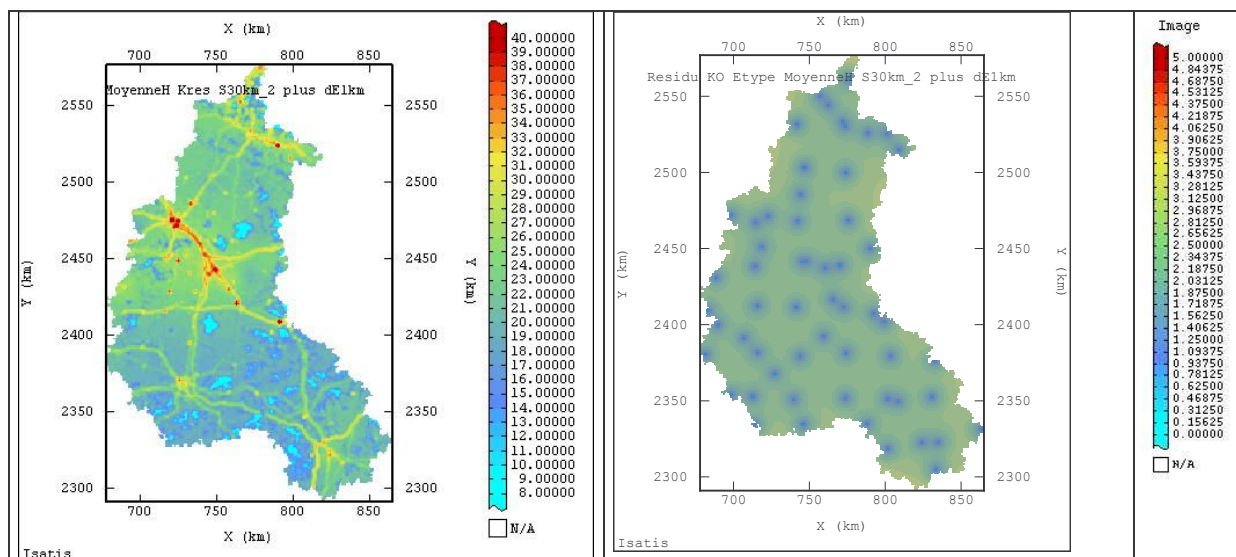


Figure 50 – Concentrations de NO₂ estimées par krigeage des résidus et écarts-types de l'erreur de krigeage [µg/m³].

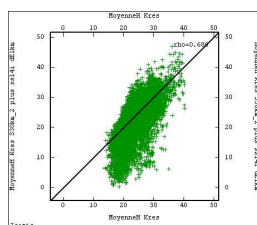


Figure 51 – Comparaison entre l'estimation obtenue pour la sélection 30km2+ avec la densité d'émissions et l'estimation de référence [µg/m³].

Les résultats de vérification obtenus d'une part, avec la densité de population à 1km et la densité d'émissions à 5km, et d'autre part, avec la densité d'émissions à 1km ont été comparés par zone géographique. La densité d'émissions à 1km permet généralement de mieux approcher la concentration au voisinage des plus grands axes (concentration que le krigeage avec population tend à sous-estimer). En revanche, elle induit localement des surestimations (certains petits points jaunes ou rouges sur la carte). Statistiquement, le krigeage avec la seule densité d'émissions est moins précis.

Ces résultats mettent en exergue l'influence directe de la variable auxiliaire sur l'estimation, notamment dans les zones dépourvues de mesures. Ainsi la contrepartie d'un échantillonnage restreint, même si la validation croisée se révèle satisfaisante, est la nécessité d'accorder un large crédit à la variable auxiliaire.

Remarque : Il serait intéressant de comparer le krigeage des résidus au cokrigeage colocalisé. Néanmoins, un échantillonnage réduit ne favorise pas l'application de cette seconde méthode parce qu'il rend malaisée la détermination des structures spatiales (simples et croisée) des variables. Dans cet exemple, les variogrammes simples et le variogramme croisé de la concentration et des émissions ont un aspect pépétique.

ETUDE DE L'ETE

L'analyse géostatistique du jeu de données a été conduite en liaison avec ATMO Champagne-Ardenne. Nous en présentons ci-après les principaux résultats, en complétant l'étude par une dégradation de l'échantillonnage.

Etude du jeu de données initial

HISTOGRAMMES

Les niveaux de concentration estivaux sont nettement inférieurs à ceux de l'hiver, avec une assez large proportion de données proches de la limite de quantification ($\sim 6-7 \mu\text{g}/\text{m}^3$). Quelques valeurs plus élevées que les autres, qui ne sont pas toujours situées aux mêmes points, se détachent des histogrammes associés à chaque période.

- période 1 : points 16, 143 ;
- période 2 : point 204 ;
- période 3 : point 88
- période 4 : points 204, 143

(Rem. : en été, le site 235, exclus de l'analyse hivernale, n'a pas fait l'objet d'un échantillonnage.) En revanche, les concentrations hivernales mesurées en ces points appartiennent bien à la distribution statistique des données de l'hiver.

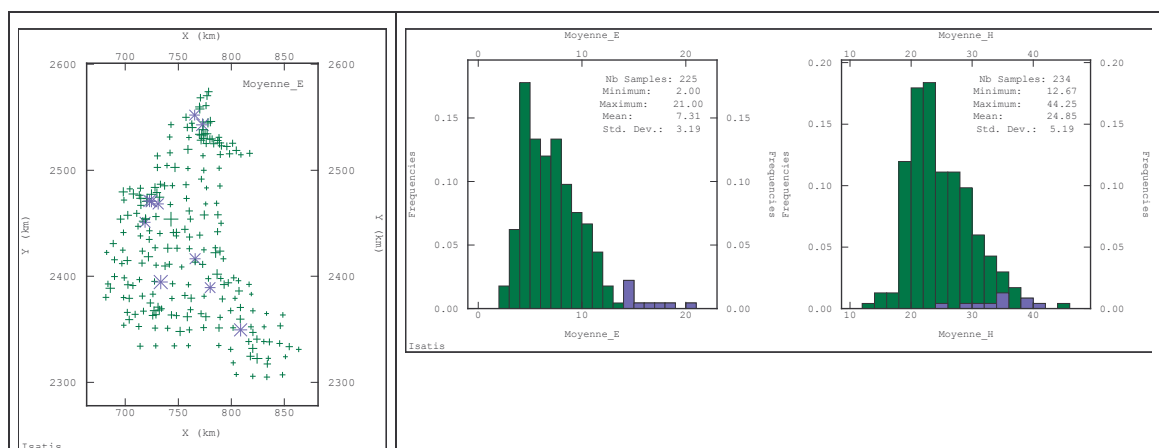


Figure 52 – Histogramme des concentrations moyennes sur l'été (figure de gauche) et sur l'hiver (figure de droite)

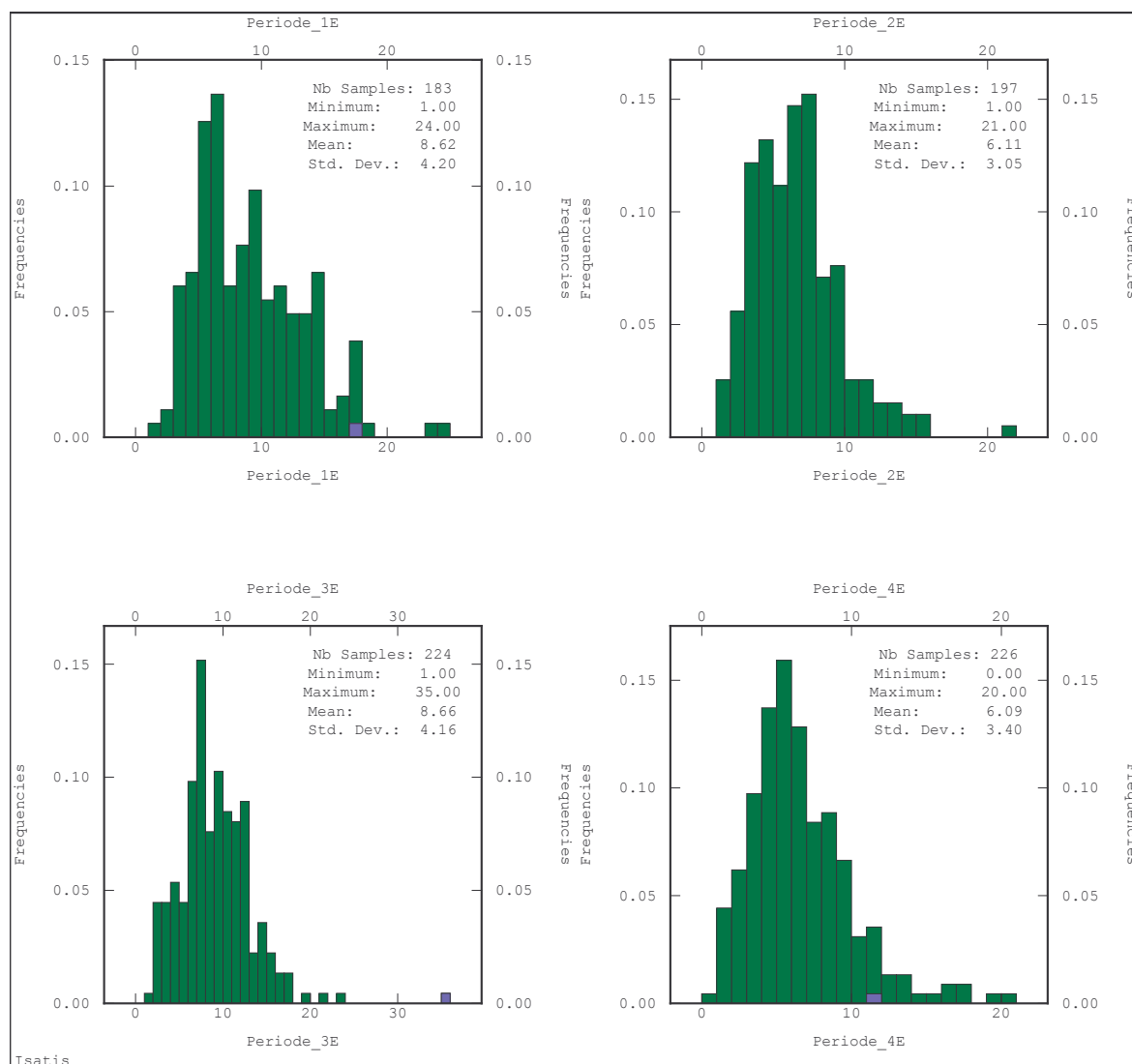


Figure 53 – Histogramme des concentrations mensuelles de l'été

Le point 88, qui compte une donnée manquante (2nde période) et qui, à la 3^e période, apparaît très éloigné de la distribution, est retiré de l'analyse. Ce point augmente fortement la variabilité spatiale des concentrations, comme a pu le montrer un premier examen de la nuée variographique.

Le point 16, dont ATMO CA considère la valeur mesurée comme trop élevée par rapport à l'environnement du site, est également retiré.

Le nouveau jeu de travail ne comprend désormais plus ces deux points.

CAS MONOVARIABLE : MODELISATION DU VARIOGRAMME – VALIDATION CROISEE

Plusieurs modèles monovariés sont ajustés. Dans tous les cas, l'estimation par validation croisée est peu précise, avec un effet de lissage marqué. Ce résultat peut s'expliquer par les caractéristiques spatiales du phénomène, moins structuré en été qu'en hiver. Les portées des structures qui composent le variogramme sont plus courtes (9km et 40km aux incertitudes d'ajustement près). En particulier, la portée de la première structure est inférieure à la taille moyenne de la maille d'échantillonnage (15 km). Par suite, l'estimation en un point est moins influencée par son voisinage et tend à s'approcher de la moyenne expérimentale des données.

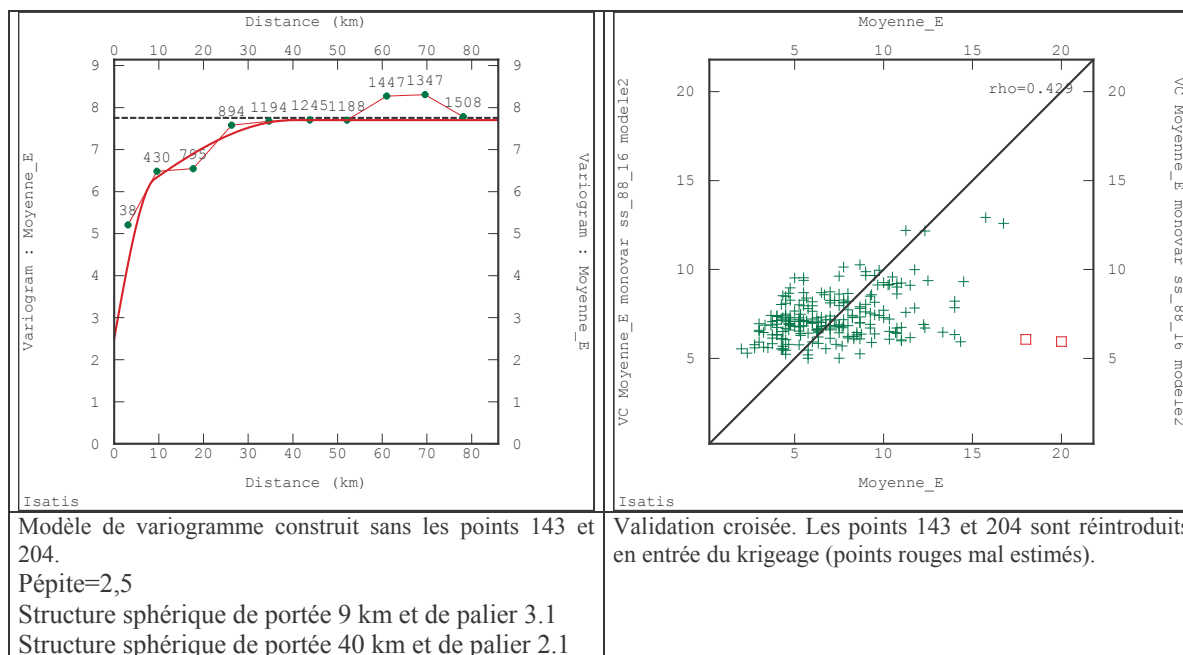


Figure 54 – Estimation par krigeage ordinaire de la concentration estivale [$\mu\text{g}/\text{m}^3$] : modélisation du variogramme et validation croisée

CAS BIVARIABLE : UTILISATION DES DONNEES HIVERNALES

La corrélation observée entre les concentrations d'été et d'hiver suggère de réaliser une modélisation bivariable. Le but est de pouvoir estimer la moyenne estivale de NO_2 par cokrigeage ordinaire avec la moyenne hivernale. Deux validations croisées sont mises en œuvre (Figure 55) :

- dans la première, la concentration d'été en un point expérimental x_k est estimée uniquement à l'aide données estivales et hivernales mesurées aux autres points (points x_i , $i \neq k$) ; la mesure hivernale en x_k est ignorée. C'est dans ce type de configuration que se trouvent la plupart des points de la grille d'estimation.
- dans la seconde, la concentration d'été en un point expérimental x_k est estimée à l'aide des données estivales et hivernales mesurées aux autres points (points x_i , $i \neq k$) et de la mesure hivernale en x_k .

Dans le premier cas, le cokrigeage améliore faiblement les résultats.

Dans le second cas, l'estimation gagne nettement en précision. Ce résultat signifie qu'en un point où la concentration estivale est inconnue, il est possible de fournir une estimation plus fiable de cette dernière si la donnée hivernale est disponible. Ce résultat peut avoir son intérêt en cas de données manquantes.

En neuf sites, la concentration moyenne de l'été n'a pu être mesurée ou a dû être invalidée par ATMO Champagne-Ardenne ; en revanche la mesure hivernale existe. Cette concentration estivale est estimée soit par krigeage ordinaire, soit par cokrigeage ordinaire à l'aide de tous les autres points. La Figure 56 montre l'influence du cokrigeage. (Rem. : aucune comparaison à la mesure n'est possible pour ces sites.)

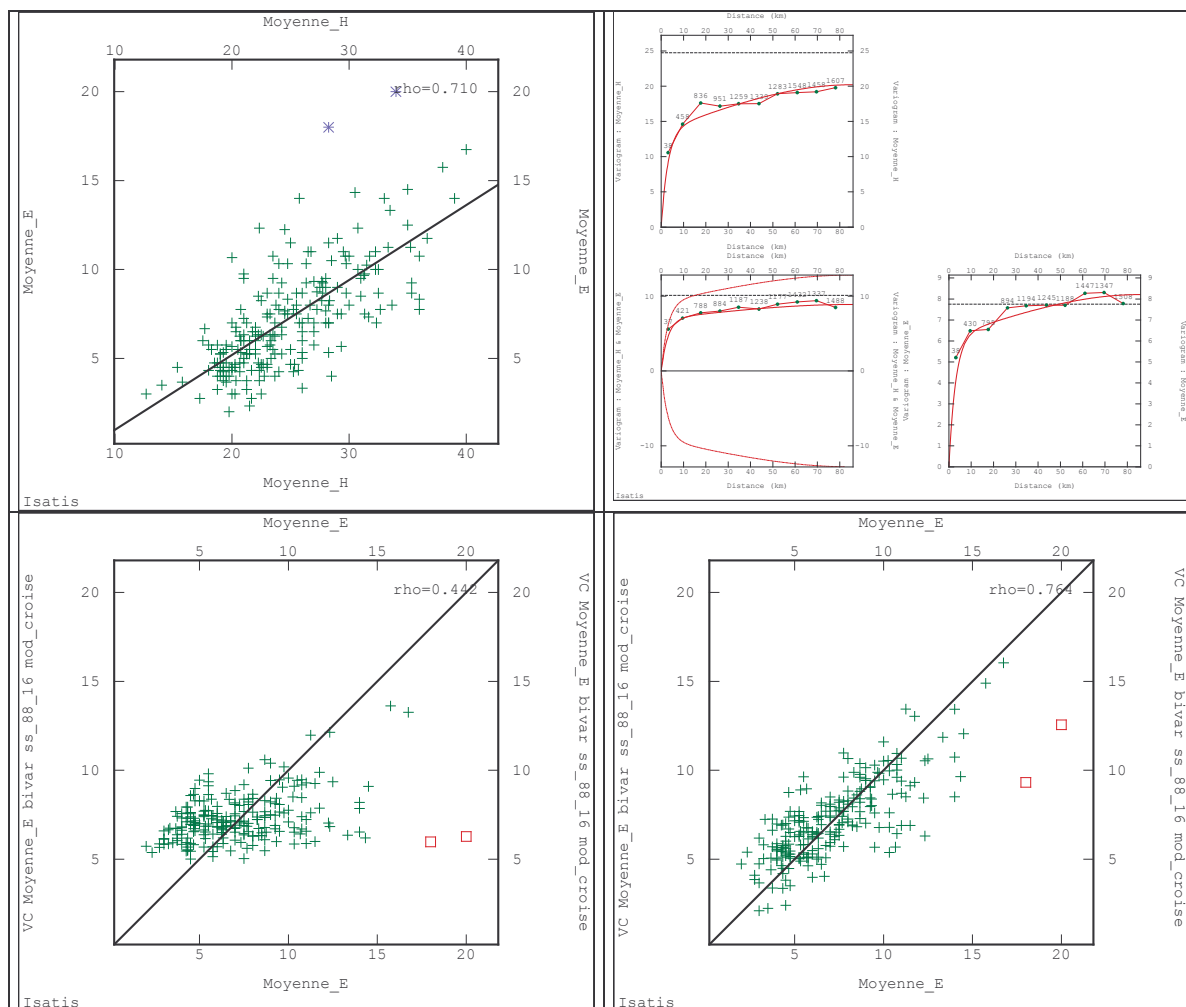


Figure 55 – Estimation par cokrigage ordinaire de la concentration estivale [µg/m³].
Ligne du haut : corrélation été-hiver (en bleu : points 143 et 204) et modélisation du variogramme.

Ligne du bas : validation croisée en utilisant ou non la concentration hivernale au point cible. Les points 143 et 204, mal réestimés du fait qu'ils se trouvent dans un environnement moins pollué que la valeur qu'ils mesurent, sont masqués dans le calcul du coefficient de corrélation.

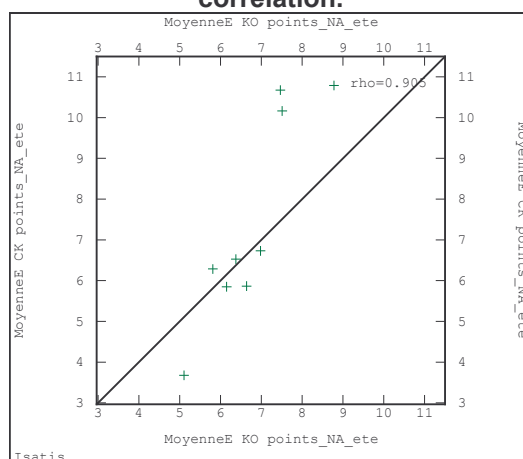


Figure 56 – Comparaison entre les valeurs estimées par krigeage ordinaire (en abscisse) et par cokrigage ordinaire (en ordonnée) aux points d'échantillonnage où la donnée estivale est manquante et la donnée hivernale disponible.

CAS MULTIVARIABLE : MODELISATION DU VARIOGRAMME – VALIDATION CROISEE

Par rapport à l'hiver, la corrélation avec la densité d'émissions ne change quasiment pas pour les petits rayons (1 ou 2km) tandis qu'elle diminue à plus grande échelle (5km). La corrélation avec la densité de population, notamment dans un rayon de 5km, décroît plus sensiblement ; le nuage de corrélation est un peu plus dispersé. Ces différences entre saisons pourraient avoir leur cause 1) dans le fait qu'un inventaire d'émissions moyen annuel est employé, or l'importance et le poids relatif des différentes sources de NOx ne sont pas les mêmes en été et en hiver. En outre, la contribution du chauffage, reflétée en partie par la densité de population, diminue ; 2) dans les conditions de dispersion, dépendantes de la météorologie donc liées à la saison.

Tableau 15 – Coefficient de corrélation entre la variable de concentration et les variables auxiliaires. Le coefficient de corrélation de l'hiver est rappelé entre parenthèses.

	Corrélation avec :		Corrélation avec :
$\ln(1+dE1km)$	0.67 (0.65)	$\ln(1+dpop1km)$	0.71 (0.75)
$\ln(1+dE2km)$	0.68 (0.66)	$\ln(1+dpop2km)$	0.69 (0.74)
$\ln(1+dE5km)$	0.62 (0.68)	$\ln(1+dpop5km)$	0.59 (0.73)

Deux régressions linéaires sont établies :

- l'une en fonction de la densité d'émissions à 2km et de la densité de population à 1km ;
- l'autre en fonction de la densité d'émissions à 2km.

Dans les deux cas, l'hypothèse, sous-jacente au krigeage des résidus, d'absence de corrélation spatiale entre la ou les variables denses et le résidu de la régression n'est pas vérifiée. Une analyse plus approfondie serait donc nécessaire. Malgré cela, afin d'aboutir à des résultats préliminaires, des modèles sont ajustés sur les variogrammes expérimentaux des résidus et testés par validation croisée.

Les nuages de corrélation entre estimations et observations paraissent plus dispersés qu'en hiver ; en particulier, le gain de précision par rapport à la régression seule est peu perceptible. Relativement à l'estimation monovariée, il est plus marqué. Le krigeage des résidus réduit l'effet de lissage.

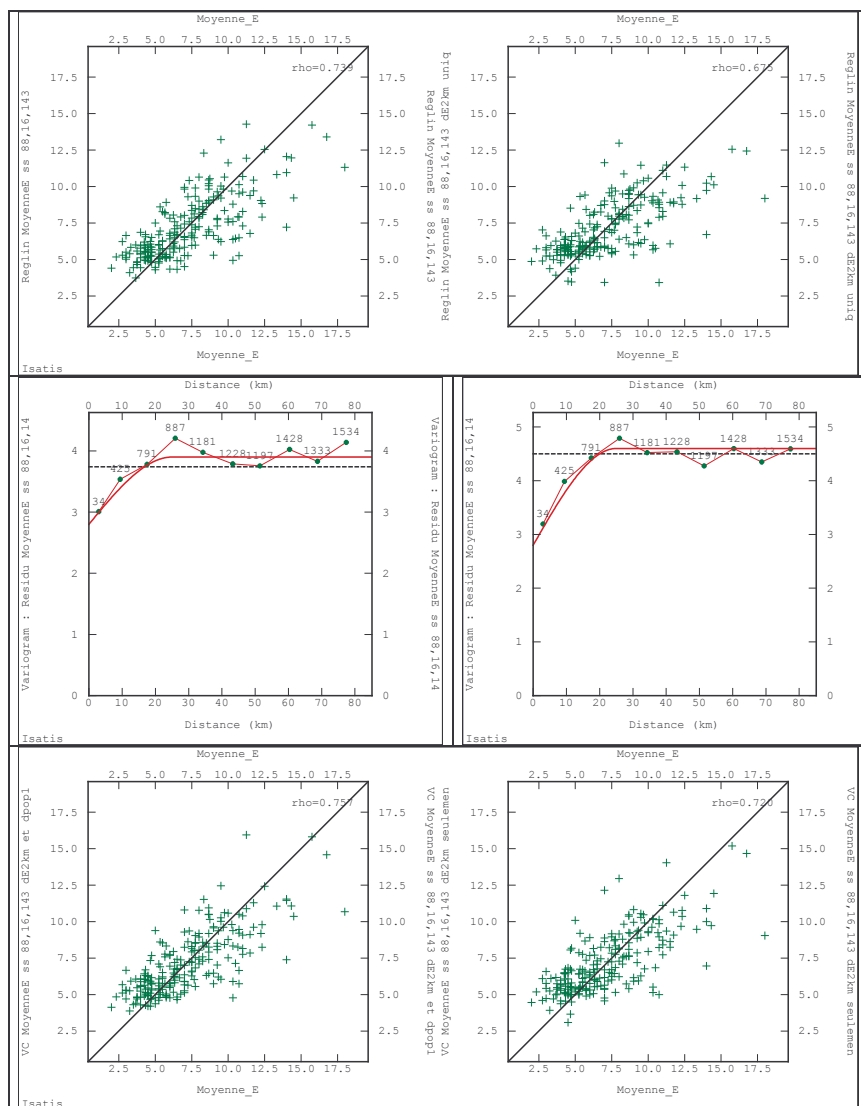


Figure 57 – Ligne du haut : nuages de corrélation entre les régressions linéaires et les mesures estivales. Ligne du milieu : variogrammes des résidus. Ligne du bas : nuages de corrélation entre les valeurs estimées par validation croisée et les valeurs observées.

**Figures de gauche : avec la densité d'émissions à 2km et la densité de population ;
figures de droite : avec seulement la densité d'émissions à 2km**

ESTIMATION SUR LA GRILLE

Nous nous limitons à la présentation des cartes de concentration. Celles de l'écart-type de krigeage sont cohérentes avec la théorie : le krigeage ordinaire, suivi du cokrigeage ordinaire avec la moyenne hivernale, conduisent en moyenne aux écarts-types d'estimation les plus élevés. Les écarts-types les plus faibles sont fournis par le krigeage des résidus, avec la densité d'émissions et la densité de population comme variables explicatives.

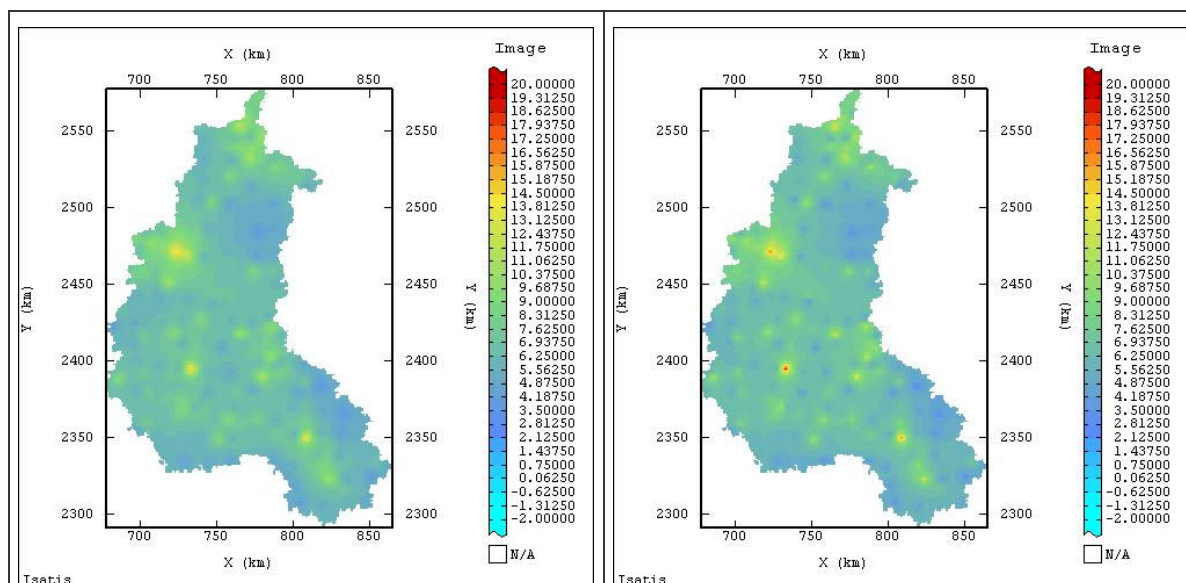


Figure 58 – Estimation par krigeage ordinaire ponctuel (à gauche) et par cokrigeage ordinaire avec la moyenne hivernale (à droite) [$\mu\text{g}/\text{m}^3$], sur une grille de pas égal à 1km.

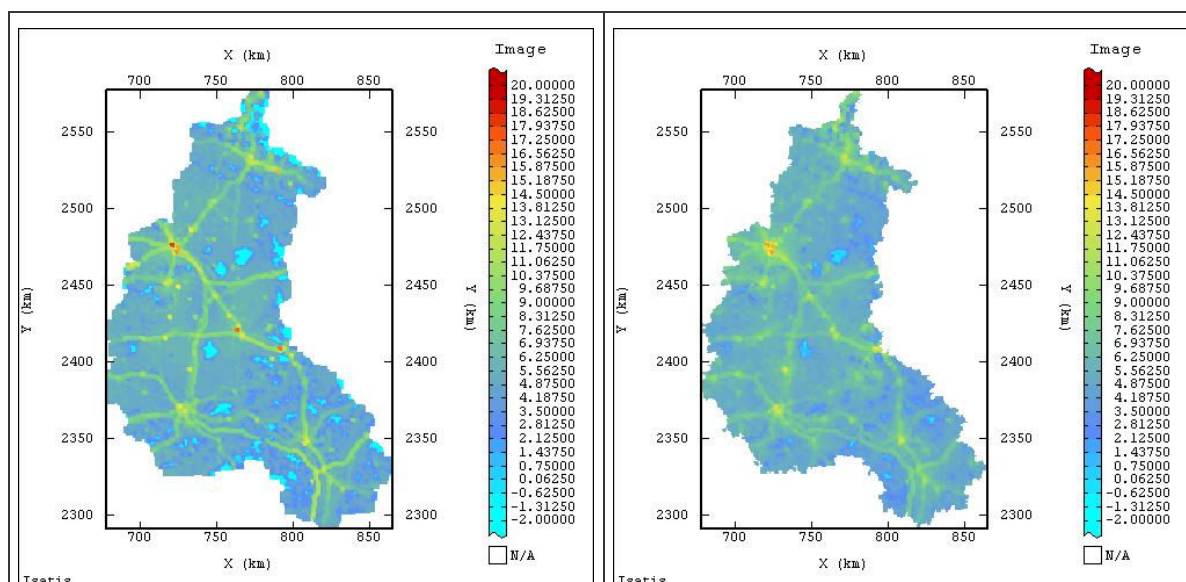


Figure 59 – Estimation par krigeage des résidus avec la densité d'émissions à 2km (à gauche), la densité d'émissions à 2km et la densité de population à 1km (à droite) [$\mu\text{g}/\text{m}^3$], sur une grille de pas égal à 1km.

Le cokrigeage avec l'hiver rehausse légèrement les concentrations urbaines mais dans l'ensemble, il diffère peu du krigeage ordinaire (Figure 58).

L'estimation avec la seule densité d'émissions n'est pas satisfaisante car elle fait apparaître localement des valeurs négatives, y compris à l'intérieur du domaine. Ce problème est beaucoup plus circonscrit lorsque la densité de population est introduite dans le krigeage. Comme pour l'hiver, les cartographies montrent l'influence des variables auxiliaires (Figure 59).

Remarque : on aurait pu examiner la corrélation entre les résidus de l'été et de l'hiver, en vue d'effectuer un cokrigeage des résidus. Cette analyse approfondie, plus coûteuse en temps, n'a pas été conduite.

Etude d'une sélection

La sélection étudiée est S30km2+, décrite au paragraphe 0.

ESTIMATION PAR KRIGEAGE DES RESIDUS

Compte tenu des corrélations avec la concentration, deux choix de variables explicatives sont considérés :

- la densité d'émissions à 1km et la densité de population à 1km ;
- la densité d'émissions à 5km et la densité de population à 1km (même choix que pour l'hiver).

Bien qu'elle ait un poids relativement faible, la densité d'émissions est conservée dans les deux régressions.

Les calculs obéissent toujours à la même démarche :

- modélisation du variogramme des résidus (Figure 60);
- validation croisée (Figure 61) ;
- vérification sur le jeu de données complémentaire (Figure 62) ;
- estimation sur la grille (Figure 63).

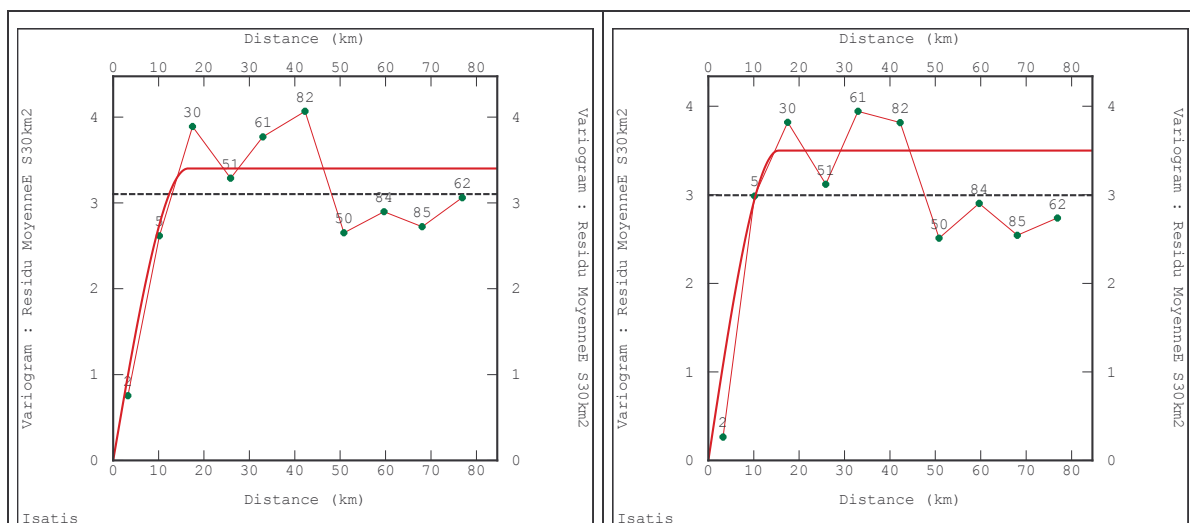


Figure 60 –Variogrammes des résidus (fig. de gauche : les variables auxiliaires sont la densité d'émissions à 1km et densité de population à 1km ; figure de droite : les variables auxiliaires sont la densité d'émissions à 5km et densité de population à 1km)

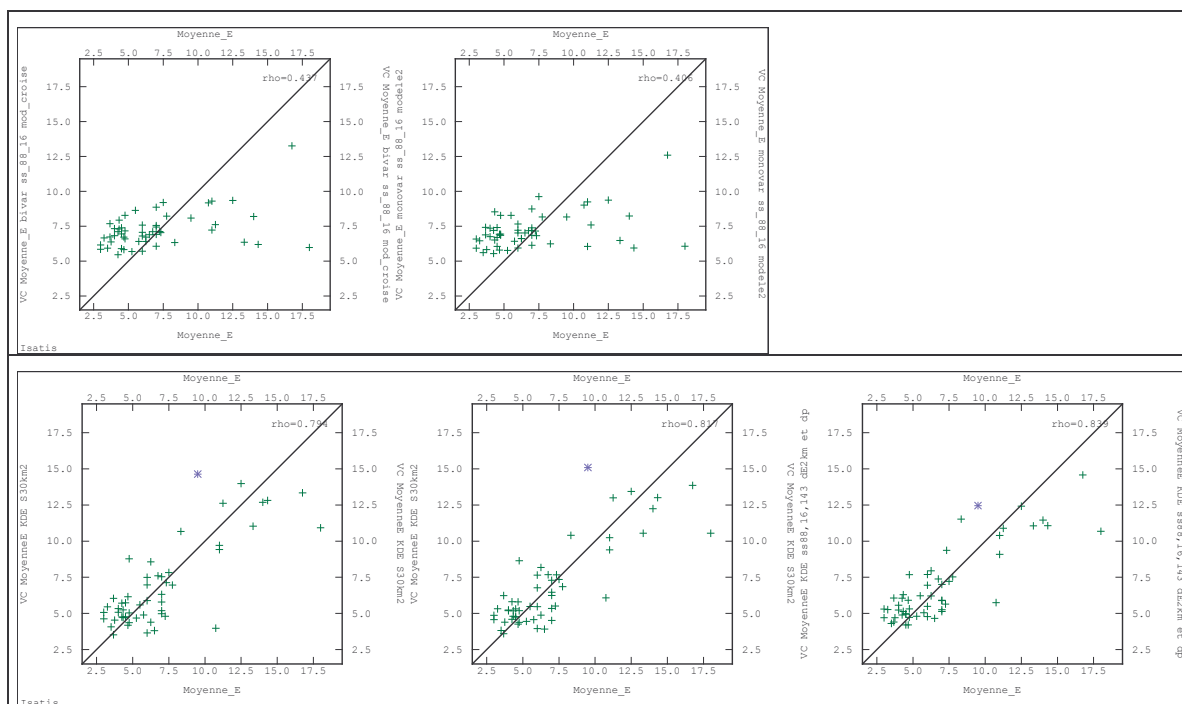


Figure 61 – Ligne du haut : restriction à la sélection S30km+ des résultats de validation croisée obtenus, à partir du jeu de données complet, par cokrigage ordinaire avec l'hiver (sans la donnée d'hiver au point cible) et par krigeage ordinaire.

Ligne du bas, de gauche à droite : restriction à la sélection S30km+ des résultats de validation croisée obtenus, à partir du jeu de données complet, par krigeage des résidus (émissions + population); validation croisée sur la sélection réalisée par krigeage des résidus (avec soit la densité d'émissions et de population à 1km, soit la densité d'émissions à 5km et la densité de population à 1km)

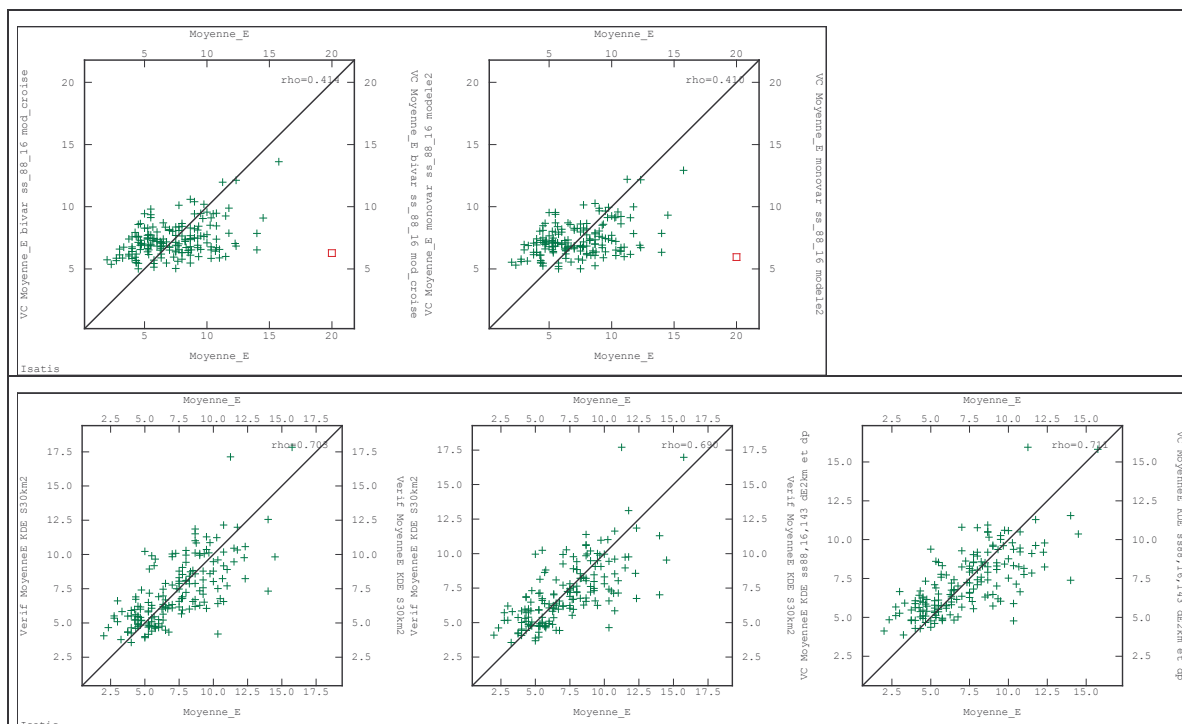


Figure 62 – Ligne du haut : restriction au complémentaire de la sélection S30km+ des résultats de validation croisée obtenus, à partir du jeu de données complet, par cokrigage ordinaire avec l'hiver (sans la donnée d'hiver au point cible) et par krigeage ordinaire.

Ligne du bas, de gauche à droite : restriction au complémentaire de la sélection S30km+ des résultats de validation croisée obtenus, à partir du jeu de données complet, par

krigeage des résidus (émissions + population); vérification sur le complémentaire réalisée par krigeage des résidus (avec soit la densité d'émissions et de population à 1km, soit la densité d'émissions à 5km et la densité de population à 1km)

Dans la Figure 61 et la Figure 62, on a également représenté les résultats de validation croisée associés au jeu de données complet (cf. 0, 0) en les restreignant d'une part, à la sélection S30km2+ et d'autre part, au complémentaire de celle-ci. La dégradation de l'échantillonnage (les deux graphiques en bas à droite, à comparer au graphique en bas à gauche) ne disperse pas sensiblement les nuages de corrélation.

La densité d'émissions ayant peu de poids dans la régression, les cartographies obtenues sont principalement influencées par la densité de population (Figure 63).

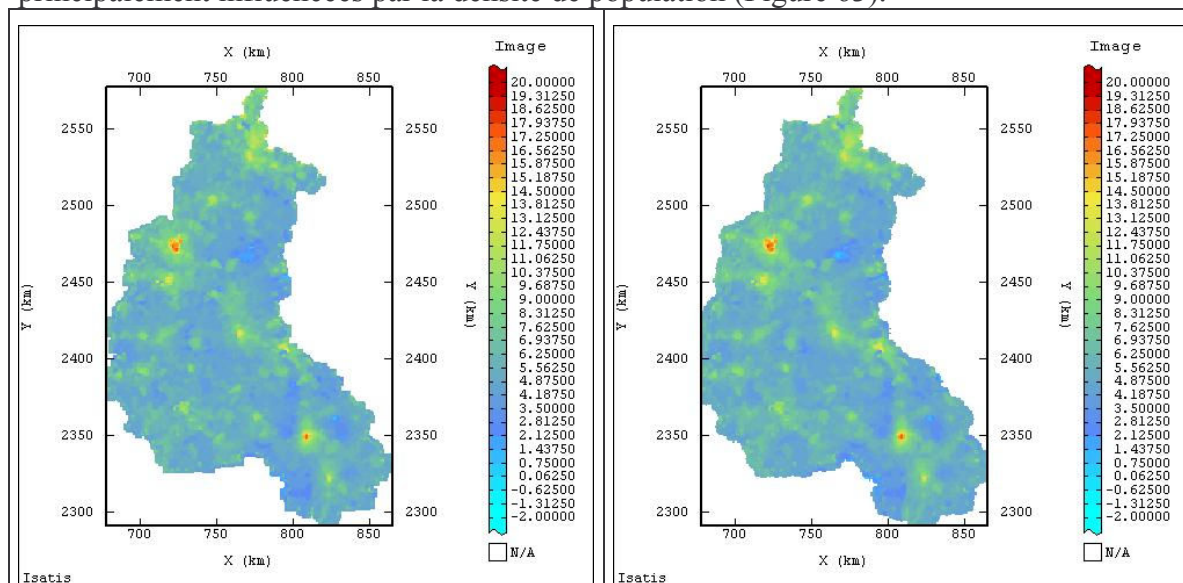


Figure 63 – Estimation par krigeage des résidus avec la densité d'émissions à 1km et la densité de population à 1km (à gauche), avec la densité d'émissions à 5km et la densité de population à 1km (à droite) [$\mu\text{g}/\text{m}^3$], sur une grille de pas égal à 1km.

REMARQUE SUBSIDIAIRE SUR LE COKRIGEAGE AVEC LA MOYENNE HIVERNALE

Le test suivant a pour but de confirmer la remarque faite en 0, à savoir qu'en un point où la donnée estivale est indisponible, mais non pas la donnée hivernale, le cokrigeage de l'été et de l'hiver rend plus précise l'estimation de la concentration de l'été.

Le jeu de données utilisé pour l'été est la **sélection S30km2+**.

Le jeu de données utilisé pour l'hiver est le **jeu de données complet** (234 points).

Un modèle bivariable est ajusté sur les variogrammes simples et croisé de l'hiver et de l'été.

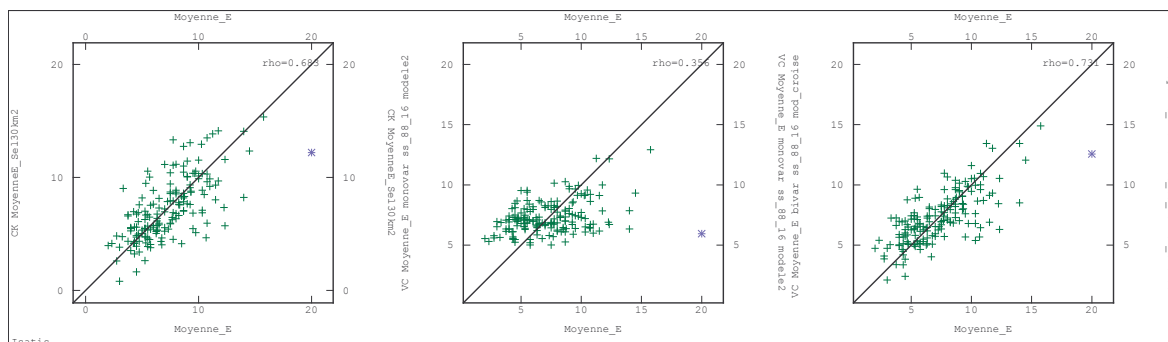


Figure 64 – Nuages de corrélation entre les mesures du complémentaire de la sélection S30km2+ et les valeurs estimées par 1) cokrigage entre les données estivales et hivernales de S30km2+; 2) krigeage ordinaire sur les données estivales de S30km2+ ; 3)) cokrigage entre les données estivales de S30km2+ et la totalité des données hivernales.

Commentaires sur l'estimation des concentrations estivales

En été, la pollution régionale de NO₂ est bien plus faible qu'en hiver (avec des concentrations inférieures à 7µg/m³ sur une large partie du domaine) ; elle se révèle aussi moins structurée dans l'espace. Ces caractéristiques, déjà constatées dans d'autres études et pour d'autres régions, rendent la modélisation et l'estimation géostatistiques plus délicates.

Sur environ un quart du domaine, l'erreur relative est supérieure à 30% (Tableau 16, Tableau 17). Si les plus grandes valeurs de cette erreur correspondent aux niveaux de pollution les plus bas, en revanche, sur toute la gamme de concentration, elle peut ponctuellement excéder 30% (Figure 65). Pour le krigeage des résidus, la réduction de l'échantillonnage ne dégrade pas les résultats de validation, ce qui confirme la remarque sur les nuages de corrélation (0). Comme c'était le cas en hiver, elle modifie néanmoins les relations avec les variables auxiliaires et atténue le poids des émissions.

Remarque : dans les statistiques et les graphiques suivants, le point 143 est ignoré. En effet, la concentration en ce point est nettement plus élevée que dans le voisinage. Quel que soit le modèle, elle est fortement sous-estimée.

Tableau 16 – Statistique des erreurs relatives (en valeur absolue) [%] de validation croisée – jeu de données complet

	Krigeage ordinaire	Cokrigeage ordinaire sans la mesure hivernale au point cible	Cokrigeage ordinaire avec la mesure hivernale au point cible	Krigeage des résidus avec la densité d'émissions à 2km	Krigeage des résidus avec la densité d'émissions et de population à 5km et 1km respectivement
Moyenne	34.60	34.63	23.04	24.14	22.31
Ecart-type	30.01	30.05	22.77	23.30	21.20
Minimum	1.17	0.42	0.65	0.00	0.00
Maximum	177.00	186.50	136.50	123.27	122.00
P75	48.33	47.32	32.27	32.41	29.23

Tableau 17 – Statistique des erreurs relatives (en valeur absolue)) [%] de validation croisée et de vérification – Sélection S30km2+

	Validation croisée sur la sélection S30km2+		Vérification sur le complémentaire de la sélection S30km2+	
	Krigeage des résidus avec la densité d'émissions et de population à 1km	Krigeage des résidus avec la densité d'émissions et de population à 5km et 1km respectivement	Krigeage des résidus avec la densité d'émissions et de population à 1km	Krigeage des résidus avec la densité d'émissions et de population à 5km et 1km respectivement
Moyenne	24.06	22.58	22.33	21.54
Ecart-type	19.76	18.89	21.31	21.12
Minimum	1.64	0.36	0.27	0.00
Maximum	84.84	82.11	120.67	108.00
P75	33.25	30.88	27.89	27.45

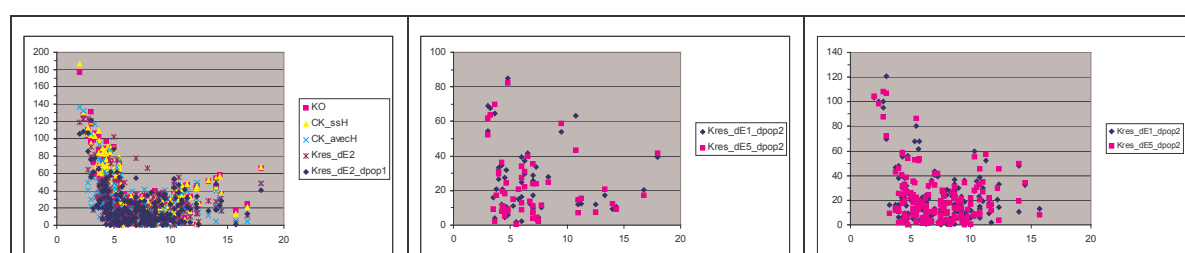


Figure 65 – Erreur relative d'estimation (en valeur absolue) [%] en fonction de la concentration mesurée [$\mu\text{g}/\text{m}^3$]. De gauche à droite : validation croisée sur le jeu de données complet, validation croisée sur la sélection S30km2+, vérification sur le complémentaire de la sélection S30km2+.

REFERENCES

Cori A., 2005. Utilisation de données environnementales pour la cartographie du NO₂ sur la région rouennaise. Première partie. AIR Normand, Ecole des Mines de Paris.

Annexe 6 – Région : Centre

Etude à l'échelle régionale : cas de la région Centre (NO₂)

PARTIE I. Echantillonnage régulier sans variable auxiliaire

DESCRIPTION DU CAS

Présentation du jeu de données initial

Le jeu de données étudié a été élaboré au cours de l'étude LCSQA 2003 « Cartographie de la pollution atmosphérique en zone rurale influencée au moyen d'un modèle déterministe de dispersion des polluants »⁴. Il est constitué des données de concentration en dioxyde d'azote obtenues par échantillonnage passif lors de deux campagnes de mesure, d'une durée de 14 jours chacune, ainsi que des résultats de modélisation déterministe correspondant à la simulation de ces campagnes sur une partie nord de la région Centre.

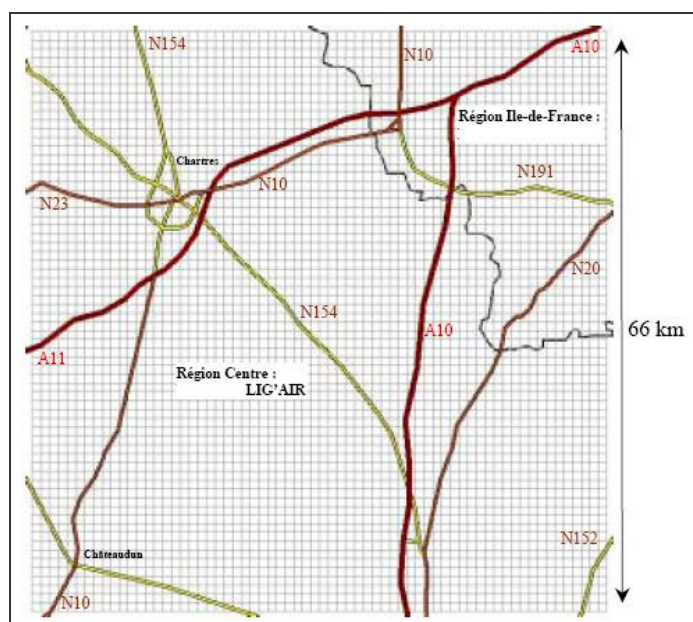


Figure 1 – Zone d'étude, nord de la région centre, avec maillage de modélisation.

Une campagne de mesure des teneurs en NO₂ par échantillonnage passif sur une période de quinze jours entre les 10 et 25 juillet 2002 a été effectuée par l'EMD en partenariat avec LIG'Air. 63 tubes ont été répartis sur un domaine composé de 36 mailles de 11 km sur 11 km. Les simulations déterministes ont été réalisées avec Transchim (CORIA) sur un domaine de 66x66 km² avec une maille de 1km.

⁴ LCSQA-EMD, 2003. *Cartographie de la pollution atmosphérique en zone rurale influencée au moyen d'un modèle déterministe de dispersion des polluants*. Convention 115/2003.

Principales caractéristiques du jeu de données étudié :

	Campagne du 10 au 25 juillet 2002	Campagne du 3 au 17 avril 2003
Moyenne (en $\mu\text{g}/\text{m}^3$)	12,48	18.25
Maximum (en $\mu\text{g}/\text{m}^3$)	40.56	29.77
Minimum (en $\mu\text{g}/\text{m}^3$)	5.91	10.76

Tableau 1 : statistiques des données de campagne

Les données exploitées dans cette partie sont les résultats des simulations calées préalablement sur les campagnes de mesure. Complétant les données de ces campagnes, ils représentent en effet une plus grande densité d'information spatiale. On suppose qu'ils fournissent une image fidèle de la réalité.

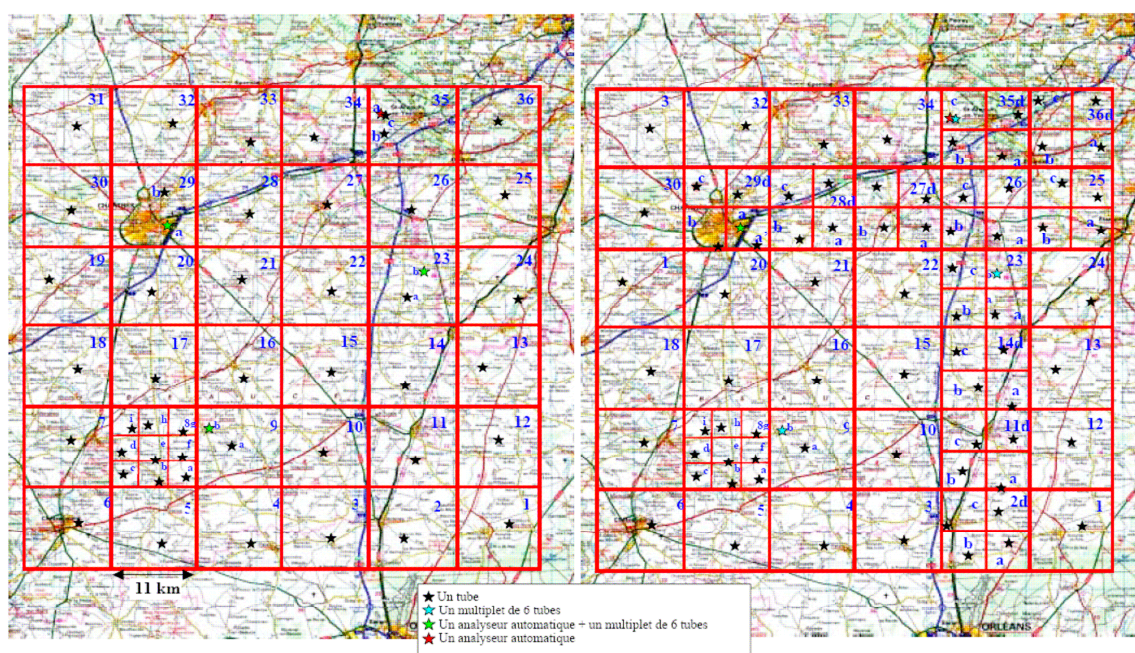


Figure 2 : échantillonnage spatial de la première (à gauche) et la deuxième (à droite) campagne de mesures

Méthodologie de sélection des données

A partir du maillage de référence nous avons élaboré 8 maillages dégradés. Les quatre premiers maillages ont été dégradés de manière régulière à partir du maillage de référence :

- maillage de référence 33x33 (1089 pts) au pas de 2km,
- maillage de 17x17 carré (289 pts) au pas de 4km,
- maillage de 11x11 carré (121 pts) au pas de 6km,
- maillage de 9x9 carré (81 pts) au pas de 8km,
- maillage de 7x7 carré (49 pts) au pas de 10km,

Les données de chaque maillage sont exploitées selon une démarche de géostatistique, comme s'il s'agissait d'un jeu de données expérimental. Les étapes classiques d'une cartographie géostatistique sont ainsi mises en œuvre : analyse

exploratoire des données et modélisation de la structure spatiale, validation du modèle, estimation par krigeage.

Pour chacun des maillages testés, les variogrammes expérimentaux ont été ainsi déterminés. Nous avons ensuite modélisé un variogramme puis effectué la validation croisée. L'utilisation d'un jeu de données de référence a permis de procéder à une validation absolue. Enfin pour chaque maillage nous avons effectué un krigeage aux points des données de validation absolue en utilisant le variogramme de référence.

ETUDE DU JEU DE DONNEES INITIAL ET ELABORATION DE LA CARTE DE REFERENCE

1) Campagne 1 (10 et 25 juillet 2002)

Etude du variogramme expérimental

Le variogramme retenu est omnidirectionnel avec 10 pas de 4000m chacun. Aucune donnée supplémentaire n'a été écartée.

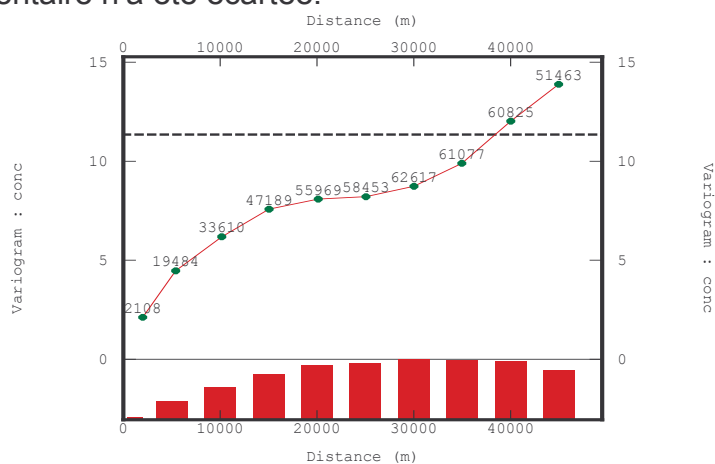


Figure 3 – Variogramme expérimental

Modélisation du variogramme et validation croisée

Le modèle de variogramme ajusté est composé d'une structure pépitique (palier=2.6) additionnée d'une structure sinus cardinal et d'une structure linéaire.

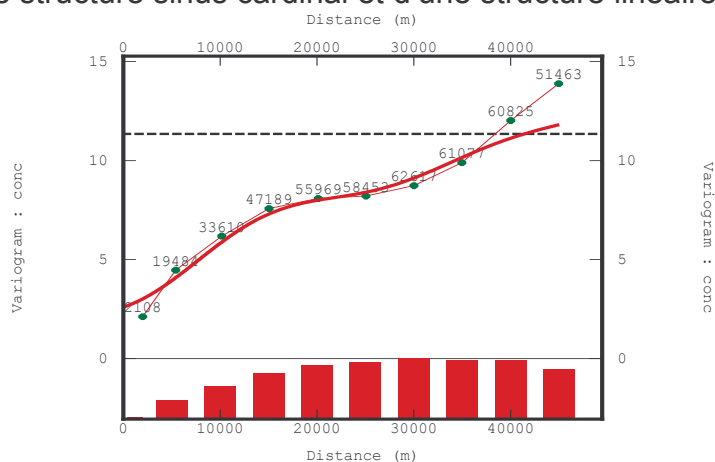


Figure n°4 : modèle : Pépité(2.6)+ sincard(75000 ;1.9)+lineaire(41500 ;6.7)

Afin d'apprécier la qualité du modèle, nous l'avons soumis au test de validation croisée. Ce test consiste à calculer, au moyen du modèle, une estimation de la concentration en polluant, pour chaque point de l'espace où celle-ci a été mesurée.

<i>MODELE REF</i>	moyenne	variance
Erreur	-0,0096	3,23
Erreur standardisée	-0,0064	1,02

Tableau 2 : résultats de validation croisée

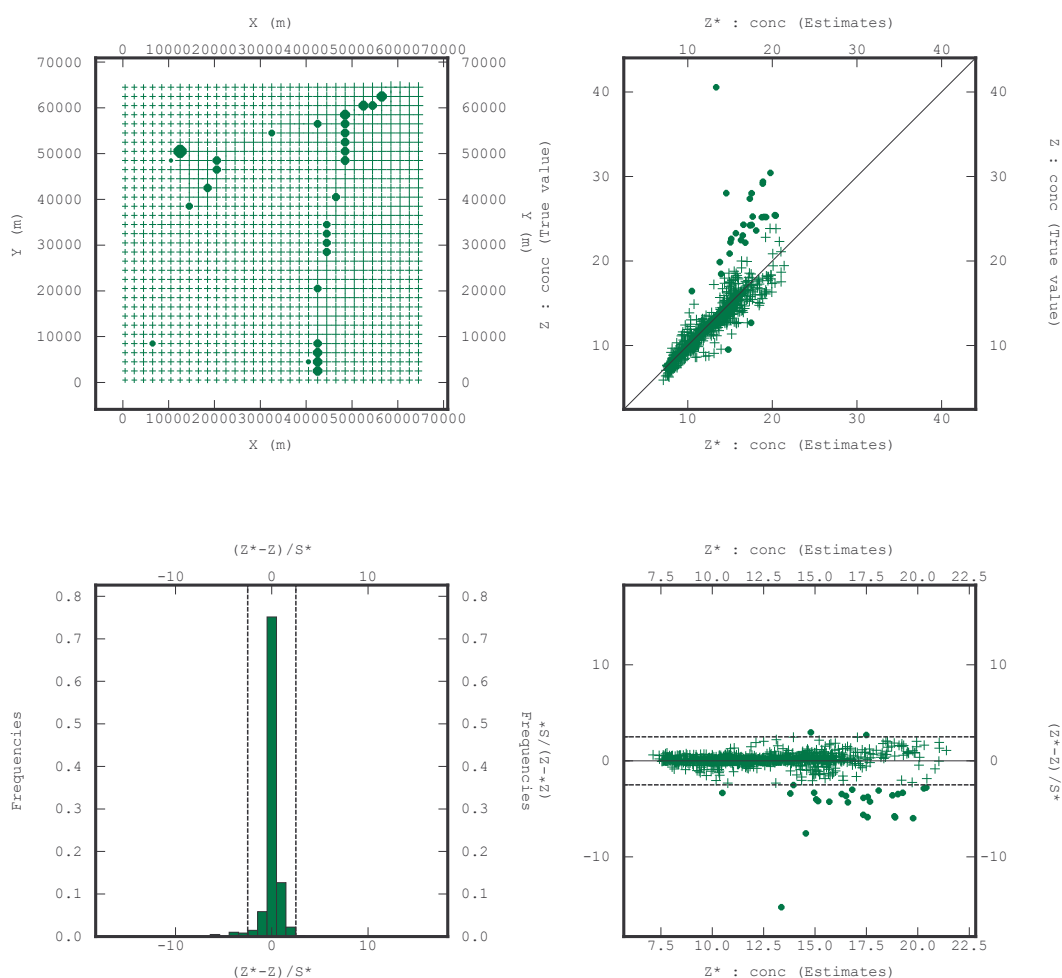


Figure n°5 : Résultats de validation croisée pour le modèle de référence

La plupart des données sont correctement réestimées. La sous-estimation de certaines valeurs s'explique par leur situation de proximité.

Etablissement de la carte de référence pour la campagne 1

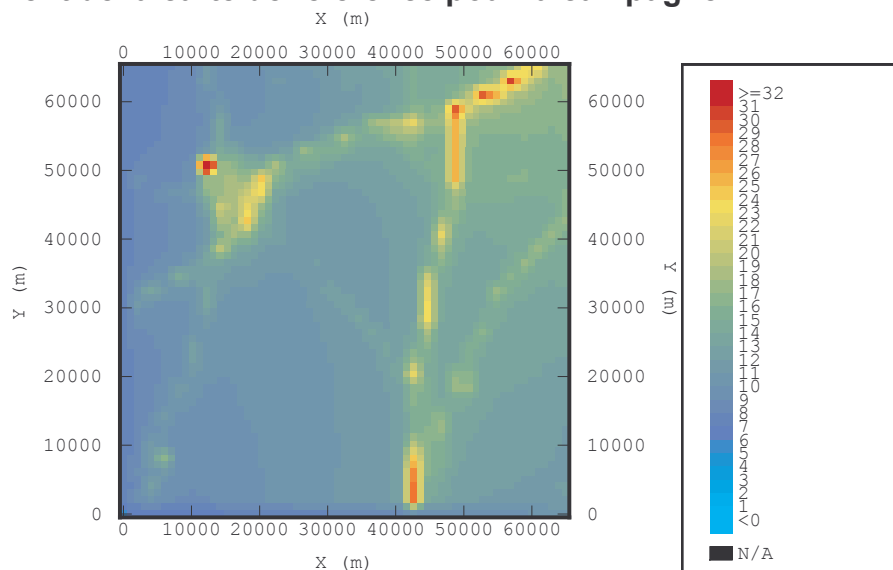


Figure n°6 : Estimation par krigeage ponctuel, à partir de toutes les données, sur une grille de 2km de côté.

A partir de ce cas de référence l'impact de la dégradation du maillage d'échantillonnage a été étudié.

COMPARAISON DES ESTIMATIONS

Un tableau de synthèse des résultats des validations, croisée à partir de la sélection, et absolue à partir du complémentaire, des modèles peut être établi à partir des différents jeux de données étudiés (Tableau 3).

La moyenne et la variance de l'erreur de validation croisée évoluent peu au cours du processus de dégradation. Il n'y a pas de variation monotone des paramètres statistiques avec le nombre de points.

maillages		ref	17x17 carré	9x9 carré	7x7 carré
nombre de points		1089	289	81	49
Min ($\mu\text{g}/\text{m}^3$)		5.91	5.91	5.91	5.91
Max ($\mu\text{g}/\text{m}^3$)		40.56	29.16	25.42	19.43
Moyenne ($\mu\text{g}/\text{m}^3$)		12.43	12.25	11.93	11.3
ecart-type ($\mu\text{g}/\text{m}^3$)		3.47	3.5	3.73	2.99
Variance ($\mu\text{g}/\text{m}^3$)		12.06	12.24	13.9	8.94
Validation croisée					
erreur					
	moy	-0.0096	0.00473	0.0166	0.03745
	var	3.23	3.298	3.019	0.987
erreur std.					
	moy	-0.0064	-0.00055	0.0042	0.019
	var	1.02	0.985	1.335	0.718
validation absolue					
error					
	moy	0.044	-0.066	-0.1807	-0.6581
	var	1.83	2.76	4.712	4.49
erreur std.					
	moy	0.025	-0.055	-0.145	-0.8641
	var	0.58	1.98	2.984	7.5306

Tableau 3 : synthèse de validation des différents schémas de dégradation

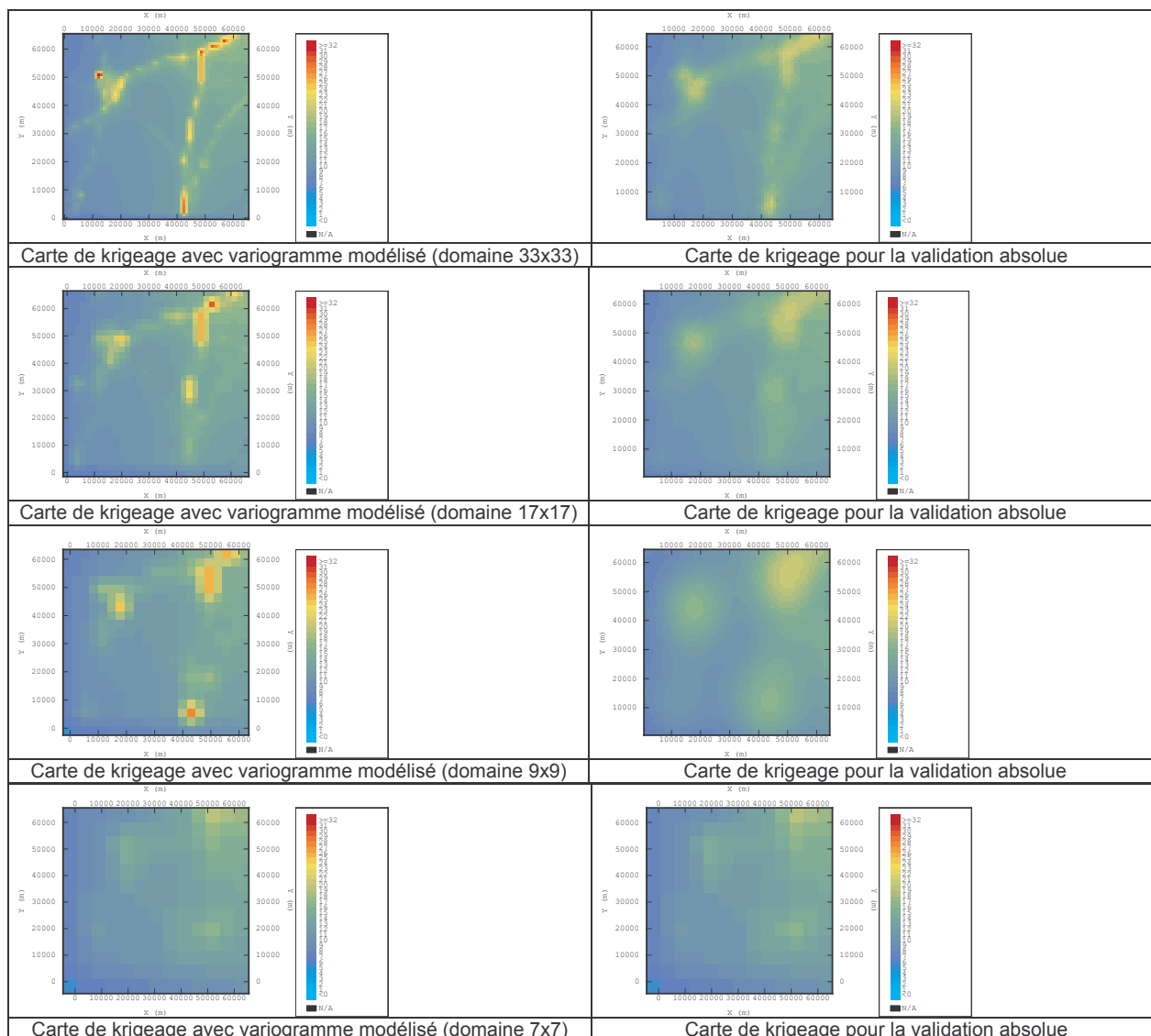


Figure 7 – Comparaison des cartes d'écarts-types de krigeage établies par krigeage ponctuel sur la grille de référence pour les différentes sélections

En valeur absolue, l'écart avec la carte de référence peut atteindre $21 \mu\text{g m}^{-3}$ pour le maillage 7X7. On observe un net effet de lissage : sous-estimation en zones éloignées de toute pollution par les oxydes d'azote, surestimation dans les zones influencées par la pollution de proximité (centres-villes, axes routiers).

2) Campagne 2 (3 au 17 avril 2003)

Etude du variogramme expérimental

Le variogramme retenu est omnidirectionnel avec 10 pas de 4000m chacun. Aucune donnée supplémentaire n'a été écartée.

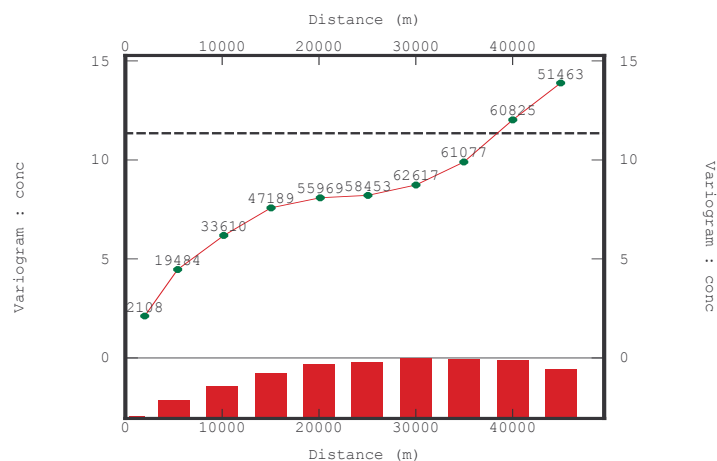


Figure 8 – Variogramme expérimental

Modélisation du variogramme et validation croisée

Le modèle de variogramme ajusté est composé d'une structure pépitique (palier=2.6) additionnée d'une structure sinus cardinal et d'une structure linéaire.

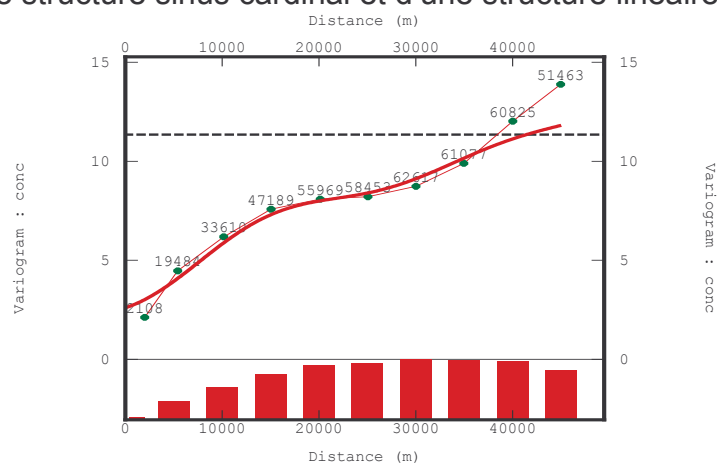


Figure n°9 : modèle : Modèle : Pépite(2.6)+ sincard(75000 ;1.9)+lineaire(41500 ;6.7)

Afin d'apprécier la qualité du modèle, nous l'avons soumis au test de validation croisée. Ce test consiste à calculer, au moyen du modèle, une estimation de la concentration en polluant, pour chaque point de l'espace où celle-ci a été mesurée.

MODELE REF	moyenne	variance
Erreur	-0,0096	3,23
Erreur standardisée	-0,0064	1,02

Tableau 4 : résultats de validation croisée

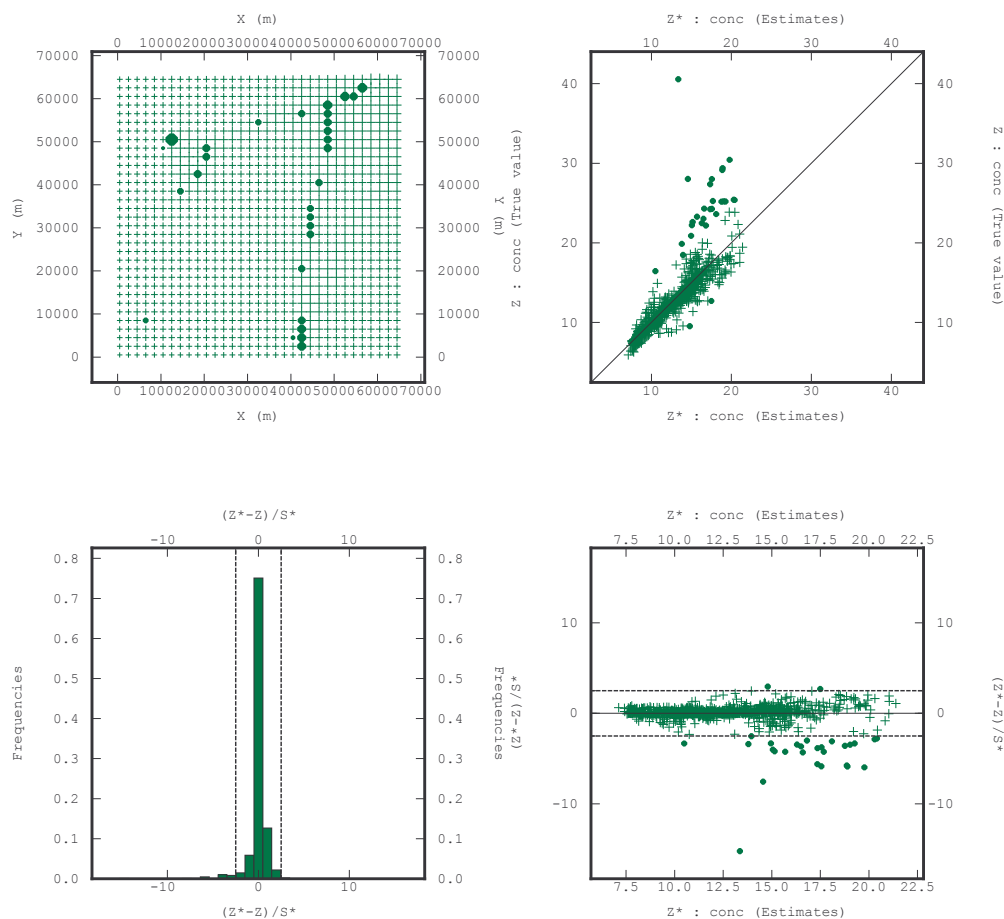


Figure n°10 : Résultats de validation croisée pour le modèle de référence

La plupart des données sont correctement réestimées. La sous-estimation de certaines valeurs s'explique par leur situation de proximité.

Etablissement de la carte de référence pour la campagne 2

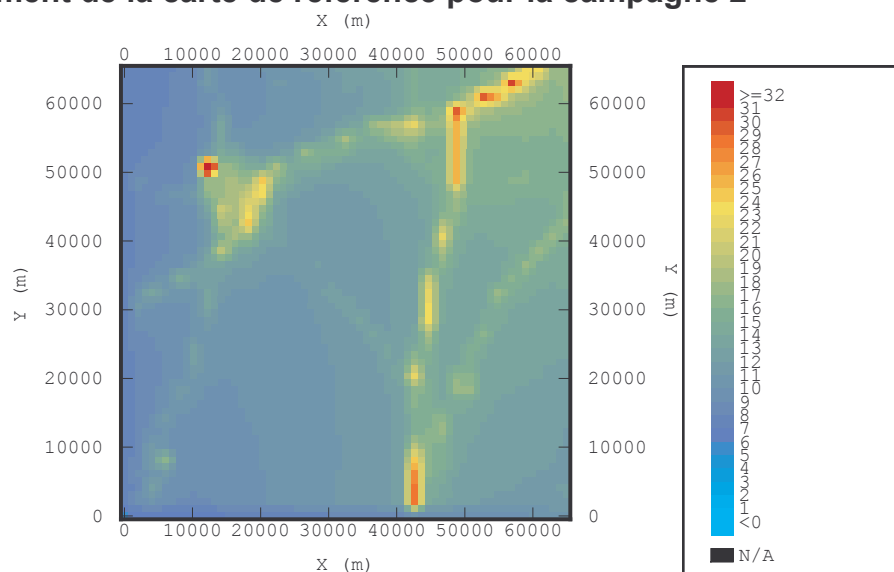


Figure n°11 : Estimation par krigeage ponctuel, à partir de toutes les données, sur une grille de 2km de côté.

A partir de ce cas de référence l'impact de la dégradation du maillage d'échantillonnage a été étudié.

COMPARAISON DES ESTIMATIONS

Un tableau de synthèse des résultats des validations, croisée à partir de la sélection, et absolue à partir du complémentaire, des modèles peut être établi à partir des différents jeux de données étudiés (Tableau 5).

La moyenne et la variance de l'erreur de validation croisée évoluent peu au cours du processus de dégradation. Il n'y pas de variation monotone des paramètres statistiques avec le nombre de points.

	ref	car_17	car_9	car_7
nombre de points	1089	289	81	49
Min ($\mu\text{g}/\text{m}^3$)	10.76	10.76	10.76	11.24
Max ($\mu\text{g}/\text{m}^3$)	32.31	27.5	20.99	29.4
Moyenne ($\mu\text{g}/\text{m}^3$)	18.56	18.14	16.99	17.92
ecart-type ($\mu\text{g}/\text{m}^3$)	3.65	3.66	3.83	3.15
Variance ($\mu\text{g}/\text{m}^3$)	11.38	9.3	5.0583	9.93
Validation croisée				
Erreur				
moy	0.00182	0.00614	0.00345	0.02481
var	2.7167	3.02322	3.18741	3.27754
Erreur std.				
moy	0.00063	0.0022	0.0022	0.00729
var	1.3536	1.64412	3.33237	1.13414
validation absolue				
Erreur				
moy	0.0422	-0.0326	-0.1059	-0.41
var	1.3658	2.0012	3.3061	4.3761
Erreur std.				
moy	0.0304	-0.0261	-0.1286	-0.258
var	0.706	1.2921	5.0503	1.6828

Tableau 5 : synthèse de validation des différents schémas de dégradation

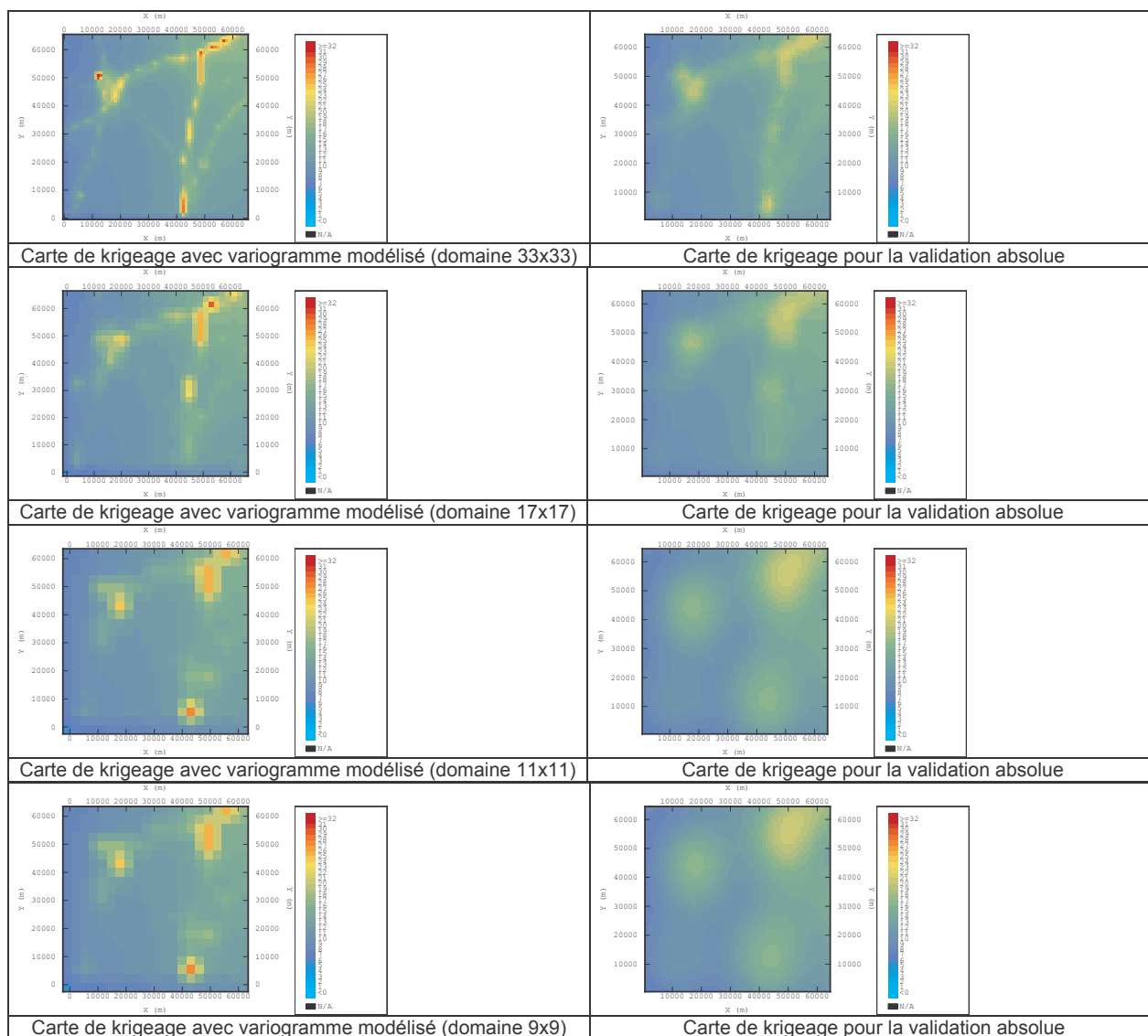


Figure 12 – Comparaison des cartes d'écarts-types de krigeage établies par krigeage ponctuel sur la grille de référence pour les différentes sélections

En valeur absolue, l'écart avec la carte de référence peut atteindre $13 \mu\text{g m}^{-3}$ pour le maillage 7X7. On observe un net effet de lissage : sous-estimation en zones éloignées de toute pollution par les oxydes d'azote, surestimation dans les zones influencées par la pollution de proximité (centres-villes, axes routiers). Cependant les effets de lissage et les écarts sont moindres sur cette deuxième campagne.

Les travaux menés en 2005 montre que la dégradation du maillage intervient pour une structure de maille comprise entre 17x17 et 9x9.

PARTIE II. Echantillonnage régulier avec variable auxiliaire

En vue d'améliorer la qualité de l'estimation, on va conserver les mailles régulières d'échantillonnage pour déterminer les jeux de données sélectionnés et leurs complémentaires, et on va intégrer une variable auxiliaire corrélée avec le polluant. La connaissance d'une variable auxiliaire en un point où la concentration en polluant ne l'est pas permet d'apporter une meilleure connaissance pour l'estimation.

1 DESCRIPTION DE LA VARIABLE AUXILIAIRE : EMISSION

Pour cette étude, nous disposons d'une variable auxiliaire relativement classique, les émissions, connue sur des mailles de 1 km de côté.

Les émissions jouent un rôle primordiale dans le phénomène pollution atmosphérique et leurs connaissances devraient permettre de modéliser la distribution spatiale des données .

CALCUL DU COFACTEUR

On cherche à obtenir par des transformations mathématiques appliquées à la variable auxiliaire (de manière empirique), une nouvelle variable présentant la meilleure corrélation linéaire possible avec la concentration en polluant. On choisira d'appliquer la méthode de cokrigage colocalisé (méthode stationnaire). La figure 13 présente le diagramme de corrélation entre les émissions et la concentration en NO_2

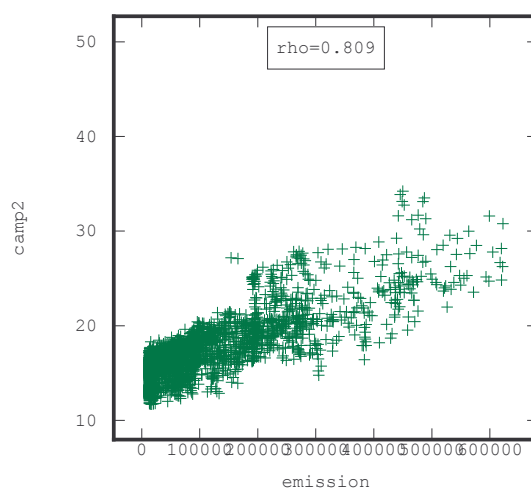


Figure 13 – Diagramme de corrélation émissions / concentration en NO_2

La corrélation est de très bonne qualité entre émissions/ concentrations, nous avons cependant essayé d'améliorer celle-ci en prenant une forme classique de variable auxiliaire, à savoir $\text{var_auxiliaire} = \log(\text{émissions}+1)$. On calcule donc la variable auxiliaire en tous points du domaine. On obtient le diagramme de corrélation suivant.

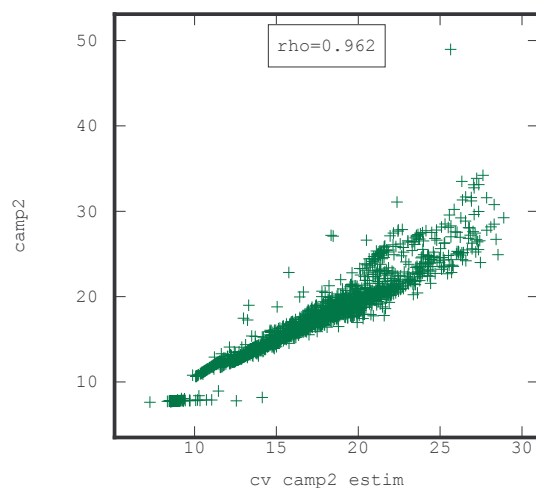


Figure 14 – Diagramme de corrélation entre la valeur auxiliaire $\log(\text{émission}+1)$ et la concentration en NO_2

ETUDE DU KRIGEAGE COLOCALISE

En utilisant le krigeage colocalisé, on fait l'hypothèse que les deux variables (concentration de NO_2 et variable auxiliaire ($\log(\text{émission}+1)$)) sont liées par un modèle variographique. On va établir le modèle bivariable à partir du jeu de données et des valeurs de variable auxiliaire définies aux mêmes coordonnées.

Etude des variogrammes expérimentaux

On calcule pour toutes les directions les deux variogrammes simples et le variogramme croisé avec 8 pas de 6000 m.

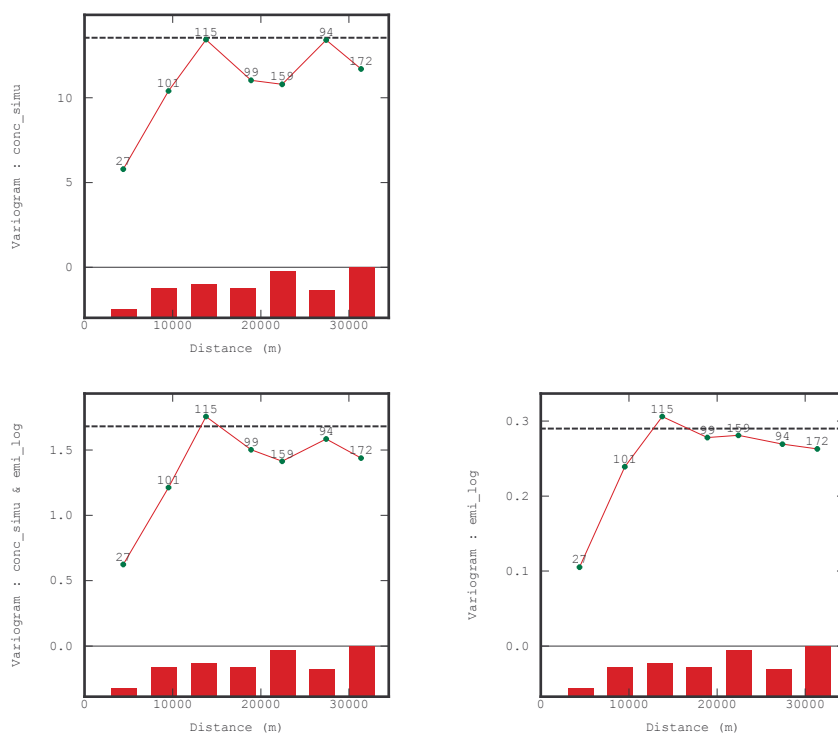
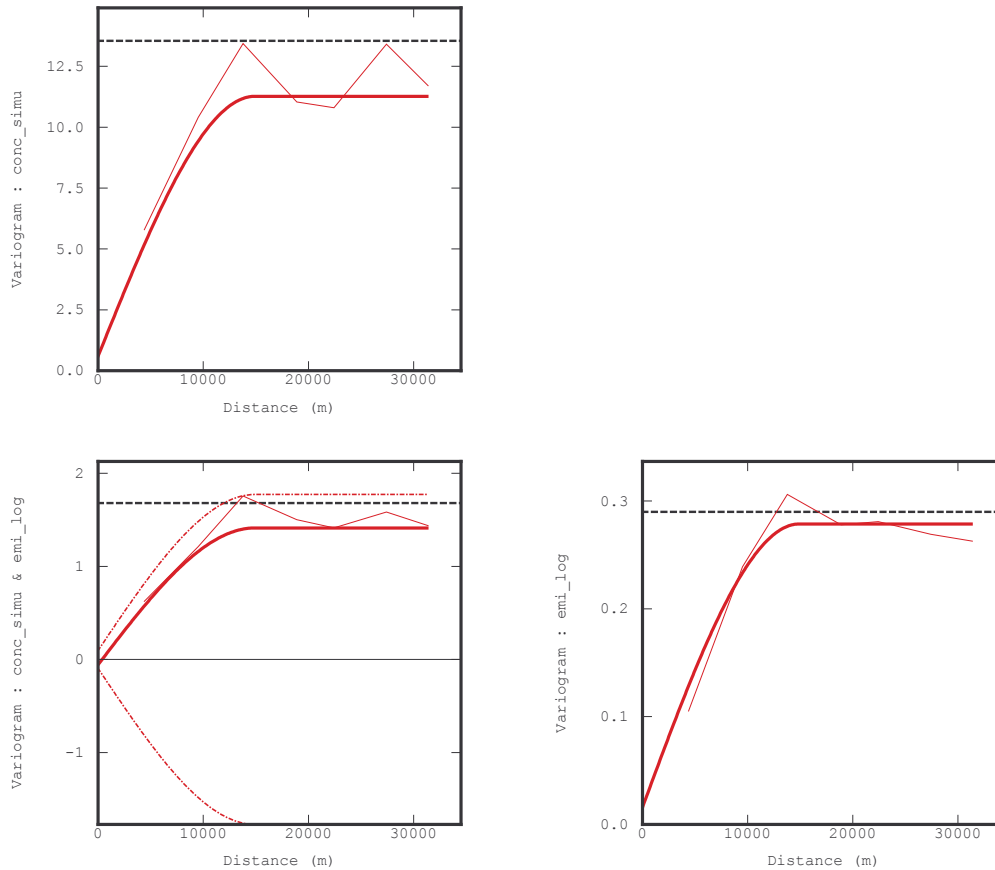


Figure 15 – Variogrammes expérimentaux simples et croisé

Modélisation des variogrammes et validation croisée

A partir de plusieurs structures simples, le logiciel définit les structures permettant de simuler les différents variogrammes.



**Figure 16 – Variogrammes simulés simples et croisé :
pépité (0,6) + sphérique (14 km)**

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 10 km.

(1089 points)	Moyenne	Variance
Erreur	-0,18	2,3
Erreur standardisée	0.10	0.67

4 données sur 1089 non robuste.

Tableau 6 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les concentrations

ELABORATION DE LA CARTE

On estime les concentrations en tous points de la carte de référence à l'aide du jeu de données, de la variable colocalisée {log (émission +1)} et du modèle bivariable, en considérant un voisinage de 10 km.

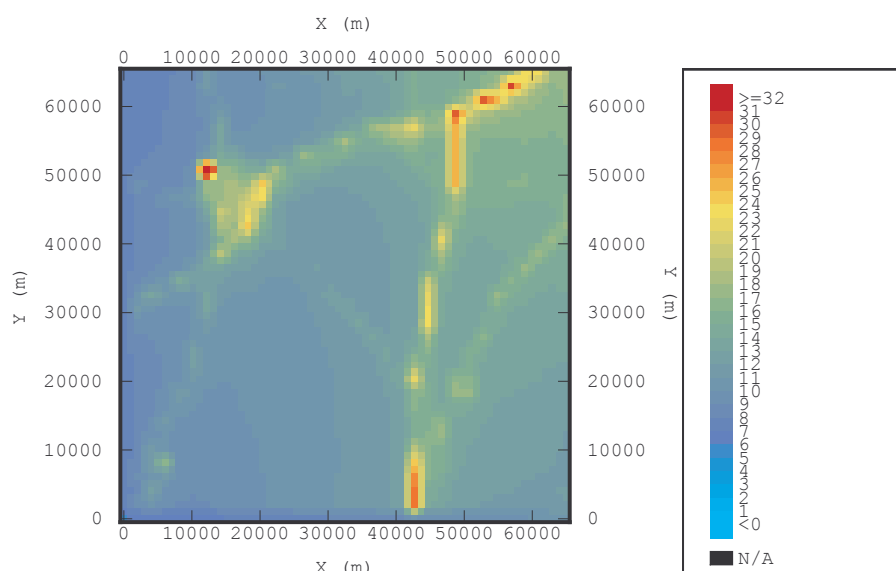


Figure 18 - Estimation par krigeage ponctuel colocalisé, à partir du jeu de données

Par rapport au maillage de référence l'apport de la variable auxiliaire est très faible (écarts relatifs inférieurs à 3% par rapport à la carte de référence). Nous avons donc étudié l'apport de la variable auxiliaire pour des structures dégradées.

COMPARAISON DES ESTIMATIONS ET DISCUSSION

L'étude à porter sur un ensemble de maillage régulier :

- maillage de 17x17 carré (289 pts) au pas de 4km,
- maillage de 10x10 carré (100 pts) au pas de 7km,
- maillage de 9x9 carré (81 pts) au pas de 8km,

Pour chacun de ces maillages nous avons étudiés la carte krigée sans et avec variables auxiliaires. L'objectif étant d'estimer l'apport de la variable auxiliaire sur le résultat. La comparaison, résumée tableau 7, s'effectue au moyen outils statistiques de comparaison entre simulation. Nous avons utilisés deux indicateurs, FB (fractional Bias) et NMSE (Normalized Mean Square Error).

$$\text{Fractional Bias (FB)} : \text{FB} = 2 \times \frac{\overline{C_{cal}} - \overline{C_{mes}}}{\overline{C_{cal}} + \overline{C_{mes}}}$$

$$\text{Normalized Mean Square Error (NMSE)} : \text{NMSE} = \frac{(C_{cal} - C_{mes})^2}{C_{cal} \times C_{mes}}$$

En considérant :

C_{cal} la concentration obtenue en chaque point pour chaque maillage dégradé, et
 C_{mes} la concentration en chaque point du maillage de référence .

FB fournit une information sur la tendance du modèle à surestimer (FB positif) ou sous-estimer (FB négatif) les concentrations. FB est un nombre sans dimension, compris entre -2 et $+2$. La valeur idéale de FB est 0, et il est considéré qu'elle doit être comprise entre $-0,5$ et $+0,5$ pour qu'une simulation soit acceptable. NMSE donne une information sur la dispersion globale des données. La valeur idéale de NMSE est 0, et il est considéré qu'elle doit être inférieure à 0,5 pour qu'une simulation soit acceptable.

	FB		NMSE	
	sans var auxiliaire	avec var auxiliaire	sans var auxiliaire	avec var auxiliaire
Maillage 17x17	-0.13	-0.12	0.14	0.10
Maillage 10x10	-0.18	-0.18	0.32	0.13
Maillage 9x9	-0.56	-0.54	0.98	0.56

Tableau 7 : performances des différents maillages

Les performances des différentes simulations montrent très clairement l'impact positif de la variable auxiliaire à mesure que le maillage se dégrade. Globalement les simulations sous-estiment la concentration par rapport à la simulation de référence, et la dispersion de l'erreur augmente avec la dégradation du maillage. On peut noter l'amélioration apportée par la variable auxiliaire et la dégradation nette des performances sur le maillage 9x9.

PARTIE III. Echantillonnage stratifié avec variable auxiliaire

Afin de mieux décrire des singularités du domaine d'étude, il apparaît souvent pertinent de resserrer le maillage au droit d'un émetteur particulier. Nous nous proposons d'étudier cette stratification du maillage sur le domaine de référence de notre étude. A partir de cette expérience nous tenterons d'établir des règles de bonne pratique applicable à toutes études.

Nous avons choisi de tester l'impact de ce maillage sur une simulation dégradé en tenant compte des indication du chapitre précédent. Nous avons montré qu'un maillage régulier de 10x10 apporte une réponse cohérente entre précision et limitation de coût d'échantillonnage. Avec cet objectif d'environ 100 points de mesures nous allons tenter d'améliorer la précision, sans trop ajouter de points, au moyen d'un maillage stratifié représentatif des émissions du domaine.

ELABORATION DU MAILLAGE

1) Etablissement du maillage régulier

Pour bâtir le maillage stratifié nous avons tout d'abord utilisée les points représentatifs des campagnes de mesures présent sur un maillage régulier de 9X9 (figure n°19) ; Nous avons retenus 34 points.

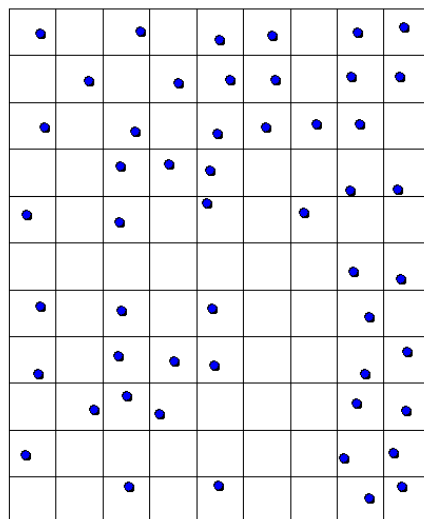


figure n°19 : répartition des points de campagne sur le maillage 10x10

A partir de ces points, nous avons construit un jeu complémentaire provenant des données de simulations pour compléter le maillage régulier de 9x9. Nous avons ajouté 47 points (figure n°20)

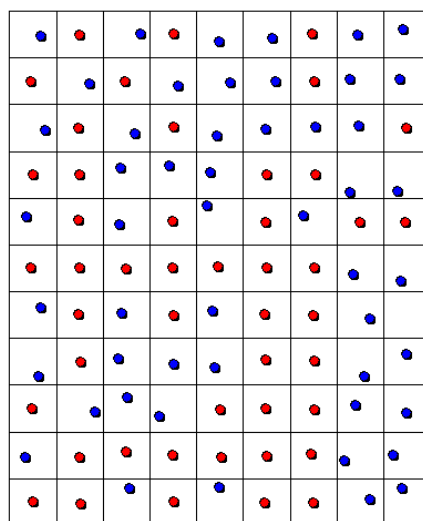


figure n°20 : répartition des nouveaux points (en rouge) sur le maillage

2) Représentativité de la variable auxiliaire

La démarche consiste ensuite à contrôler la bonne répartition des valeurs échantillonnées par rapport à la gamme de valeur de la variable auxiliaire. Dans la

cadre de notre étude nous avons considéré les émissions de NO₂ comme variable auxiliaire. Sur l'ensemble du domaine nous avons établi les statistiques de la distribution des émissions sur le domaine. Pour établir ces statistiques nous avons considéré la répartition des points de prélèvements suivant le centile des émissions sur le domaines (tableau n°8)

percentiles	Nombre de points				
	grille 9x9	Complément	grille 9x9 régulière	grille 9x9 améliorée	modifications
0.05	2	4	6	4	-2
0.1	3	3	6	4	-2
0.15	2	1	3	4	1
0.2	1	2	3	4	1
0.25	3	1	4	4	0
0.3	1	4	5	4	-1
0.35	1	2	3	4	1
0.4	0	1	1	4	3
0.45	5	3	8	4	-4
0.5	0	1	1	4	3
0.55	0	3	3	4	1
0.6	1	2	3	4	1
0.65	2	1	3	4	1
0.7	3	2	5	4	-1
0.75	1	3	4	4	0
0.8	1	3	4	4	0
0.85	4	2	6	4	-2
0.9	1	3	4	4	0
0.95	1	1	2	4	2
0.97	0	2	2	4	2
0.98	1	2	3	4	1
0.99	1	1	2	4	2
1	0	0	0	4	4
Nbre points	34	47	81	92	11

Tableau n°8 : Analyse de la répartition des points d'échantillonnage par rapport à la représentativité de la variable auxiliaire

Le premier groupe de point (34) provenant de l'échantillonnage des campagnes de mesures renseigne partiellement la distribution des émissions (5 classes de percentiles ne sont pas renseignées). Le complément du fichier par 47 points permet de renseigner la totalité des percentiles à l'exception du percentile 99. Nous avons ensuite amélioré la distribution des points en terme de représentativité des émissions. L'objectif est d'obtenir un nombre de points à peu près identique suivant les différents classes de percentile. Nous avons choisi de tester une répartition égale des points suivant les percentiles. Ce choix conduit à ajouter globalement 11 points, (ajout de 23 points et suppression de 12 points).

La distribution finale comprend 92 points répartis équitablement suivant la distribution des émissions sur le domaine (figure n°21).

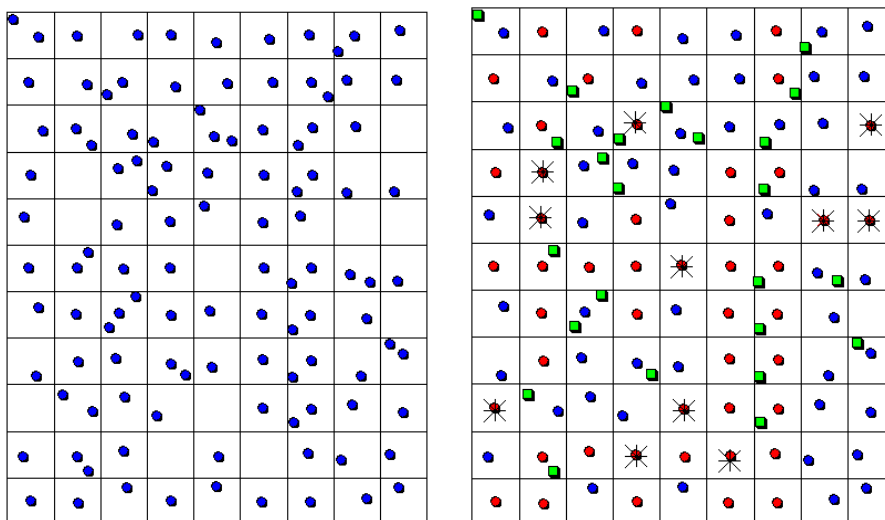


figure n°21 : distribution finale du maillage stratifié :
à droite avec les points supplémentaire (carré) et les points supprimés (croix),
à gauche maillage stratifié

ETUDE DU KRIGEAGE COLOCALISE

En utilisant le krigeage colocalisé, on fait l'hypothèse que les deux variables (concentration de NO_2 et variable auxiliaire ($\log(\text{émission}+1)$)) sont liées par un modèle variographique. On va établir le modèle bivariable à partir du jeu de données et des valeurs de variable auxiliaire définies aux mêmes coordonnées.

Etude des variogrammes expérimentaux

On calcule pour toutes les directions les deux variogrammes simples et le variogramme croisé avec 8 pas de 6000 m.

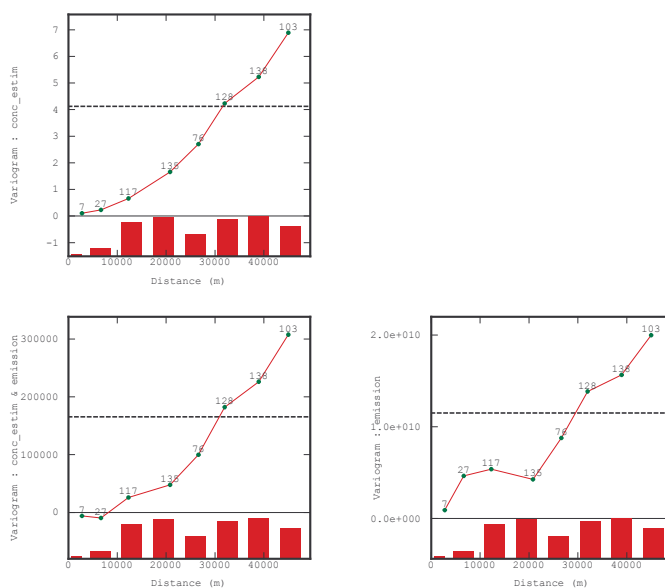


Figure 22 – Variogrammes expérimentaux simples et croisé

Modélisation des variogrammes et validation croisée

A partir de plusieurs structures simples, le logiciel définit les structures permettant de simuler les différents variogrammes. On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 10 km.

(92 points)	Moyenne	Variance
Erreur	-0,45	6,3
Erreur standardisée	0.23	0.78

18 données sur 92 non robuste.

Tableau 9 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les concentrations

ELABORATION DE LA CARTE

On estime les concentrations en tous points de la carte de référence à l'aide du jeu de données, de la variable colocalisée {log (émission +1)} et du modèle bivariable, en considérant un voisinage de 10 km.

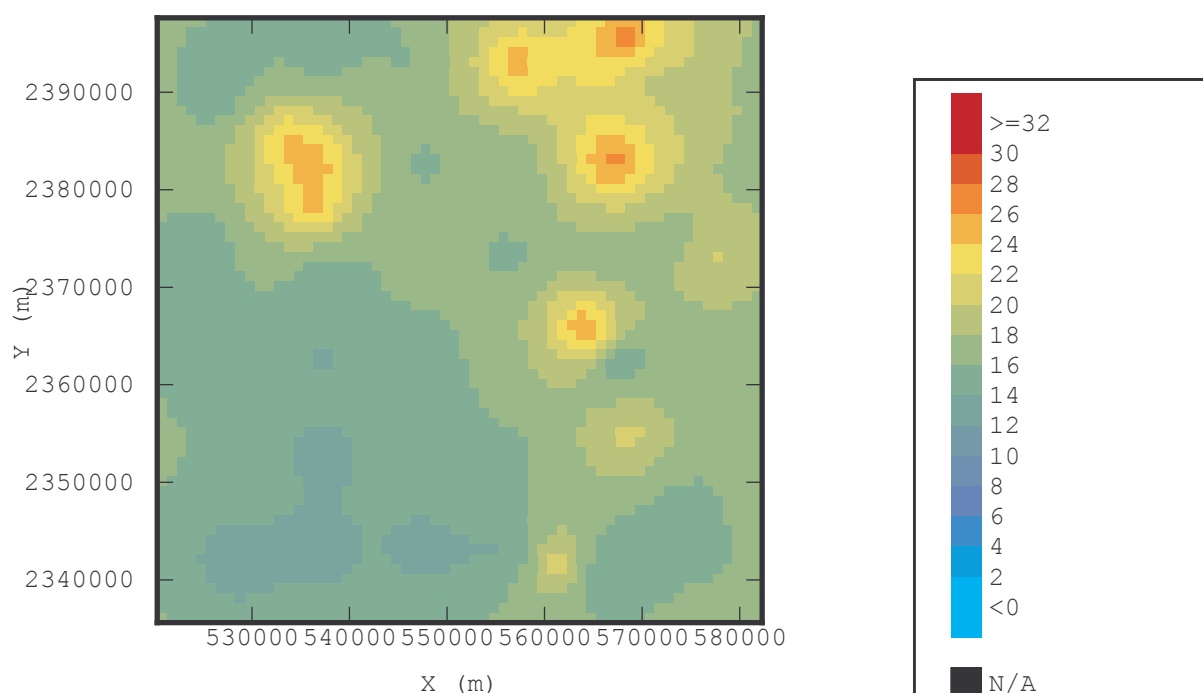


Figure 23: Estimation par krigeage ponctuel colocalisé, de la concentration en NO₂ pour l'utilisation d'un maillage stratifié

Le calcul des scores statistiques permet d'établir une comparaison avec les simulations précédentes (tableau 10)

	FB		NMSE	
	sans var auxiliaire	avec var auxiliaire	sans var auxiliaire	avec var auxiliaire
Maillage 17x17	-0.13	-0.12	0.14	0.10
Maillage 10x10	-0.18	-0.18	0.32	0.13
Maillage 9x9	-0.56	-0.54	0.98	0.56
Maillage stratifié		-0.20		0.14

Tableau n°10 : comparaison des performances des différents maillages

On remarque que le maillage stratifié n'apporte pas un gain déterminant par rapport à un maillage régulier, les performances sont semblables à un maillage 10x10 (100 points) avec une distribution de 92 points.

PARTIE IV. Conclusions et perspectives

Cette partie des travaux concernant l'adaptation des plans d'échantillonnage aux objectifs de campagne s'est attachée à l'étude du dioxyde d'azote l'ozone à l'échelle régionale en prenant le cas de la Centre.

Les simulations et les campagnes de mesures par échantillonnage passif effectuées par l'EMD en collaboration avec LIG'AIR, ont permis d'établir une carte de référence et d'étudier la dégradation des estimations de concentrations du dioxyde d'azote en utilisant un maillage régulier.

On a pu observer que la structure des variogrammes expérimentaux se dégrade nettement lorsque le nombre de points diminue, à partir du maillage 9×9 km (81 points).

Ce travail a également montré qu'une maille de 6 km minimum de côté était nécessaire pour vérifier les critères de qualité, soit environ 100 points, ce qui, compte tenu de la superficie du domaine, correspond à une densité d'échantillonnage de 1 point pour environ 44 km².

Dans un second temps, nous avons cherché à utiliser des variables explicatives du comportement spatial du dioxyde d'azote pour améliorer les estimations, et éventuellement réduire la densité d'échantillonnage.

Nous avons utilisé le cokrigeage colocalisé pour prendre en compte le cofacteur émission dans l'estimation des concentrations. La prise en compte de cette variable auxiliaire a permis d'améliorer de manière significative les cartes d'estimation lorsque le maillage était fortement dégradé, en revanche l'impact est faible pour le maillage de référence.

Nous avons ensuite envisagé le recours à un maillage stratifié, devant permettre une meilleure description des zones à fort contraste d'émission (proximité d'axe de trafic, de site industriel). La méthode employée, basée sur un maillage régulier complété de points représentatifs de l'étendue des émissions. Cette apport n'a pas été significatif par rapport à l'utilisation d'un maillage régulier.