

ECHANTILLONNAGE ET CARTOGRAPHIE DU BENZENE

ANNEXE 10 – Urbain : Agen

ANNEXE 11 – Urbain : Lille

ANNEXE 12 – Urbain : Troyes, Orléans, Bourges (étude variographique)

Annexe 10 – Urbain : Agen

1. TOUTES LES DONNÉES	3
1.1 Analyse préliminaire	3
1.2 Modélisation variographique et validation croisée	4
1.3 Estimation	8
2. JEU DE DONNÉES DE BENZÈNE RÉDUIT	9
3. REMARQUES	12

Agen – Benzène – Etude exploratoire

La campagne de mesure de benzène traitée dans cette étude a été conduite par AIRAQ du 6 juin au 4 juillet 2002. 57 sites, dont 43 sites de fond, ont fait l'objet d'un échantillonnage. Des mesures de NO₂ ont été réalisées parallèlement. L'analyse porte sur les moyennes mensuelles obtenues.

1. TOUTES LES DONNEES

1.1 Analyse préliminaire

Les échantillons se répartissent selon un maillage irrégulier, fonction de la densité de population. La taille de la maille varie entre 500m -plus forte densité- et 3 km -plus faible densité- (Figure 1).

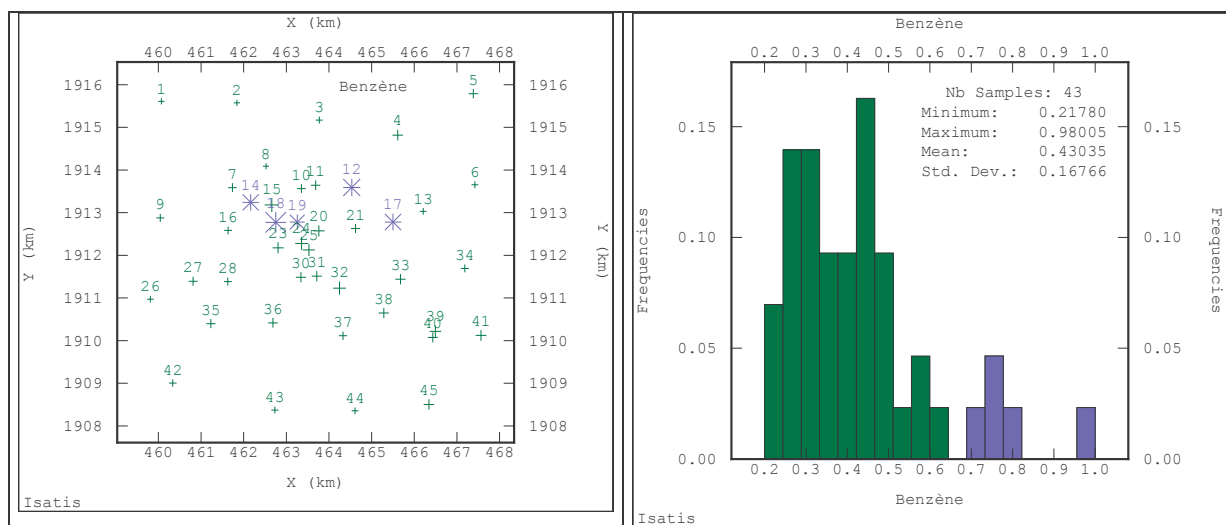


Figure 1 – Implantation des sites d'échantillonnage et histogramme des concentrations moyennes sur la période. En bleu : points de concentration plus élevée.

La concentration de benzène est corrélée linéairement à celle du dioxyde d'azote malgré une certaine dispersion du nuage de corrélation (Figure 2). Les données de NO₂ seront exploitées dans un cokrigage.

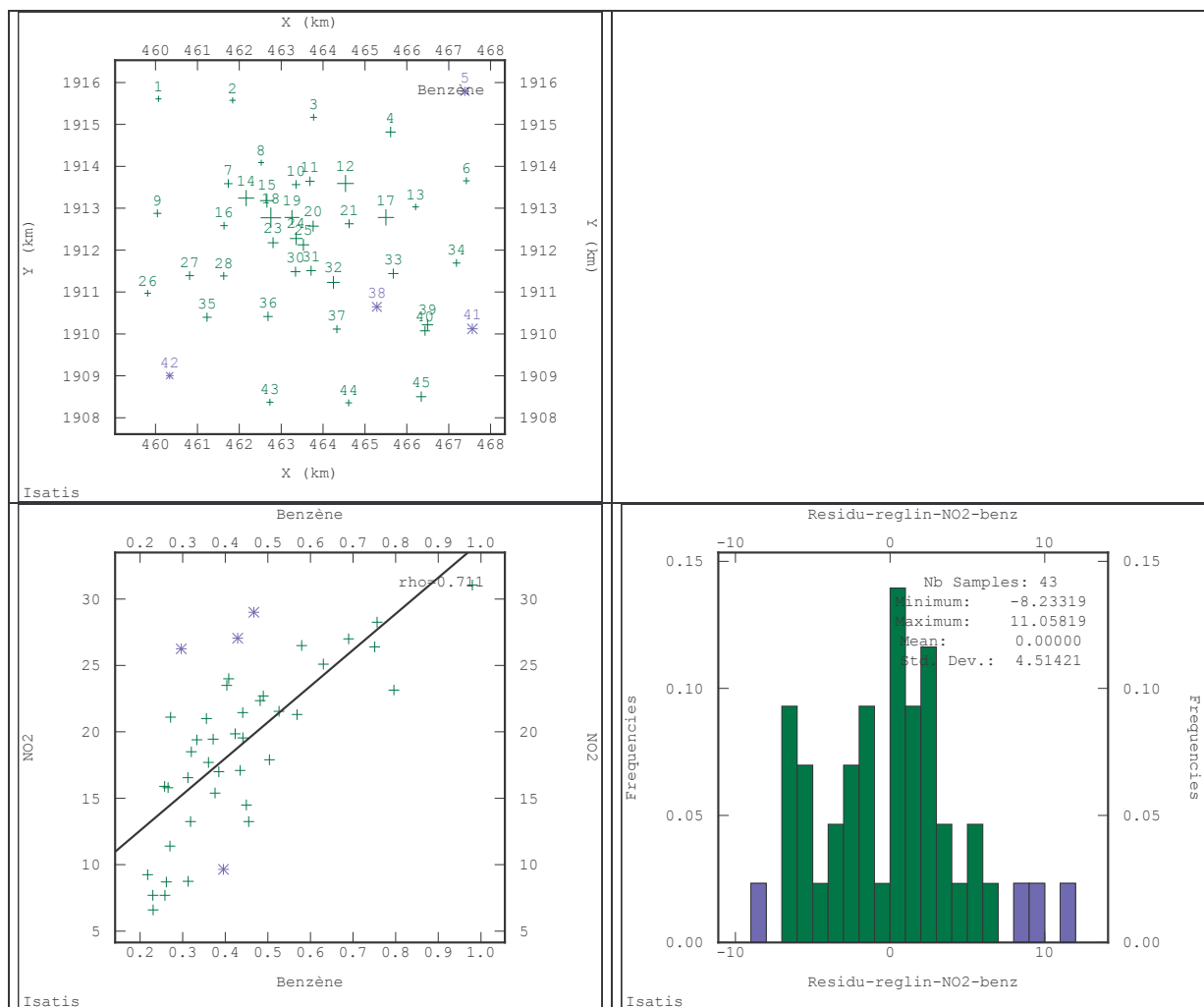


Figure 2 – Etude de la corrélation avec les concentrations de dioxyde d’azote. En bleu : points où les résidus de la régression $\text{NO}_2 = f(\text{Benzène})$ sont les plus élevés en valeur absolue (points 5, 38, 41, 42)

1.2 Modélisation variographique et validation croisée

Le point 18 (de concentration maximale) est supprimé du calcul du variogramme dont il masque la structure.

Deux modélisations sont réalisées (Figure 3) :

- dans le cas monovarié,

Modèle : $\text{pépite} = 0.003 + \text{sphérique}(\text{portée} = 1\text{km}, \text{palier} = 0.0068) + \text{sphérique}(\text{portée} = 3.2\text{km}, \text{palier} = 0.0151)$

- dans le cas bivariable, avec NO_2 comme seconde variable.

Modèle :

$\text{sphérique}(\text{portée} = 1\text{km}) + \text{sphérique}(\text{portée} = 3.2\text{km})$ pour les deux variogrammes simples et le variogramme croisé. NB : Les paliers sont ajustés automatiquement afin d’obtenir plus aisément un modèle autorisé. Les paliers ajustés pour le benzène diffèrent du cas monovarié, en particulier celui de la première structure est plus élevé.

NB : le modèle monovarié est très similaire au modèle ajusté par AIRAQ [1].

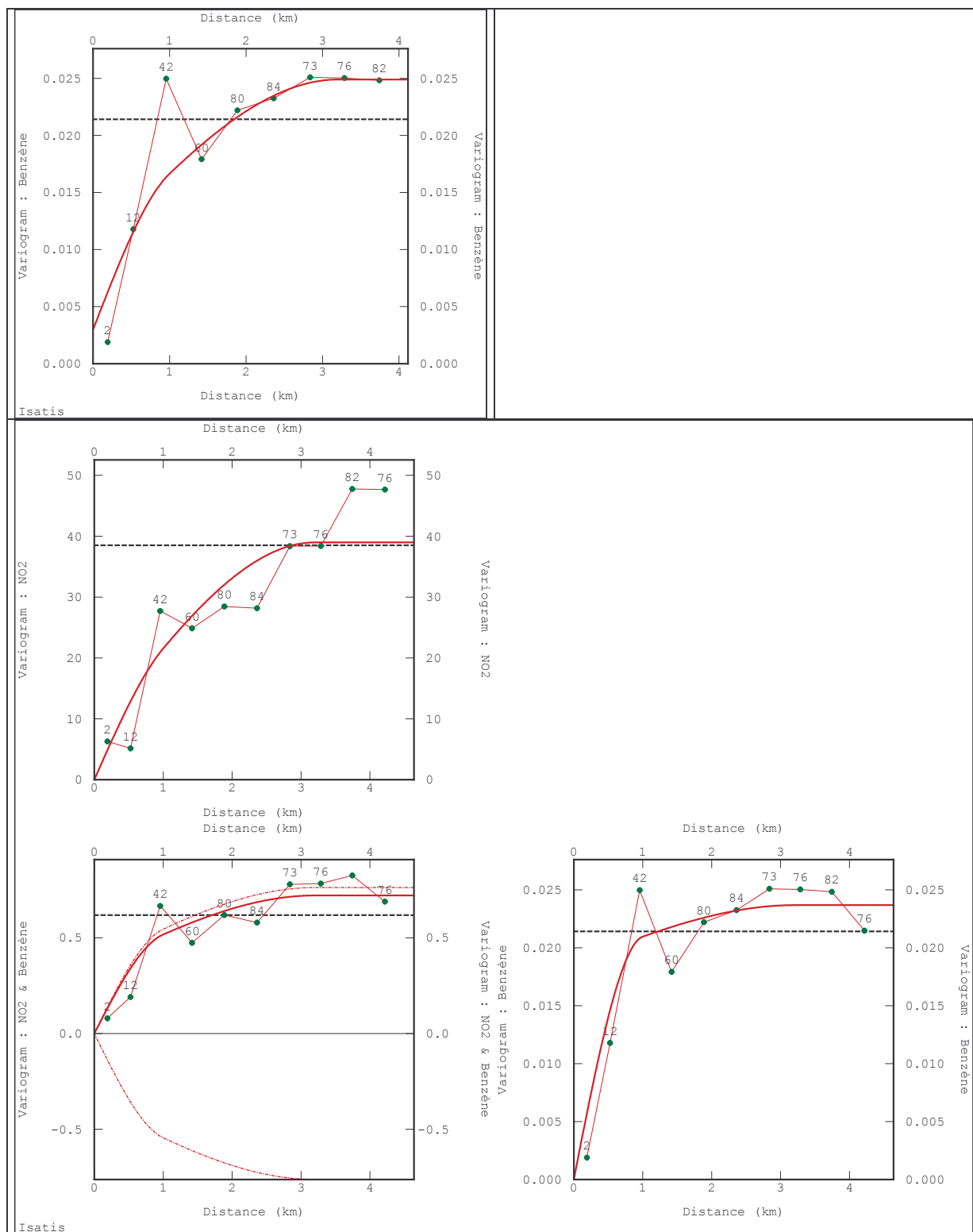


Figure 3 – Ajustement d'un modèle de variogramme monovariante (ligne du haut) ou bivariable (ligne du bas).

Validation croisée :

Le point 18, retiré de l'analyse variographique, est réintroduit dans les données d'entrée du krigeage car il améliore les résultats de validation croisée.

En cokrigeage, deux validations croisées sont effectuées : dans l'une, la valeur de NO₂ au point cible est ignorée en même temps que celle de benzène (on est alors dans le cas homotopique : les mesures de NO₂ et de benzène sont connues exactement aux mêmes points), dans l'autre, elle est conservée (cas hétérotopique).

Du fait du nombre limité de données, un voisinage d'estimation unique est défini.

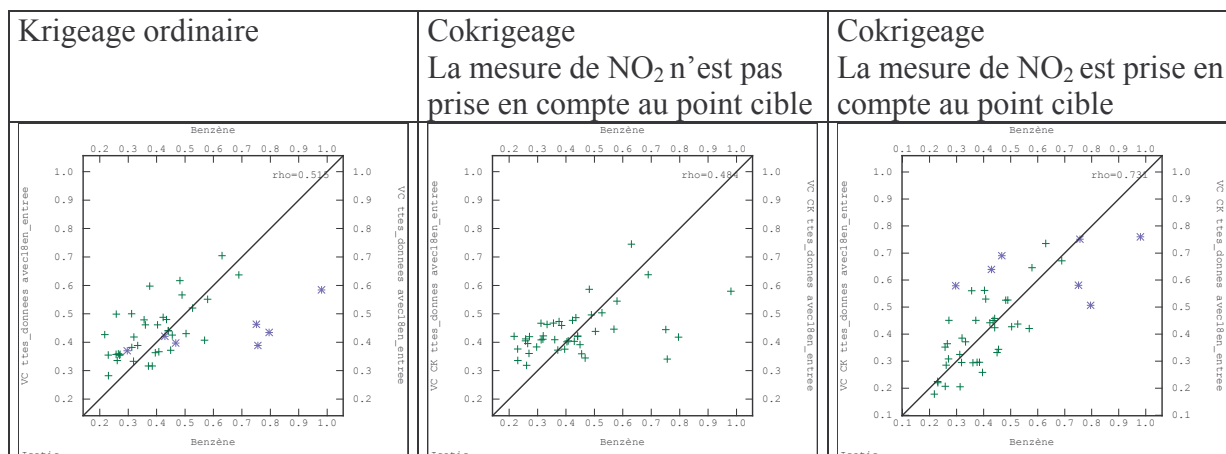


Figure 4 – Corrélation entre valeurs mesurées et estimées par validation croisée.

Le cokrigeage homotopique n'a pas d'effet notable. En revanche le cokrigeage hétérotopique -dans le sens où la concentration de NO₂ au point cible est prise en compte- réduit en moyenne l'erreur d'estimation (Figure 4, Tableau 1).

Tableau 1 - Statistiques des erreurs de validation croisée

	Krigeage ordinaire		Cokrigeage (homotopique)		Cokrigeage (hétérotopique)	
	Erreur (avec 18)	Erreur relative (avec 18)	Erreur (avec 18)	Erreur relative (avec 18)	Erreur (avec 18)	Erreur relative (avec 18)
N	43	43	43	43	43	43
Min	-0.40	0.01	-0.42	0.28	-0.29	0.67
P25	-0.05	11.1	-0.03	9.70	-0.08	8.26
P50	0.05	22.1	0.05	21.8	0.01	19.9
P75	0.09	35.0	0.10	40.8	0.07	32.2
Max	0.24	96.0	0.20	93.2	0.28	95.2
Moy	0.008	26.2	0.008	27.0	0.006	22.7
Etype	0.146	22.0	0.148	21.2	0.121	19.3
Var	0.021	483.6	0.022	450.2	0.015	372.1
Corr Est-Mes	0.515		0.484		0.731	

En krigeage ordinaire, et en cokrigeage homotopique, les concentrations mesurées aux points 12, 14, 17 et 18 sont nettement sous-estimées (points bleus sous la bissectrice). Le cokrigeage hétérotopique avec le NO₂ corrige partiellement cette sous-estimation. En revanche, il dégrade l'estimation aux points 38, 41 et 42 : proches de la bissectrice en krigeage ordinaire, ces points s'en éloignent en cokrigeage (surestimation). Ce résultat s'accorde avec les remarques faites sur le nuage de corrélation entre benzène et NO₂ (les points 38, 41 et 42, qui présentent une concentration relativement élevée de NO₂ mais une concentration faible ou moyenne de benzène, s'écartent du nuage).

Pour compléter cette évaluation, quatre nouvelles validations croisées sont réalisées. Pour chacune, 12 points sélectionnés aléatoirement sont retirés du jeu initial de benzène (Figure 5).

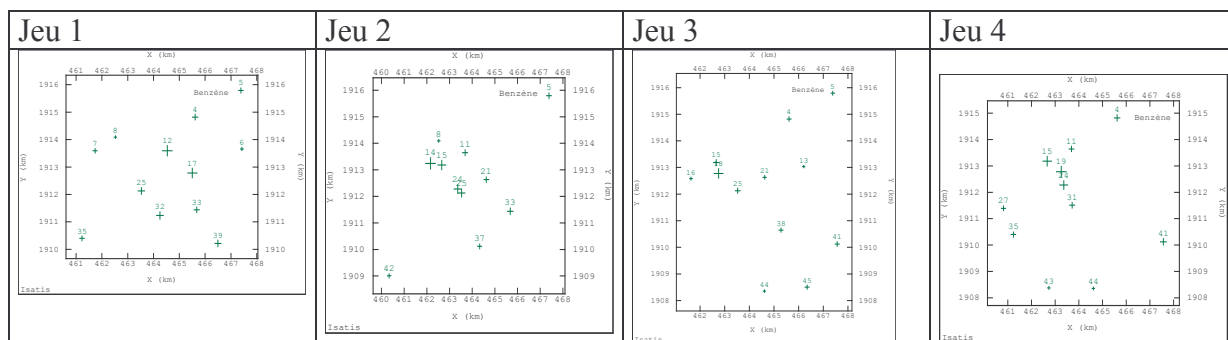


Figure 5 – Implantation des sites de validation

Sur ces quatre jeux, on estime les concentrations de benzène :

- en krigeage ordinaire, à l'aide des 33 mesures restantes ;
- en cokrigeage hétérotopique, à l'aide des 33 mesures restantes de benzène et des 45 mesures de NO₂.

Le modèle variographique établi sur l'ensemble des données de benzène et de NO₂ est utilisé.

Les statistiques d'erreur sont généralement plus favorables au cokrigeage hétérotopique, sauf dans le cas du jeu 4. Une dégradation de l'estimation par rapport au krigeage ordinaire (surestimation) est observée pour les points suivants :

Jeu 2 : points 11 et 42 ; Jeu 3 : points 38 et 41 ; Jeu 4 : points 11, 41 et 43.

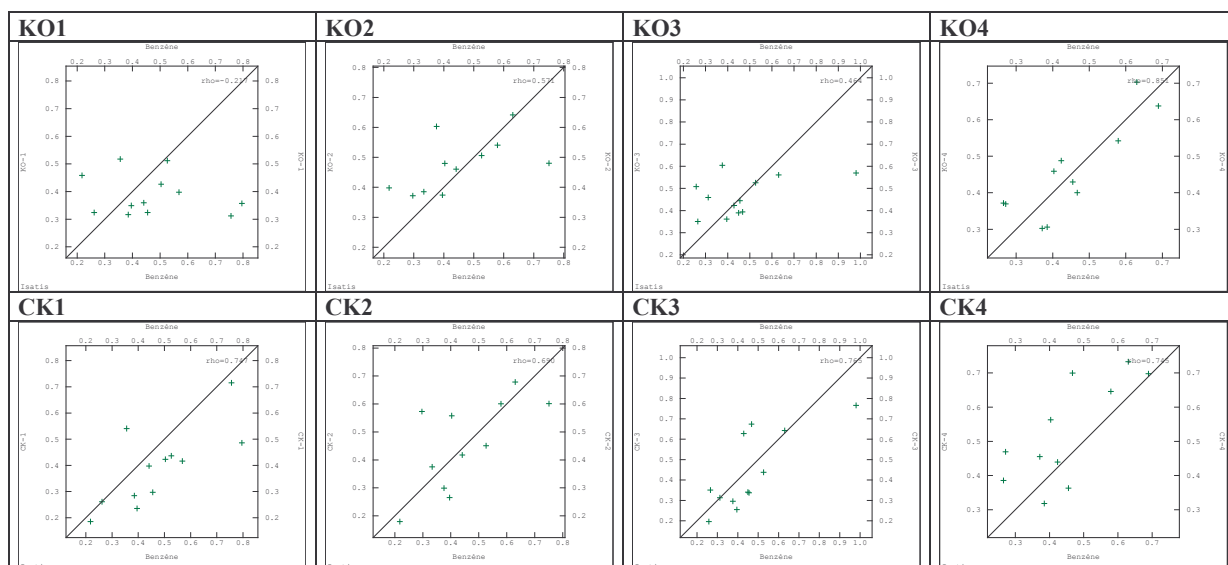


Figure 6 - Corrélation entre valeurs mesurées et estimées pour chaque jeu de données de validation.

Tableau 2 - Statistiques des erreurs de validation

	KO1	KO2	KO3	KO4	CK1	CK2	CK3	CK4
N données	12	12	12	12	12	12	12	12
Moy erreur	-0.084	0.027	0.004	0.006	-0.082	0.004	-0.026	0.075
Var erreur	0.042	0.017	0.030	0.005	0.014	0.016	0.017	0.011
Moy erreur relative	34.9	23.8	27.6	17.2	23.5	24.2	24.6	27.3
Var erreur relative	862.1	696.3	866.7	128.4	246.3	637.9	201.2	482.3
RMSE	0.213	0.126	0.165	0.070	0.139	0.120	0.129	0.124
Coef. cor	-0.217	0.571	0.464	0.851	0.747	0.690	0.765	0.745

Les écarts-types de cokrigage sont plus faibles que les écarts-types de krigeage ordinaire du fait que les mesures de NO₂ aux 12 points sont prises en compte :

	KO1	KO2	KO3	KO4	CK1	CK2	CK3	CK4
N	12	12	12	12	12	12	12	12
Min	0.098	0.112	0.100	0.099	0.056	0.075	0.062	0.061
Max	0.163	0.160	0.162	0.161	0.104	0.141	0.104	0.141
Moy	0.136	0.131	0.138	0.131	0.089	0.093	0.092	0.091
Var	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

1.3 Estimation

La grille d'estimation a une résolution spatiale de 500m. Un voisinage unique est utilisé. Sauf aux points de données, les écarts-types de cokrigage sont plus élevés. Ce résultat s'explique par le modèle de variogramme du benzène : comme il a été mentionné, le palier de la première structure est supérieur au palier ajusté manuellement en krigeage ordinaire. Les données d'entrée de benzène et de NO₂ étant disponibles aux mêmes points, seul un cokrigage homotopique est possible.

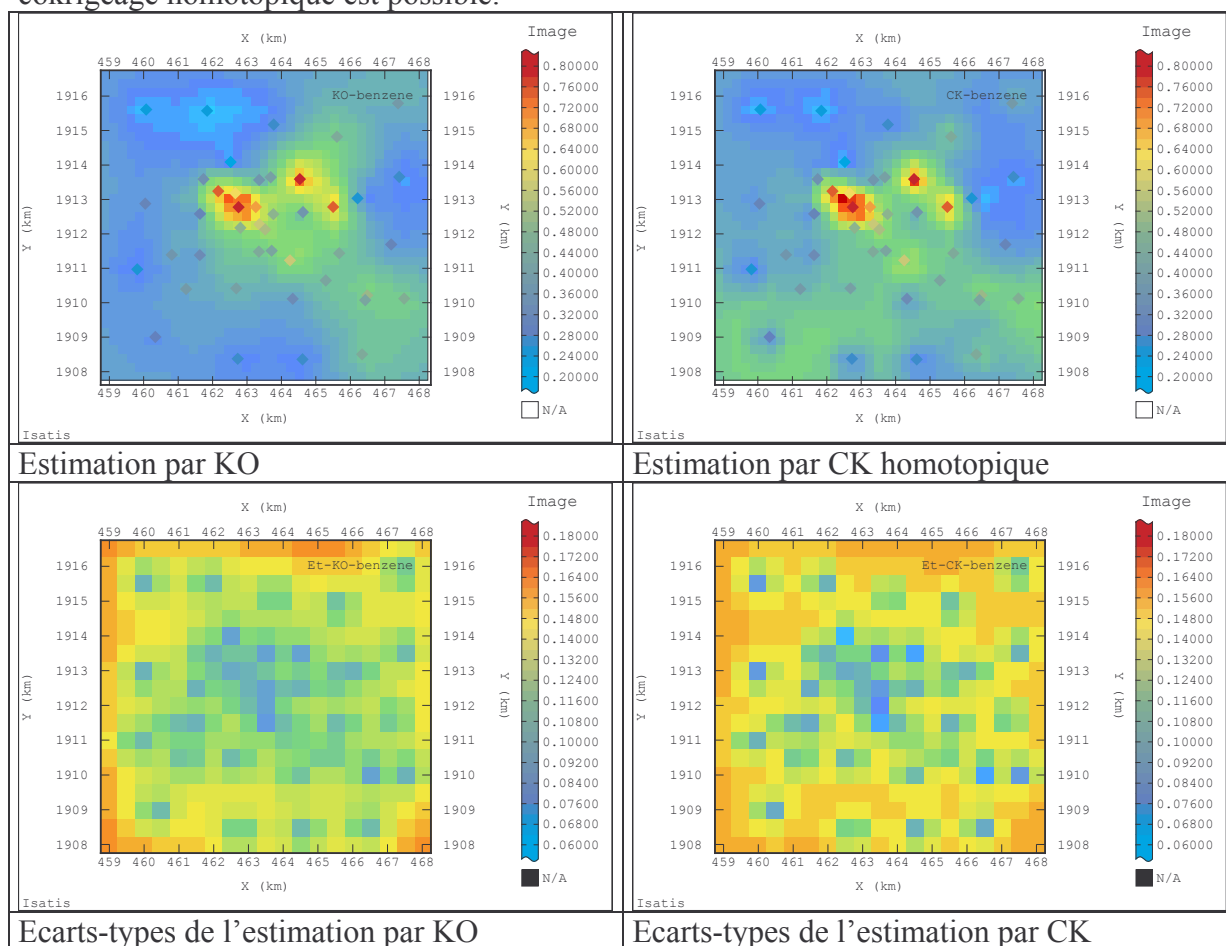


Figure 7 – Estimation des concentrations de benzène par krigeage (fig. de gauche) et cokrigage (fig. de droite) ordinaires. Les losanges représentent les points de mesure. L'échelle de couleurs est la même pour la grille et les points.

Les cartes diffèrent dans la zone sud-ouest où les concentrations estimées par cokrigage sont plus élevées. Or c'est précisément là que se situe le point 42, qui s'écarterait du nuage de corrélation entre benzène et NO₂ (Figure 2). Il est possible que dans cette zone, le cokrigage surestime les niveaux de benzène. Nous ne disposons pas de point complémentaire pour le vérifier.

Les cartes de cokrigage de la Figure 7 sont présentées à titre d'illustration. Comme on l'a mentionné, c'est plutôt dans le cas hétérotopique que le cokrigage devrait révéler son efficacité.

2. JEU DE DONNEES DE BENZENE REDUIT

L'effectif initial de données, inférieur à 50, ne peut être réduit de manière importante. Dans cette expérience, quatre points seulement du centre-ville sont éliminés afin de tester la sensibilité des résultats au resserrement urbain de l'échantillonnage.

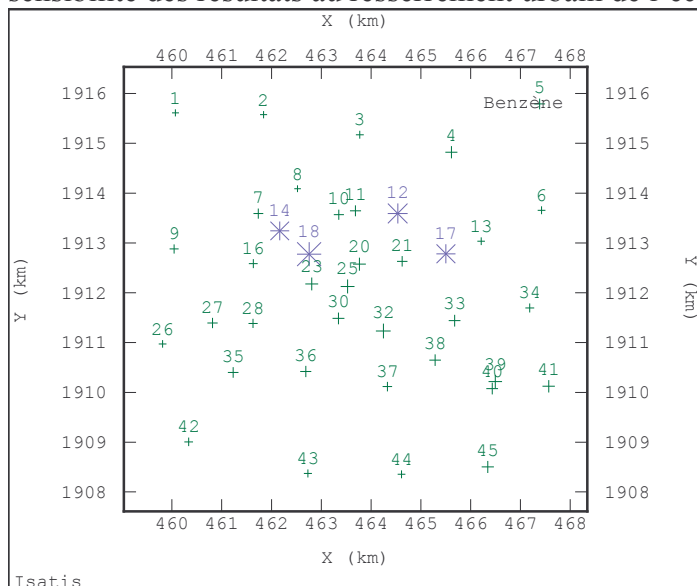


Figure 8 – Jeu de données réduit : implantation des sites d'échantillonnage

Les points 14 et 18, qui destructurent le variogramme, sont retirés de l'étape de modélisation et réintroduits pour la validation croisée et l'estimation. Deux nouveaux modèles mono et bidimensionnels sont définis (pas d'effet de pépite, modèle emboîté comprenant une structure sphérique de portée 1km et une structure sphérique de portée 3.4km). Ils ne diffèrent que légèrement des modèles ajustés précédemment.

Variogramme simple :

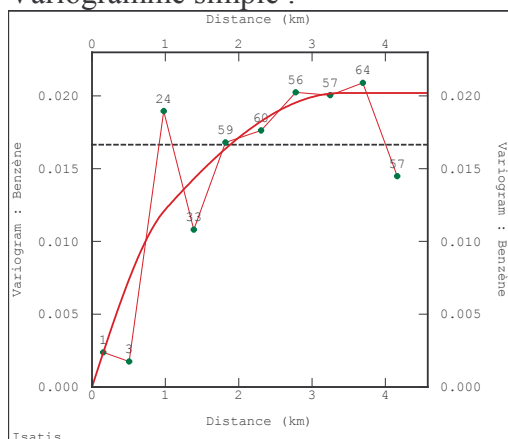


Figure 9 – Jeu réduit : ajustement d'un modèle de variogramme

Les résultats de validation croisée sont assez peu affectés par la suppression des quatre points ; la sous-estimation au point 18 est néanmoins amplifiée. Là encore, la prise en compte des données de NO₂ aux points où l'on ignore la mesure de benzène améliore l'estimation.

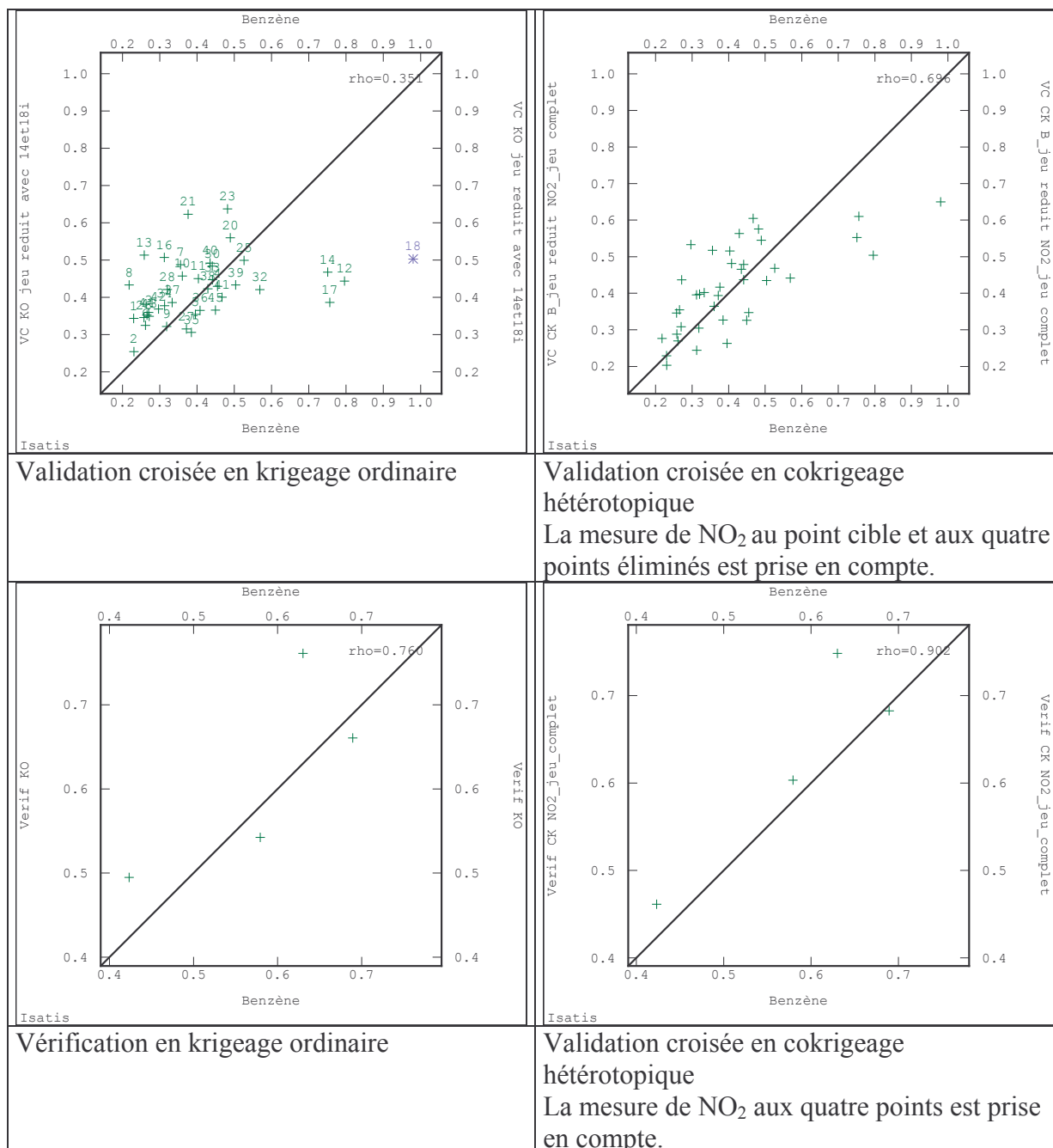


Figure 10 - Corrélation entre valeurs mesurées et estimées en validation croisée.(ligne du haut) et absolue (ligne du bas)

Tableau 3 – Jeu réduit : statistiques des erreurs de validation croisée :

	Krigeage ordinaire		Cokrigage (homotopique)		Cokrigage (hétérotopique)	
	Erreur (avec 18)	Erreur relative (avec 18)	Erreur (avec 18)	Erreur relative (avec 18)	Erreur (avec 18)	Erreur relative (avec 18)
N	39	39	39	39	39	39
Min	-0.48	1.04	-0.49	1.30	-0.33	0.32
P25	-0.05	12.4	-0.04	10.6	-0.06	11.2
P50	0.05	24.3	0.05	25.2	0.03	20.6
P75	0.09	35.6	0.10	40.6	0.08	28.6
Max	0.26	98.9	0.21	98.5	0.24	79.7
Moy	0.006	28.2	0.005	28.0	0.002	21.9
Etype	0.159	23.3	0.157	21.3	0.121	16.1

	Krigeage ordinaire		Cokrigeage (homotopique)		Cokrigeage (hétérotopique)	
	Erreur (avec 18)	Erreur relative (avec 18)	Erreur (avec 18)	Erreur relative (avec 18)	Erreur (avec 18)	Erreur relative (avec 18)
Var	0.025	540.6	0.025	452.2	0.015	260.1
Corr Est-Mes	0.351		0.356		0.696	

Dans l'estimation par cokrigeage, la totalité du jeu de NO₂ est conservé.

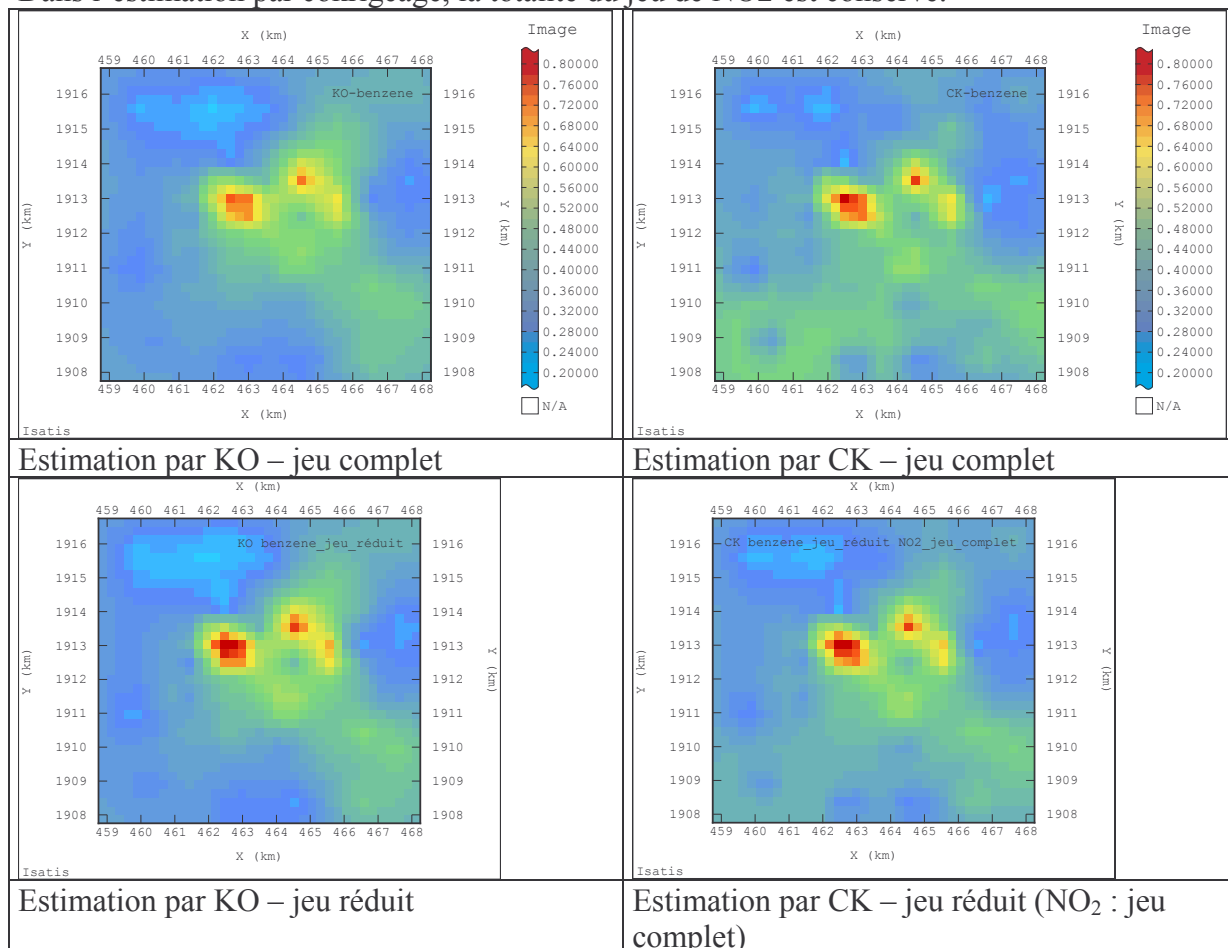


Figure 11 – Estimation des concentrations de benzène par krigeage (fig. de gauche) et cokrigeage (fig. de droite) ordinaires.

3. REMARQUES

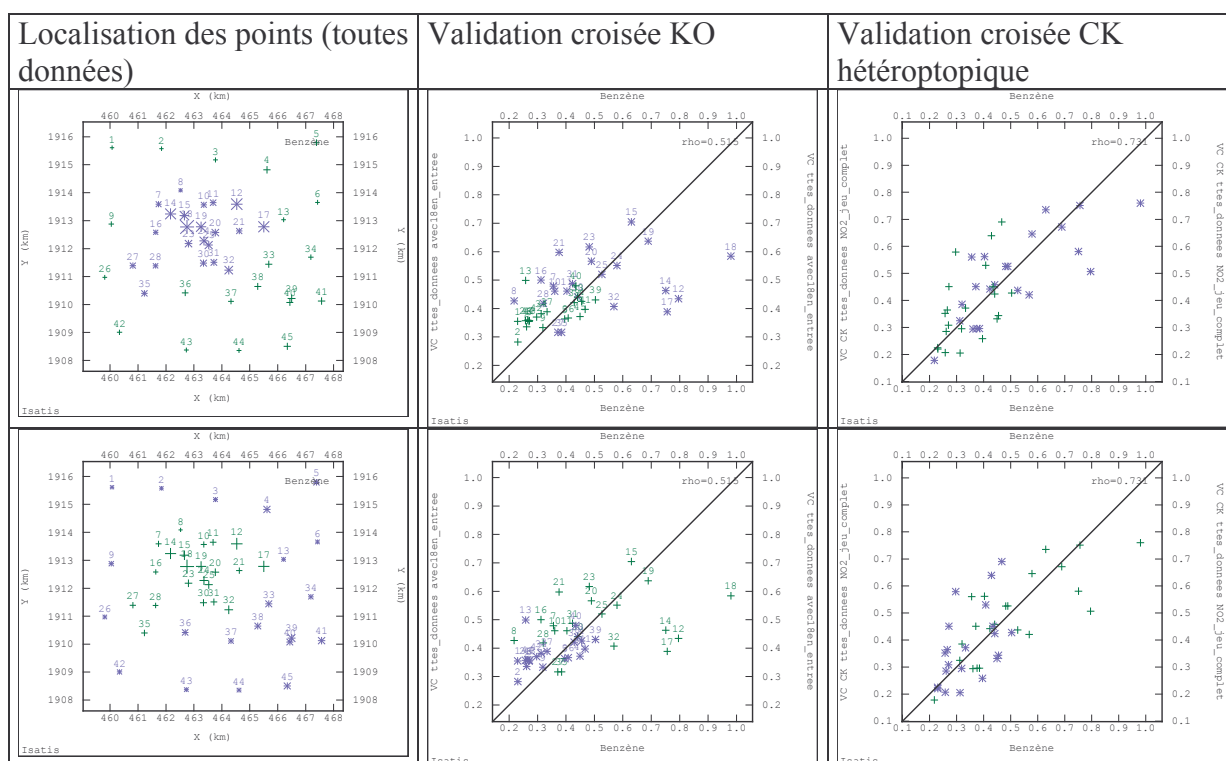


Figure 12 – Etude des résultats de la validation croisée selon la localisation des points de mesure

- La variabilité des concentrations en centre-ville fait qu'en dépit du resserrement de l'échantillonnage, certaines valeurs peuvent être mal réestimées par validation croisée (cette moindre précision n'apparaît pas sur les cartes d'écart-type de krigeage puisque cet écart-type dépend du modèle de variogramme et de la configuration géographique des données mais non pas de la variabilité locale des concentrations).
- En périphérie, l'échantillonnage adopté (un point par maille d'1,5 à 2,5km environ) permet une estimation satisfaisante des niveaux de concentration.
- Un cokrigeage homotopique n'agit pas sur la qualité de l'estimation. Le résultat est autre si l'on dispose de données de NO₂ en des points n'ayant pas fait l'objet de mesures de benzène. L'introduction de ces données dans le cokrigeage accroît la précision de l'estimation, notamment aux points de plus forte concentration de benzène. Dans le cas présent, les mesures de NO₂ et de benzène sont concomitantes ; cependant l'utilisation de données antérieures ou postérieures de NO₂ est envisageable si les concentrations des deux polluants sont bien corrélées.

4. REFERENCES

AIRAQ, 2002. Campagnes de mesure par échantillonnage passif. Agglomération d'Agen. Rapport n° ET/CA/02/05, www.airaq.asso.fr.

Annexe 11 – Urbain : Lille

BENZENE / Etude à l'échelle d'une agglomération : cas de l'arrondissement de Lille

PARTIE I. ECHANTILLONNAGE RÉGULIER SANS VARIABLE AUXILIAIRE 15

I.1.	Description du cas	15
• I.1.1.	Présentation du jeu de données initial	15
• I.1.2.	Méthodologie de sélection des données.....	15
I.2.	Etude des jeux de données initiaux et élaboration des cartes de référence	18
• I.2.1.	Etude des variogrammes expérimentaux.....	18
• I.2.2.	Modélisation du variogramme et validation croisée	19
• I.2.3.	Etablissement de la carte de référence	21
I.3.	Etude de la sélection de données S3 (maille de 3 × 3 km).....	24
• I.3.1.	Etude du variogramme expérimental	24
• I.3.2.	Modélisation du variogramme et test de validation croisée.....	25
• I.3.3.	Test de validation absolue sur le complémentaire C3	27
• I.3.4.	Etablissement de la carte pour la sélection S3	28
I.4.	Etude de la sélection de données S5 (maille de 5 × 5 km).....	30
• I.4.1.	Etude du variogramme expérimental	30
• I.4.2.	Modélisation du variogramme et test de validation croisée.....	31
• I.4.3.	Test de validation absolue sur le complémentaire C5	33
• I.4.4.	Etablissement de la carte pour la sélection S5	35
I.5.	Etude de la sélection de données S7 (maille de 7 × 7 km).....	35
• I.5.1.	Etude du variogramme expérimental	36
• I.5.2.	Modélisation du variogramme et test de validation croisée.....	37
• I.5.3.	Test de validation absolue sur le complémentaire C7	39
• I.5.4.	Etablissement de la carte pour la sélection S7	41
I.6.	Comparaison des estimations et discussion.....	41
• I.6.1.	Cas de l'été 2003.....	41
• I.6.2.	Cas de l'hiver 2004	46

PARTIE II. ECHANTILLONNAGE RÉGULIER AVEC VARIABLE(S) AUXILIAIRE(S)

50

II.1. Description des variables explicatives	50
II.2. Calcul des corrélations	51
• II.2.1. Analyse des variables explicatives	51
• II.2.2. Elaboration d'un cofacteur pertinent	52
II.3. Etude du krigeage colocalisé sur les sélections S3	54
• II.3.1. Cas de l'été 2003	54
• II.3.2. Cas de l'hiver 2004	57
II.4. Etude du jeu de données initial et élaboration de la carte de référence ...	61
• II.4.1. Cas de l'été 2003	61
• II.4.2. Cas de l'hiver 2004	64
II.5. Etude de la sélection de données S5 (maille de 5 × 5 km)	68
• II.5.1. Cas de l'été 2003	68
• II.5.2. Cas de l'hiver 2004	72
II.6. Etude de la sélection de données S7 (maille de 7 × 7 km)	76
• II.6.1. Cas de l'été 2003	76
• II.6.2. Cas de l'hiver 2004	80
II.7. Comparaison des estimations et discussion	84
• II.7.1. Cas de l'été 2003	84
• II.7.2. Cas de l'hiver 2004	87

PARTIE III. CONCLUSIONS ET PERSPECTIVES

91

PARTIE IV. RÉFÉRENCES

92

5. DESCRIPTION DU CAS

5.1 Présentation du jeu de données initial

Les mesures proviennent de campagnes menées par l'AREMA-LM au cours de l'été 2003 et l'hiver 2004 sur l'arrondissement lillois, sur un domaine d'environ 42×38 km [1], et d'une superficie d'environ 1580 km². Au cours de ces campagnes, les concentrations en benzène ont été mesurées sur une semaine à l'aide de tubes passifs.

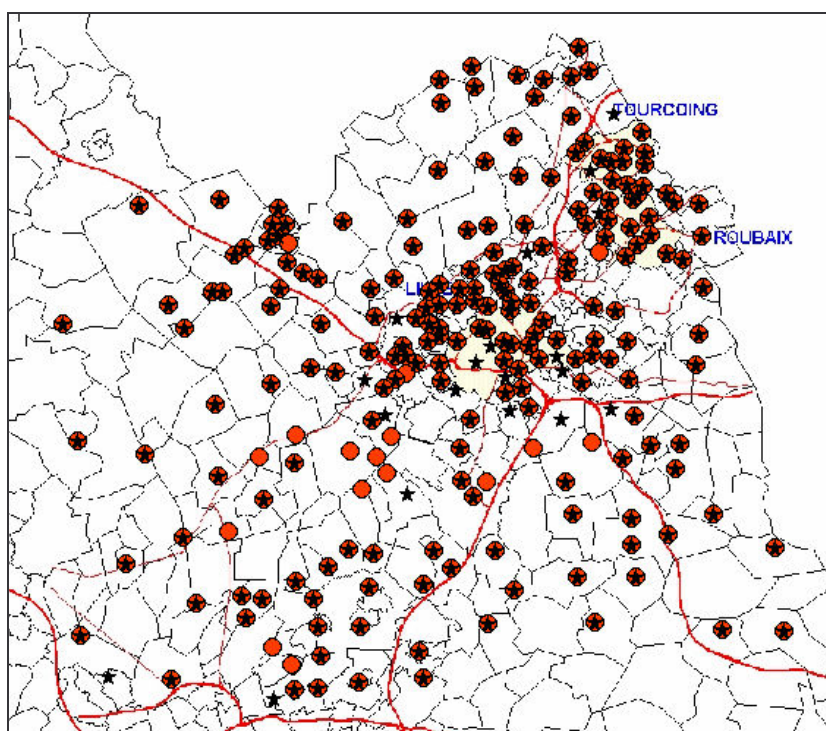
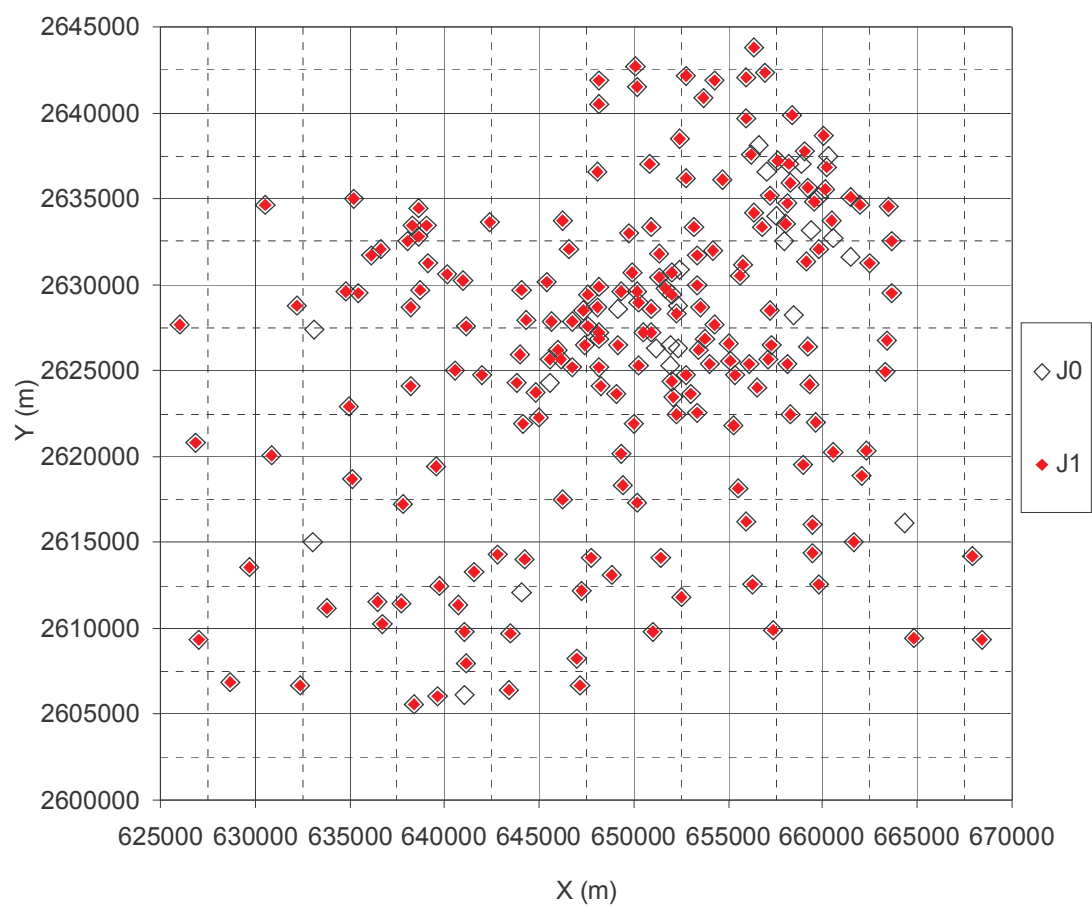


Figure 1 – Zone de l'étude et répartition des sites de mesure pour l'été 2003 (étoiles) et l'hiver 2004 (cercles)

5.2 Méthodologie de sélection des données

Nous avons utilisé les valeurs de concentrations en benzène obtenues sur les campagnes de l'été 2003 et l'hiver 2004, présentant respectivement 218 et 215 données initiales et notés $J_{0,e}$ et $J_{0,h}$ pour constituer 2 jeux de référence pour chaque saison. Pour chaque jeu, nous avons écarté les points entraînant des discontinuités spatiales, notamment ceux présentant les plus fortes concentrations. Les jeux de données notés $J_{1,e}$ et $J_{1,h}$ sont constitués de 194 et 208 points. La totalité des données J_0 sera en revanche utilisée pour la définition de la carte de krigeage de référence. La figure suivante présente la répartition de ces jeux de données pour chaque saison.



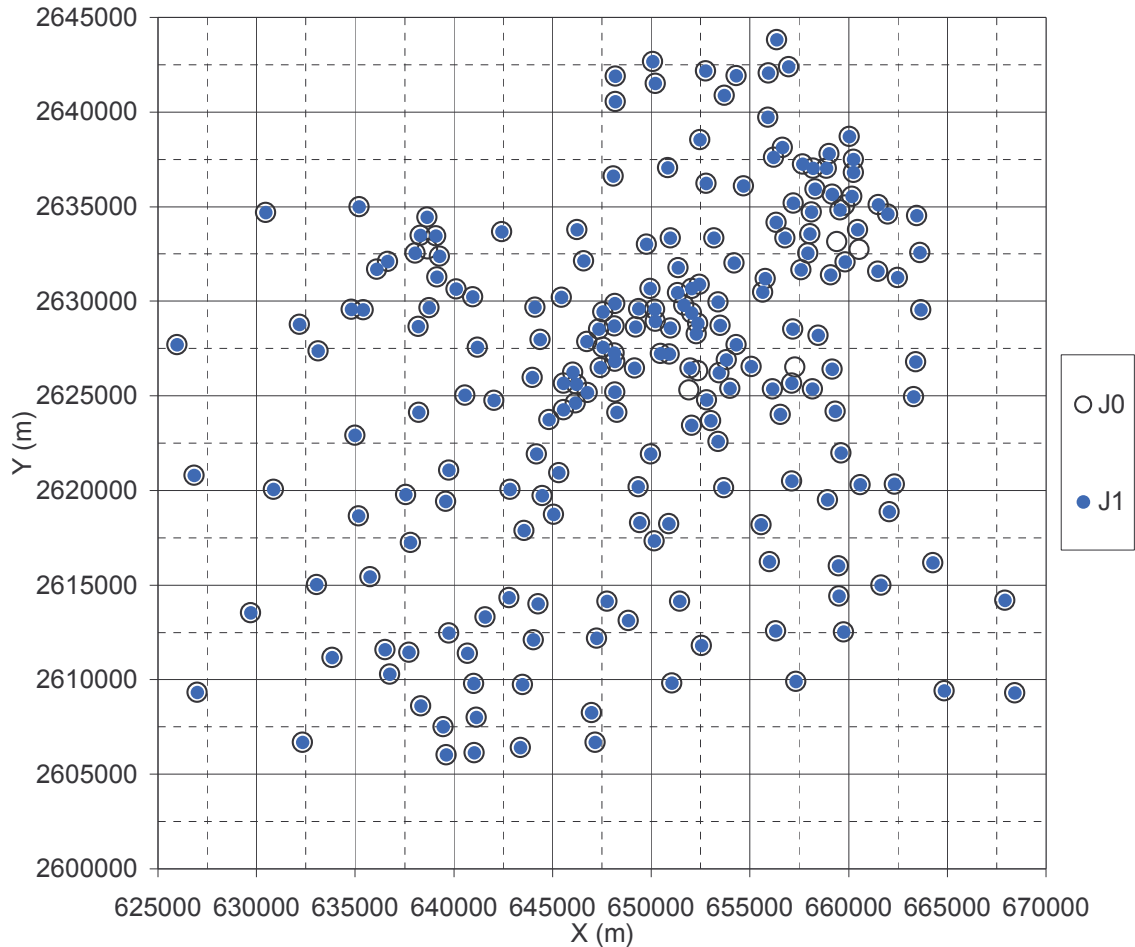


Figure 2 – Répartition des jeux de données J₀ et J₁ pour (haut) l'été 2003 et (bas) l'hiver 2004

Les données ont été sélectionnées selon trois maillages réguliers à partir des J_{1,i} (i = e,h) :

- mailles carrées de 3 km de côté (sélection S3i)
- mailles carrées de 5 km de côté (sélection S5i)
- mailles carrées de 7 km de côté (sélection S7i)

A chaque sélection correspond un jeu de données complémentaires noté C3i, C5i et C7i, respectivement. Pour constituer chaque sélection, le maillage a été calé sur le domaine d'étude et le point le plus proche du centre de la maille a été choisi. Les caractéristiques statistiques des différents jeux de données sont rassemblées dans les tableaux suivants.

Tableau 1 – Caractéristiques statistiques des jeux de données étudiés pour l'été 2003

	J _{0,e}	J _{1,e}	S3e	C3e	S5e	C5e	S7e	C7e
Nombre de données	218	194	100	94	53	141	35	159
Minimum	0,49	0,61	0,61	0,64	0,61	0,64	0,64	0,61
Percentile 25	0,81	0,79	0,76	0,88	0,71	0,82	0,77	0,82
Médiane	0,94	0,91	0,84	0,96	0,84	0,95	0,83	0,94
Percentile 75	1,10	1,05	1,02	1,06	0,96	1,07	1,02	1,07
Maximum	1,99	1,31	1,26	1,31	1,31	1,25	1,26	1,31
Etendue	1,50	0,70	0,65	0,67	0,70	0,61	0,62	0,70
Moyenne	0,98	0,92	0,88	0,97	0,86	0,94	0,87	0,93
Ecart-type	0,065	0,027	0,029	0,022	0,032	0,024	0,027	0,027
Variance	0,254	0,165	0,169	0,148	0,178	0,154	0,163	0,164
CV%	26,0	18,0	19,2	15,3	20,7	16,3	18,8	17,6

Tableau 2 – Caractéristiques statistiques des jeux de données étudiés pour l'hiver 2004

	J_{0,h}	J_{1,h}	S3h	C3h	S5h	C5h	S7h	C7h
Nombre de données	215	208	104	104	55	153	34	174
Minimum	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75
Percentile 25	0,99	0,99	0,94	1,05	0,95	1,01	0,93	0,99
Médiane	1,13	1,12	1,07	1,245	1,03	1,17	1,06	1,14
Percentile 75	1,37	1,33	1,21	1,46	1,19	1,38	1,24	1,37
Maximum	2,68	1,92	1,77	1,92	1,90	1,92	1,59	1,92
Etendue	1,93	1,17	1,02	1,17	1,15	1,17	0,84	1,17
Moyenne	1,22	1,18	1,09	1,28	1,09	1,22	1,08	1,20
Ecart-type	0,103	0,073	0,042	0,086	0,051	0,077	0,041	0,077
Variance	0,321	0,27	0,205	0,293	0,226	0,278	0,202	0,278
CV%	26,4	22,8	18,9	22,8	20,7	22,8	18,6	23,1

6. ETUDE DES JEUX DE DONNEES INITIAUX ET ELABORATION DES CARTES DE REFERENCE

6.1 Etude des variogrammes expérimentaux

Après sélection des données J₁, une anisotropie est observable selon un axe de 40° par rapport à la direction Nord pour la saison estivale. Elle ne correspond pas à la direction des vents observée sur la période de campagne (plutôt Est-Ouest en été), mais traduit plutôt une répartition inhomogène des points selon un axe privilégié. De plus, cette anisotropie n'est pas observée en hiver, où les vents sont dirigés Ouest et Sud-Ouest. On a donc choisi de conserver des variogrammes omnidirectionnels.

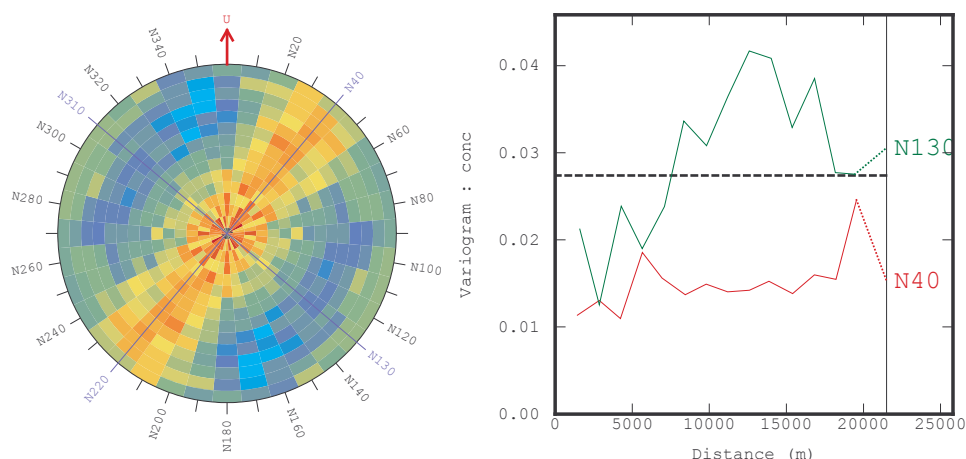


Figure 3 – Carte de variogramme pour le benzène estival

Le variogramme retenu pour l'été présente 15 pas de 1400 m chacun ; celui pour l'hiver 12 pas de 1800 m chacun. Aucune donnée supplémentaire n'a été écartée des jeux de données J₁.

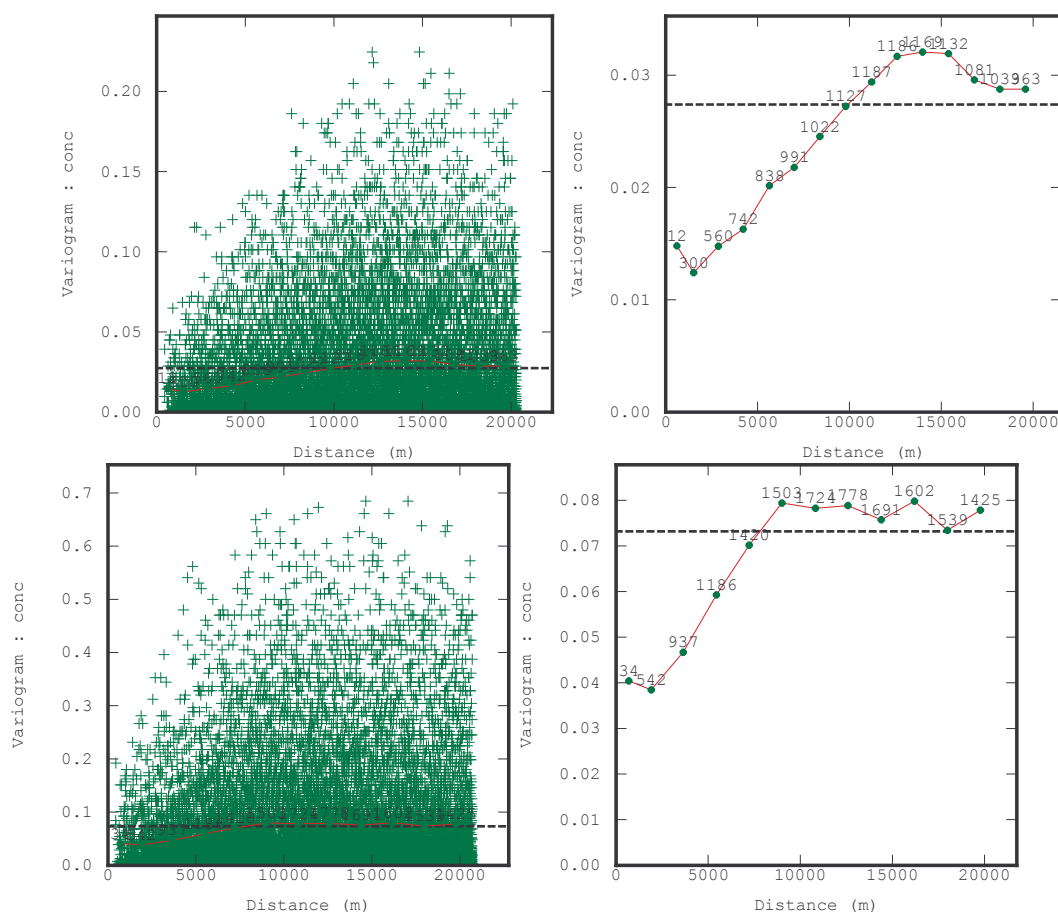


Figure 4 – Variogrammes expérimentaux (haut) estival (194 sites ; 1400 m ; 15 pas) et (bas) hivernal (208 sites ; 12 pas ; 1800 m)

6.2 Modélisation du variogramme et validation croisée

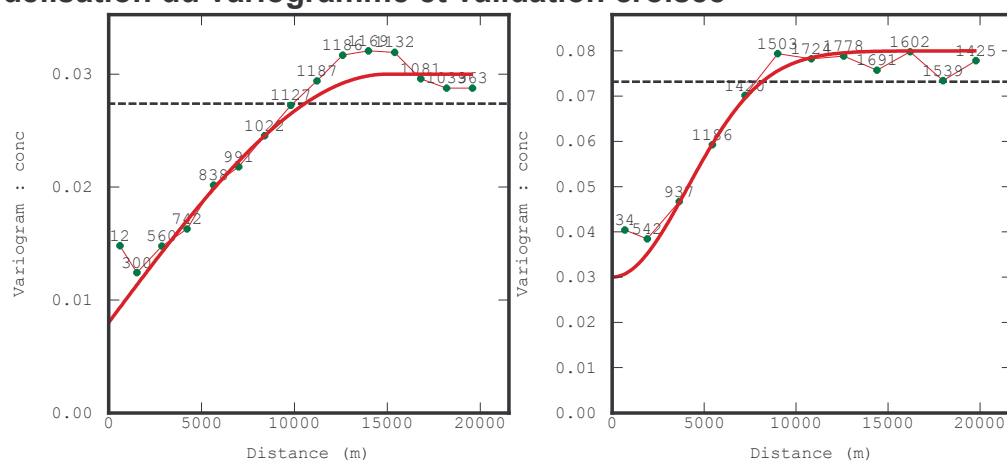


Figure 5 – Variogrammes simulés pour (gauche) l'été : pépite (0,008) + sphérique (15 km ; 0,022) et (droite) l'hiver : pépite (0,03) + gaussien (10 km ; 0,05)

On a considéré un voisinage circulaire de 20 km (1 seul secteur, 30 points). Ce voisinage sera celui utilisé par la suite lors de l'étude de dégradation du maillage.

6.2.1 VALIDATION CROISEE SUR LE JEU DE DONNEES $J_{1,E}$

Tableau 3 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur le jeu de données $J_{1,e}$

$J_{1,e}$	Moyenne	Variance
Erreur	-0,00146	0,01291
Erreur standardisée	-0,01595	1,00166

Toutes les données sont robustes.

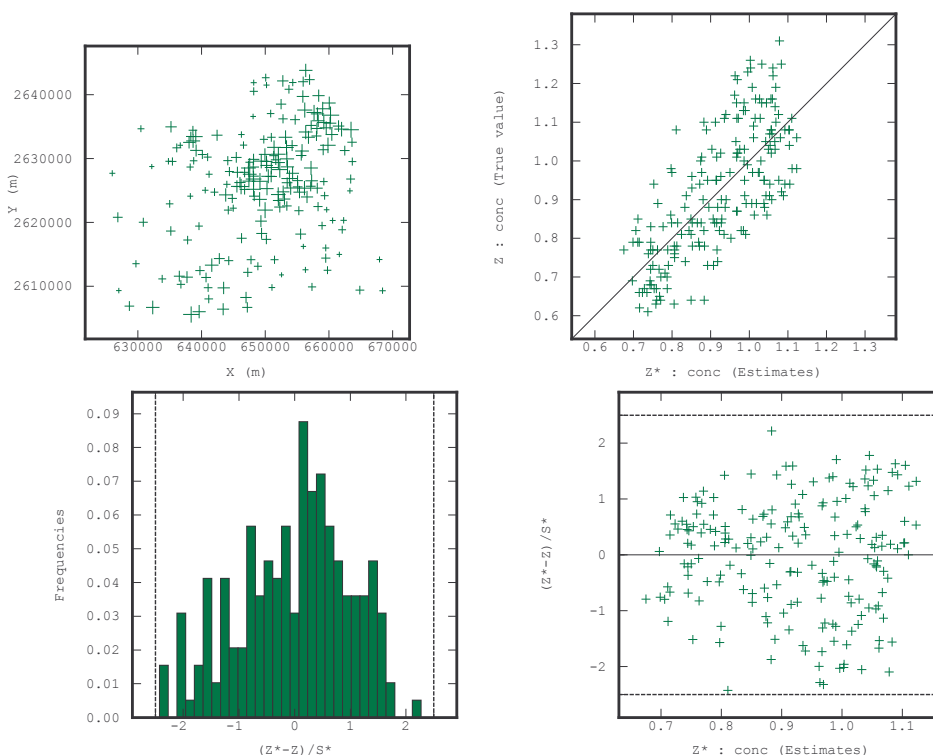


Figure 6 – Résultats de la validation croisée pour le jeu de données $J_{1,e}$

La moyenne de l'erreur est proche de 0 et la variance de l'erreur standardisée est proche de 1 : la plupart des données sont correctement réestimées. La corrélation entre concentrations estimées et observées est de : $r = 0,727$.

6.2.2 VALIDATION CROISEE SUR LE JEU DE DONNEES $J_{1,H}$

Tableau 4 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur le jeu de données $J_{1,h}$

$J_{1,h}$ (208 points)	Moyenne	Variance
Erreur	-0,00545	0,03200
Erreur standardisée	-0,02812	0,86232

1 donnée sur 208 non robuste.

$J_{1,h}$ (207 points)	Moyenne	Variance
Erreur	-0,00312	0,03104
Erreur standardisée	-0,01528	0,83215

Toutes les données sont robustes.

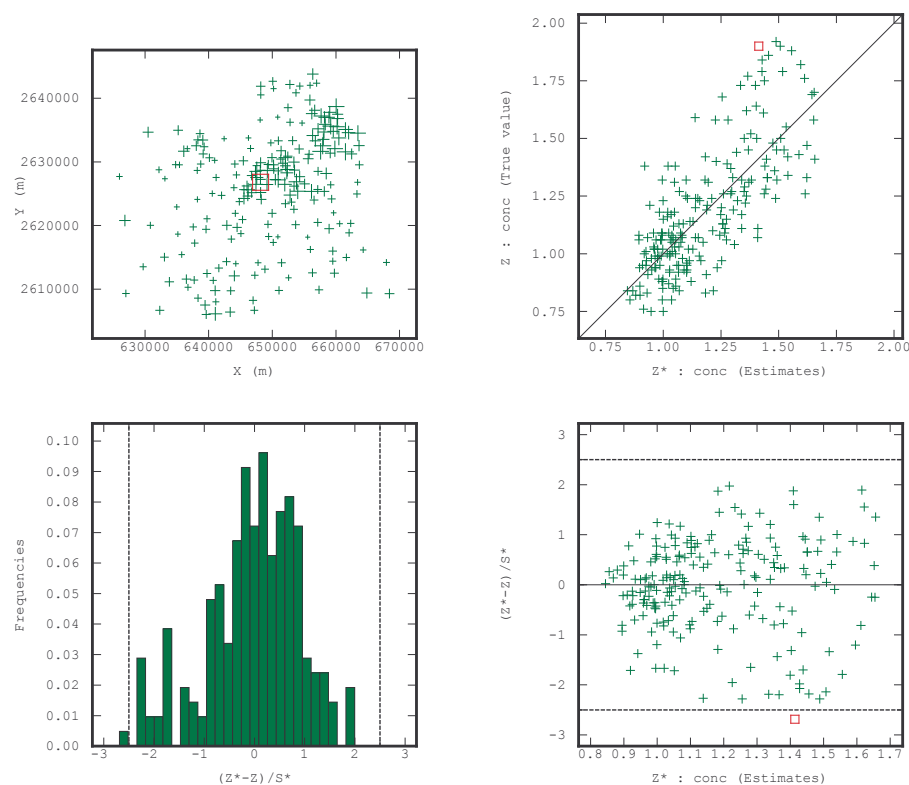


Figure 7 – Résultats de la validation croisée pour le jeu de données $J_{1,h}$

La corrélation entre concentrations estimées et observées est de : $r = 0,750$.

6.3 Etablissement de la carte de référence

Pour établir la carte de référence, les concentrations en benzène ont été estimées aux nœuds d'une grille régulière de 250 m de côté. Les caractéristiques de cette maille d'estimation sont les suivantes :

$X_{\min} = 623000 \text{ m}$	$X_{\max} = 672750 \text{ m}$	$\Delta X = 250 \text{ m}$	200 nœuds
$Y_{\min} = 2604000 \text{ m}$	$Y_{\max} = 2645000 \text{ m}$	$\Delta Y = 250 \text{ m}$	165 nœuds

On a considéré le même voisinage que celui utilisé pour la validation croisée, et utilisé les modèles de référence établis avec les jeux de données J_1 pour kriger les jeux de données J_0 .

6.3.1 KRIGEAGE DES CONCENTRATIONS POUR LE JEU DE DONNEES $J_{0,E}$

Tableau 5 – Statistiques sur l'estimation sur une grille régulière de 250 m de côté, par krigeage ordinaire ponctuel pour l'été 2003

$J_{0,e}$	Mesures ($\mu\text{g m}^{-3}$)	Estimations par krigeage ponctuel ($\mu\text{g m}^{-3}$)	Ecart-types de krigeage ($\mu\text{g m}^{-3}$)
Nombre de données	218	25467	25467
Minimum	0,49	0,59	0,10
Percentile 25	0,81	0,74	0,11
Médiane	0,94	0,81	0,12
Percentile 75	1,10	0,93	0,13
Maximum	1,99	1,47	0,23
Etendue	1,50	0,88	0,13
Moyenne	0,98	0,85	0,13
Ecart-type	0,065	0,15	0,02
Variance	0,254	0,02	0,00
CV%	26,0	17,0	17,0

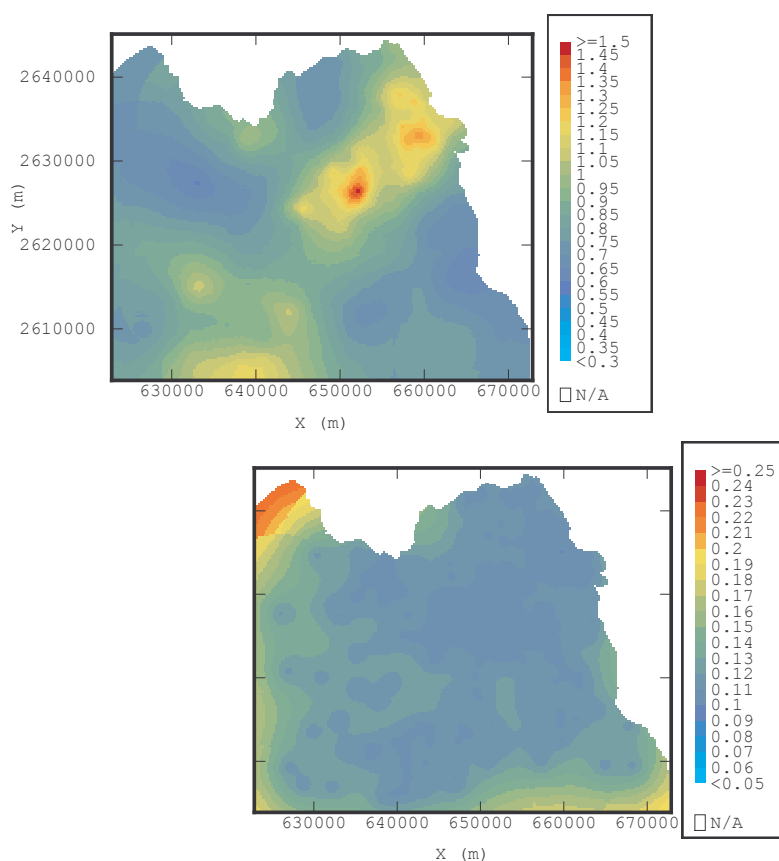


Figure 8 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données $J_{0,e}$, avec le modèle de référence et (droite) écarts-types de krigeage

La carte de référence estivale fait bien apparaître les zones de plus forte concentration que sont les agglomérations de Lille, Roubaix et Tourcoing. La zone de concentration moyenne au sud du domaine peut s'expliquer par la proximité d'axes routiers.

6.3.2 KRIGEAGE DES CONCENTRATIONS POUR LE JEU DE DONNEES $J_{0,h}$

Tableau 6 – Statistiques sur l'estimation sur une grille régulière de 250 m de côté, par krigeage ordinaire ponctuel pour l'hiver 2004

$J_{0,h}$	Mesures ($\mu\text{g m}^{-3}$)	Estimations par krigeage ponctuel ($\mu\text{g m}^{-3}$)	Ecart-types de krigeage ($\mu\text{g m}^{-3}$)
Nombre de données	215	25467	25467
Minimum	0,75	0,82	0,18
Percentile 25	0,99	0,98	0,19
Médiane	1,13	1,05	0,20
Percentile 75	1,37	1,17	0,23
Maximum	2,68	1,86	0,39
Etendue	1,93	1,04	0,21
Moyenne	1,22	1,09	0,22
Ecart-type	0,103	0,17	0,04
Variance	0,321	0,03	0,00
CV%	26,4	15,0	18,0

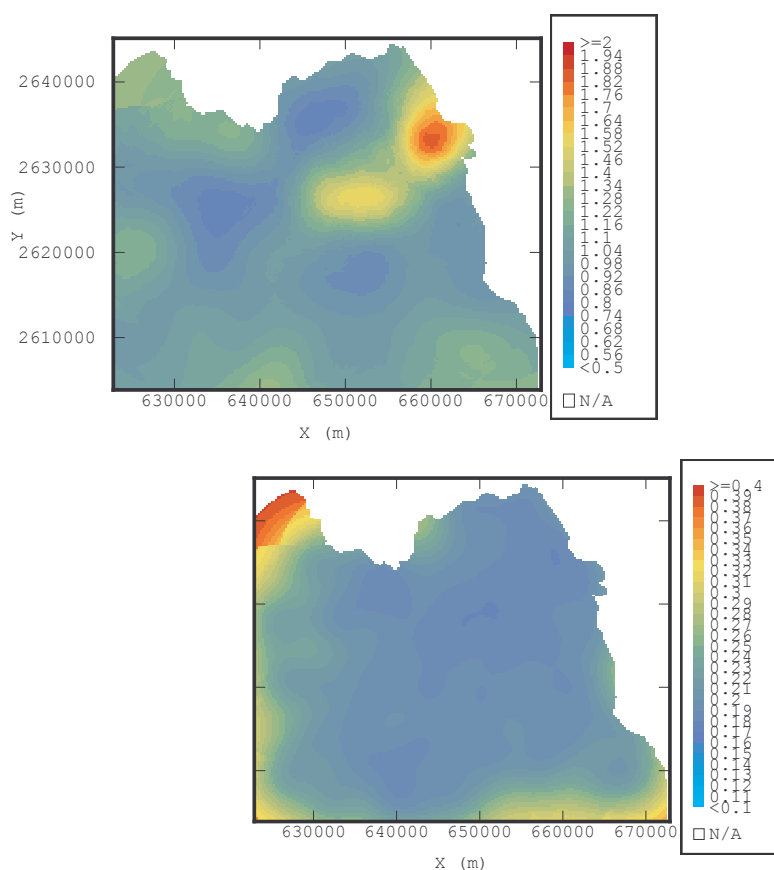


Figure 9 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données $J_{0,h}$, avec le modèle de référence et (droite) écarts-types de krigeage

Les valeurs sont plus fortes qu'en été et le Nord-Est de la zone urbaine semble le plus soumis aux fortes concentrations de benzène. A partir de ces jeux de référence, l'impact de la dégradation du maillage a été étudié sur des maillages carrés réguliers.

7. ETUDE DE LA SELECTION DE DONNEES S3 (MAILLE DE 3 × 3 KM)

Les sélections S3e et S3h contiennent 100 et 104 données, respectivement, correspondant à 1 point de mesure par maille carrée de 3 km de côté.

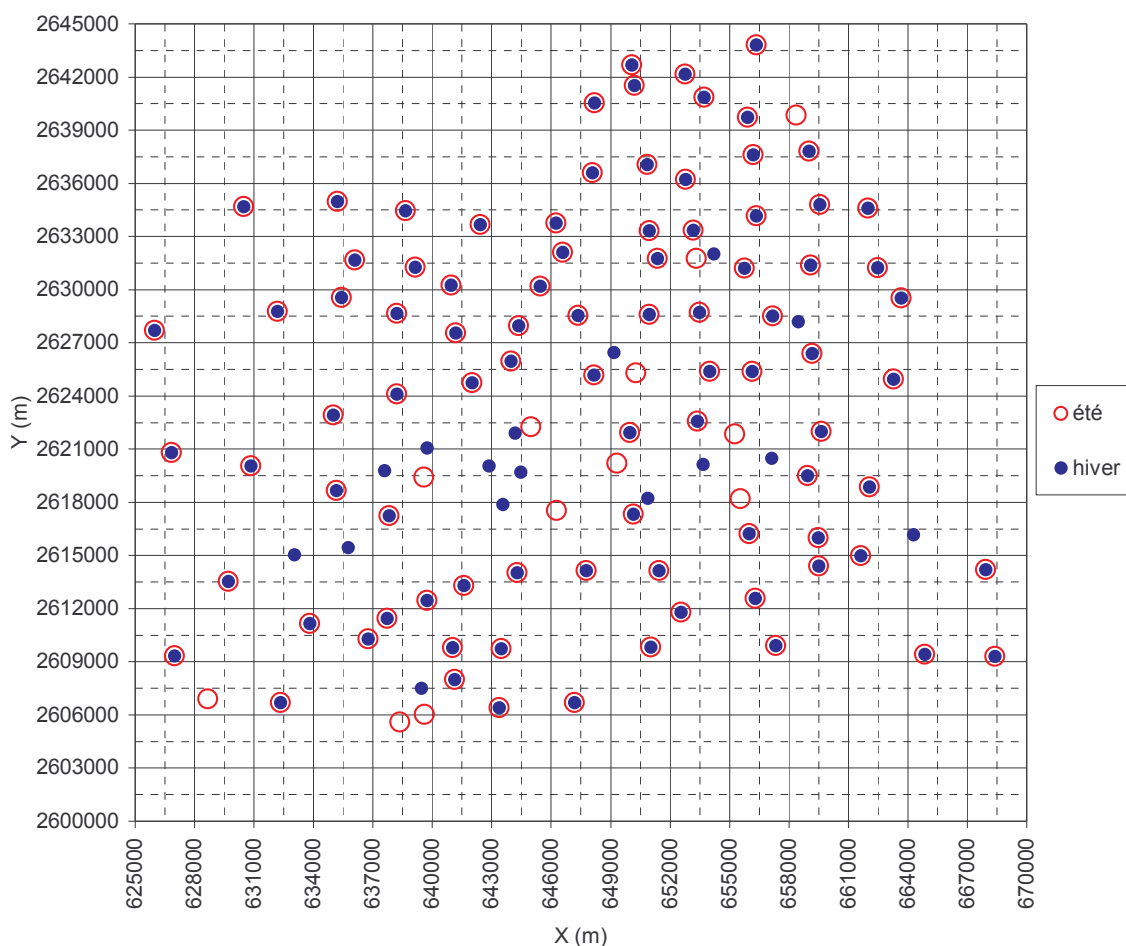


Figure 10 - Répartition des jeux de données S3e et S3h

7.1 Etude du variogramme expérimental

Les variogrammes choisis sont omnidirectionnels et comprennent pour l'été 20 pas de 1500 m chacun, et pour l'hiver 12 pas de 1800 m chacun. Aucun point n'a été éliminé.

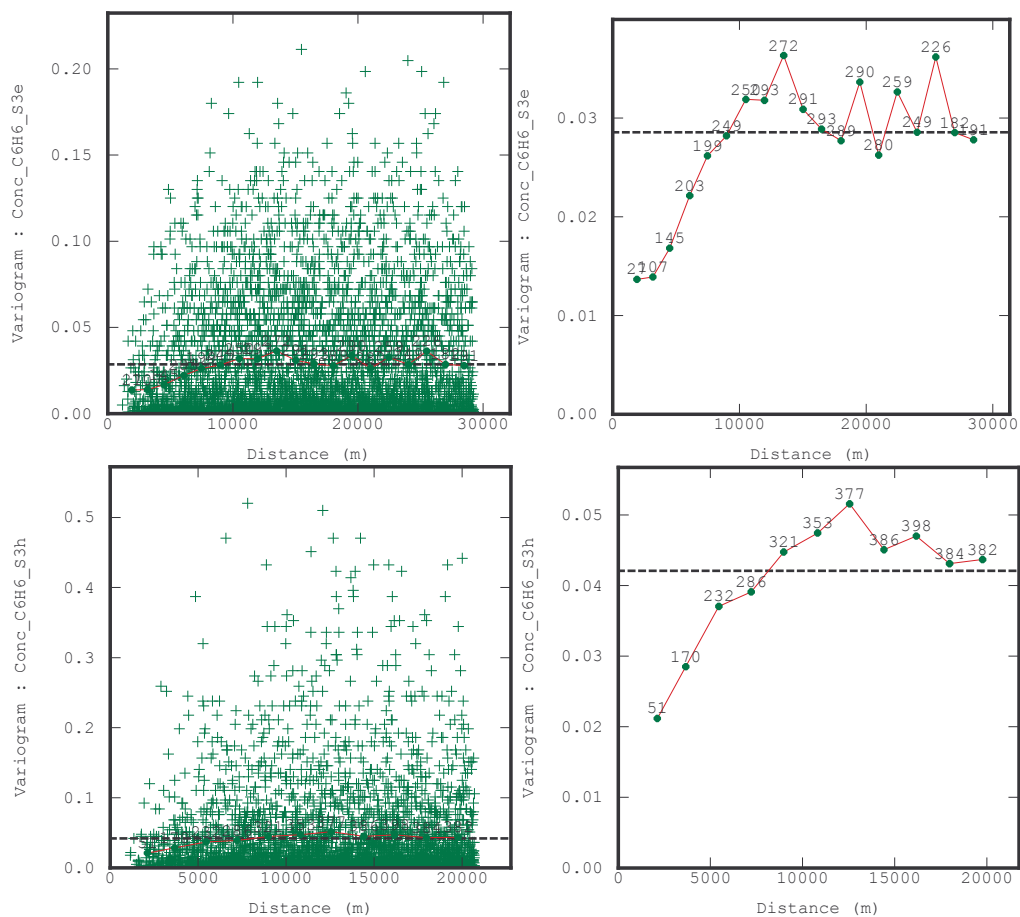


Figure 11 – Variogrammes expérimentaux (haut) estival (100 sites ; 20 pas ; 1500 m) et (bas) hivernal (104 sites ; 12 pas ; 1800 m)

7.2 Modélisation du variogramme et test de validation croisée

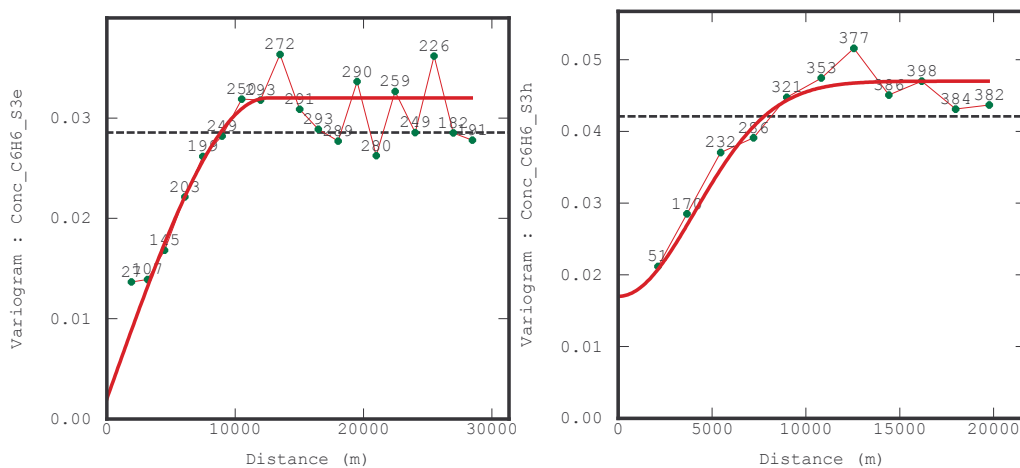


Figure 12 – Variogrammes simulés pour (gauche) l'été : pépite (0,002) + sphérique (12,5 km ; 0,03) et (droite) l'hiver : pépite (0,017) + gaussien (10 km ; 0,03)

7.2.1 VALIDATION CROISEE SUR LE JEU DE DONNEES S3E

Tableau 7 – Statistiques sur l’erreur d’estimation, obtenues par validation croisée sur le jeu de données S3e

S3e	Moyenne	Variance
Erreur	-0,00047	0,01295
Erreur standardisée	-0,00446	1,09039

Toutes les données sont robustes.

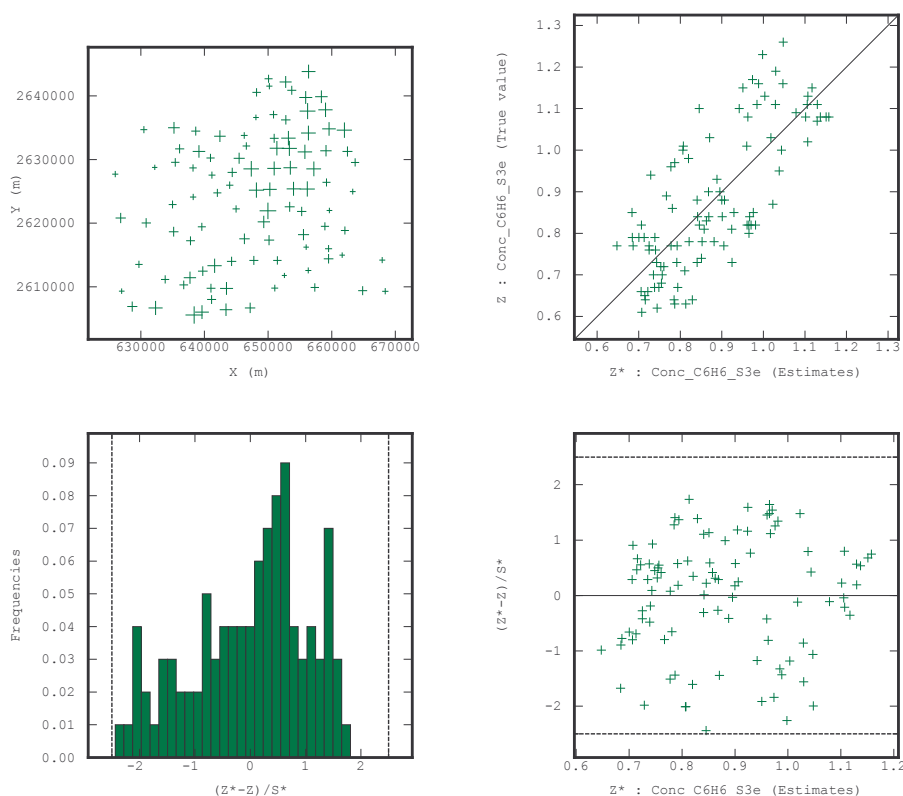


Figure 13 – Résultats de la validation croisée pour le jeu de données S3e

La corrélation entre valeurs estimées et mesurées est : $r = 0,741$.

7.2.2 VALIDATION CROISEE SUR LE JEU DE DONNEES S3H

Tableau 8 – Statistiques sur l’erreur d’estimation, obtenues par validation croisée sur le jeu de données S3h

S3h (104 points)	Moyenne	Variance
Erreur	-0,00347	0,02639
Erreur standardisée	-0,01652	1,00707

2 données sur 104 non robustes.

S3h (102 points)	Moyenne	Variance
Erreur	0,00472	0,02341
Erreur standardisée	0,03660	0,87974

Toutes les données sont robustes.

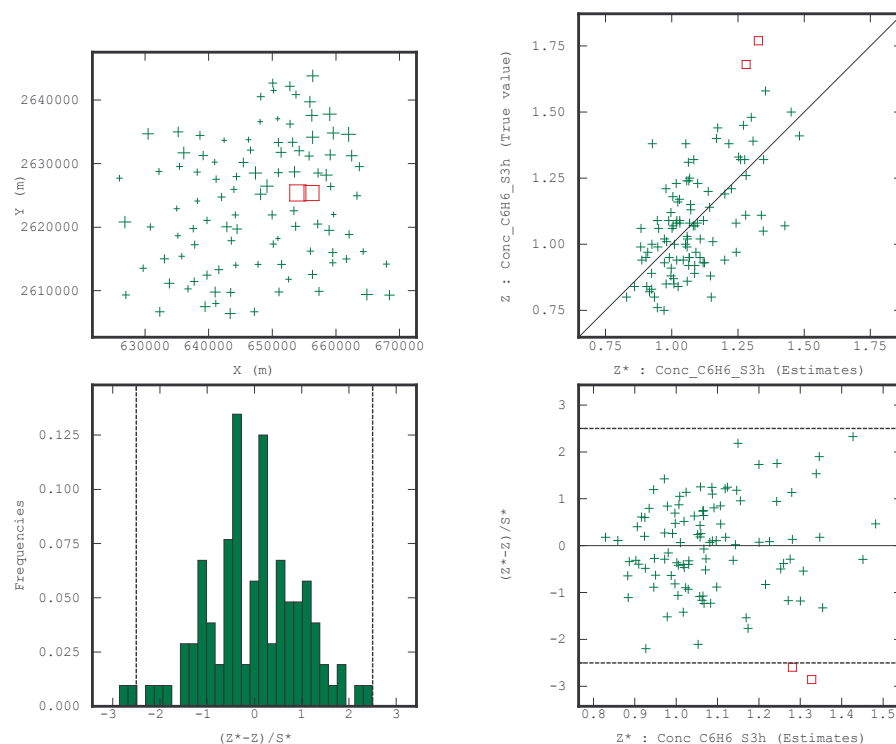


Figure 14 – Résultats de la validation croisée pour le jeu de données S3h

Les deux points non robustes présentent les plus fortes concentrations de la sélection. Le coefficient de corrélation entre valeurs observées et estimées est : $r = 0,614$.

7.3 Test de validation absolue sur le complémentaire C3

Les concentrations de benzène sont estimées aux points n'ayant pas été sélectionnés dans les jeux de données S3, puis comparées aux valeurs réelles.

7.3.1 VALIDATION ABSOLUE SUR LE JEU DE DONNEES C3E

Tableau 9 – Résultats de la validation absolue sur le complémentaire C3e pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S3e

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S3e	
C3e	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	94	94	94
Minimum	0,64	0,70	0,08
Percentile 25	0,88	0,89	0,09
Médiane	0,96	1,04	0,09
Percentile 75	1,06	1,11	0,10
Maximum	1,31	1,14	0,11
Etendue	0,67	0,44	0,03
Moyenne	0,97	0,99	0,09
Ecart-type	0,022	0,13	0,01
Variance	0,148	0,02	0,00
CV%	15	13	6
Corrélation mesures/estimations		$r = 0,637$	

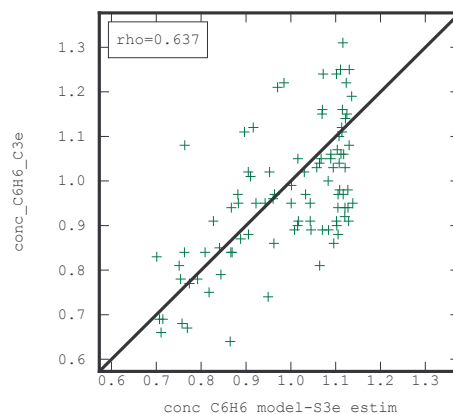


Figure 15 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C3e

7.3.2 VALIDATION ABSOLUE SUR LE JEU DE DONNEES C3H

Tableau 10 – Résultats de la validation absolue sur le complémentaire C3h pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S3h

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S3h	
C3h	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	104	104	104
Minimum	0,75	0,90	0,14
Percentile 25	1,05	1,06	0,15
Médiane	1,245	1,23	0,15
Percentile 75	1,46	1,35	0,15
Maximum	1,92	1,50	0,17
Etendue	1,17	0,60	0,03
Moyenne	1,28	1,22	0,15
Ecart-type	0,086	0,17	0,00
Variance	0,293	0,03	0,00
CV%	23	14	3
Corrélation mesures/estimations		$r = 0,686$	

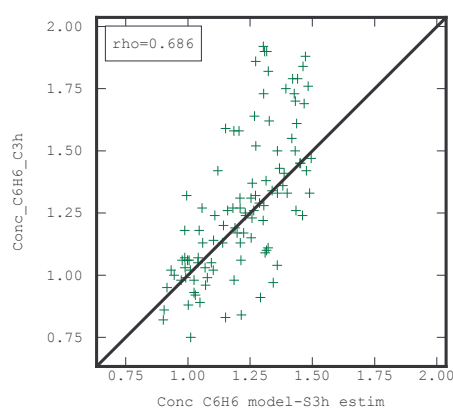


Figure 16 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C3h

7.4 Etablissement de la carte pour la sélection S3

Les cartes de krigeage correspondante ont été établies.

7.4.1 KRIGEAGE DES CONCENTRATIONS POUR LE JEU DE DONNEES S3E

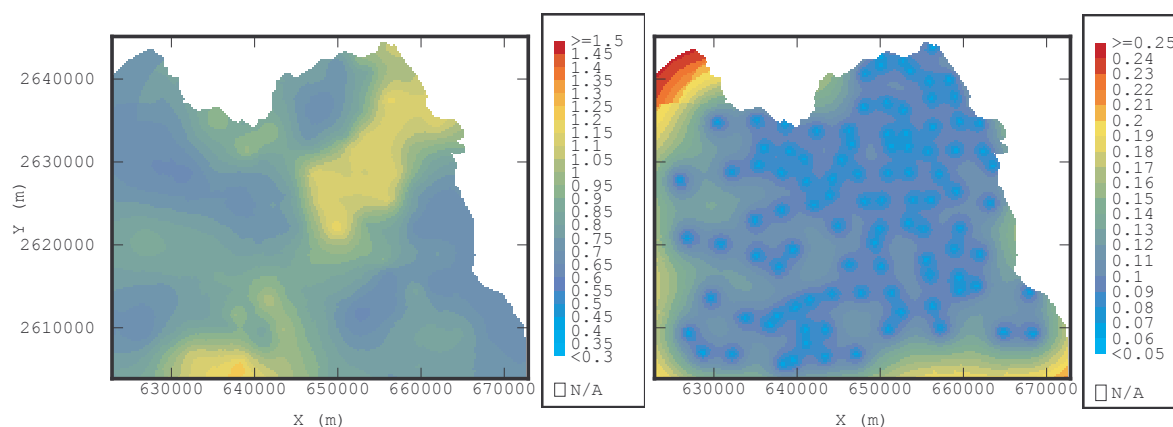


Figure 17 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S3e et (droite) écarts-types de krigeage

7.4.2 KRIGEAGE DES CONCENTRATIONS POUR LE JEU DE DONNEES S3H

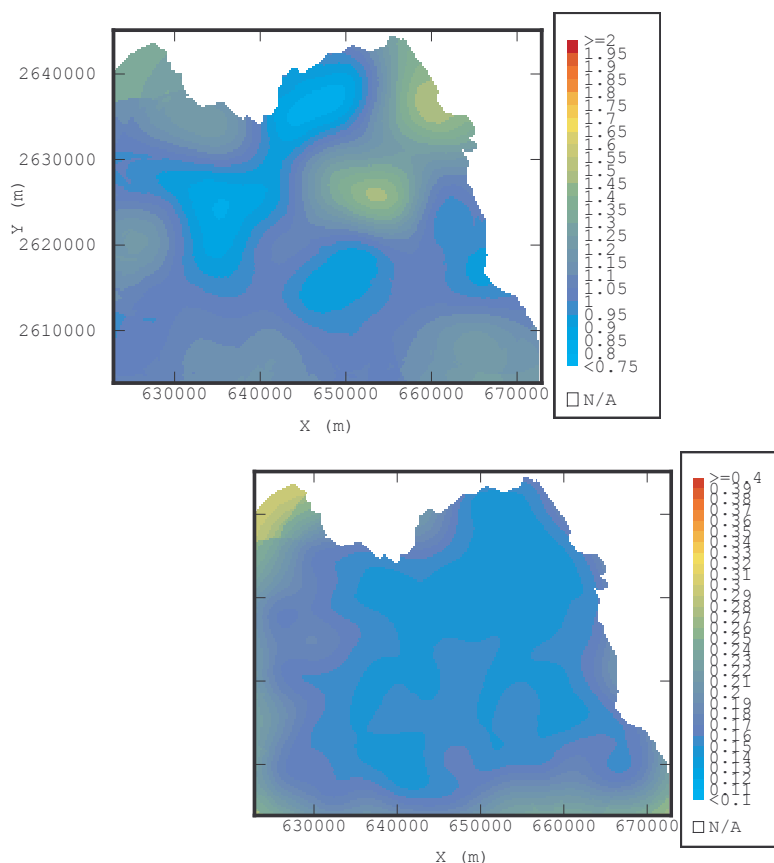


Figure 18 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S3h et (droite) écarts-types de krigeage

8. ETUDE DE LA SELECTION DE DONNEES S5 (MAILLE DE 5 × 5 KM)

Les sélections S5e et S5h contiennent 53 et 55 données, respectivement, correspondant à 1 point de mesure par maille carrée de 5 km de côté.

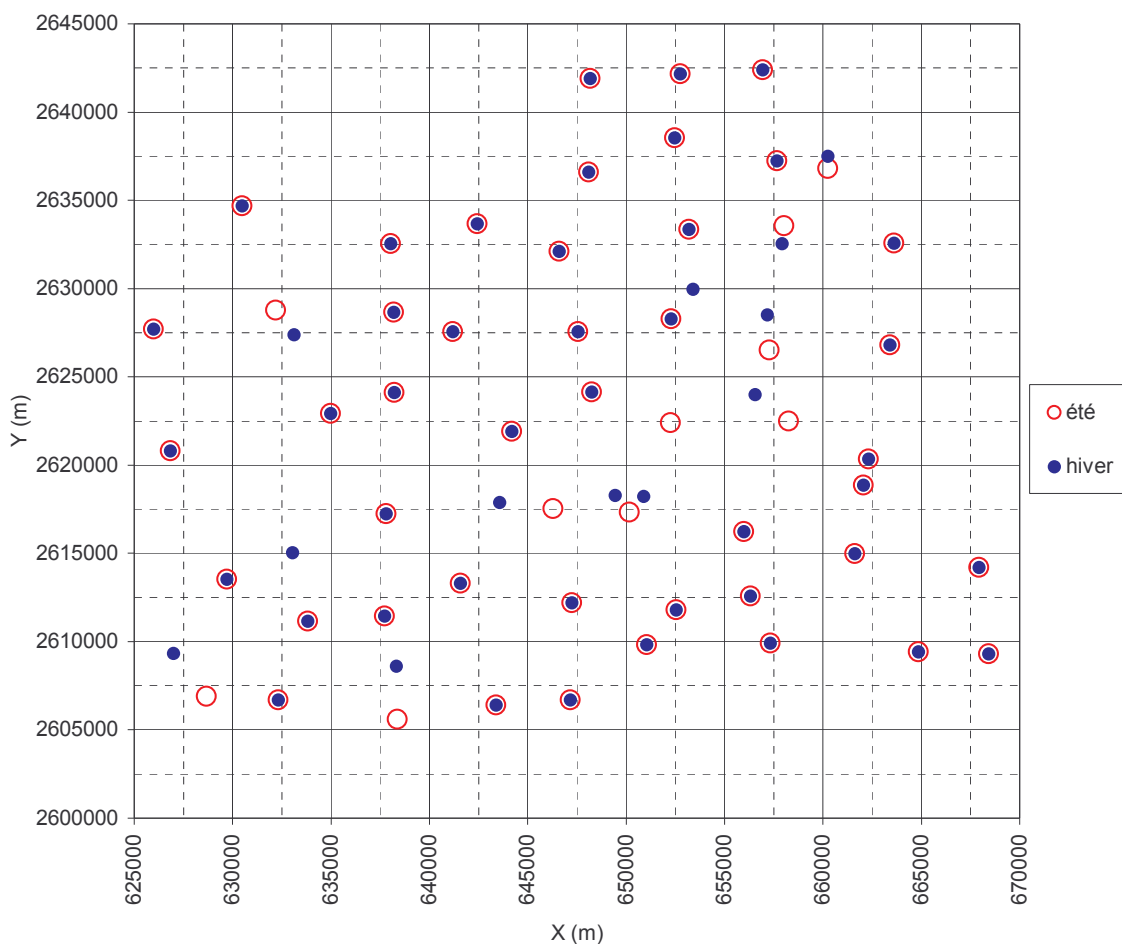


Figure 19 - Répartition des jeux de données S5e et S5h

8.1 Etude du variogramme expérimental

Les variogrammes choisis sont omnidirectionnels et comprennent pour l'été 19 pas de 1300 m chacun, et pour l'hiver 16 pas de 1300 m chacun. Aucun point n'a été éliminé.

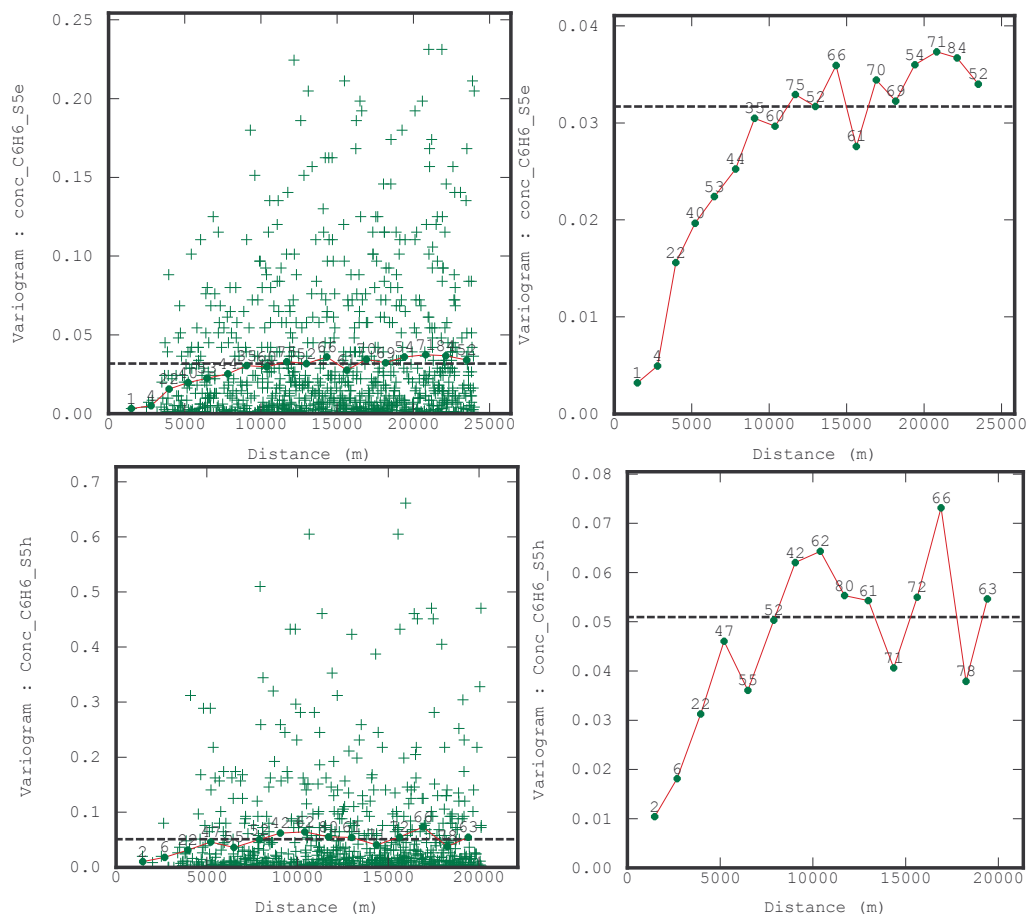


Figure 20 – Variogrammes expérimentaux (haut) estival (53 sites ; 19 pas ; 1300 m) et (bas) hivernal (55 sites ; 16 pas ; 1300 m)

8.2 Modélisation du variogramme et test de validation croisée

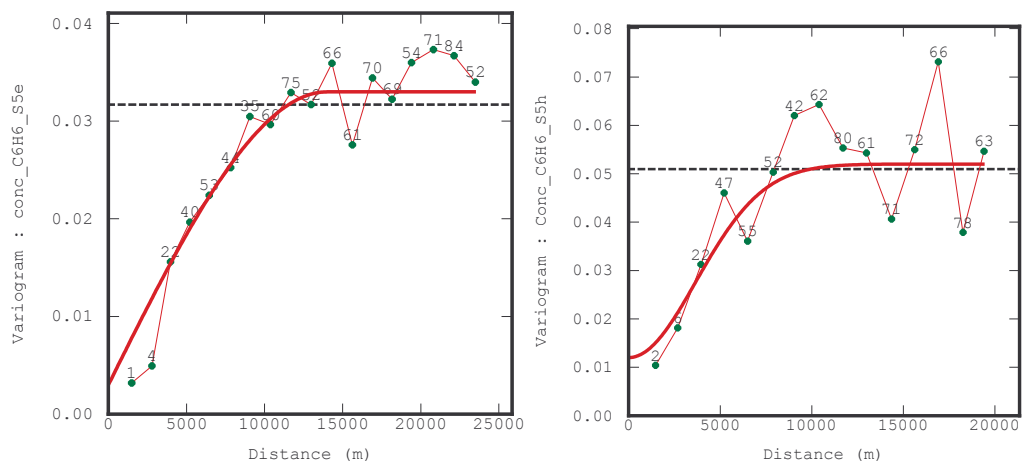


Figure 21 – Variogrammes simulés pour (gauche) l'été : pépite (0,003) + sphérique (14 km ; 0,03) et (droite) l'hiver : pépite (0,012) + gaussien (9 km ; 0,04)

8.2.1 VALIDATION CROISEE SUR LE JEU DE DONNEES S5E

Tableau 11 – Statistiques sur l’erreur d’estimation, obtenues par validation croisée sur le jeu de données S5e

S5e	Moyenne	Variance
Erreur	0,00255	0,01571
Erreur standardisée	0,01119	0,86449

Toutes les données sont robustes.

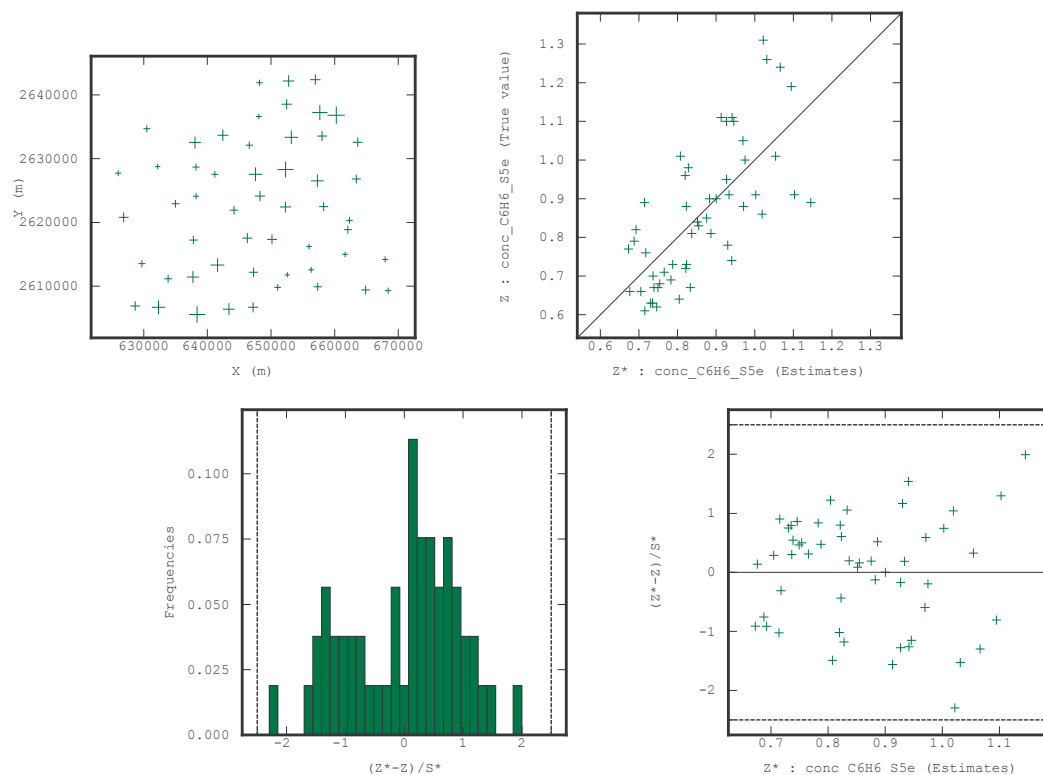


Figure 22 – Résultats de la validation croisée pour le jeu de données S5e

La corrélation obtenue a pour coefficient : $r = 0,710$.

8.2.2 VALIDATION CROISEE SUR LE JEU DE DONNEES S5H

Tableau 12 – Statistiques sur l’erreur d’estimation, obtenues par validation croisée sur le jeu de données S5h

S5h (55 points)	Moyenne	Variance
Erreur	0,00073	0,03903
Erreur standardisée	0,00036	1,04591

1 donnée sur 55 non robuste.

S5h (54 points)	Moyenne	Variance
Erreur	0,01144	0,03346
Erreur standardisée	0,05888	0,87692

Toutes les données sont robustes.

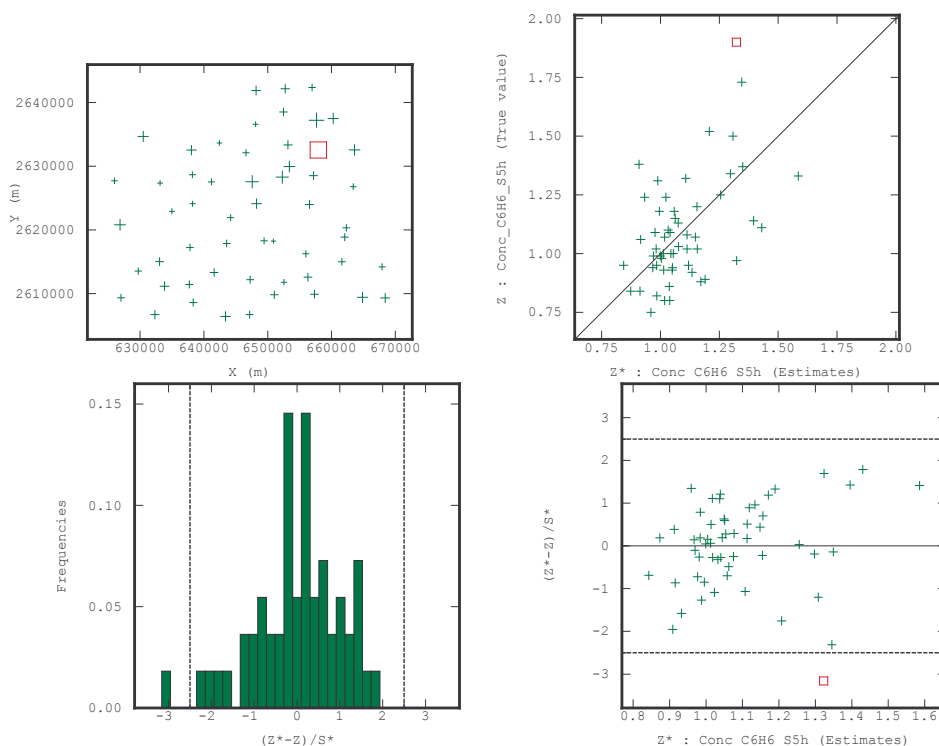


Figure 23 – Résultats de la validation croisée pour le jeu de données S5h

Le point non robuste présente la plus forte concentration de la sélection. Le coefficient de corrélation entre valeurs observées et estimées est : $r = 0,510$.

8.3 Test de validation absolue sur le complémentaire C5

8.3.1 VALIDATION ABSOLUE SUR LE JEU DE DONNEES C5E

Tableau 13 – Résultats de la validation absolue sur le complémentaire C5e pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S5e

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S5e	
C5e	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	141	141	141
Minimum	0,64	0,67	0,09
Percentile 25	0,82	0,86	0,10
Médiane	0,95	0,98	0,11
Percentile 75	1,07	1,06	0,11
Maximum	1,25	1,20	0,14
Etendue	0,61	0,53	0,05
Moyenne	0,94	0,96	0,11
Ecart-type	0,024	0,13	0,01
Variance	0,154	0,02	0,00
CV%	16	13	8
Corrélation mesures/estimations		$r = 0,642$	

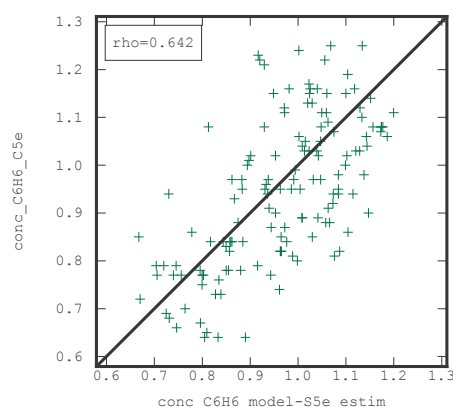


Figure 24 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C5e

8.3.2 VALIDATION ABSOLUE SUR LE JEU DE DONNEES C5H

Tableau 14 – Résultats de la validation absolue sur le complémentaire C5h pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S5h

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S5h	
C5h	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	153	153	153
Minimum	0,75	0,84	0,13
Percentile 25	1,01	1,04	0,14
Médiane	1,17	1,14	0,15
Percentile 75	1,38	1,38	0,16
Maximum	1,92	1,76	0,19
Etendue	1,17	0,92	0,06
Moyenne	1,22	1,21	0,15
Ecart-type	0,077	0,22	0,01
Variance	0,278	0,05	0,00
CV%	23	18	7
Corrélation mesures/estimations		$r = 0,707$	

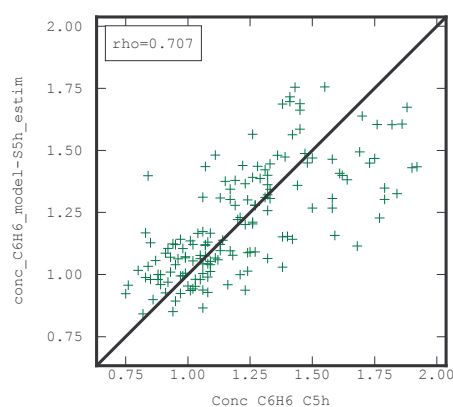


Figure 25 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C5h

8.4 Etablissement de la carte pour la sélection S5

8.4.1 KRIGEAGE DES CONCENTRATIONS POUR LE JEU DE DONNEES S5E

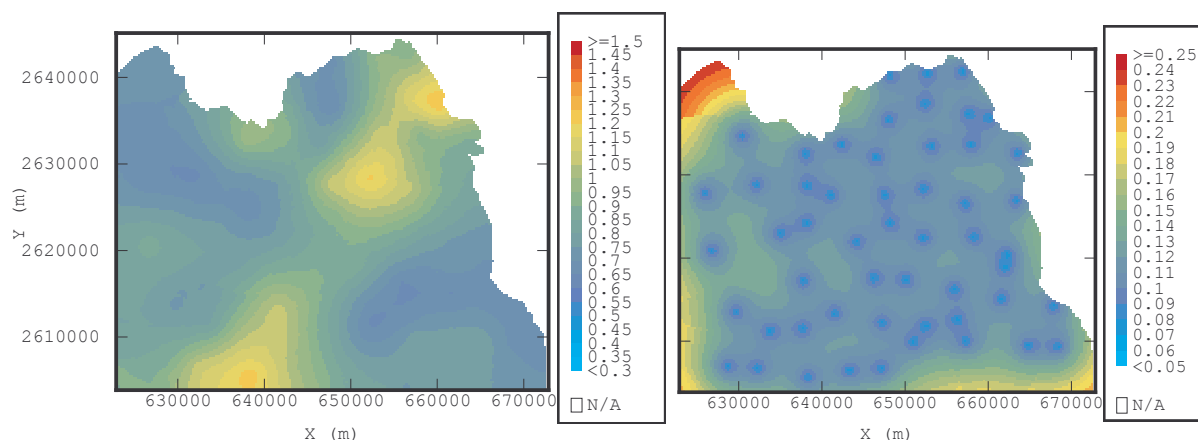


Figure 26 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S5e et (droite) écarts-types de krigeage

8.4.2 KRIGEAGE DES CONCENTRATIONS POUR LE JEU DE DONNEES S5H

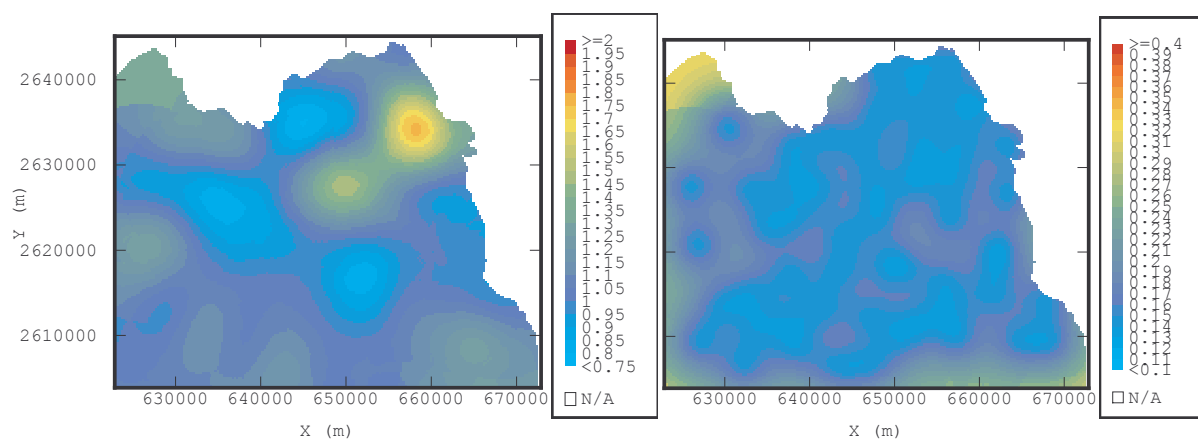


Figure 27 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S5h et (droite) écarts-types de krigeage

9. ETUDE DE LA SELECTION DE DONNEES S7 (MAILLE DE 7 × 7 KM)

Les sélections S7e et S7h contiennent 35 et 34 données, respectivement, correspondant à 1 point de mesure par maille carrée de 7 km de côté.

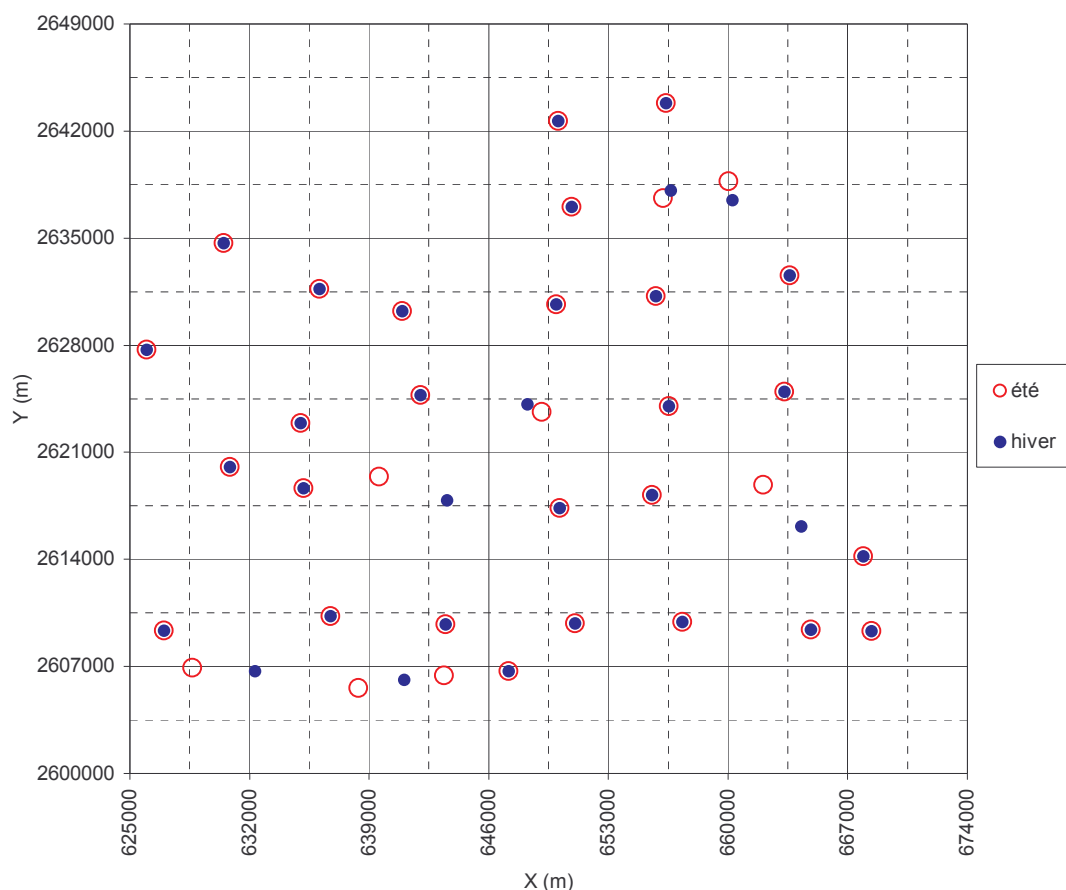
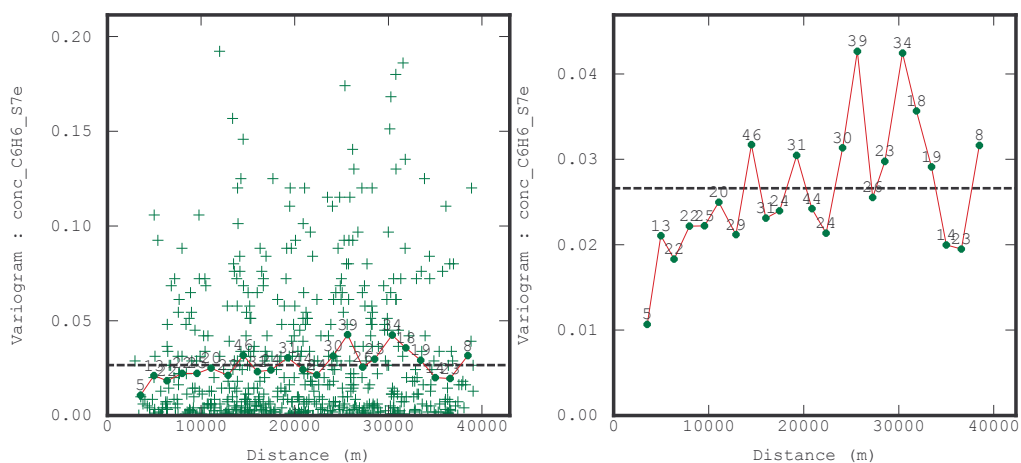


Figure 28 - Répartition des jeux de données S7e et S7h

9.1 Etude du variogramme expérimental

Les variogrammes choisis sont omnidirectionnels et comprennent pour l'été 25 pas de 1600 m chacun, et pour l'hiver 16 pas de 1800 m chacun. Aucun point n'a été éliminé de la sélection estivale ; 3 points ont été écartés de la sélection hivernale.



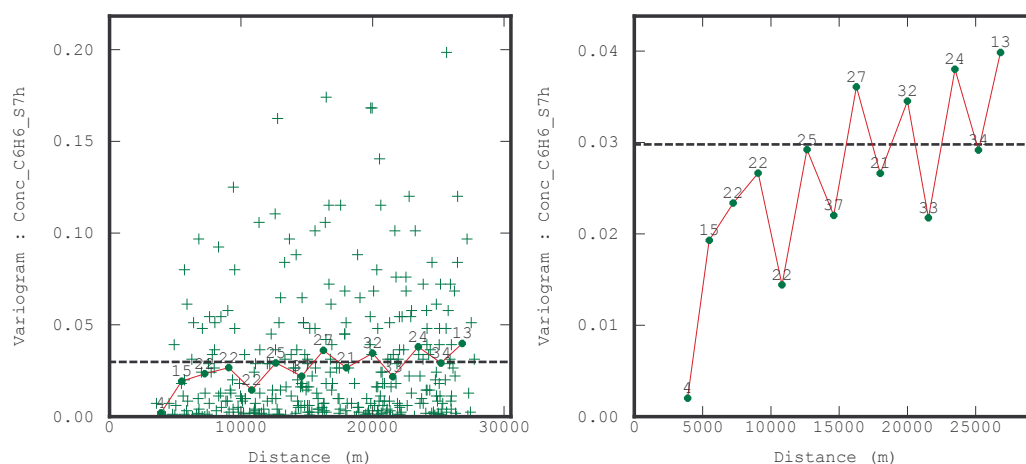


Figure 29 – Variogrammes expérimentaux (haut) estival (35 sites ; 25 pas ; 1600 m) et (bas) hivernal (31 sites ; 16 pas ; 1800 m)

9.2 Modélisation du variogramme et test de validation croisée

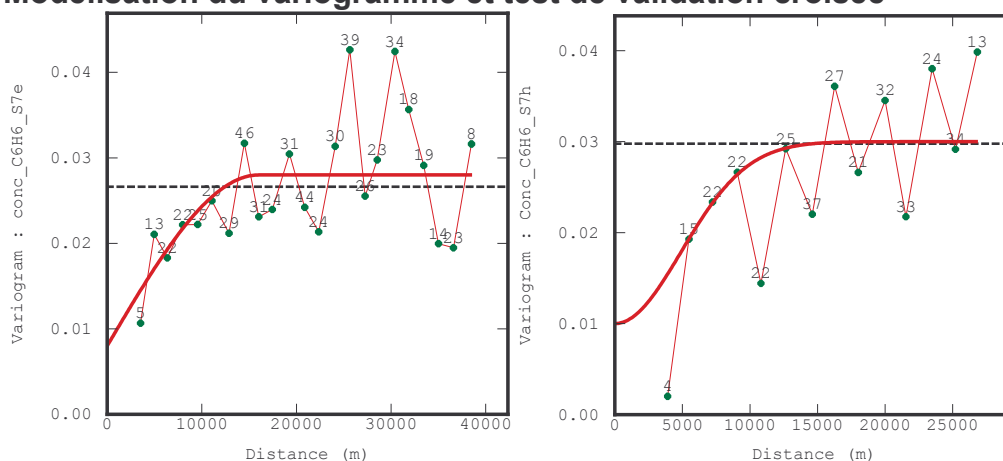


Figure 30 – Variogrammes simulés pour (gauche) l'été : pépite (0,008) + sphérique (16 km ; 0,02) et (droite) l'hiver : pépite (0,01) + gaussien (12 km ; 0,02)

9.2.1 VALIDATION CROISEE SUR LE JEU DE DONNEES S7E

Tableau 15 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur le jeu de données S7e

S7e	Moyenne	Variance
Erreur	0,00706	0,01942
Erreur standardisée	0,03068	0,87716

Toutes les données sont robustes.

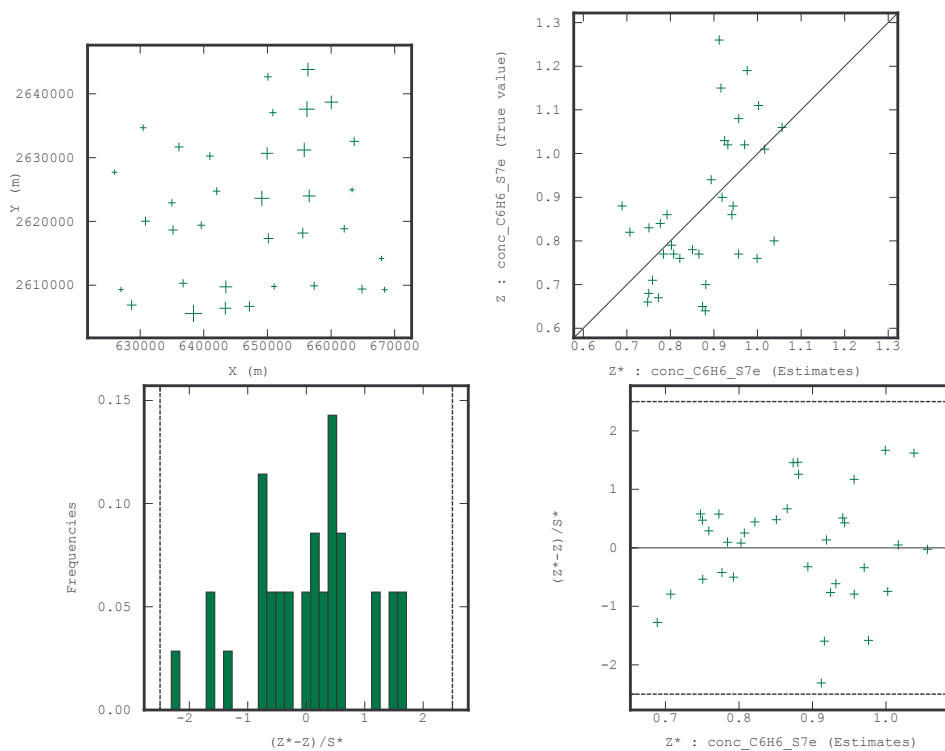


Figure 31 – Résultats de la validation croisée pour le jeu de données S7e

Le coefficient de corrélation est de : $r = 0,525$.

9.2.2 VALIDATION CROISEE SUR LE JEU DE DONNEES S7H

Tableau 16 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur le jeu de données S7h

S7h (34 points)	Moyenne	Variance
Erreur	0,00097	0,03938
Erreur standardisée	-0,00021	1,66275

2 données sur 34 non robustes.

S7h (32 points)	Moyenne	Variance
Erreur	0,00550	0,02656
Erreur standardisée	0,02943	1,09543

Toutes les données sont robustes.

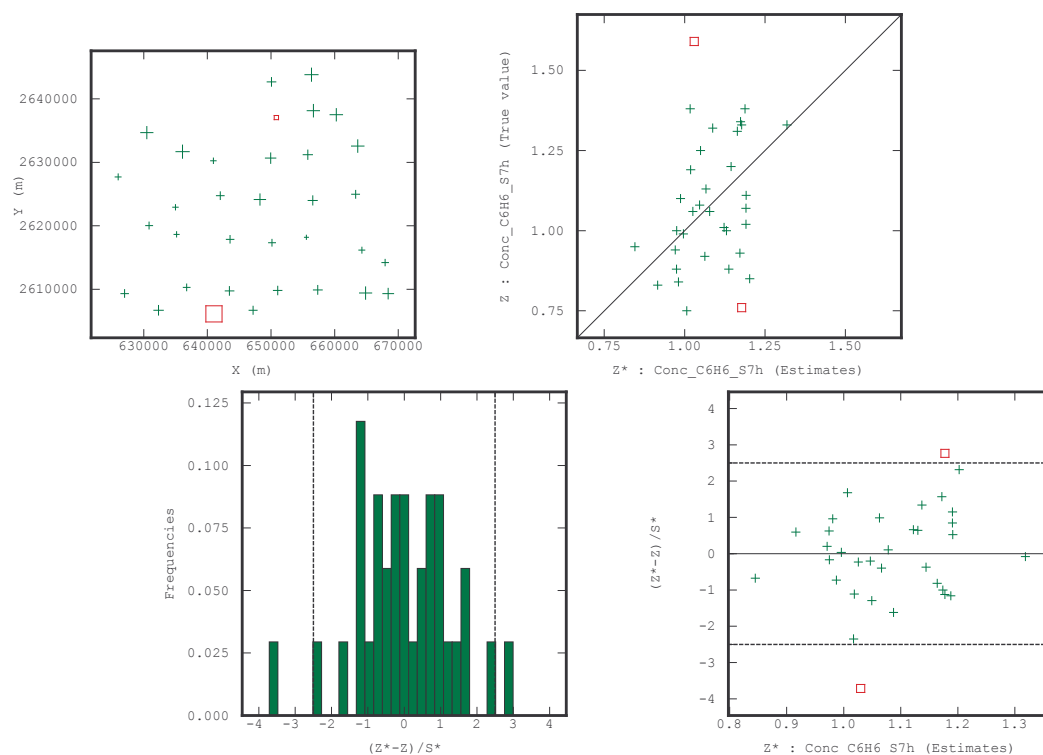


Figure 32 – Résultats de la validation croisée pour le jeu de données S7h

Le coefficient de corrélation est de : $r = 0,278$.

9.3 Test de validation absolue sur le complémentaire C7

9.3.1 VALIDATION ABSOLUE SUR LE JEU DE DONNEES C7E

Tableau 17 – Résultats de la validation absolue sur le complémentaire C7e pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S7e

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S7e	
C7e	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	159	159	159
Minimum	0,61	0,76	0,12
Percentile 25	0,82	0,86	0,13
Médiane	0,94	0,96	0,13
Percentile 75	1,07	1,03	0,13
Maximum	1,31	1,10	0,15
Etendue	0,70	0,34	0,03
Moyenne	0,93	0,94	0,13
Ecart-type	0,027	0,10	0,01
Variance	0,164	0,01	0,00
CV%	17,6	10,0	4,0
Corrélation mesures/estimations		$r = 0,652$	

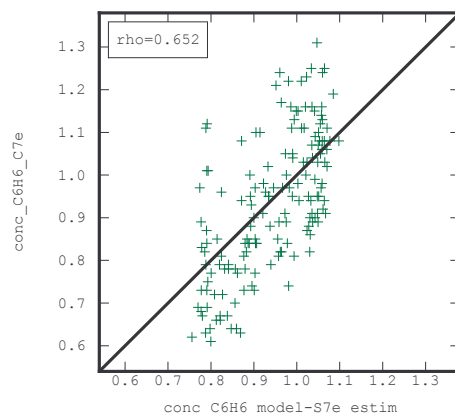


Figure 33 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C7e

9.3.2 VALIDATION ABSOLUE SUR LE JEU DE DONNEES C7H

Tableau 18 – Résultats de la validation absolue sur le complémentaire C7h pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S7h

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S7h	
C7h	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	174	174	174
Minimum	0,75	0,87	0,12
Percentile 25	0,99	1,01	0,13
Médiane	1,14	1,18	0,13
Percentile 75	1,37	1,18	0,13
Maximum	1,92	1,34	0,16
Etendue	1,17	0,47	0,04
Moyenne	1,20	1,10	0,13
Ecart-type	0,077	0,12	0,01
Variance	0,278	0,01	0,00
CV%	23,1	10,5	4,7
Corrélation mesures/estimations		$r = 0,607$	

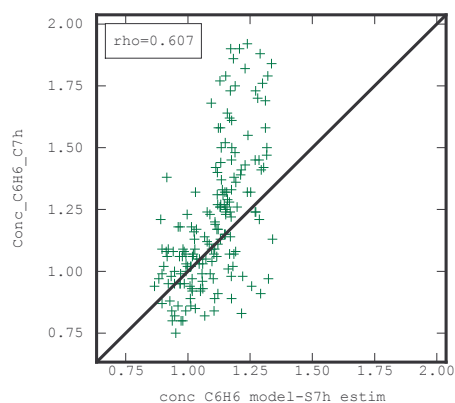


Figure 34 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C7h

9.4 Etablissement de la carte pour la sélection S7

9.4.1 KRIGEAGE DES CONCENTRATIONS POUR LE JEU DE DONNEES S7E

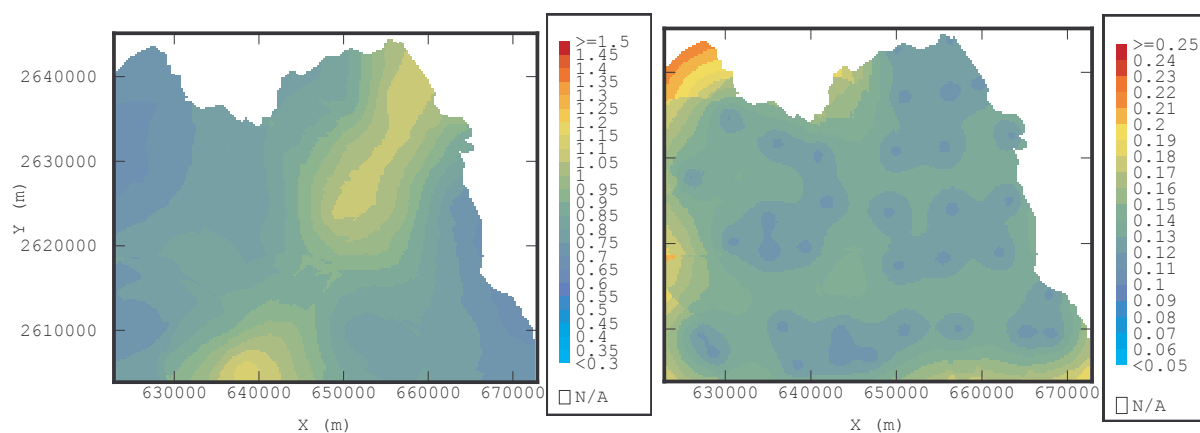


Figure 35 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S7e et (droite) écarts-types de krigeage

9.4.2 KRIGEAGE DES CONCENTRATIONS POUR LE JEU DE DONNEES S7H

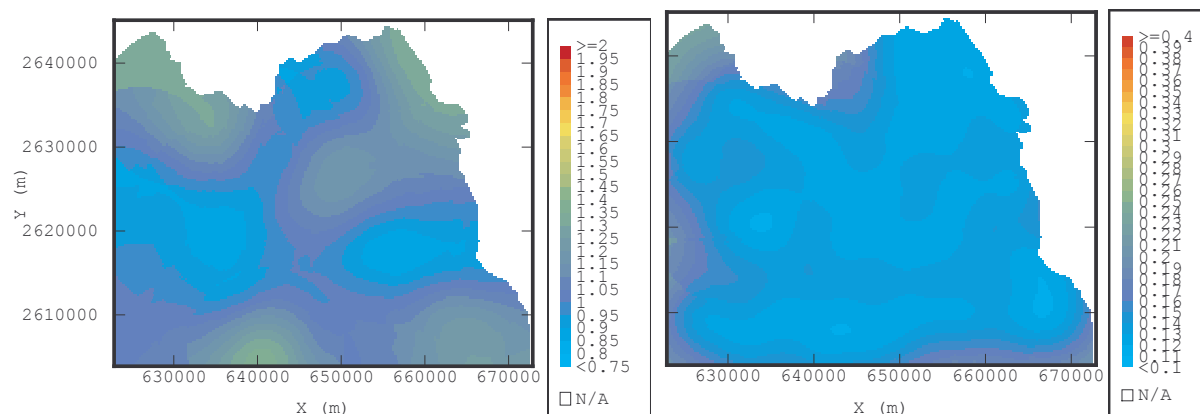


Figure 36 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S7h et (droite) écarts-types de krigeage

10. COMPARAISON DES ESTIMATIONS ET DISCUSSION

10.1 Cas de l'été 2003

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.

Tableau 19 - Synthèse des résultats de validation pour l'été 2003

Valeurs	J1e	S3e	S5e	S7e
Nombre de données	194	100	53	35
Minimum ($\mu\text{g m}^{-3}$)	0,61	0,61	0,61	0,64
Maximum ($\mu\text{g m}^{-3}$)	1,31	1,26	1,31	1,26
Moyenne ($\mu\text{g m}^{-3}$)	0,92	0,88	0,86	0,87
Ecart-type ($\mu\text{g m}^{-3}$)	0,027	0,029	0,032	0,027
Variance ($\mu\text{g m}^{-3}$)	0,165	0,169	0,178	0,163

Validation croisée					
Erreur					
	Moyenne	-0,00146	-0,00047	0,00255	0,00706
	Variance	0,01291	0,01295	0,01571	0,01942
Erreur standardisée					
	Moyenne	-0,01595	-0,00446	0,01119	0,03068
	Variance	1,00166	1,09039	0,86449	0,87716
Coefficient de corrélation $r =$		0,727	0,741	0,710	0,525

Validation absolue					
Erreur					
	Moyenne	-	0,09	0,11	0,13
	Variance	-	0,00	0,00	0,00
Coefficient de corrélation $r =$		-	0,637	0,642	0,652

Les différents variogrammes expérimentaux peuvent également être comparés.

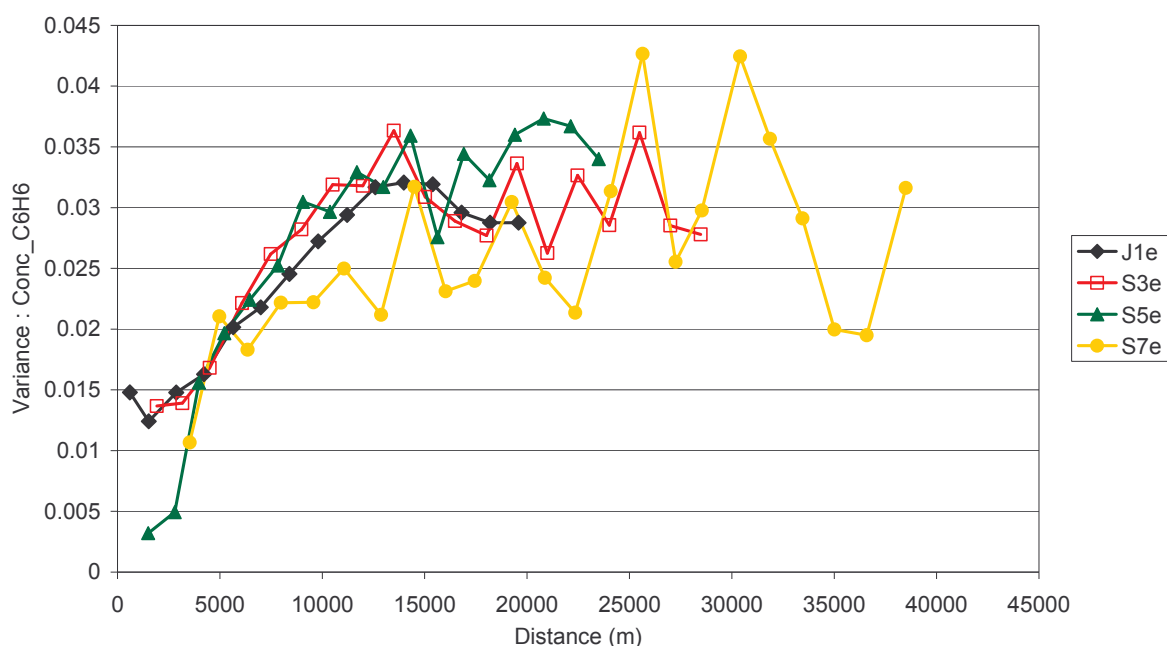


Figure 37 – Comparaison des variogrammes expérimentaux pour l'été 2003

Dans le cas des données estivales, les variogrammes expérimentaux des sélections S3e et S5e reproduisent fidèlement l'allure du variogramme de référence, avec une portée équivalente. Le variogramme expérimental de la sélection S7e présente un comportement plus erratique et une portée plus grande.

Si on représente les corrélations en tout point de la grille de référence entre les estimations obtenues avec le modèle de référence et avec les différentes sélections, on observe une dégradation progressive, notamment pour les concentrations les plus élevées.

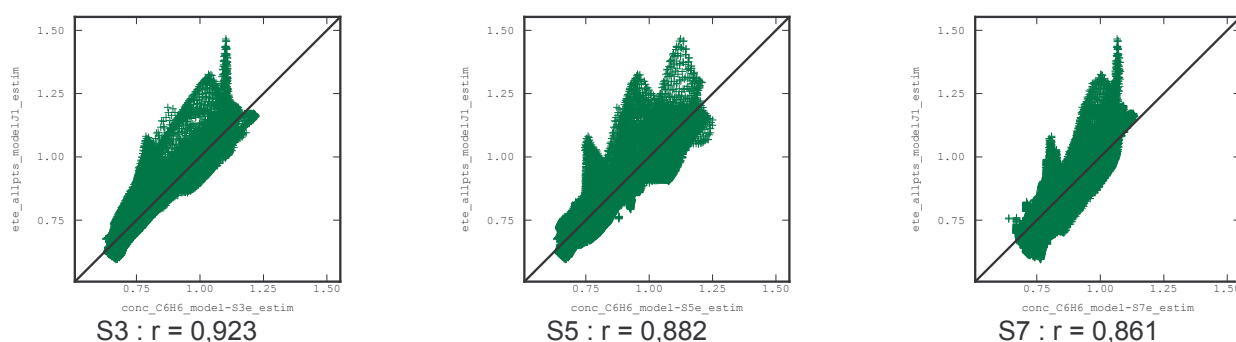
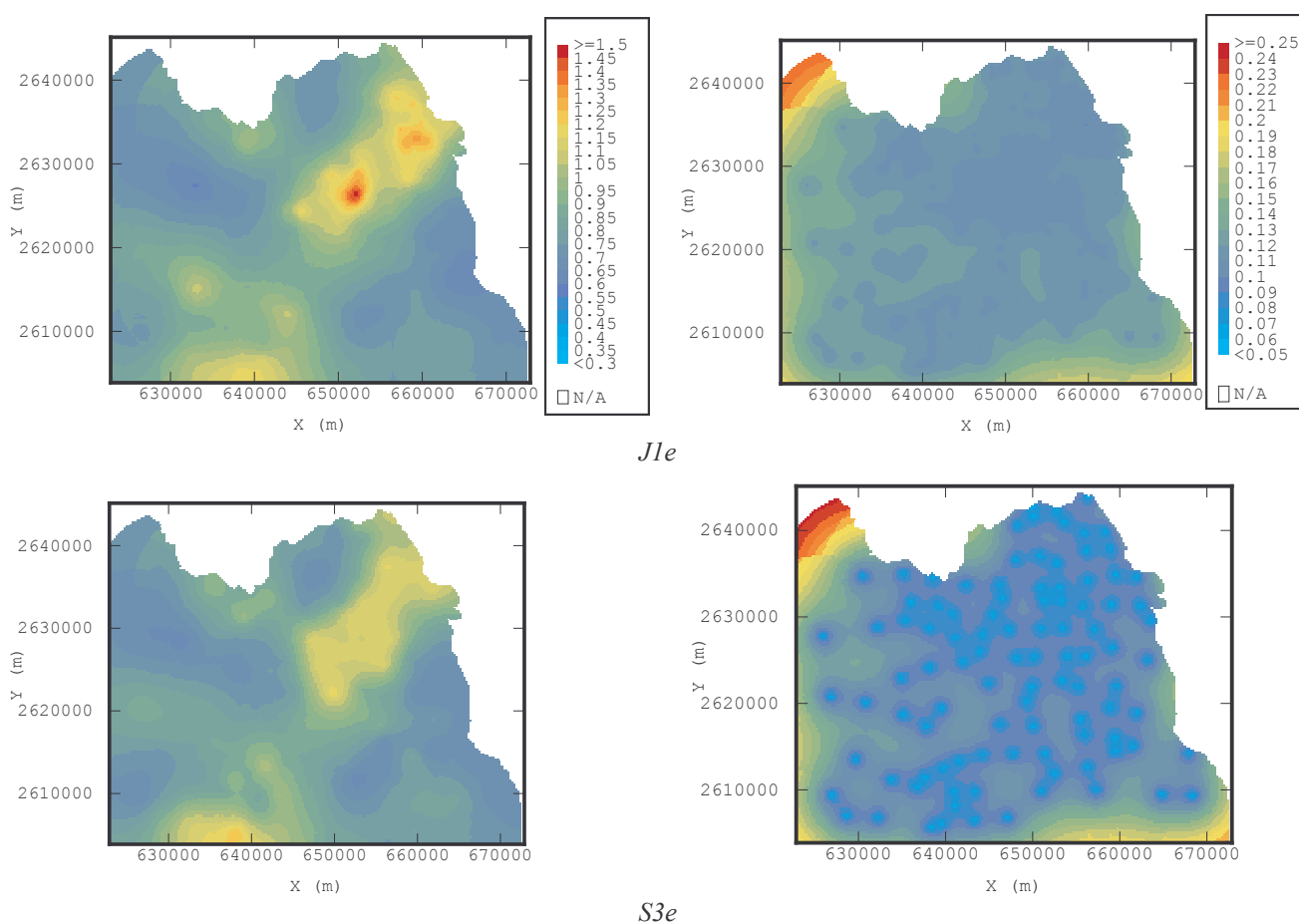


Figure 38 – Comparaison entre l'estimation de référence et les estimations obtenues à partir des différentes sélections pour l'été 2003

De même, la figure ci-dessous qui présente les différentes cartes d'estimation et d'écarts-types de krigeage montre qu'il existe un net effet de lissage au niveau des centres urbains, même si globalement les zones de faibles et fortes concentration sont correctement reproduites.



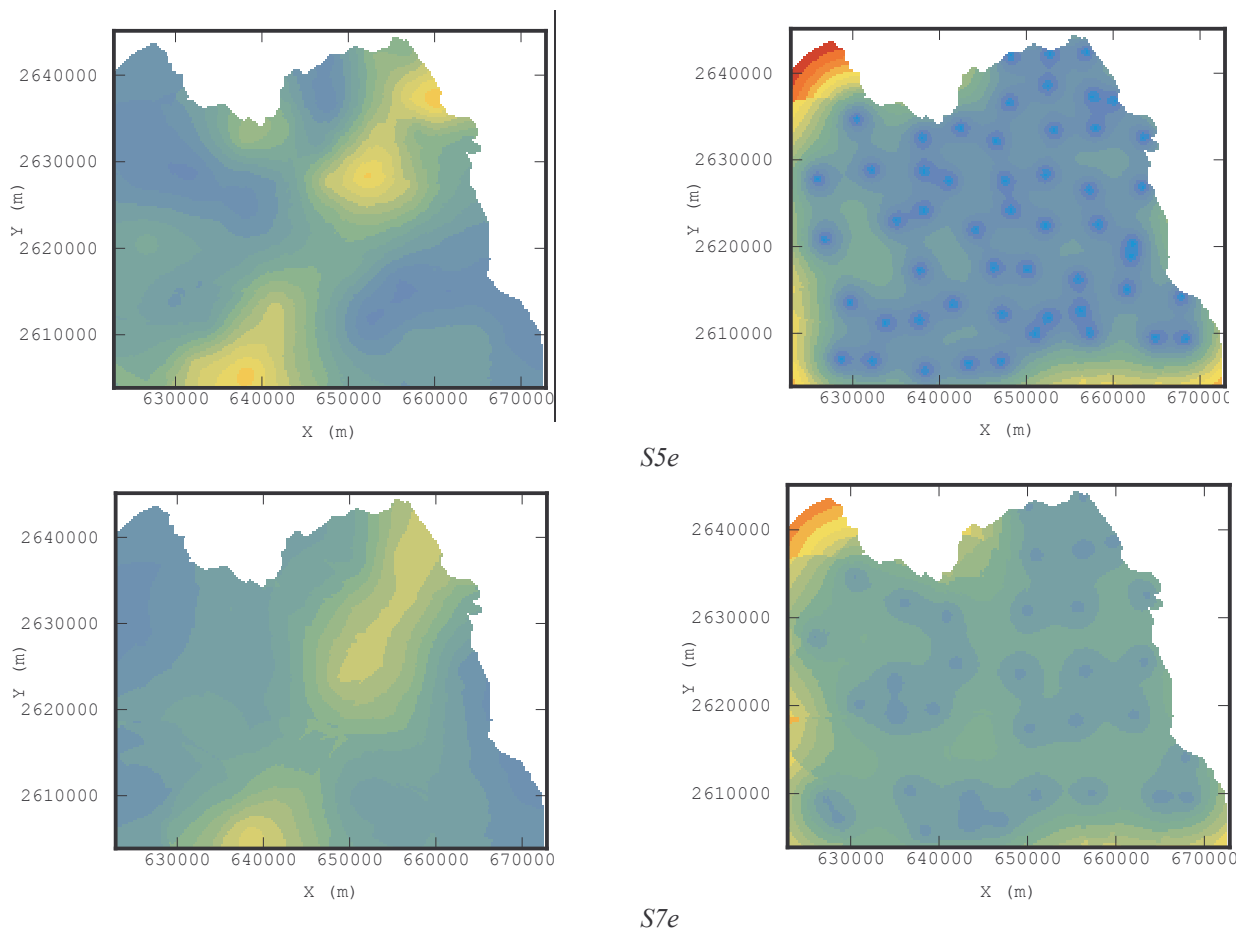


Figure 39 – Comparaison des cartes (gauche) d'estimation établies par krigeage ponctuel sur la grille de référence et (droite) d'écarts-types de krigeage pour les différentes sélections de l'été 2003

Pour évaluer la robustesse du modèle de référence, on représente également les corrélations entre les estimations calculées à partir du modèle de la sélection et à partir du modèle de référence pour le même jeu de données.

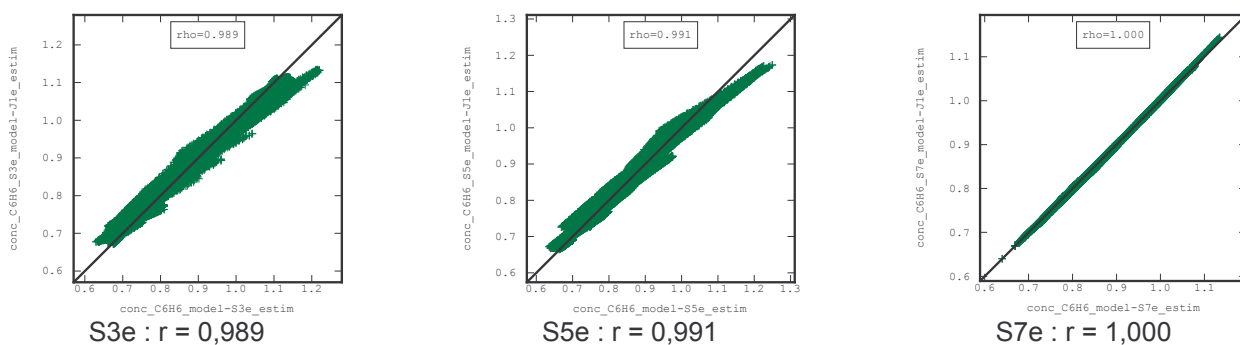


Figure 40 – Comparaison entre les estimations obtenues pour chaque sélection avec le modèle de référence (ordonnée) et avec le modèle de la sélection (abscisse) pour l'été 2003

Finalement, à partir de la carte de référence et des cartes de krigeage des différentes sélections, on peut évaluer l'effet du nombre de données expérimentales sur la qualité des cartes, en calculant les différences absolue et relative avec la carte de référence.

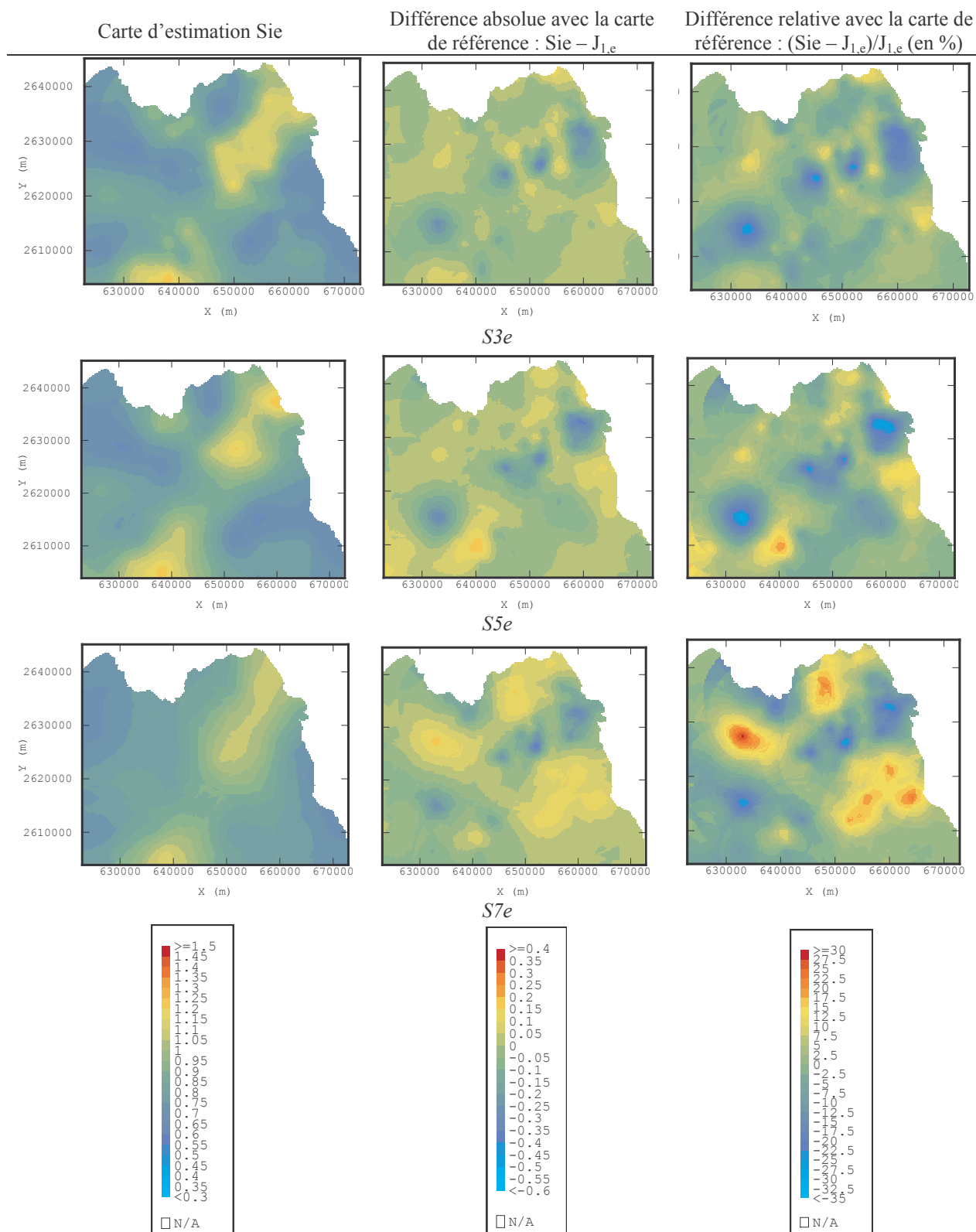


Figure 41 – Comparaison entre les cartes d'estimation obtenues avec le modèle de référence et avec les différentes sélections pour l'hiver 2003 (de haut en bas : S3e, S5e, S7e)

Les sélections S3e à S7e respectent le critère de qualité pour les estimations (différence inférieure à 30% avec la carte de référence). En revanche nous avons vu que la validation

croisée pour la sélection S7e était peu satisfaisante, le coefficient de corrélation de 0,7 environ pour les sélections précédentes chutant brusquement à 0,5.

10.2 Cas de l'hiver 2004

Tableau 20 - Synthèse des résultats de validation pour l'hiver 2004

Valeurs	J _{1h}	S3h	S5h	S7h
Nombre de données	208	104	55	34
Minimum ($\mu\text{g m}^{-3}$)	0,75	0,75	0,75	0,75
Maximum ($\mu\text{g m}^{-3}$)	1,92	1,77	1,90	1,59
Moyenne ($\mu\text{g m}^{-3}$)	1,18	1,09	1,09	1,08
Ecart-type ($\mu\text{g m}^{-3}$)	0,073	0,042	0,051	0,041
Variance ($\mu\text{g m}^{-3}$)	0,270	0,205	0,226	0,202

Validation croisée					
Erreur					
	Moyenne	-0,00312	-0,00347	0,00073	0,00550
	Variance	0,03104	0,02639	0,03903	0,02656
Erreur standardisée					
	Moyenne	-0,01528	-0,01652	0,00036	0,02943
	Variance	0,83215	1,00707	1,04591	1,09543
Coefficient de corrélation $r =$		0,750	0,614	0,510	0,278

Validation absolue					
Erreur					
	Moyenne	-	0,15	0,15	0,13
	Variance	-	0,00	0,00	0,00
Coefficient de corrélation $r =$		-	0,686	0,707	0,607

Les différents variogrammes expérimentaux peuvent également être comparés.

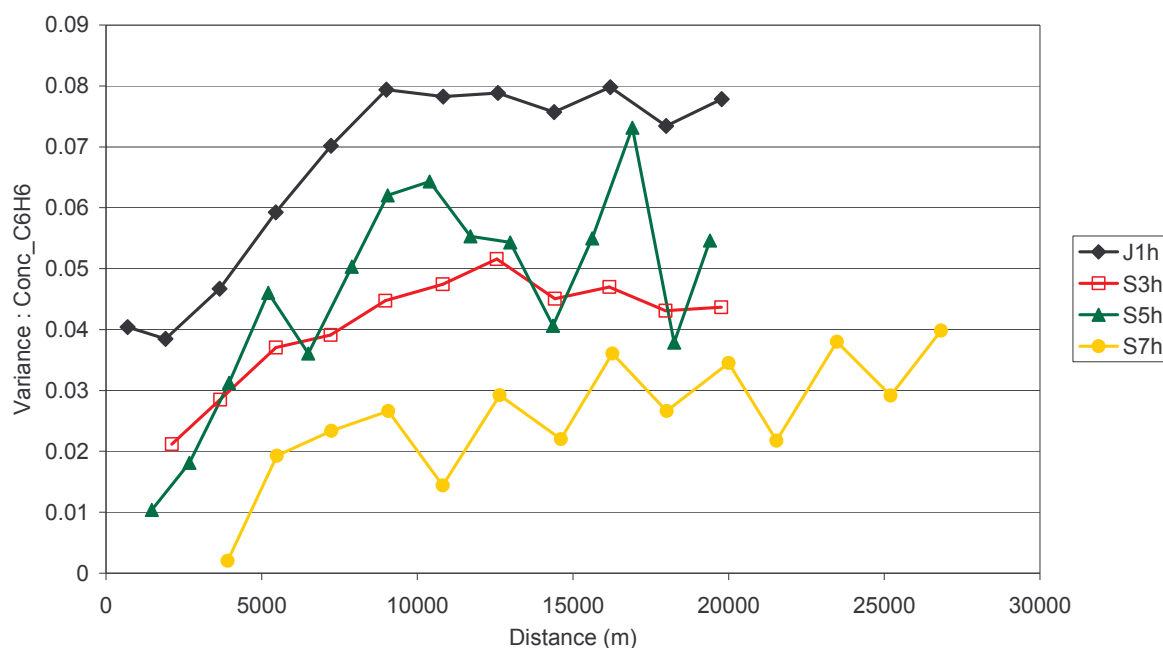


Figure 42 – Comparaison des variogrammes expérimentaux pour l'hiver 2004

La dégradation du maillage semble entraîner un allongement de la portée du variogramme. On note également que pour les sélections S5h et S7h, le variogramme expérimental est beaucoup plus erratique.

La figure ci-après représente les corrélations en tout point de la grille de référence entre les estimations obtenues avec le modèle de référence et avec les différentes sélections : le coefficient de corrélation diminue fortement à partir de la sélection S7h.

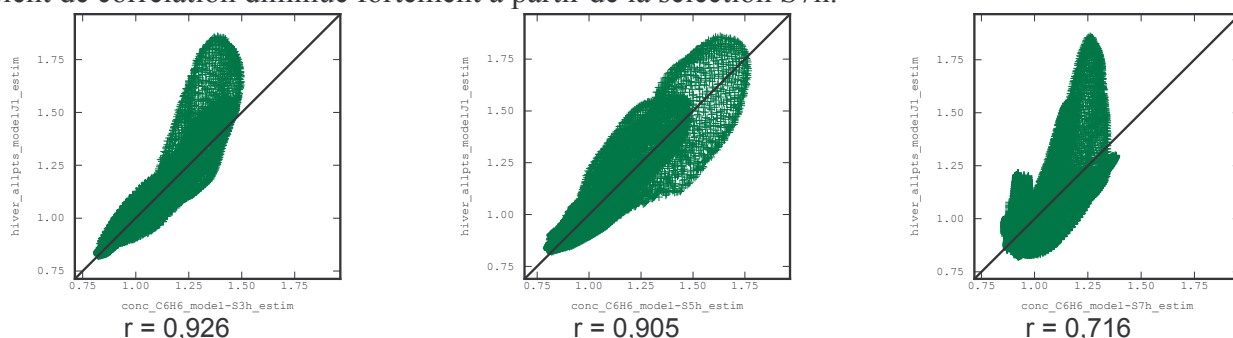
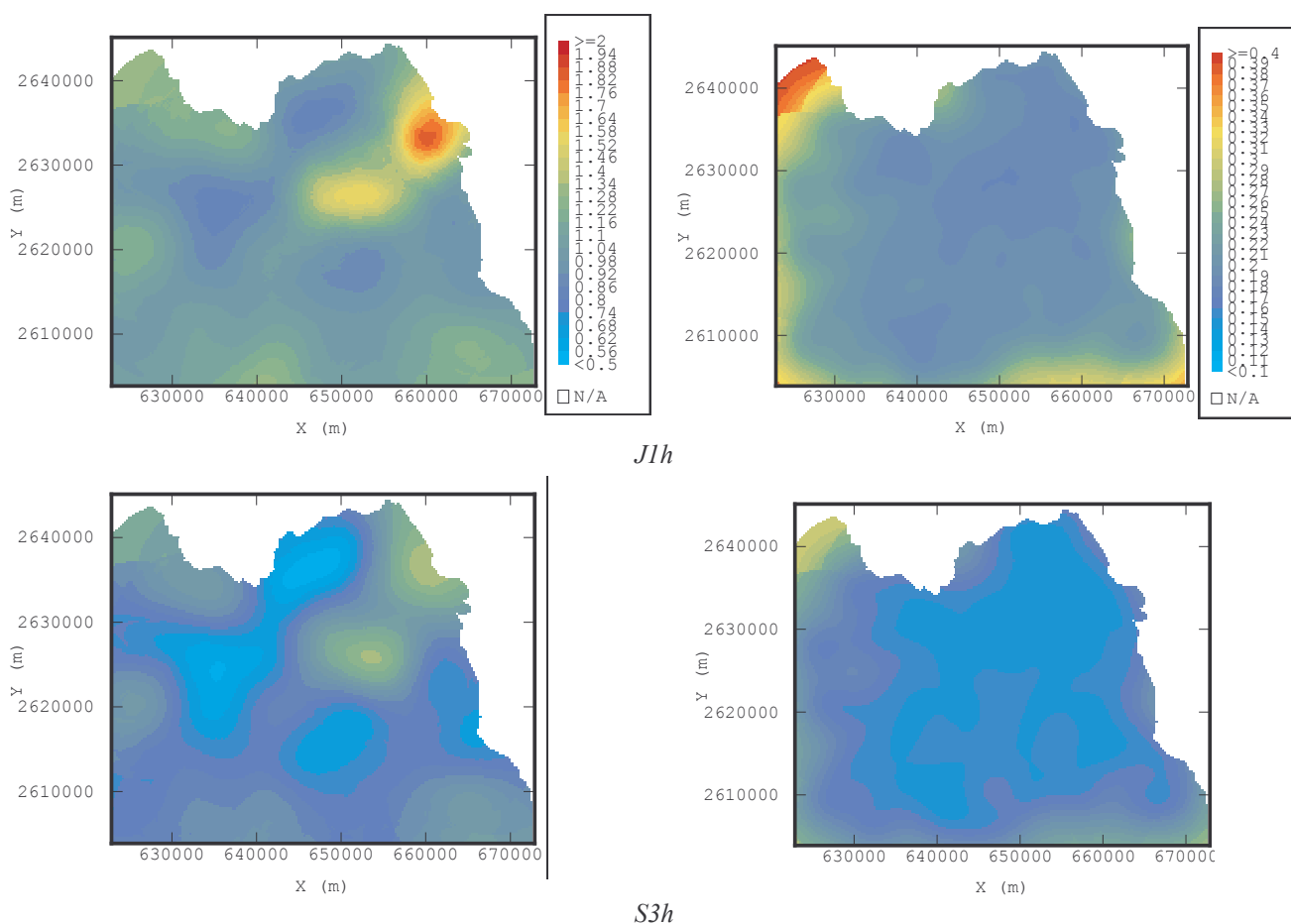


Figure 43 – Comparaison entre l'estimation de référence et les estimations obtenues à partir des différentes sélections pour l'hiver 2004



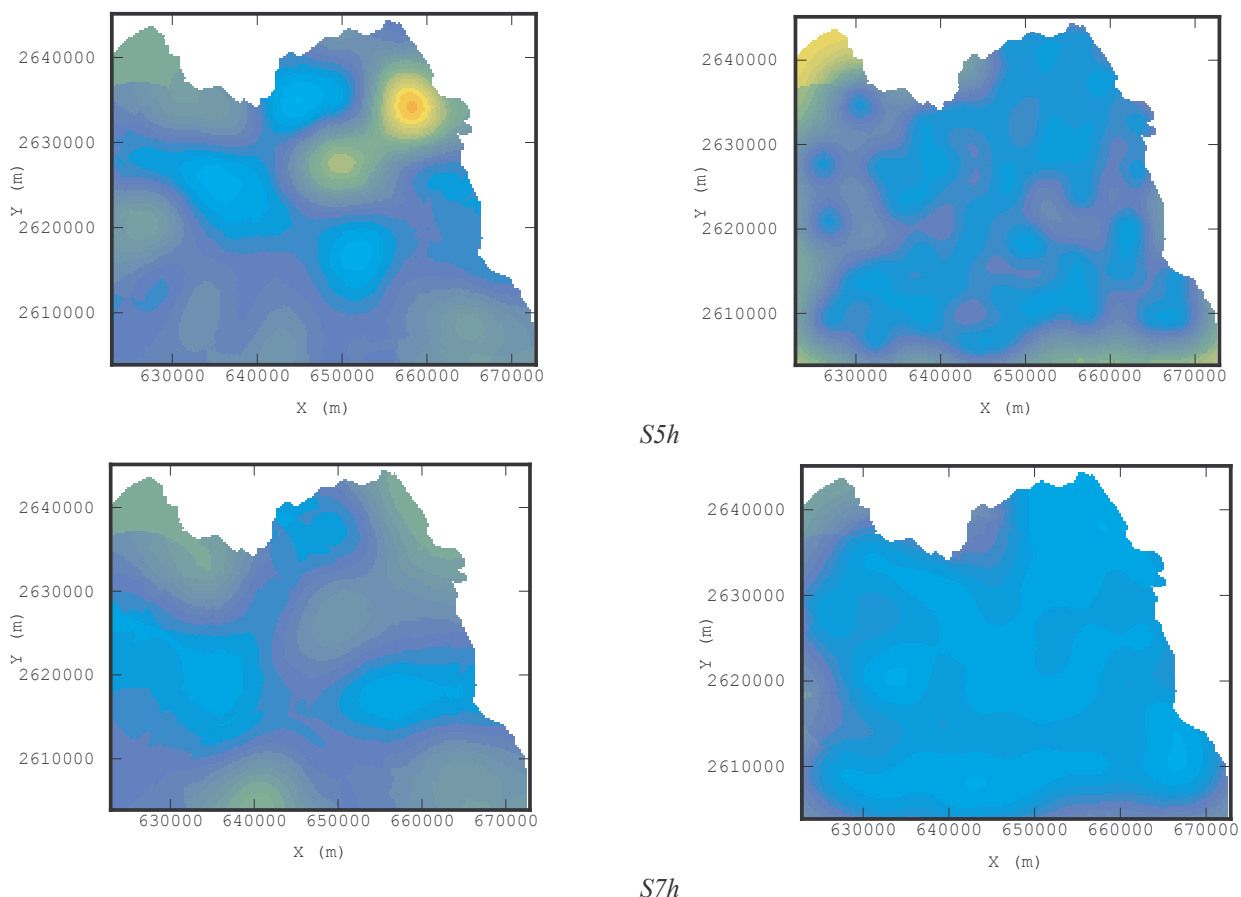


Figure 44 – Comparaison des cartes (gauche) d'estimation établies par krigeage ponctuel sur la grille de référence et (droite) d'écarts-types de krigeage pour les différentes sélections de l'hiver 2004

Pour évaluer la robustesse du modèle de référence, on représente également les corrélations entre les estimations calculées à partir du modèle de la sélection et à partir du modèle de référence pour le même jeu de données. Les résultats sont excellents pour tous les modèles.

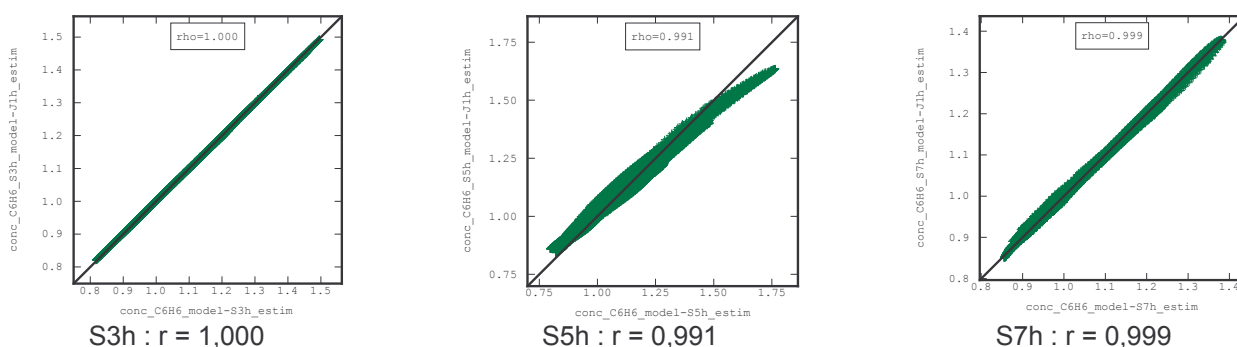


Figure 45 – Comparaison entre les estimations obtenues pour chaque sélection avec le modèle de référence (ordonnée) et avec le modèle de la sélection (abscisse) pour l'hiver 2004

Finalement, à partir de la carte de référence et des cartes de krigeage des différentes sélections, on peut évaluer l'effet du nombre de données expérimentales sur la qualité des cartes, en calculant les différences absolue et relative avec la carte de référence.

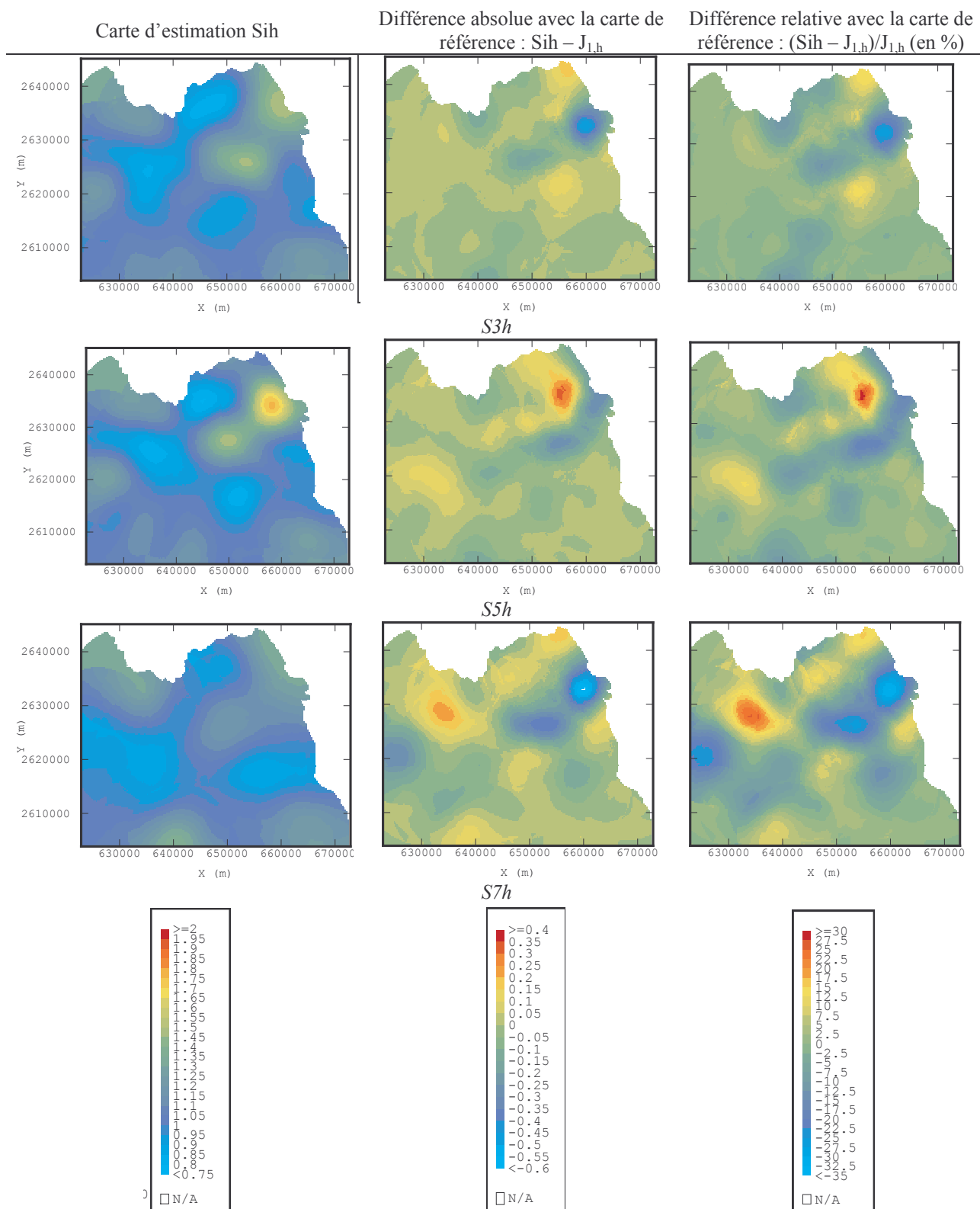


Figure 46 – Comparaison entre les cartes d'estimation obtenues avec le modèle de référence et avec les différentes sélections pour l'hiver 2003 (de haut en bas : S3h, S5h, S7h)

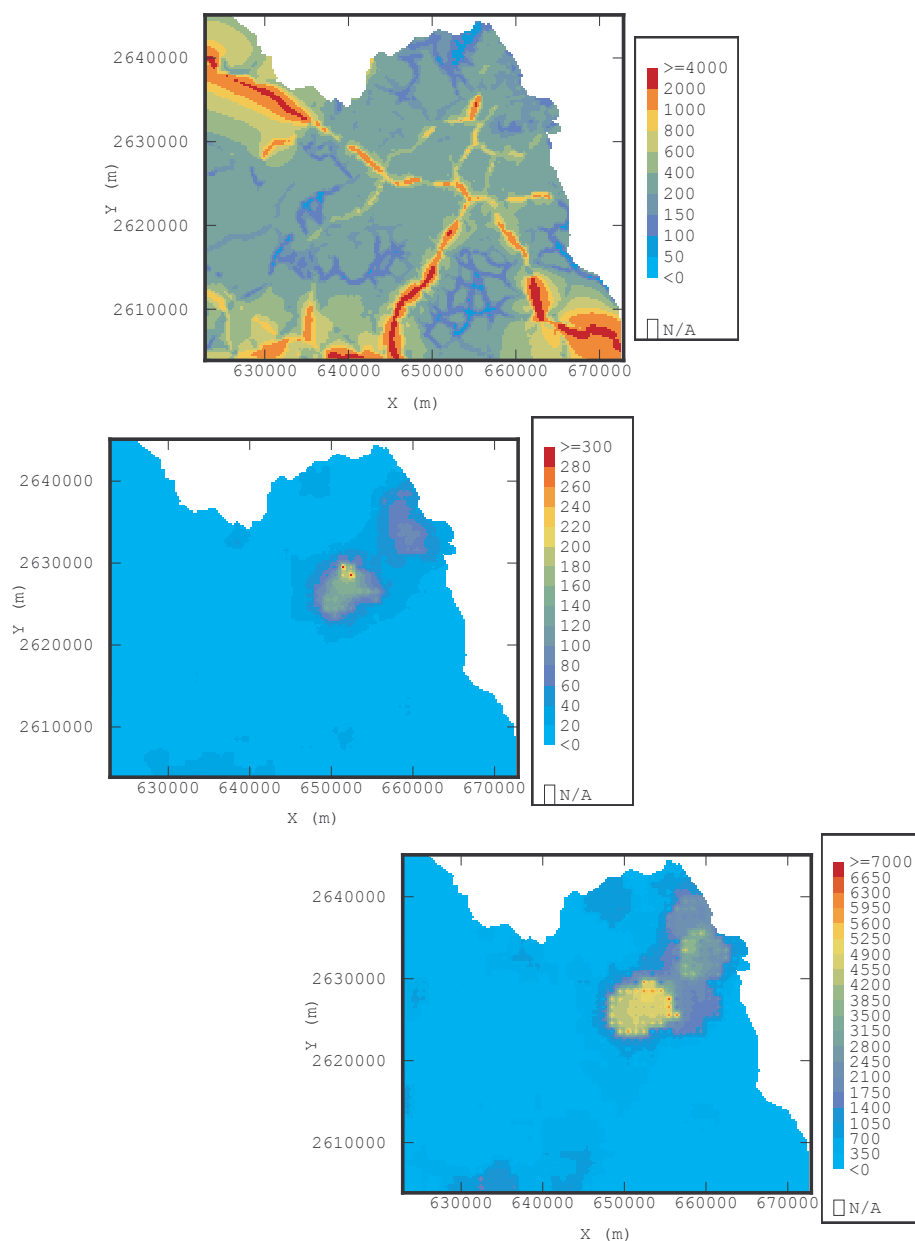
La sous-estimation des concentrations est là encore observée au niveau des centres urbains, où elle atteint 33% pour la sélection S7h.

PARTIE II. Echantillonnage régulier avec variable(s) auxiliaire(s)

En vue d'améliorer la qualité de l'estimation, on conserve les mailles régulières d'échantillonnage (mêmes sélections et complémentaires), et on va intégrer une ou plusieurs variables auxiliaires corrélées 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 information pour l'estimation.

11. DESCRIPTION DES VARIABLES EXPLICATIVES

On dispose des données d'émissions sur l'arrondissement lillois pour le benzène en g/km, les NO_x et les COV en tonnes/an, fournies par Atmo-Nord-Pas-de-Calais. Ces émissions sont interpolées aux nœuds d'une grille de maille régulière de 250 m de côté.



**Figure 47 – Carte des émissions interpolées dans l'arrondissement lillois :
de haut en bas, C_6H_6 (échelle non linéaire), NO_x et COV**

12. CALCUL DES CORRELATIONS

12.1 Analyse des variables explicatives

On cherche de manière empirique la meilleure corrélation possible entre les concentrations observées de benzène et les valeurs d'émissions. Pour cela, on détermine les valeurs des émissions aux points des jeux de données de référence J_0 , ce qui permet de tracer des nuages de corrélation. La figure ci-après présente ces nuages de corrélation pour l'hiver 2004 ; des allures similaires sont obtenues pour l'été 2003.

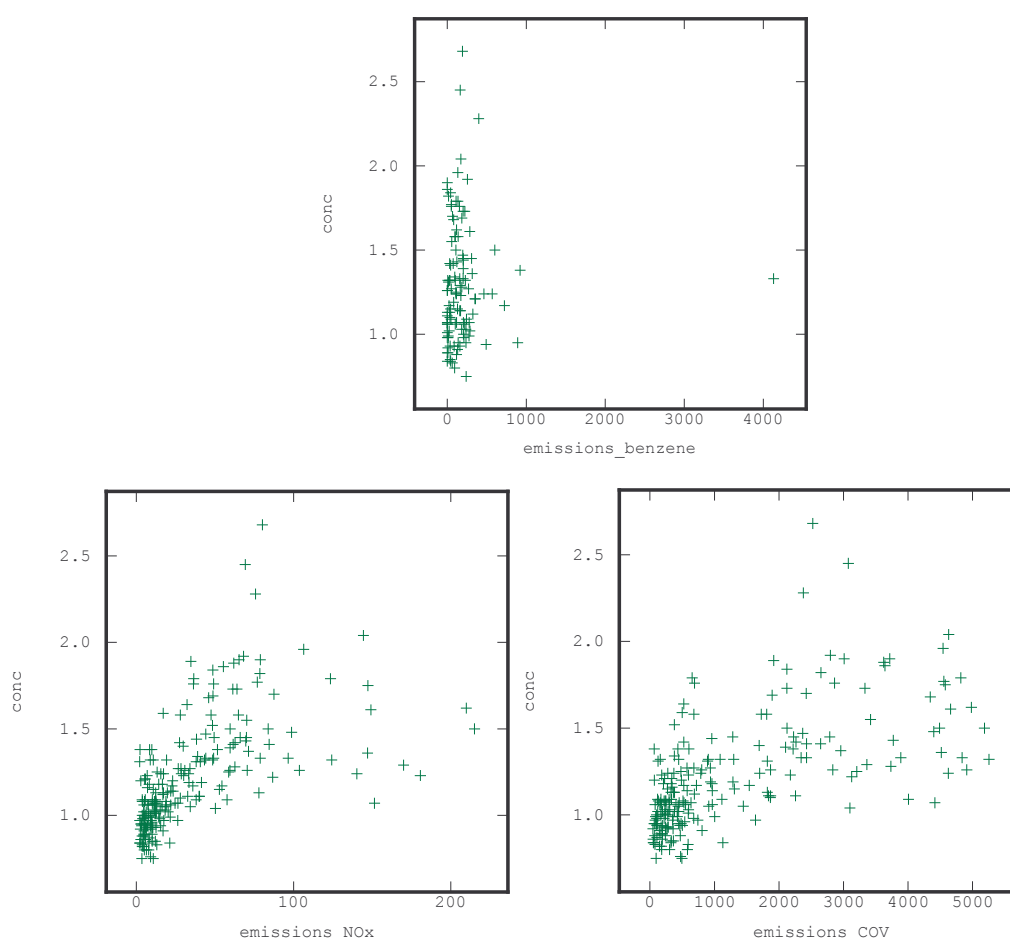


Figure 48 – Nuages de corrélation entre C_6H_6 et les trois variables d'émission pour l'hiver 2004

Les deux nuages de corrélation entre les concentrations de benzène et les valeurs d'émission de NO_x et de COV semblent indiquer que les deux variables sont liées d'une façon non linéaire. En revanche, il n'y pas de lien évident entre les concentrations en benzène mesurées et l'inventaire d'émissions du benzène ! Ceci peut toutefois s'expliquer par la nature des deux campagnes de mesures, destinées à mesurer les concentrations de fond de C_6H_6 et non la pollution de proximité en bordure des axes routiers.

**Tableau 21 - Coefficients de corrélation des différentes variables
avec les concentrations de C_6H_6 pour l'été 2003 et l'hiver 2004**

	Emissions C ₆ H ₆	Emissions NO _x	Emissions COV
Eté 2003	0,078	0,579	0,496
Hiver 2004	0,024	0,581	0,624

Pour déterminer un cofacteur pertinent, on n'utilisera donc dans la partie suivante que les émissions de NO_x et de COV.

12.2 Elaboration d'un cofacteur pertinent

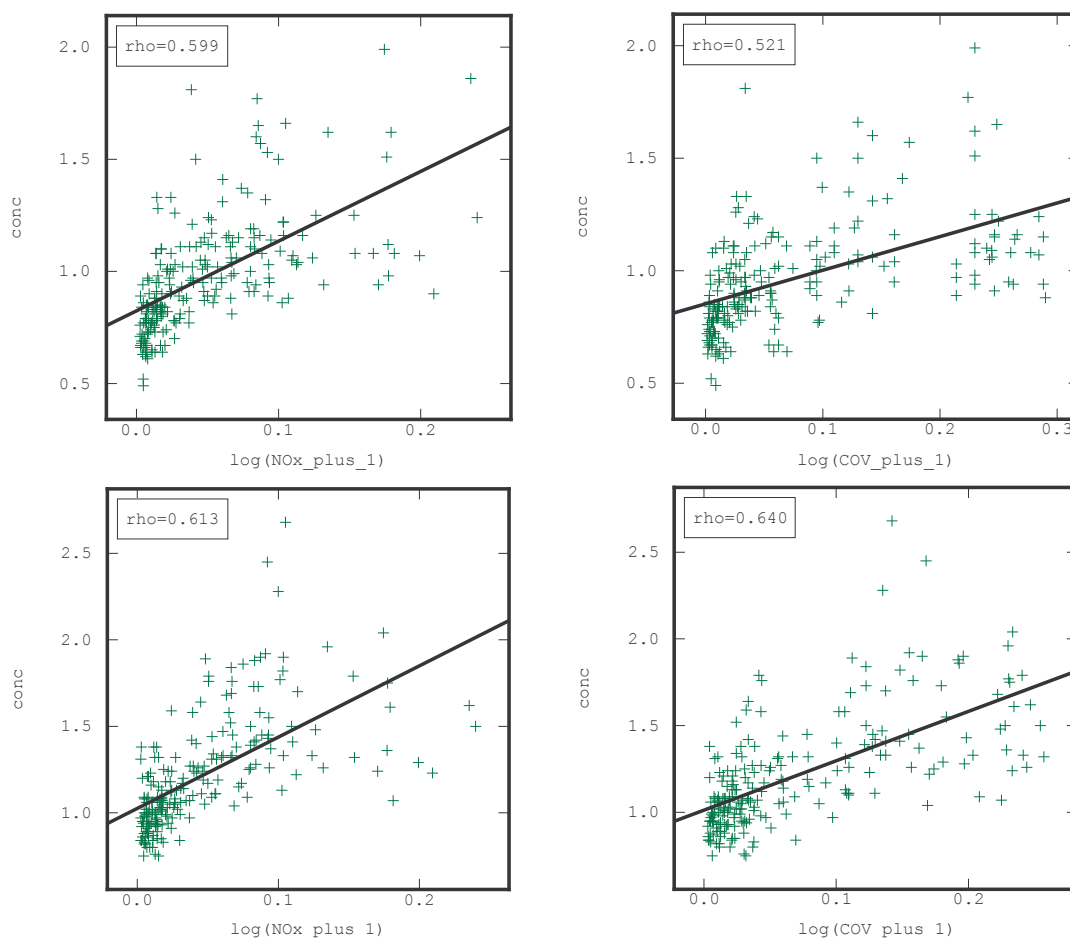
On va chercher une combinaison linéaire des deux variables d'émission. On normalise d'abord les variables explicatives E suivant la relation :

$$E_{\text{norm}} = \frac{E_i - E_{\text{min}}}{E_{\text{max}} - E_{\text{min}}}$$

Vu le profil des nuages de corrélation, on calcule ensuite le logarithme décimal de chaque valeur (le « +1 » permettant de ne pas générer d'erreur en cas de valeur nulle d'émission) :

$$\{\text{Log } E_{\text{plus}_1}\} = \log (E_{\text{norm}} + 1)$$

Les corrélations sont légèrement améliorées comme le montre la figure ci-après.



Figures 49 – Nuages de corrélation entre C₆H₆ et les variables {Log E_{plus_1}} pour (haut) l'été 2003 et (bas) l'hiver 2004

Pour trouver un cofacteur pertinent, on calcule la régression linéaire multiple pour les deux saisons. Les résultats sont présentés dans le tableau ci-après.

Tableau 22 – Résultats de la régression linéaire multiple avec les variables explicatives d'émissions de COV et de NO_x pour l'été 2003 et l'hiver 2004.

	Coefficients pondérateurs		Coefficient de corrélation
	{Log NO _x _plus_1}	{Log COV_plus_1}	
Eté 2003	9,707	1,323	0,564
Hiver 2004	5,579	6,393	0,646

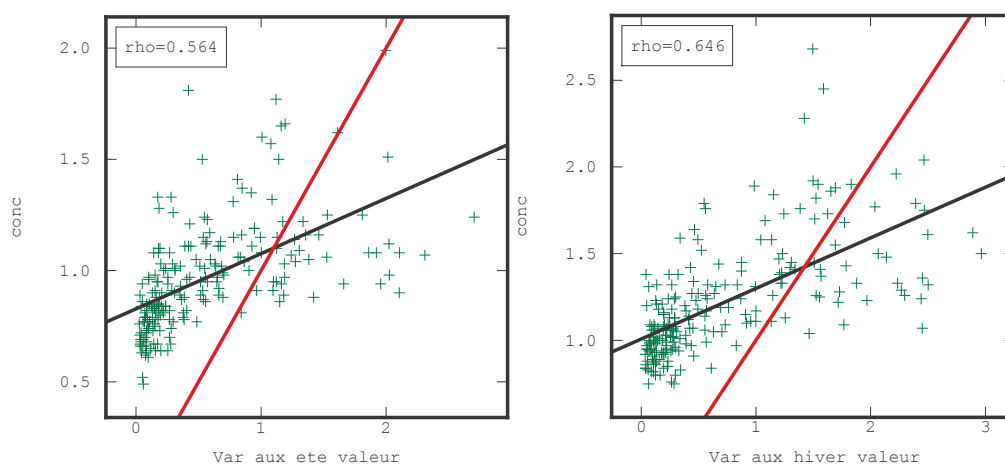


Figure 50 – Nuages de corrélation entre les valeurs trouvées par régression linéaire multiple et les concentrations observées pour (gauche) l'été 2003 et (droite) l'hiver 2004 ; ligne noire : droite de régression ; ligne rouge : 1^{ère} bissectrice

On constate que le coefficient de corrélation estival est plus faible avec les deux variables explicatives qu'avec la variable {Log NO_x_plus_1} seule ; on choisira donc de ne conserver que cette variable comme variable auxiliaire. En revanche, pour l'hiver, la prise en compte des deux variables améliore la corrélation : ces deux variables seront donc conservées.

La figure suivante présente les valeurs des variables auxiliaires calculées à partir des coefficients de la régression linéaire multiple sur la grille de référence.

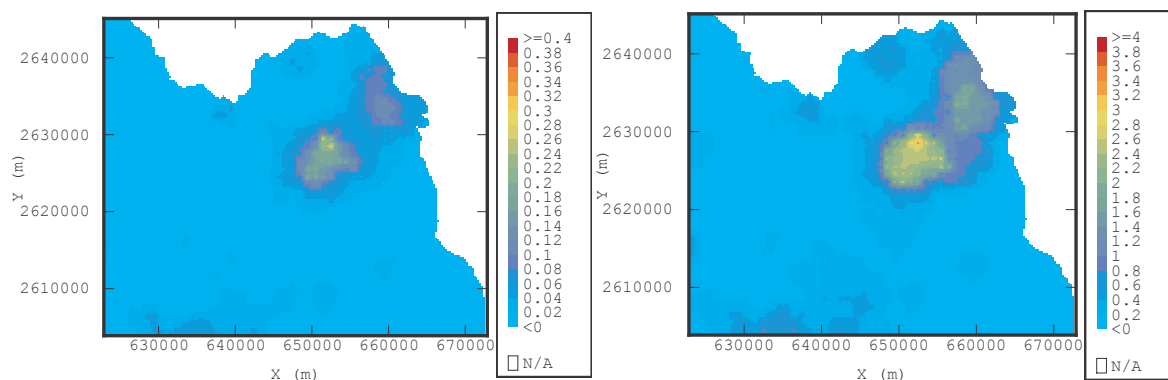


Figure 51 – Carte des variables auxiliaires d'émissions pour (gauche) l'été 2003 et (droite) l'hiver 2004

En l'absence de corrélation forte, on va utiliser le cokrigage colocalisé qui suppose que les concentrations de benzène, connues aux points de mesure, et la variable auxiliaire, connue en tout point de la grille de référence, sont liées par un modèle variographique.

13. ETUDE DU KRIGEAGE COLOCALISE SUR LES SELECTIONS S3

On va établir le modèle bivariable à partir des jeux de données S3 et des valeurs de variable auxiliaire définies aux mêmes coordonnées, de manière à pouvoir tester les modèles sur les complémentaires C3.

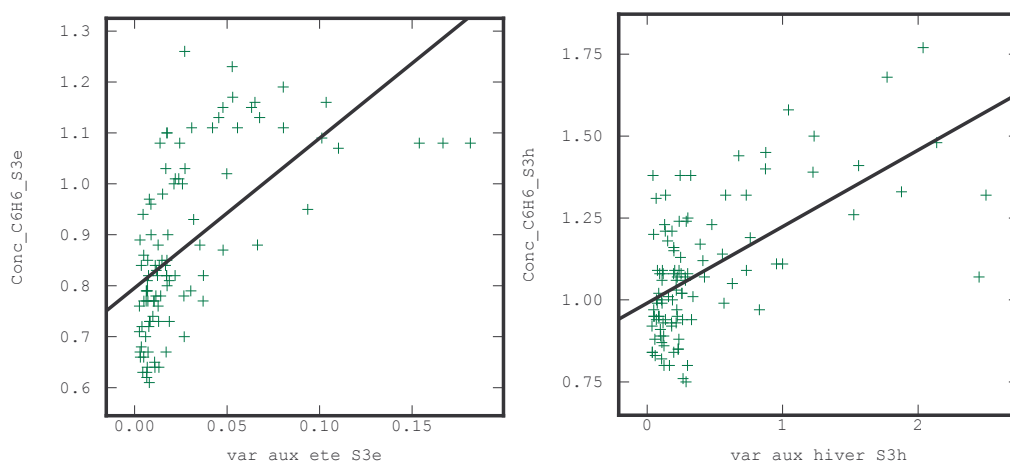


Figure 52 – Diagramme de corrélation entre les valeurs auxiliaires et les concentrations en benzène pour (gauche) S3e ($r = 0,598$) et (droite) S3h ($r = 0,604$)

13.1 Cas de l'été 2003

13.1.1 ETUDE DES VARIOGRAMMES EXPERIMENTAUX

Le variogramme est omnidirectionnel avec 12 pas de 1500 m chacun.

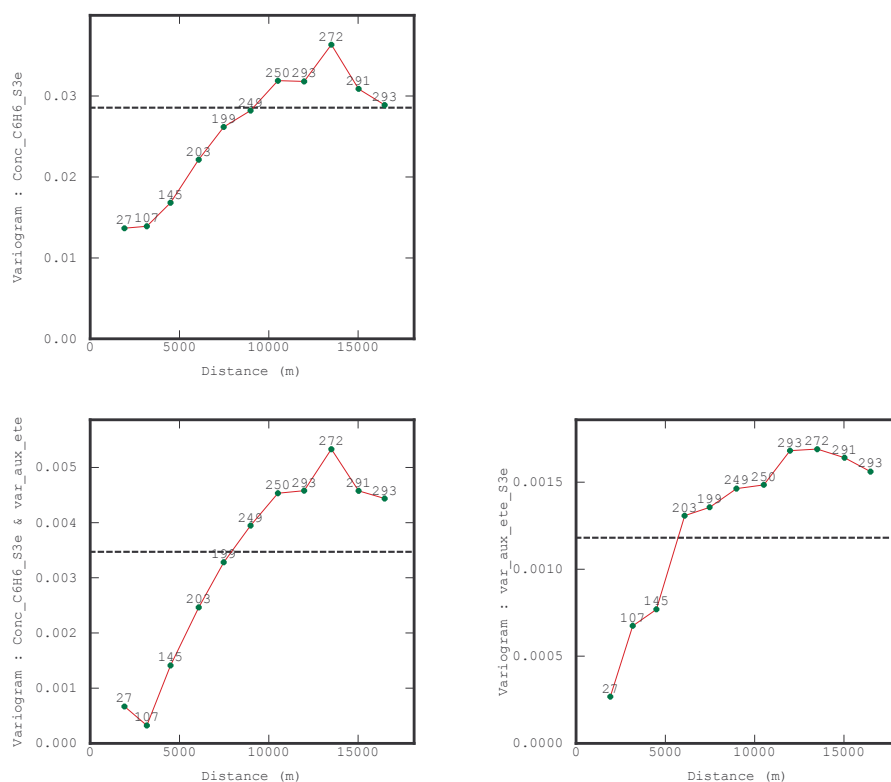


Figure 53 – Variogrammes expérimentaux simples et croisé (100 sites ; 12 pas ; 1500 m)

13.1.2 MODELISATION DES VARIOGRAMMES ET VALIDATION CROISEE

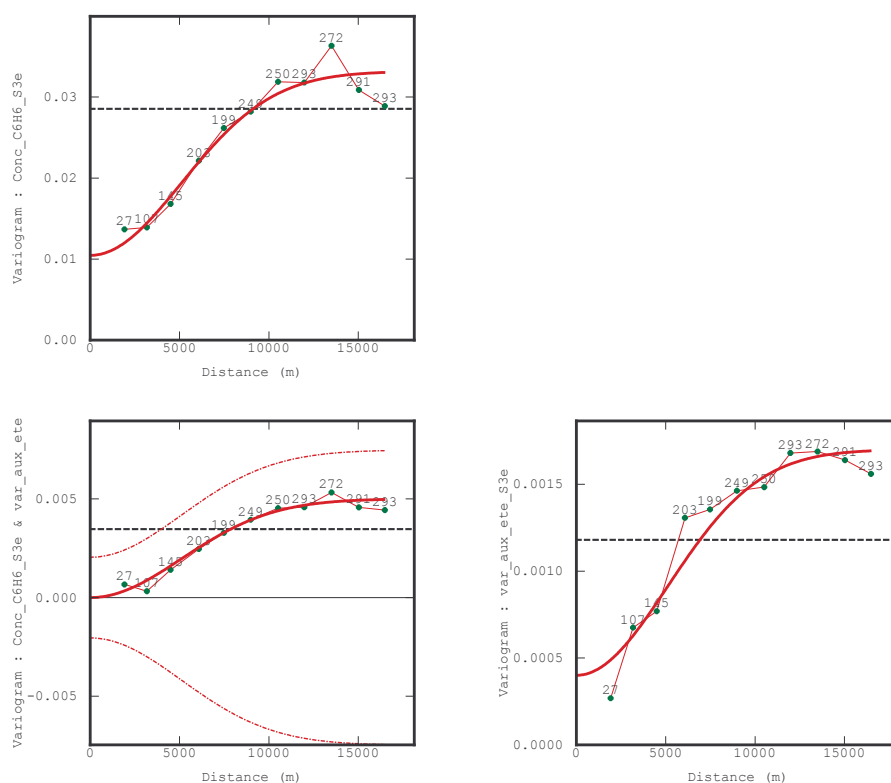


Figure 54 – Variogrammes simulés simples et croisé

Number of variables = 2
 - Variable 1 : var_aux_ete_S3e
 - Variable 2 : Conc_C6H6_S3e
 Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix:			Decomposition into factors			Decomposition into eigen vectors (whose variance is eigen values):				
	Variable 1	Variable 2		Variable 1	Variable 2		Variable 1	Variable 2	Eigen Val.	Var. Perc.
Variable 1	0.0001	0.0000	Factor 1	0.0003	0.0984	Factor 1	0.0034	1.0000	0.0097	98.69
Variable 2	0.0000	0.0097	Factor 2	0.0113	-0.0000	Factor 2	1.0000	-0.0034	0.0001	1.31

S2 : Gaussian - Scale = 12500m

Variance-Covariance matrix:			Decomposition into factors			Decomposition into eigen vectors (whose variance is eigen values):				
	Variable 1	Variable 2		Variable 1	Variable 2		Variable 1	Variable 2	Eigen Val.	Var. Perc.
Variable 1	0.0012	0.0053	Factor 1	0.0343	0.1545	Factor 1	0.2168	0.9762	0.0250	100.00
Variable 2	0.0053	0.0239	Factor 2	0.0000	0.0000	Factor 2	0.9762	-0.2168	0.0000	0.00

S3 : Gaussian - Scale = 5000m

Variance-Covariance matrix :			Decomposition into factors			Decomposition into eigen vectors (whose variance is eigen values) :				
	Variable 1	Variable 2		Variable 1	Variable 2		Variable 1	Variable 2	Eigen Val.	Var. Perc.
Variable 1	0.0004	-0.0003	Factor 1	0.0204	-0.0159	Factor 1	0.7890	-0.6144	0.0007	100.00
Variable 2	-0.0003	0.0003	Factor 2	0.0000	0.0000	Factor 2	0.6144	0.7890	0.0000	0.00

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 20 km.

Tableau 23 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les concentrations

S3e (100 points)	Moyenne	Variance
Erreur	0,00038	0,01218
Erreur standardisée	0,00318	0,91693

Toutes les données sont robustes.

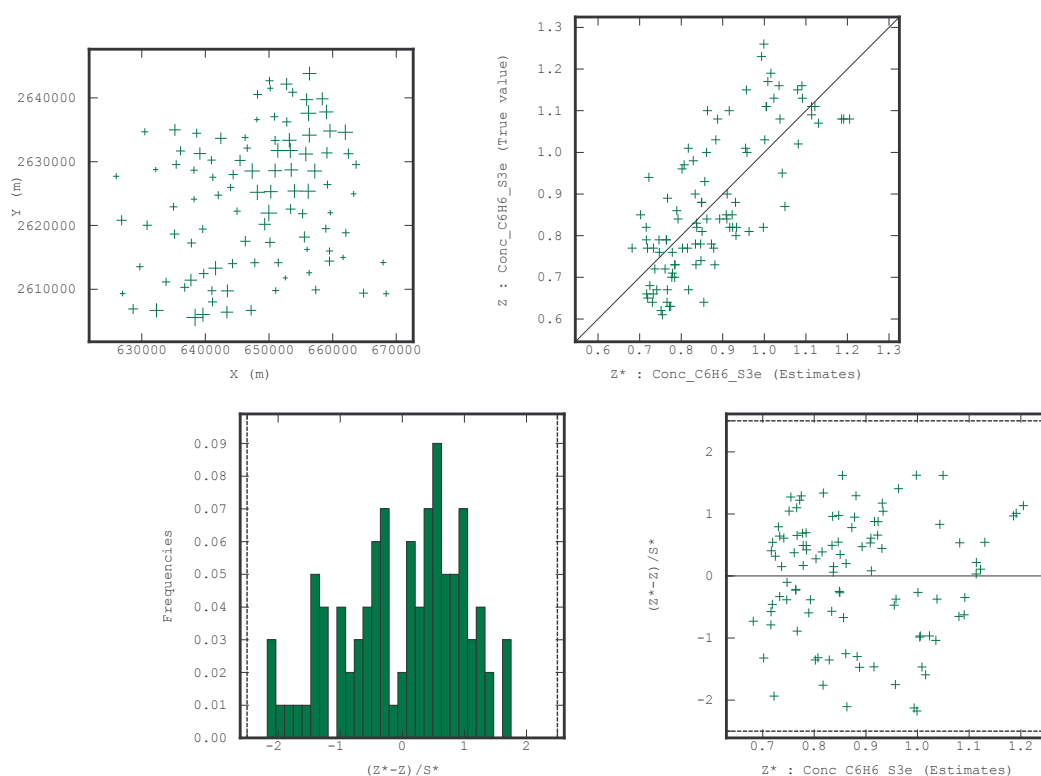


Figure 55 – Résultats de la validation croisée

13.1.3 TEST DE VALIDATION ABSOLUE SUR LE COMPLEMENTAIRE C3E

Tableau 24 – Résultats de la validation absolue sur le complémentaire C3e pour le krigeage ponctuel colocalisé

		Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S3e	
C3e	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	94	94	94
Minimum	0,64	0,70	0,106
Percentile 25	0,88	0,88	0,108
Médiane	0,96	1,02	0,108
Percentile 75	1,06	1,11	0,110
Maximum	1,31	1,19	0,121
Etendue	0,67	0,49	0,015
Moyenne	0,97	1,00	0,110
Ecart-type	0,022	0,14	0,003
Variance	0,148	0,02	0,000
CV%	15,3	13,7	2,2
Corrélation mesures/estimations		$r = 0,663$	

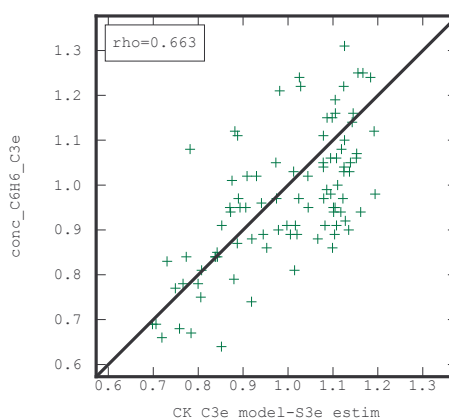


Figure 56 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C3e

13.2 Cas de l'hiver 2004

13.2.1 ETUDE DES VARIOGRAMMES EXPERIMENTAUX

Le variogramme est omnidirectionnel avec 12 pas de 1800 m chacun.

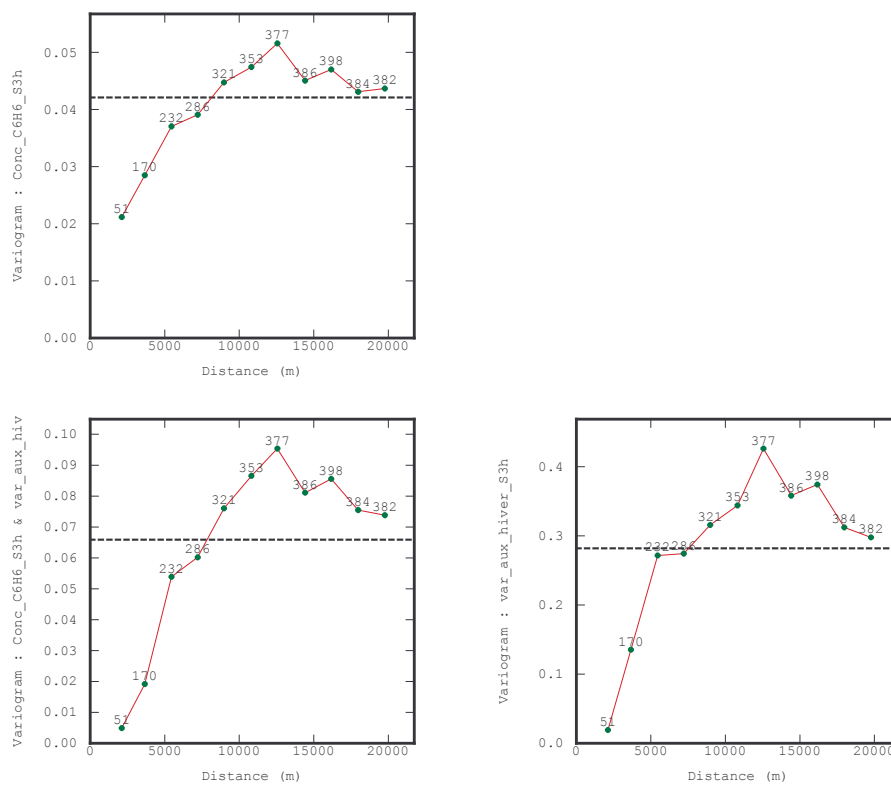


Figure 57 – Variogrammes expérimentaux simples et croisé (104 sites ; 12 pas ; 1800 m)

13.2.2 MODELISATION DES VARIOGRAMMES ET VALIDATION CROISEE

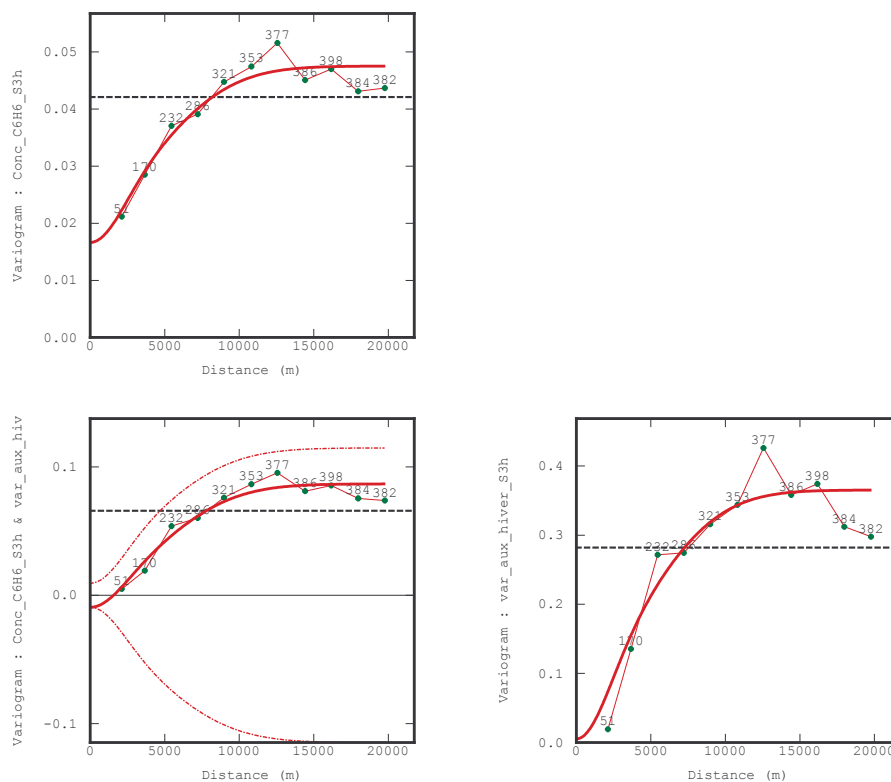


Figure 58 – Variogrammes simulés simples et croisé

Number of variables = 2
- Variable 1 : var_aux_hiver_S3h

- Variable 2 : Conc_C6H6_S3h
Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix:			Decomposition into factors		Decomposition into eigen vectors (whose variance is eigen values):			
Variable 1 Variable 2			(normalized eigen vectors):		Variable 1 Variable 2 Eigen Val. Var. Perc.			
Variable 1	0.0052	-0.0093	Factor 1	-0.0724	0.1290	Factor 1	-0.4895	0.8720
Variable 2	-0.0093	0.0166	Factor 2	0.0000	0.0000	Factor 2	0.8720	0.4895
							0.0219	100.00
							0.0000	0.00

S2 : Gaussian - Scale = 12000m

Variance-Covariance matrix:			Decomposition into factors		Decomposition into eigen vectors (whose variance is eigen values):			
Variable 1 Variable 2			(normalized eigen vectors):		Variable 1 Variable 2 Eigen Val. Var. Perc.			
Variable 1	0.2483	0.0737	Factor 1	0.4983	0.1478	Factor 1	0.9587	0.2844
Variable 2	0.0737	0.0219	Factor 2	-0.0023	0.0077	Factor 2	-0.2844	0.9587
							0.2702	99.98
							0.0001	0.02

S3 : Gaussian - Scale = 5000m

Variance-Covariance matrix :			Decomposition into factors		Decomposition into eigen vectors (whose variance is eigen values) :			
Variable 1 Variable 2			(normalized eigen vectors) :		Variable 1 Variable 2 Eigen Val. Var. Perc.			
Variable 1	0.1116	0.0224	Factor 1	0.3339	0.0696	Factor 1	0.9789	0.2042
Variable 2	0.0224	0.0090	Factor 2	-0.0134	0.0641	Factor 2	-0.2042	0.9789
							0.1163	96.45
							0.0043	3.55

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 20 km.

Tableau 25 – Statistiques sur l’erreur d’estimation, obtenues par validation croisée sur les concentrations

S3h (104 points)	Moyenne	Variance
Erreur	-0,00800	0,02685
Erreur standardisée	-0,04527	1,07717

2 données sur 104 non robustes.

S3h (102 points)	Moyenne	Variance
Erreur	-0,00781	0,02335
Erreur standardisée	-0,04719	0,94940

Toutes les données sont robustes.

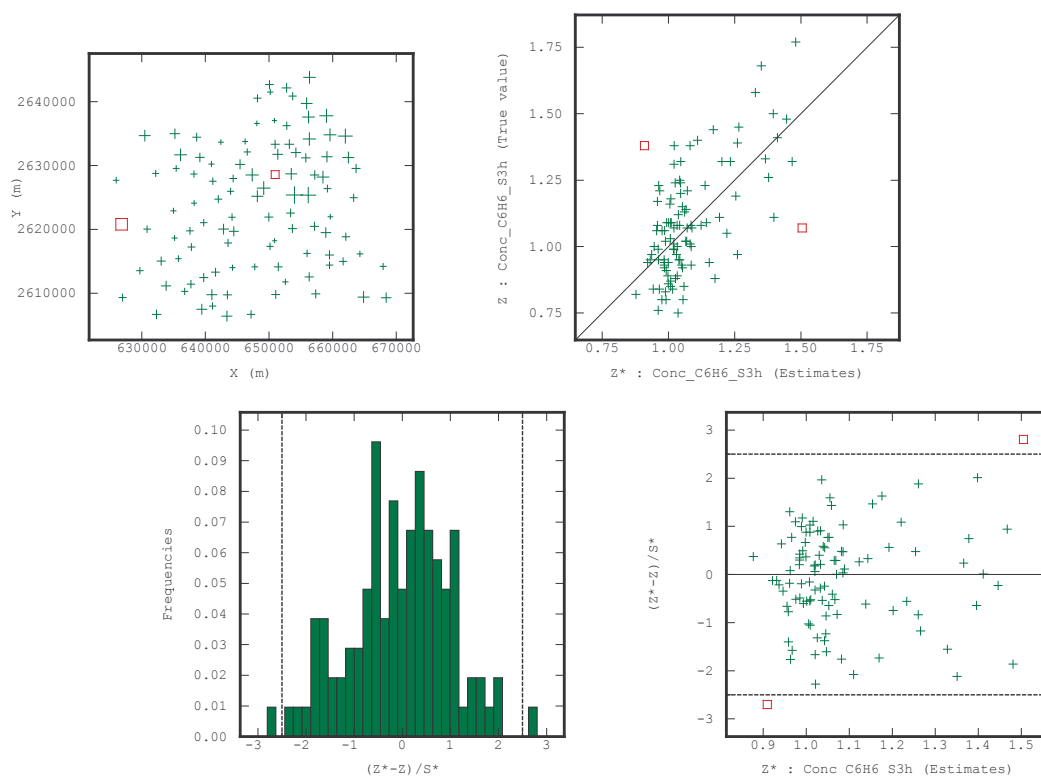


Figure 59 – Résultats de la validation croisée

13.2.3 TEST DE VALIDATION ABSOLUE SUR LE COMPLEMENTAIRE C3h

Tableau 26 – Résultats de la validation absolue sur le complémentaire C3h pour le krigeage ponctuel colocalisé

		Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S3h	
C3h	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	104	104	104
Minimum	0,75	0,90	0,08
Percentile 25	1,05	1,07	0,13
Médiane	1,245	1,20	0,14
Percentile 75	1,46	1,37	0,15
Maximum	1,92	1,68	0,16
Etendue	1,17	0,78	0,08
Moyenne	1,28	1,23	0,14
Ecart-type	0,086	0,19	0,014
Variance	0,293	0,04	0,0002
CV%	22,8	15,6	10,2
Corrélation mesures/estimations	r = 0,592		

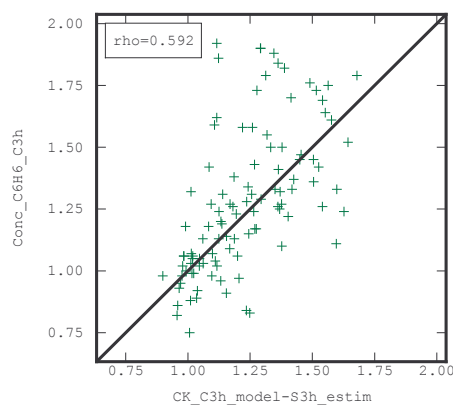


Figure 60 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C3h

Le cokrigage colocalisé donne des résultats satisfaisants tant au niveau de la validation croisée que de la validation absolue pour les deux saisons. Nous allons donc utiliser cette méthode pour établir de nouvelles cartes de référence et étudier la dégradation du maillage.

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

14.1 Cas de l'été 2003

14.1.1 ETUDE DES VARIOGRAMMES EXPERIMENTAUX

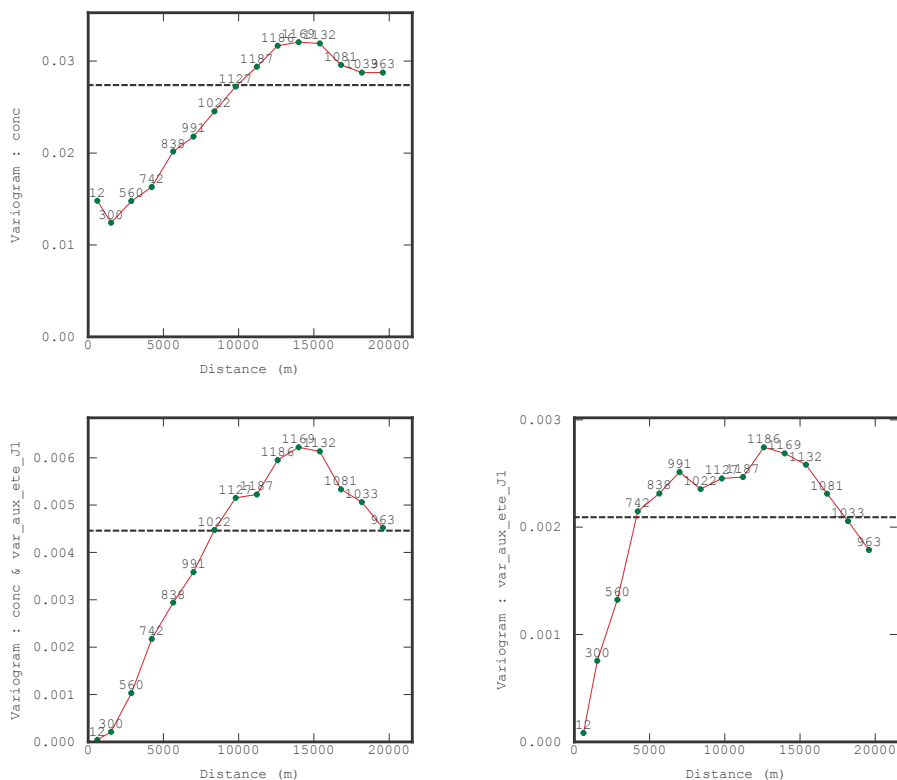


Figure 61 – Variogrammes expérimentaux simples et croisé (194 sites ; 15 pas ; 1400 m)

14.1.2 MODELISATION DES VARIOGRAMMES ET VALIDATION CROISEE

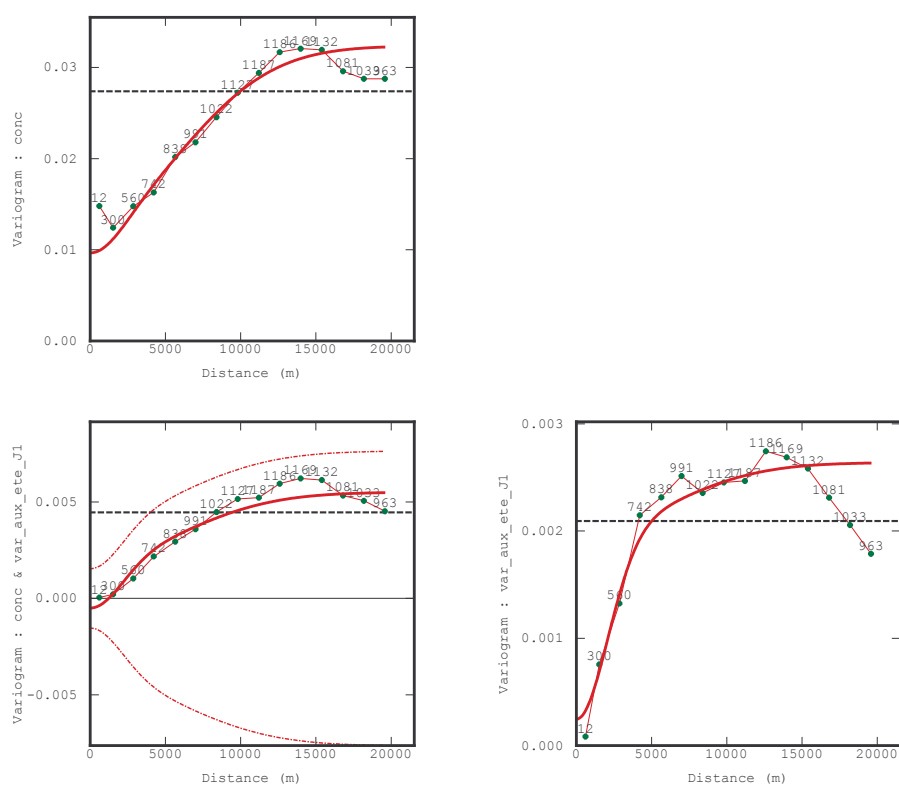


Figure 62 – Variogrammes simulés simples et croisé

Number of variables = 2
 - Variable 1 : var_aux_ete_J1
 - Variable 2 : conc
 Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix: Variable 1 Variable 2 Variable 1 0.0002 -0.0005 Variable 2 -0.0005 0.0097	Decomposition into factors (normalized eigen vectors): Variable 1 Variable 2 Factor 1 -0.0052 0.0983 Factor 2 0.0148 0.0008	Decomposition into eigen vectors (whose variance is eigen values): Variable 1 Variable 2 Eigen Val. Var. Perc. Factor 1 -0.0529 0.9986 0.0097 97.79 Factor 2 0.9986 0.0529 0.0002 2.21
--	---	---

S2 : Gaussian - Scale = 15000m

Variance-Covariance matrix: Variable 1 Variable 2 Variable 1 0.0006 0.0034 Variable 2 0.0034 0.0187	Decomposition into factors (normalized eigen vectors): Variable 1 Variable 2 Factor 1 0.0249 0.1367 Factor 2 0.0045 -0.0008	Decomposition into eigen vectors (whose variance is eigen values): Variable 1 Variable 2 Eigen Val. Var. Perc. Factor 1 0.1791 0.9838 0.0193 99.89 Factor 2 0.9838 -0.1791 0.0000 0.11
--	---	---

S3 : Gaussian - Scale = 5000m

Variance-Covariance matrix : Variable 1 Variable 2 Variable 1 0.0018 0.0026 Variable 2 0.0026 0.0040	Decomposition into factors (normalized eigen vectors) : Variable 1 Variable 2 Factor 1 0.0415 0.0631 Factor 2 0.0055 -0.0036	Decomposition into eigen vectors (whose variance is eigen values) : Variable 1 Variable 2 Eigen Val. Var. Perc. Factor 1 0.5495 0.8355 0.0057 99.25 Factor 2 0.8355 -0.5495 0.0000 0.75
---	--	--

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 20 km.

Tableau 27 – Statistiques sur l’erreur d’estimation, obtenues par validation croisée sur les concentrations

J1e (194 points)	Moyenne	Variance
Erreur	-0,00207	0,01279
Erreur standardisée	-0,02252	1,13271

1 donnée sur 194 non robuste.

J1e (193 points)	Moyenne	Variance
Erreur	-0,00072	0,01251
Erreur standardisée	-0,00888	1,10247

Toutes les données sont robustes.

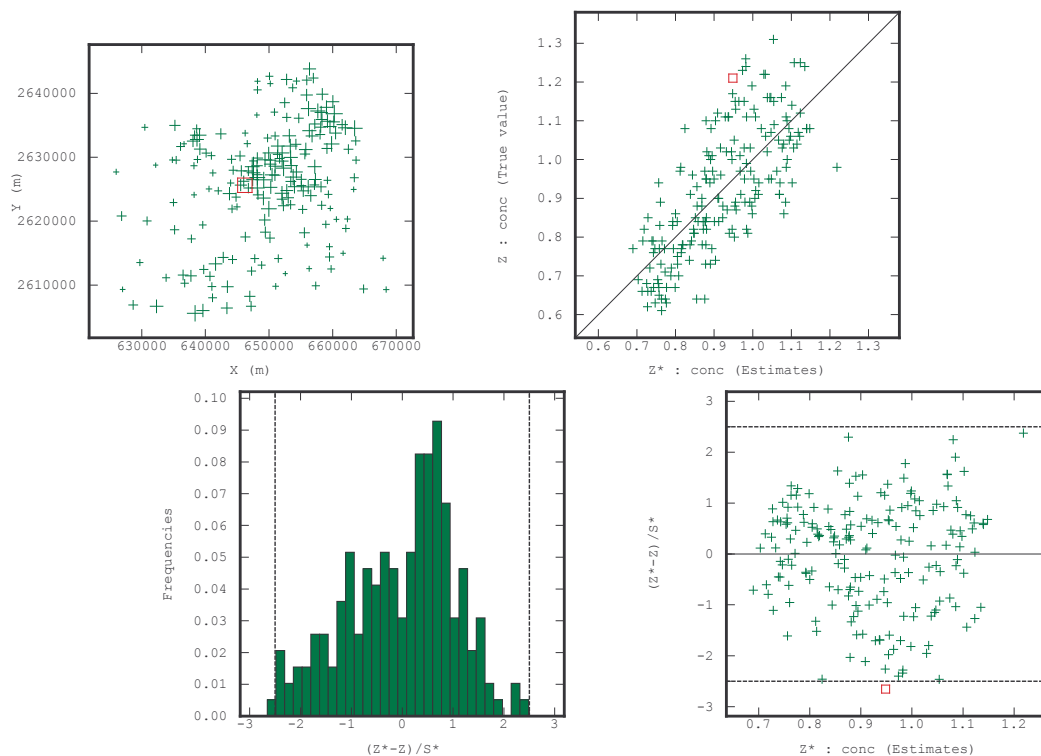


Figure 63 – Résultats de la validation croisée

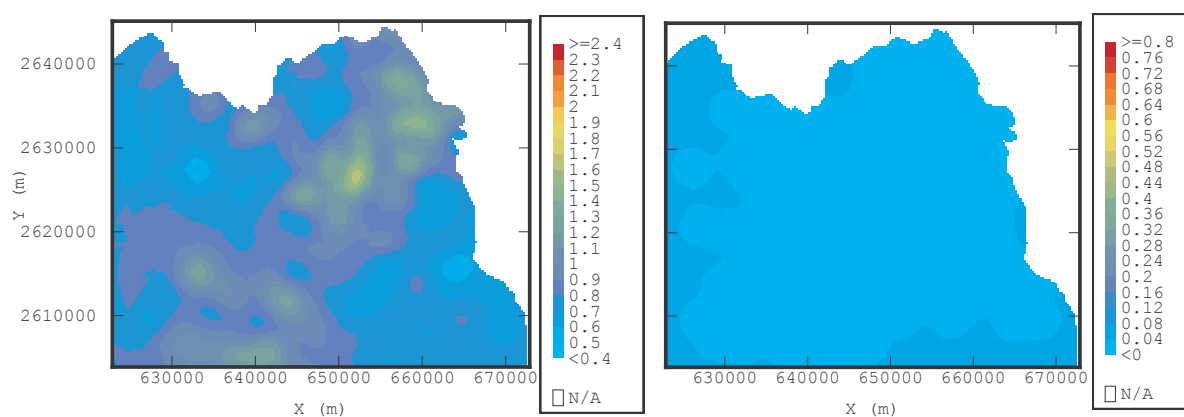
La corrélation entre données mesurées et données estimées ($r = 0,730$) est équivalente à celle obtenue pour le même jeu de données sans variable auxiliaire ($r = 0,727$).

14.1.3 ÉTABLISSEMENT DE LA CARTE DE REFERENCE

On estime les concentrations en tous points de la carte de référence à l’aide du jeu de données J_0 , de la variable colocalisée et du modèle bivariable, en considérant un voisinage de 20 km.

**Tableau 28 – Statistiques sur l'estimation sur une grille régulière de 250 m de côté,
par krigeage ponctuel colocalisé**

$J_{0,e}$	Mesures ($\mu\text{g m}^{-3}$)	Estimations par krigeage ponctuel ($\mu\text{g m}^{-3}$)	Ecart-types de krigeage ($\mu\text{g m}^{-3}$)
Nombre de données	218	25467	25467
Minimum	0,49	0,53	0,02
Percentile 25	0,81	0,72	0,02
Médiane	0,94	0,81	0,03
Percentile 75	1,10	0,92	0,04
Maximum	1,99	1,67	0,06
Etendue	1,50	1,14	0,04
Moyenne	0,98	0,85	0,03
Ecart-type	0,065	0,17	0,01
Variance	0,254	0,03	0,0001
CV%	26,0	19,8	32,9



**Figure 64 - (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données $J_{0,e}$,
et (droite) écarts-types de krigeage**

Par rapport à la carte de référence établie sans variable auxiliaire, les estimations maximales sont plus élevées. De plus, les valeurs absolues des écarts-types de krigeage sont significativement réduites.

14.2 Cas de l'hiver 2004

14.2.1 ETUDE DES VARIOGRAMMES EXPERIMENTAUX

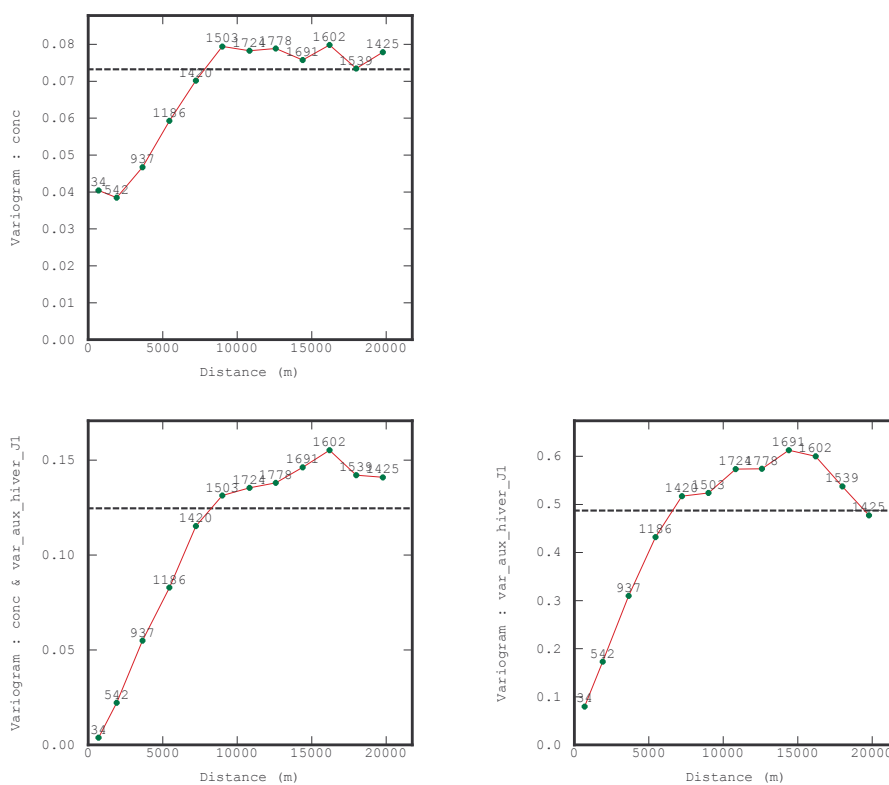


Figure 65 – Variogrammes expérimentaux simples et croisé (208 sites ; 12 pas ; 1800 m)

14.2.2 MODELISATION DES VARIOGRAMMES ET VALIDATION CROISEE

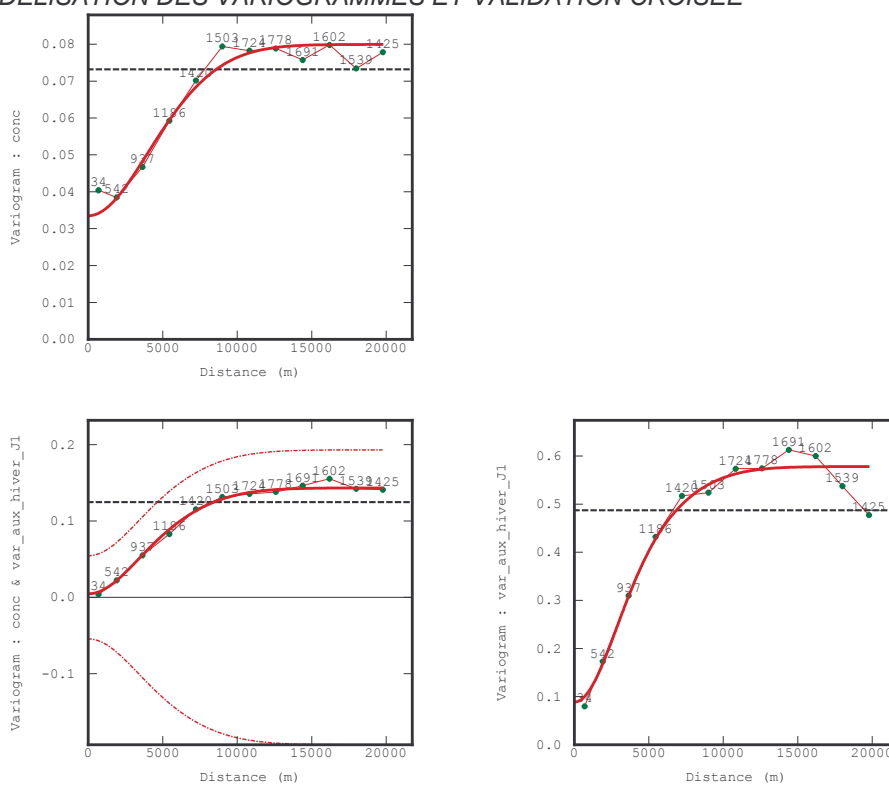


Figure 66 – Variogrammes simulés simples et croisé

Number of variables = 2
 - Variable 1 : var_aux_hiver_J1
 - Variable 2 : conc
 Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix: Variable 1 Variable 2		Decomposition into factors (normalized eigen vectors):		Decomposition into eigen vectors (whose variance is eigen values):			
Variable 1	0.0886	0.0046	Variable 1	Variable 2	Variable 1	Variable 2	Eigen Val. Var. Perc.
Variable 2	0.0046	0.0335	Factor 1	0.2974	0.0245	Factor 1	0.9966 0.0821 0.0890 72.88
			Factor 2	-0.0149	0.1814	Factor 2	-0.0821 0.9966 0.0331 27.12

S2 : Gaussian - Scale = 11000m

Variance-Covariance matrix: Variable 1 Variable 2		Decomposition into factors (normalized eigen vectors):		Decomposition into eigen vectors (whose variance is eigen values):			
Variable 1	0.2763	0.1077	Variable 1	Variable 2	Variable 1	Variable 2	Eigen Val. Var. Perc.
Variable 2	0.1077	0.0420	Factor 1	0.5257	0.2048	Factor 1	0.9318 0.3631 0.3183 100.00
			Factor 2	0.0000	0.0000	Factor 2	-0.3631 0.9318 0.0000 0.00

S3 : Gaussian - Scale = 6000m

Variance-Covariance matrix: Variable 1 Variable 2		Decomposition into factors (normalized eigen vectors):		Decomposition into eigen vectors (whose variance is eigen values):			
Variable 1	0.2129	0.0309	Variable 1	Variable 2	Variable 1	Variable 2	Eigen Val. Var. Perc.
Variable 2	0.0309	0.0045	Factor 1	0.4614	0.0670	Factor 1	0.9896 0.1437 0.2174 100.00
			Factor 2	0.0000	0.0000	Factor 2	-0.1437 0.9896 0.0000 0.00

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 20 km.

Tableau 29 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les concentrations

J1h (208 points)	Moyenne	Variance
Erreur	-0,00610	0,03066
Erreur standardisée	-0,03072	0,82268

Toutes les données sont robustes.

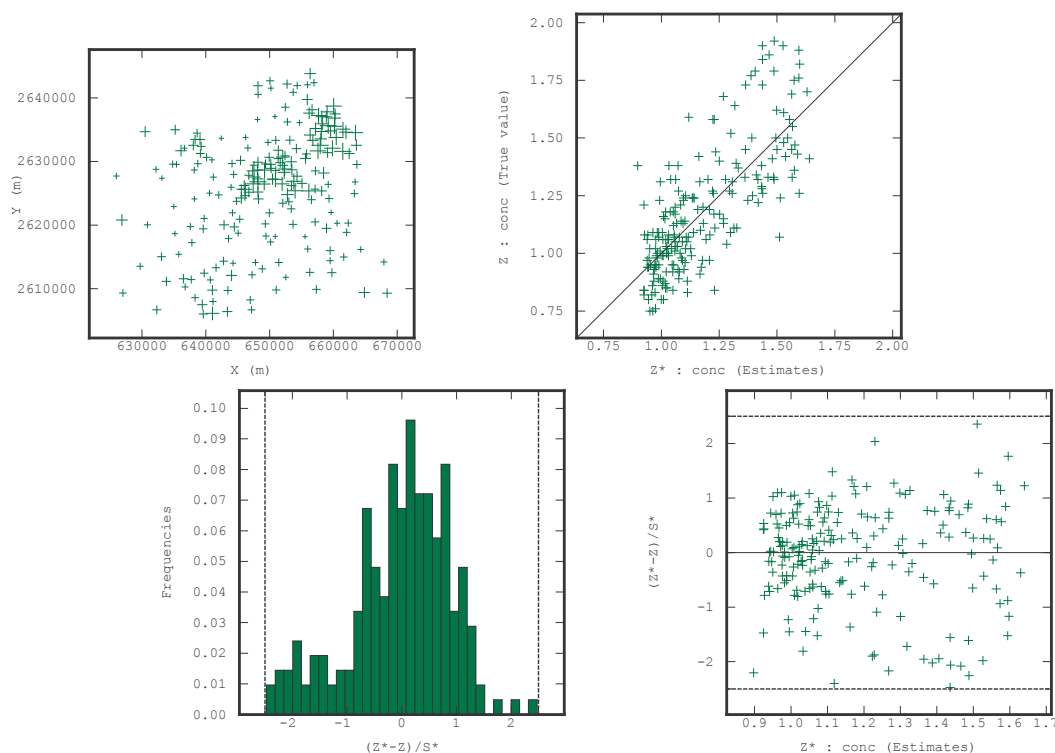


Figure 67 – Résultats de la validation croisée

14.2.3 ETABLISSEMENT DE LA CARTE DE REFERENCE

On estime les concentrations en tous points de la carte de référence à l'aide du jeu de données J_0 , de la variable colocalisée et du modèle bivariable, en considérant un voisinage de 20 km.

Tableau 30 – Statistiques sur l'estimation sur une grille régulière de 250 m de côté, par krigeage ponctuel colocalisé

$J_{0,h}$	Mesures ($\mu\text{g m}^{-3}$)	Estimations par krigeage ponctuel ($\mu\text{g m}^{-3}$)	Ecart-types de krigeage ($\mu\text{g m}^{-3}$)
Nombre de données	215	25467	25467
Minimum	0,75	0,27	0,31
Percentile 25	0,99	0,87	0,36
Médiane	1,13	1,02	0,39
Percentile 75	1,37	1,14	0,48
Maximum	2,68	2,12	0,80
Etendue	1,93	1,85	0,49
Moyenne	1,22	1,04	0,43
Ecart-type	0,103	0,25	0,10
Variance	0,321	0,06	0,01
CV%	26,4	24,2	23,3

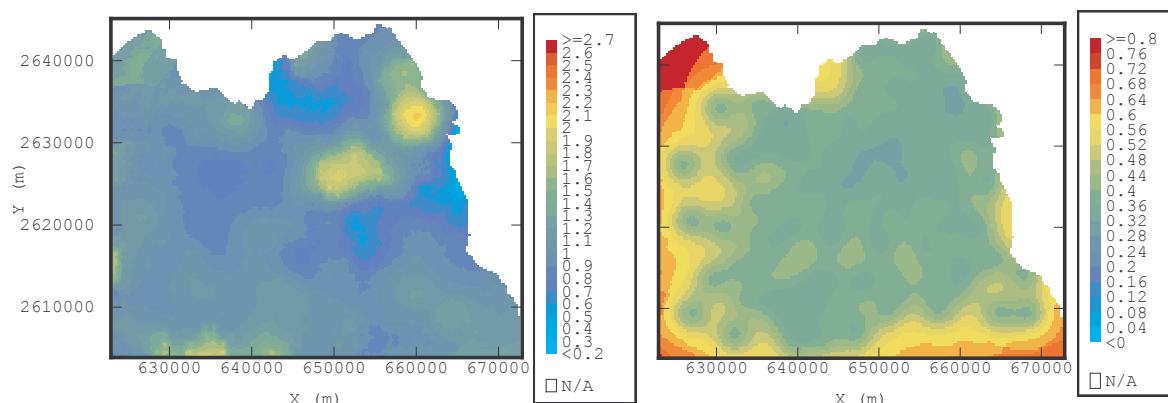


Figure 68 - (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données $J_{0,h}$, et (droite) écarts-types de krigeage

Les valeurs maximales obtenues par krigeage colocalisé sont plus élevées qu'en krigeage sans variable auxiliaire. Les concentrations de benzène sur Lille sont également plus marquées. Les écarts-types de krigeage sont plus importants qu'en krigeage ordinaire.

A partir de ces deux jeux de référence, nous allons étudier la dégradation du maillage sur les mailles de 5 et 7 km de côté.

15. ETUDE DE LA SELECTION DE DONNEES S5 (MAILLE DE 5 × 5 KM)

15.1 Cas de l'été 2003

15.1.1 ETUDE DES VARIOGRAMMES EXPERIMENTAUX

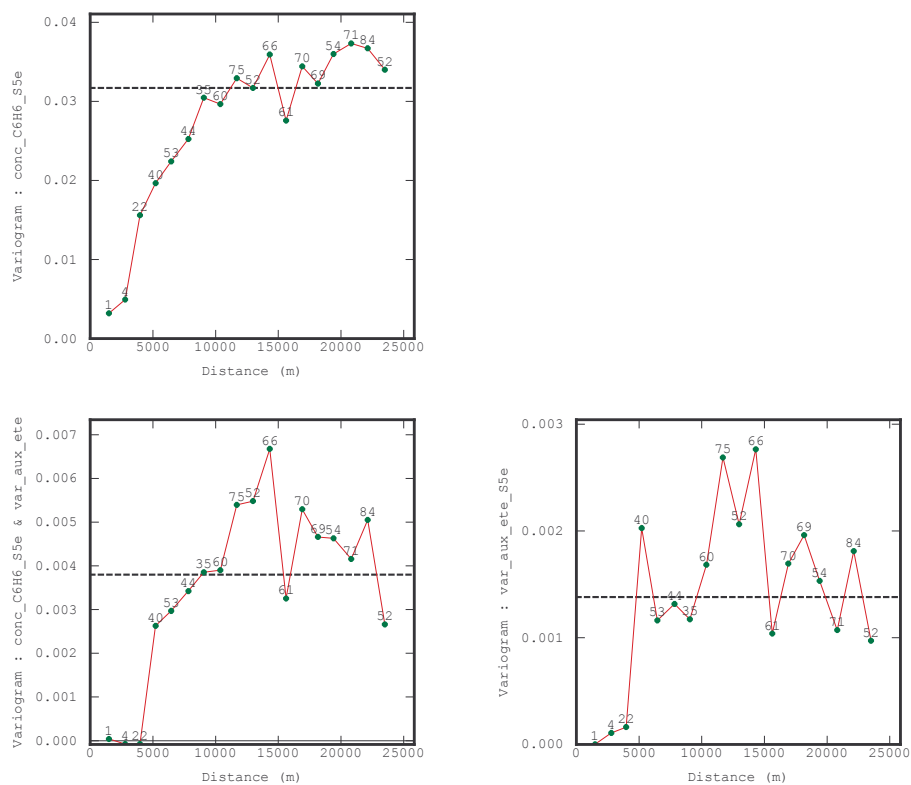


Figure 69 – Variogrammes expérimentaux simples et croisé (53 sites ; 19 pas ; 1300 m)

15.1.2 MODELISATION DES VARIOGRAMMES ET VALIDATION CROISEE

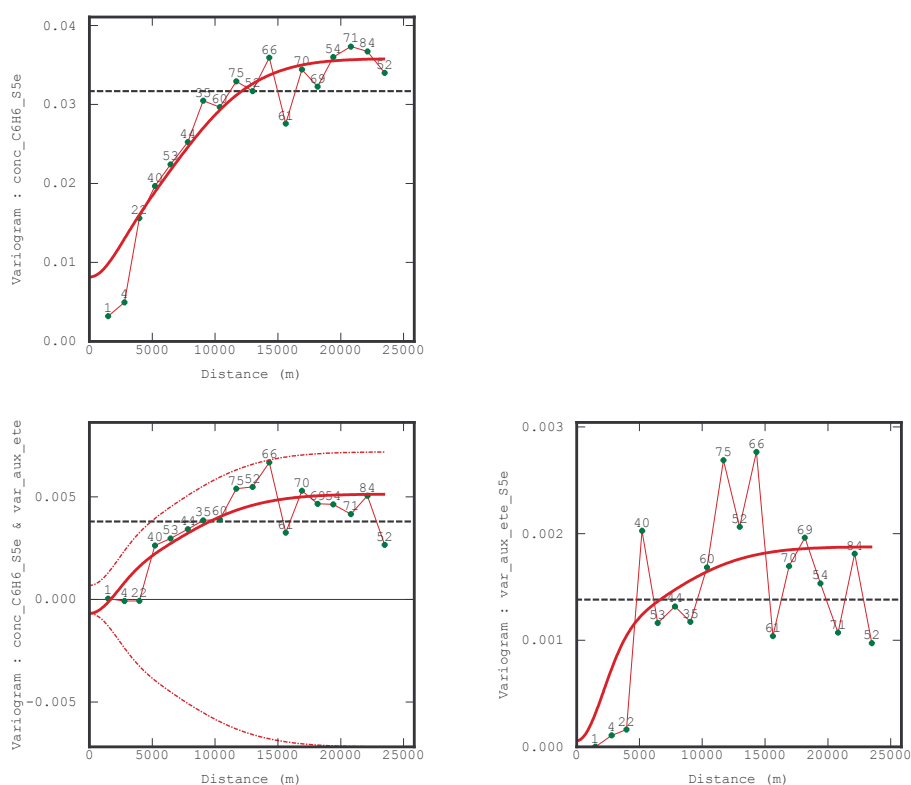


Figure 70 – Variogrammes simulés simples et croisé

Number of variables = 2
 - Variable 1 : var_aux_ete_S5e
 - Variable 2 : conc_C6H6_S5e
 Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix: Variable 1 Variable 2 Variable 1 0.0001 -0.0007 Variable 2 -0.0007 0.0081	Decomposition into factors (normalized eigen vectors):		Decomposition into eigen vectors (whose variance is eigen values):			
	Variable 1 Variable 2		Variable 1 Variable 2 Eigen Val. Var. Perc.			
	Factor 1	-0.0075 0.0903	Factor 1	-0.0833 0.9965 0.0082 100.00		
	Factor 2	0.0000 0.0000	Factor 2	0.9965 0.0833 0.0000 0.00		

S2 : Gaussian - Scale = 16000m

Variance-Covariance matrix: Variable 1 Variable 2 Variable 1 0.0008 0.0039 Variable 2 0.0039 0.0230	Decomposition into factors (normalized eigen vectors):		Decomposition into eigen vectors (whose variance is eigen values):			
	Variable 1 Variable 2		Variable 1 Variable 2 Eigen Val. Var. Perc.			
	Factor 1	0.0259 0.1515	Factor 1	0.1687 0.9857 0.0236 99.34		
	Factor 2	0.0123 -0.0021	Factor 2	0.9857 -0.1687 0.0002 0.66		

S3 : Gaussian - Scale = 5000m

Variance-Covariance matrix : Variable 1 Variable 2 Variable 1 0.0010 0.0019 Variable 2 0.0019 0.0047	Decomposition into factors (normalized eigen vectors) :		Decomposition into eigen vectors (whose variance is eigen values) :			
	Variable 1 Variable 2		Variable 1 Variable 2 Eigen Val. Var. Perc.			
	Factor 1	0.0290 0.0683	Factor 1	0.3903 0.9207 0.0055 96.75		
	Factor 2	0.0125 -0.0053	Factor 2	0.9207 -0.3903 0.0002 3.25		

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 20 km.

Tableau 31 – Statistiques sur l’erreur d’estimation, obtenues par validation croisée sur les concentrations

S5e (53 points)	Moyenne	Variance
Erreur	-0,00014	0,01451
Erreur standardisée	-0,00245	0,91385

Toutes les données sont robustes.

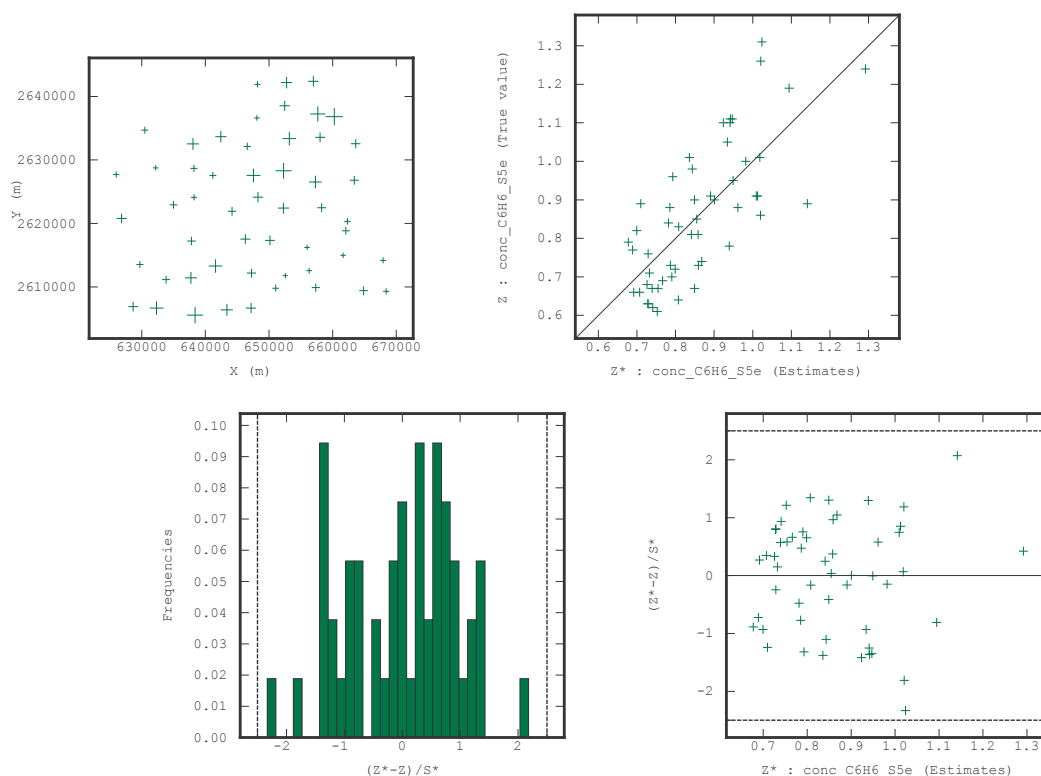


Figure 71 – Résultats de la validation croisée

La corrélation entre données mesurées et données estimées est de : $r = 0,736$.

15.1.3 TEST DE VALIDATION ABSOLUE SUR LE COMPLEMENTAIRE C5e

Tableau 32 – Résultats de la validation absolue sur le complémentaire C5e pour le krigeage ponctuel colocalisé, avec le modèle de la sélection S5e

		Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S5e	
C5e	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d’estimation ($\mu\text{g m}^{-3}$)
Nombre de données	141	141	141
Minimum	0,64	0,69	0,07
Percentile 25	0,82	0,85	0,11
Médiane	0,95	0,96	0,11
Percentile 75	1,07	1,07	0,12
Maximum	1,25	1,34	0,13
Etendue	0,61	0,65	0,06
Moyenne	0,94	0,97	0,11
Ecart-type	0,024	0,14	0,01
Variance	0,154	0,02	0,0001
CV%	16,3	15,0	6,7
Corrélation mesures/estimations		$r = 0,637$	

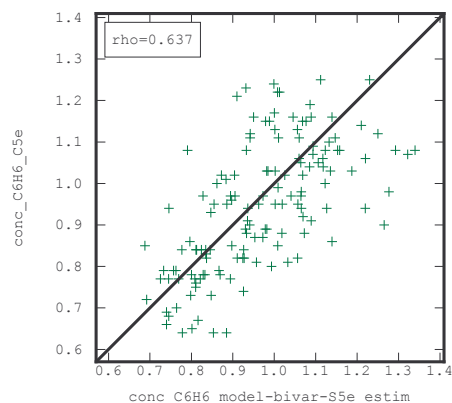


Figure 72 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C5e

15.1.4 ETABLISSEMENT DE LA CARTE DE REFERENCE

On estime les concentrations en tous points de la carte de référence à l'aide du jeu de données S5e, de la variable colocalisée et du modèle bivariable, en considérant un voisinage de 20 km.

Tableau 33 – Statistiques sur l'estimation sur une grille régulière de 250 m de côté, par krigeage ponctuel colocalisé

S5e	Mesures ($\mu\text{g m}^{-3}$)	Estimations par krigeage ponctuel ($\mu\text{g m}^{-3}$)	Ecart-types de krigeage ($\mu\text{g m}^{-3}$)
Nombre de données	53	25467	25467
Minimum	0,61	0,60	0,01
Percentile 25	0,71	0,73	0,11
Médiane	0,84	0,80	0,12
Percentile 75	0,96	0,91	0,12
Maximum	1,31	1,49	0,19
Etendue	0,70	0,89	0,18
Moyenne	0,86	0,84	0,12
Ecart-type	0,032	0,13	0,02
Variance	0,178	0,02	0,0004
CV%	20,7	15,5	16,2

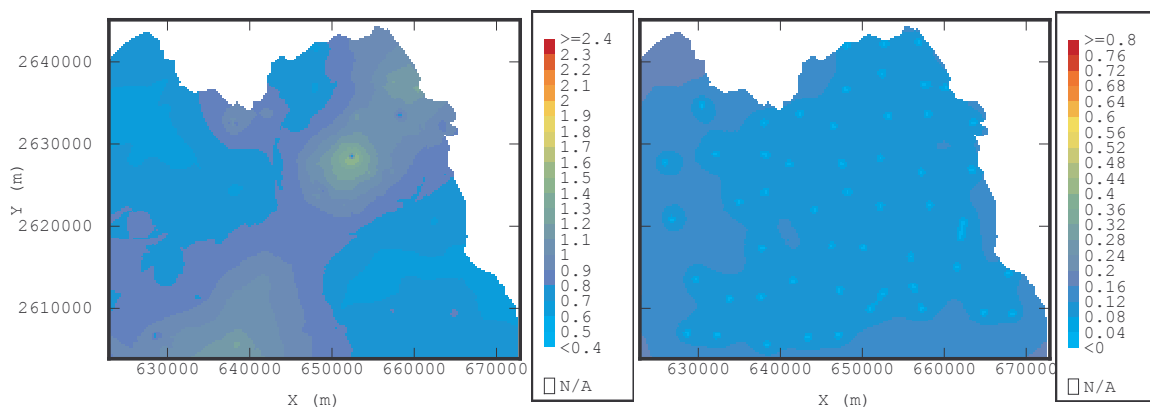


Figure 73 - (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données S5e, et (droite) écarts-types de krigeage

15.2 Cas de l'hiver 2004

15.2.1 ETUDE DES VARIOGRAMMES EXPERIMENTAUX

Les variogrammes sont omnidirectionnels avec 16 pas de 1300 m chacun.

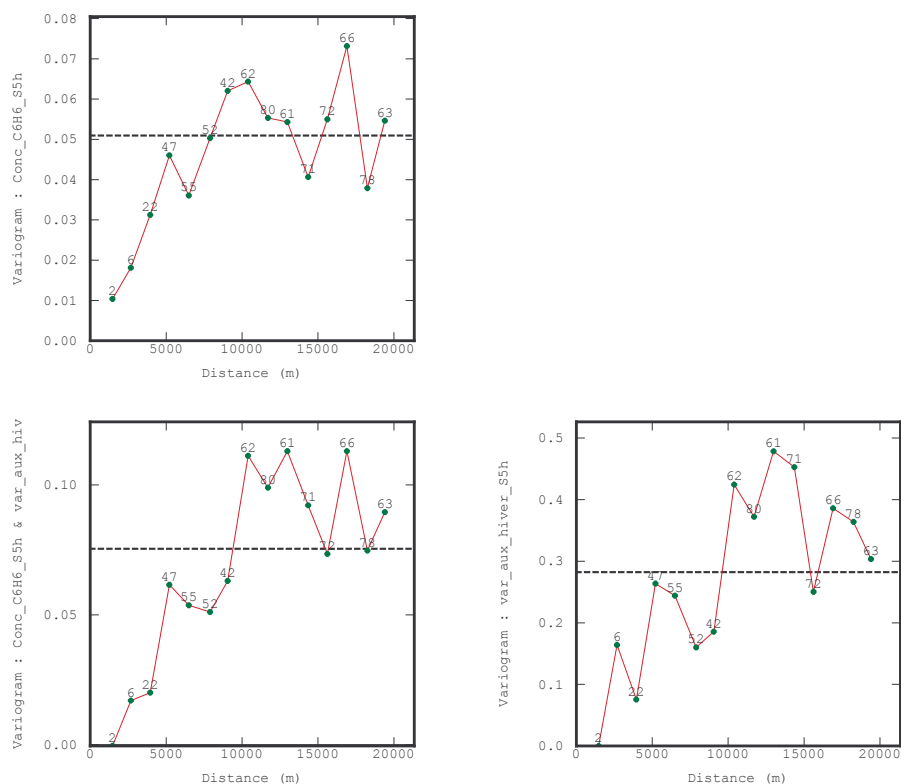


Figure 74 – Variogrammes expérimentaux simples et croisé (55 sites ; 16 pas ; 1300 m)

15.2.2 MODELISATION DES VARIOGRAMMES ET VALIDATION CROISEE

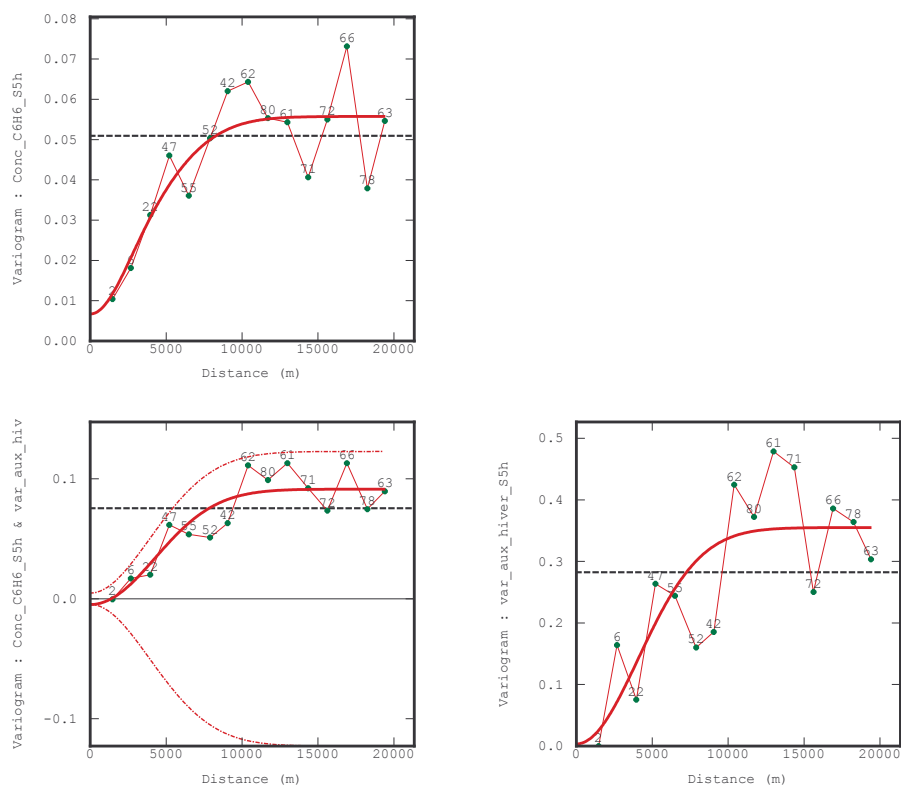


Figure 75 – Variogrammes simulés simples et croisé

Number of variables = 2
 - Variable 1 : var_aux_hiver_S5h
 - Variable 2 : Conc_C6H6_S5h
 Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix: Variable 1 Variable 2		Decomposition into factors (normalized eigen vectors): Variable 1 Variable 2		Decomposition into eigen vectors (whose variance is eigen values): Variable 1 Variable 2 Eigen Val. Var. Perc.			
Variable 1	0.0034	-0.0048	Factor 1	-0.0585	0.0818	Factor 1	-0.5818 0.8133 0.0101 100.00
Variable 2	-0.0048	0.0067	Factor 2	0.0000	0.0000	Factor 2	0.8133 0.5818 0.0000 0.00

S2 : Gaussian - Scale = 10000m

Variance-Covariance matrix: Variable 1 Variable 2		Decomposition into factors (normalized eigen vectors): Variable 1 Variable 2		Decomposition into eigen vectors (whose variance is eigen values): Variable 1 Variable 2 Eigen Val. Var. Perc.			
Variable 1	0.3503	0.0999	Factor 1	0.5913	0.1726	Factor 1	0.9599 0.2802 0.3795 97.91
Variable 2	0.0999	0.0373	Factor 2	-0.0252	0.0864	Factor 2	-0.2802 0.9599 0.0081 2.09

S3 : Gaussian - Scale = 5000m

Variance-Covariance matrix : Variable 1 Variable 2		Decomposition into factors (normalized eigen vectors) : Variable 1 Variable 2		Decomposition into eigen vectors (whose variance is eigen values) : Variable 1 Variable 2 Eigen Val. Var. Perc.			
Variable 1	0.0012	-0.0037	Factor 1	-0.0345	0.1085	Factor 1	-0.3030 0.9530 0.0130 100.00
Variable 2	-0.0037	0.0118	Factor 2	0.0000	0.0000	Factor 2	0.9530 0.3030 0.0000 0.00

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 20 km.

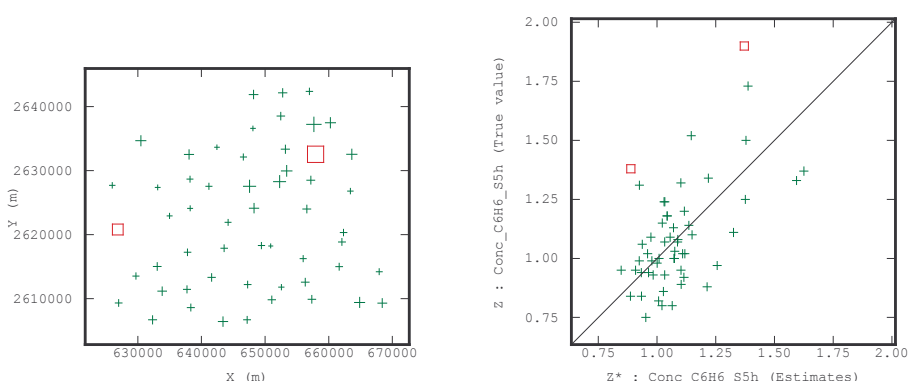
Tableau 34 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les concentrations

S5h (55 points)	Moyenne	Variance
Erreur	-0,00488	0,03606
Erreur standardisée	-0,02128	1,23180

2 données sur 55 non robustes.

S5h (53 points)	Moyenne	Variance
Erreur	0,01418	0,02742
Erreur standardisée	0,08513	0,96355

Toutes les données sont robustes.



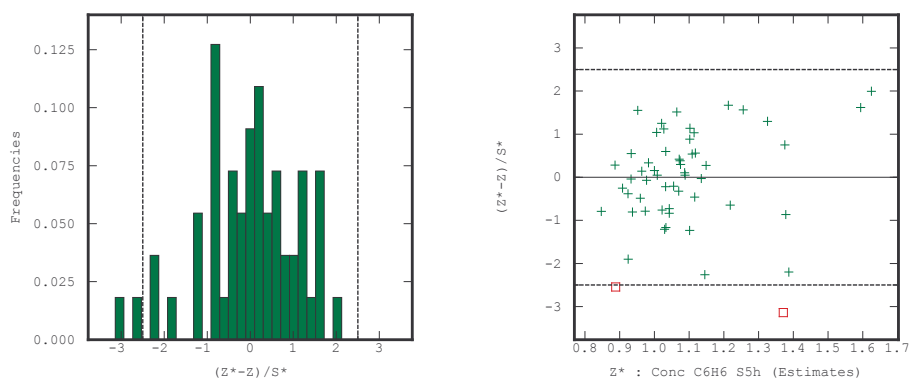


Figure 76 – Résultats de la validation croisée

La corrélation entre données mesurées et données estimées est de : $r = 0,563$.

15.2.3 TEST DE VALIDATION ABSOLUE SUR LE COMPLEMENTAIRE C5h

Tableau 35 – Résultats de la validation absolue sur le complémentaire C5h pour le krigeage ponctuel colocalisé, avec le modèle de la sélection S5h

		Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S5h	
C5h	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	153	153	153
Minimum	0,75	0,78	0,04
Percentile 25	1,01	0,99	0,11
Médiane	1,17	1,11	0,13
Percentile 75	1,38	1,42	0,14
Maximum	1,92	2,41	0,18
Etendue	1,17	1,63	0,14
Moyenne	1,22	1,23	0,13
Ecart-type	0,077	0,32	0,03
Variance	0,278	0,10	0,0007
CV%	22,8	25,7	21,0
Corrélation mesures/estimations		$r = 0,585$	

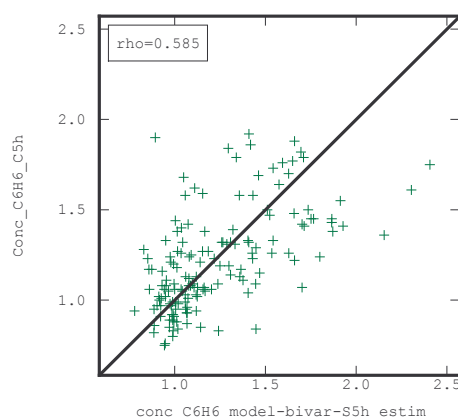


Figure 77 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C5h

15.2.4 ETABLISSEMENT DE LA CARTE DE REFERENCE

On estime les concentrations en tous points de la carte de référence à l'aide du jeu de données S5h, de la variable colocalisée et du modèle bivariable, en considérant un voisinage de 20 km.

Tableau 36 – Statistiques sur l'estimation sur une grille régulière de 250 m de côté, par krigeage ponctuel colocalisé

S5h	Mesures ($\mu\text{g m}^{-3}$)	Estimations par krigeage ponctuel ($\mu\text{g m}^{-3}$)	Ecart-types de krigeage ($\mu\text{g m}^{-3}$)
Nombre de données	55	25467	25467
Minimum	0,75	0,50	0,00
Percentile 25	0,95	0,98	0,12
Médiane	1,03	1,05	0,14
Percentile 75	1,19	1,16	0,16
Maximum	1,90	2,63	0,25
Etendue	1,15	2,13	0,25
Moyenne	1,09	1,09	0,14
Ecart-type	0,051	0,20	0,04
Variance	0,226	0,04	0,002
CV%	20,7	18,3	28,5

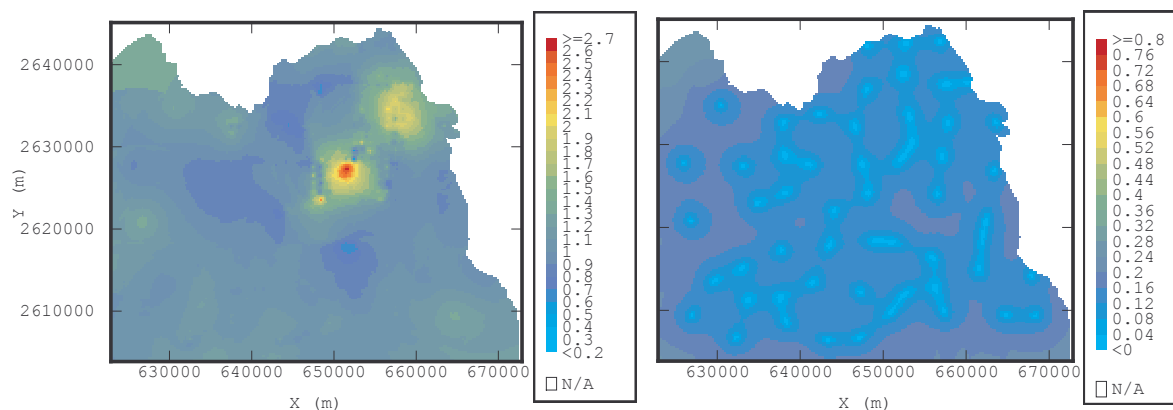


Figure 78 - (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données S5h, et (droite) écarts-types de krigeage

16. ETUDE DE LA SELECTION DE DONNEES S7 (MAILLE DE 7 × 7 KM)

16.1 Cas de l'été 2003

16.1.1 ETUDE DES VARIOGRAMMES EXPERIMENTAUX

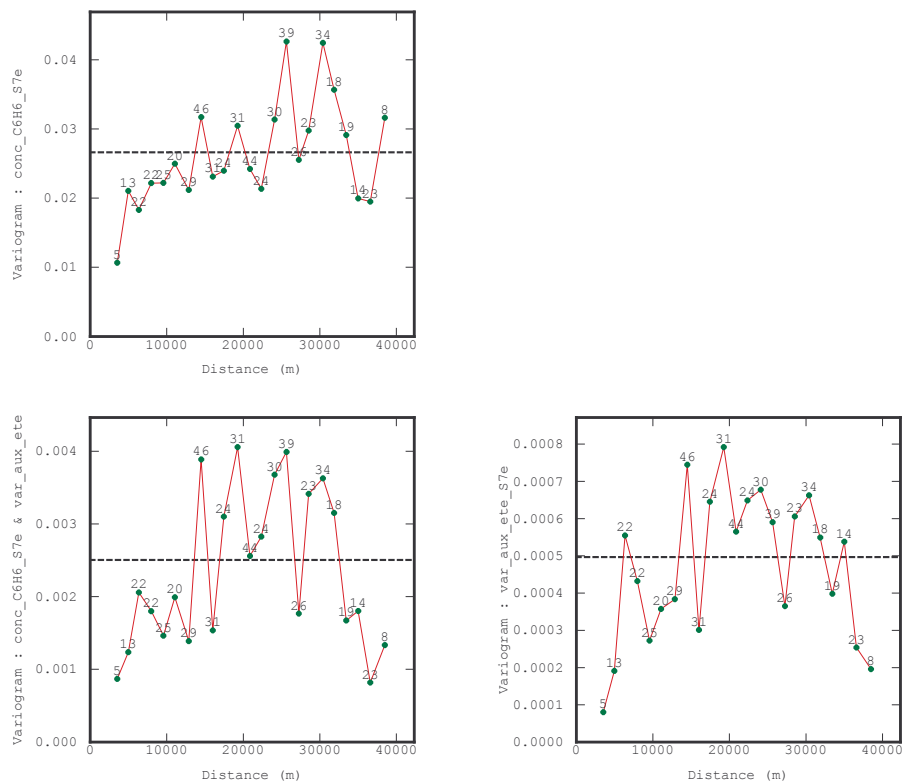


Figure 79 – Variogrammes expérimentaux simples et croisé (35 sites ; 25 pas ; 1600 m)

16.1.2 MODELISATION DES VARIOGRAMMES ET VALIDATION CROISEE

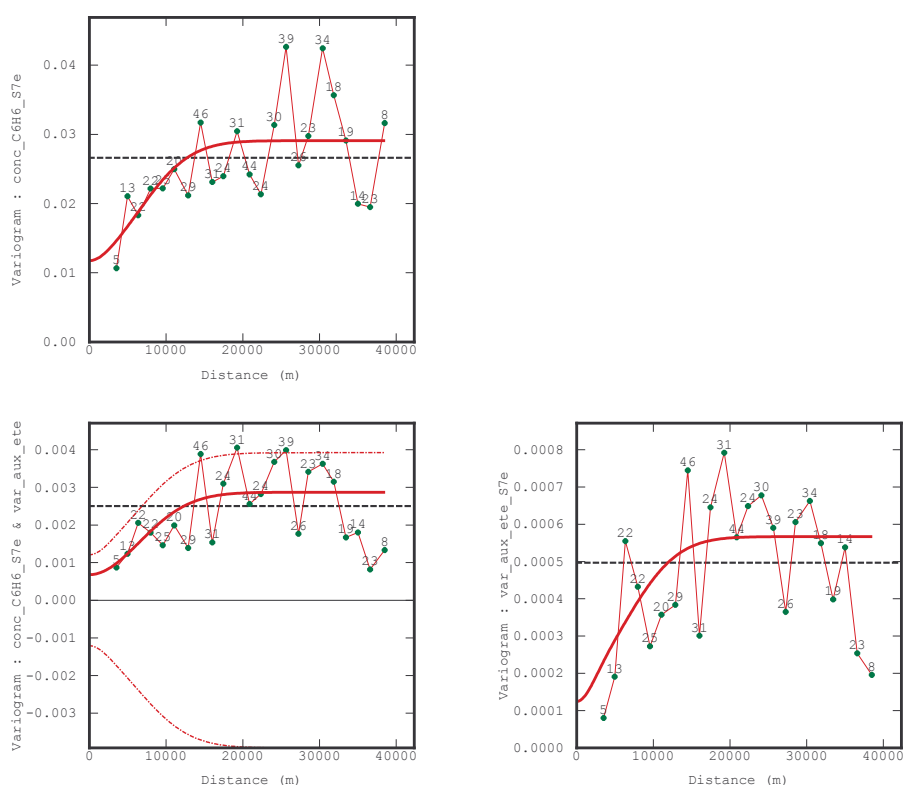


Figure 80 – Variogrammes simulés simples et croisé

Number of variables = 2
 - Variable 1 : var_aux_ete_S7e
 - Variable 2 : conc_C6H6_S7e
 Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix: Variable 1 Variable 2 Variable 1 0.0001 0.0007 Variable 2 0.0007 0.0118	Decomposition into factors (normalized eigen vectors): Variable 1 Variable 2 Factor 1 0.0063 0.1084 Factor 2 0.0092 -0.0005	Decomposition into eigen vectors (whose variance is eigen values): Variable 1 Variable 2 Eigen Val. Var. Perc. Factor 1 0.0583 0.9983 0.0118 99.29 Factor 2 0.9983 -0.0583 0.0001 0.71
--	---	---

S2 : Gaussian - Scale = 16000m

Variance-Covariance matrix: Variable 1 Variable 2 Variable 1 0.0004 0.0022 Variable 2 0.0022 0.0165	Decomposition into factors (normalized eigen vectors): Variable 1 Variable 2 Factor 1 0.0171 0.1284 Factor 2 0.0087 -0.0012	Decomposition into eigen vectors (whose variance is eigen values): Variable 1 Variable 2 Eigen Val. Var. Perc. Factor 1 0.1317 0.9913 0.0168 99.54 Factor 2 0.9913 -0.1317 0.0001 0.46
--	---	---

S3 : Gaussian - Scale = 5000m

Variance-Covariance matrix : Variable 1 Variable 2 Variable 1 0.0001 0.0000 Variable 2 0.0000 0.0008	Decomposition into factors (normalized eigen vectors) : Variable 1 Variable 2 Factor 1 0.0003 0.0291 Factor 2 0.0087 -0.0001	Decomposition into eigen vectors (whose variance is eigen values) : Variable 1 Variable 2 Eigen Val. Var. Perc. Factor 1 0.0117 0.9999 0.0008 91.82 Factor 2 0.9999 -0.0117 0.0001 8.18
---	--	--

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 20 km.

Tableau 37 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les concentrations

S7e (35 points)	Moyenne	Variance
Erreur	0,00370	0,01305
Erreur standardisée	0,01908	0,88341

Toutes les données sont robustes.

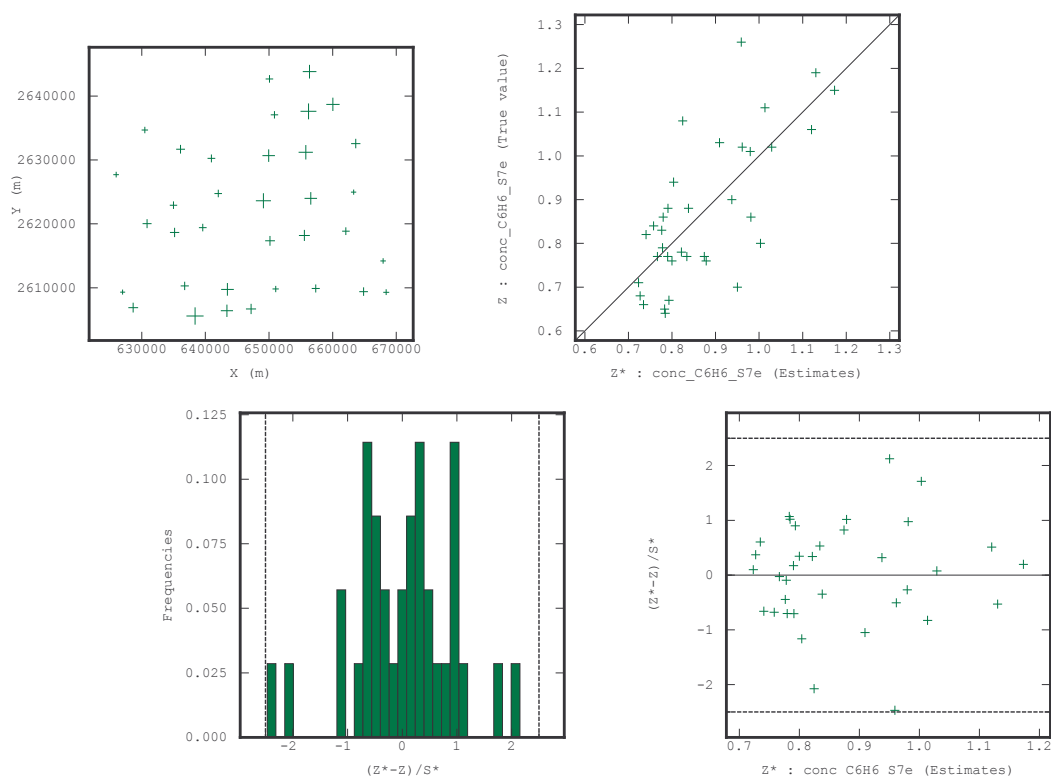


Figure 81 – Résultats de la validation croisée

La corrélation entre données mesurées et données estimées est de : $r = 0,715$.

16.1.3 TEST DE VALIDATION ABSOLUE SUR LE COMPLEMENTAIRE C7E

Tableau 38 – Résultats de la validation absolue sur le complémentaire C7e pour le krigeage ponctuel colocalisé, avec le modèle de la sélection S7e

C7e	Mesures ($\mu\text{g m}^{-3}$)	Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S7e	
		Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	159	159	159
Minimum	0,61	0,76	0,11
Percentile 25	0,82	0,84	0,11
Médiane	0,94	0,94	0,11
Percentile 75	1,07	1,12	0,11
Maximum	1,31	1,18	0,12
Etendue	0,70	0,42	0,01
Moyenne	0,93	1,00	0,11
Ecart-type	0,027	0,21	0,003
Variance	0,164	0,05	0,0000
CV%	17,6	21,4	2,4
Corrélation mesures/estimations		$r = 0,598$	

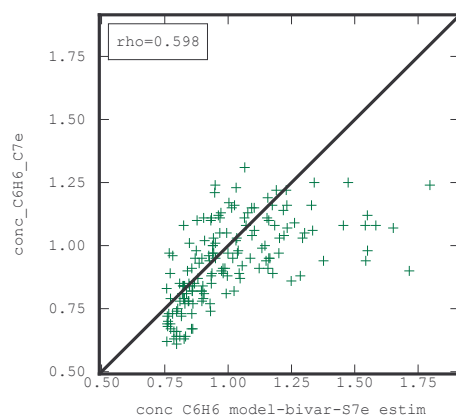


Figure 82 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C7e

16.1.4 ÉTABLISSEMENT DE LA CARTE DE REFERENCE

On estime les concentrations en tous points de la carte de référence à l'aide du jeu de données S7e, de la variable colocalisée et du modèle bivariable, en considérant un voisinage de 20 km.

Tableau 39 – Statistiques sur l'estimation sur une grille régulière de 250 m de côté, par krigeage ponctuel colocalisé

S7e	Mesures ($\mu\text{g m}^{-3}$)	Estimations par krigeage ponctuel ($\mu\text{g m}^{-3}$)	Ecart-types de krigeage ($\mu\text{g m}^{-3}$)
Nombre de données	35	25467	25467
Minimum	0,64	0,64	0,10
Percentile 25	0,77	0,75	0,11
Médiane	0,83	0,80	0,11
Percentile 75	1,02	0,89	0,12
Maximum	1,26	2,12	0,16
Etendue	0,62	1,48	0,06
Moyenne	0,87	0,84	0,12
Ecart-type	0,027	0,15	0,009
Variance	0,163	0,02	0,0001
CV%	18,8	17,3	7,9

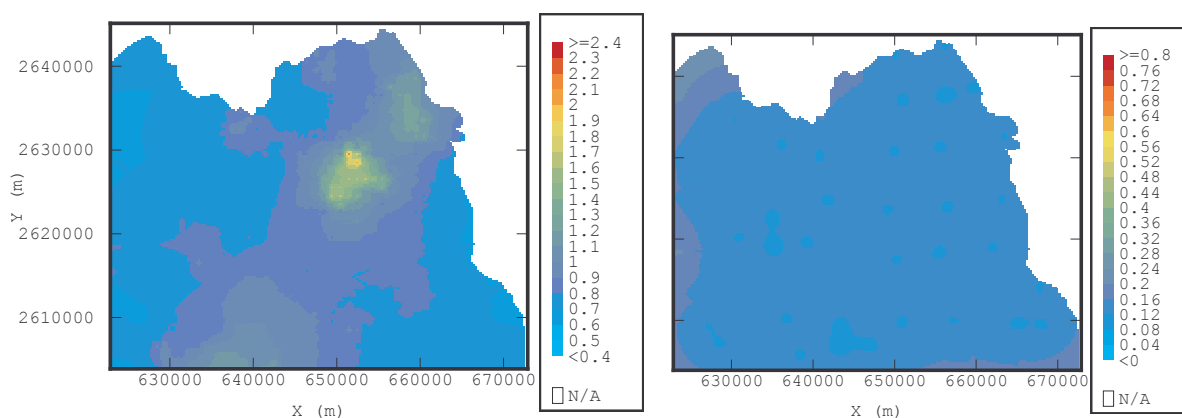


Figure 83 - (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données S7e, et (droite) écarts-types de krigeage

16.2 Cas de l'hiver 2004

16.2.1 ETUDE DES VARIOGRAMMES EXPERIMENTAUX

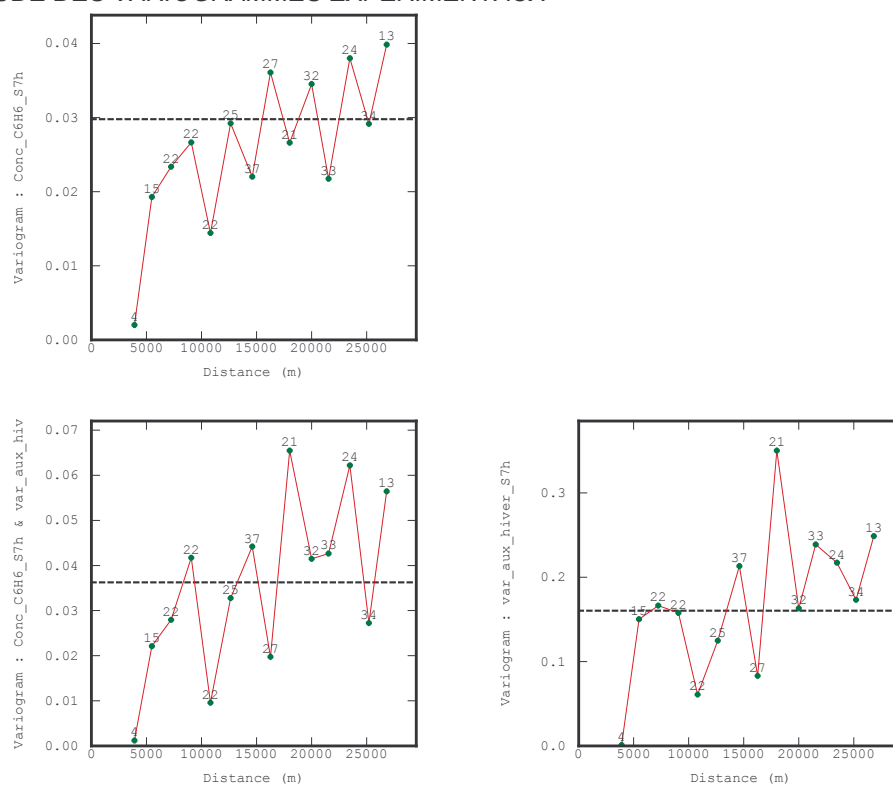


Figure 84 – Variogrammes expérimentaux simples et croisé (31 sites ; 16 pas ; 1800 m)

16.2.2 MODELISATION DES VARIOGRAMMES ET VALIDATION CROISEE

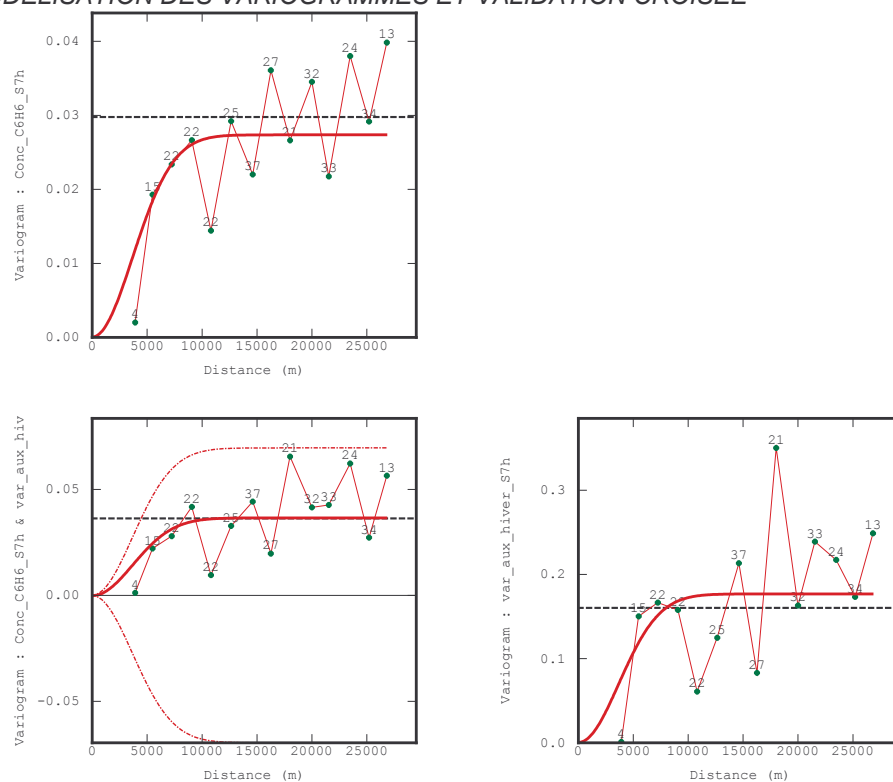


Figure 85 – Variogrammes simulés simples et croisé

Number of variables = 2

- Variable 1 : var_aux_hiver_S7h
 - Variable 2 : Conc_C6H6_S7h
 Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix: Variable 1 Variable 2 Variable 1 0.0004 -0.0001 Variable 2 -0.0001 0.0000	Decomposition into factors (normalized eigen vectors): Variable 1 Variable 2 Factor 1 0.0210 -0.0070 Factor 2 0.0000 0.0000	Decomposition into eigen vectors (whose variance is eigen values): Variable 1 Variable 2 Eigen Val. Var. Perc. Factor 1 0.9487 -0.3163 0.0005 100.00 Factor 2 0.3163 0.9487 0.0000 0.00
--	---	--

S2 : Gaussian - Scale = 9000m

Variance-Covariance matrix: Variable 1 Variable 2 Variable 1 0.1761 0.0367 Variable 2 0.0367 0.0273	Decomposition into factors (normalized eigen vectors): Variable 1 Variable 2 Factor 1 0.4185 0.0977 Factor 2 -0.0311 0.1332	Decomposition into eigen vectors (whose variance is eigen values): Variable 1 Variable 2 Eigen Val. Var. Perc. Factor 1 0.9738 0.2273 0.1847 90.80 Factor 2 -0.2273 0.9738 0.0187 9.20
--	---	---

S3 : Gaussian - Scale = 5000m

Variance-Covariance matrix : Variable 1 Variable 2 Variable 1 0.0002 -0.0001 Variable 2 -0.0001 0.0000	Decomposition into factors (normalized eigen vectors) : Variable 1 Variable 2 Factor 1 0.0144 -0.0042 Factor 2 0.0000 0.0000	Decomposition into eigen vectors (whose variance is eigen values) : Variable 1 Variable 2 Eigen Val. Var. Perc. Factor 1 0.9594 -0.2822 0.0002 100.00 Factor 2 0.2822 0.9594 0.0000 0.00
---	--	---

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 20 km.

Tableau 40 – Statistiques sur l’erreur d’estimation, obtenues par validation croisée sur les concentrations

S7h (34 points)	Moyenne	Variance
Erreur	-0,00569	0,03631
Erreur standardisée	-0,04448	2,16713

3 données sur 34 non robustes.

S7h (31 points)	Moyenne	Variance
Erreur	0,01128	0,02222
Erreur standardisée	0,09073	1,24974

Toutes les données sont robustes.

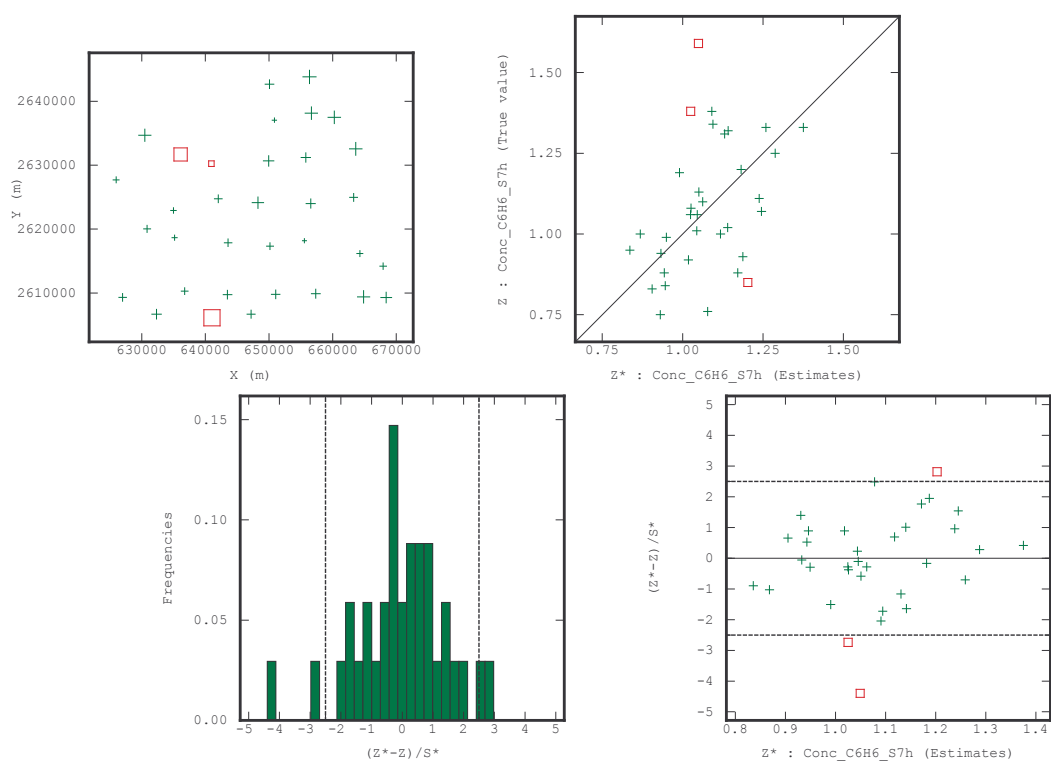


Figure 86 – Résultats de la validation croisée

La corrélation entre données mesurées et données estimées est de : $r = 0,396$.

16.2.3 TEST DE VALIDATION ABSOLUE SUR LE COMPLEMENTAIRE C7h

Tableau 41 – Résultats de la validation absolue sur le complémentaire C7h pour le krigeage ponctuel colocalisé, avec le modèle de la sélection S7h

		Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S7h	
C7h	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	174	174	174
Minimum	0,75	0,79	0,02
Percentile 25	0,99	0,97	0,06
Médiane	1,14	1,11	0,08
Percentile 75	1,37	1,33	0,09
Maximum	1,92	1,62	0,13
Etendue	1,17	0,83	0,11
Moyenne	1,20	1,14	0,07
Ecart-type	0,077	0,20	0,02
Variance	0,278	0,04	0,0005
CV%	23,1	17,7	29,8
Corrélation mesures/estimations		$r = 0,624$	

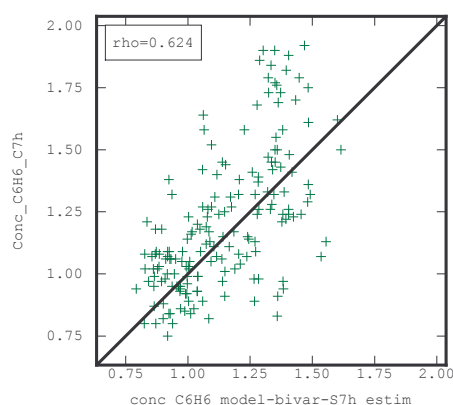


Figure 87 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C7h

16.2.4 ÉTABLISSEMENT DE LA CARTE DE REFERENCE

On estime les concentrations en tous points de la carte de référence à l'aide du jeu de données S7h, de la variable colocalisée et du modèle bivariable, en considérant un voisinage de 20 km.

Tableau 42 – Statistiques sur l'estimation sur une grille régulière de 250 m de côté, par krigeage ponctuel colocalisé

S7h	Mesures ($\mu\text{g m}^{-3}$)	Estimations par krigeage ponctuel ($\mu\text{g m}^{-3}$)	Ecart-types de krigeage ($\mu\text{g m}^{-3}$)
Nombre de données	34	25467	25467
Minimum	0,75	0,75	0,00
Percentile 25	0,93	0,93	0,06
Médiane	1,06	1,02	0,08
Percentile 75	1,24	1,18	0,10
Maximum	1,59	1,72	0,20
Etendue	0,84	0,97	0,20
Moyenne	1,08	1,06	0,08
Ecart-type	0,041	0,18	0,04
Variance	0,202	0,03	0,001
CV%	18,6	16,6	42,3

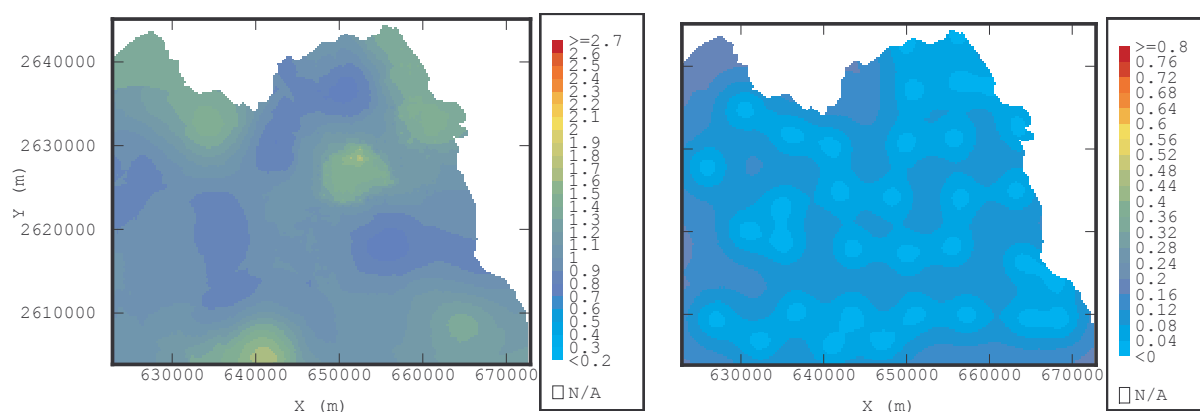


Figure 88 - (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données S7h, et (droite) écarts-types de krigeage

17. COMPARAISON DES ESTIMATIONS ET DISCUSSION

17.1 Cas de l'été 2003

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.

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.

Tableau 43 – Synthèse des résultats de validation pour l'estimation par cokrigage colocalisé

Valeurs	J _{1,e}	S3e	S5e	S7e
Nombre de données	194	100	53	35
Minimum ($\mu\text{g m}^{-3}$)	0,61	0,61	0,61	0,64
Maximum ($\mu\text{g m}^{-3}$)	1,31	1,26	1,31	1,26
Moyenne ($\mu\text{g m}^{-3}$)	0,92	0,88	0,86	0,87
Ecart-type ($\mu\text{g m}^{-3}$)	0,027	0,029	0,032	0,027
Variance ($\mu\text{g m}^{-3}$)	0,165	0,169	0,178	0,163

Validation croisée					
Erreur					
	Moyenne	-0,00072	0,00038	-0,00014	0,00370
	Variance	0,01251	0,01218	0,01451	0,01305
Erreur standardisée					
	Moyenne	-0,00888	0,00318	0,00245	0,01908
	Variance	1,10247	0,91693	0,91385	0,88341
Coefficient de corrélation $r =$		0,730	0,757	0,736	0,715

Validation absolue					
Erreur					
	Moyenne	-	0,11	0,11	0,11
	Variance	-	0,000	0,000	0,000
Coefficient de corrélation $r =$		-	0,663	0,637	0,598

La figure ci-après représente les corrélations entre les concentrations estimées sur la grille de référence à partir du jeu de données et du modèle de référence, et des différentes sélections et modèles associés. Avec la dégradation du maillage, on observe pour la maille de 5 km une légère dégradation de l'estimation; en revanche, avec la maille de 7 km, le modèle retenu améliore la corrélation même s'il tend à surestimer les concentrations les plus élevées par rapport au modèle de référence.

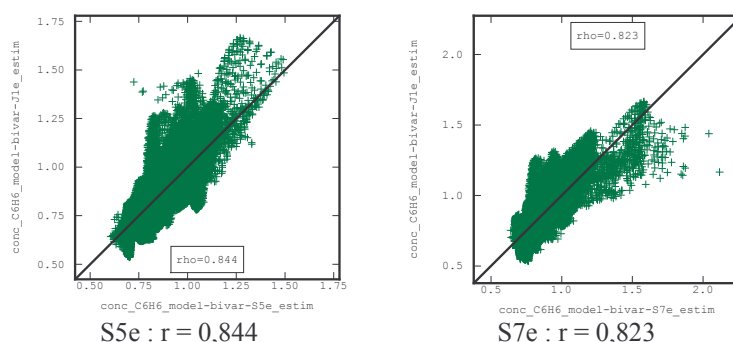


Figure 89 – Comparaison entre l'estimation de référence et les estimations obtenues à partir des différentes sélections

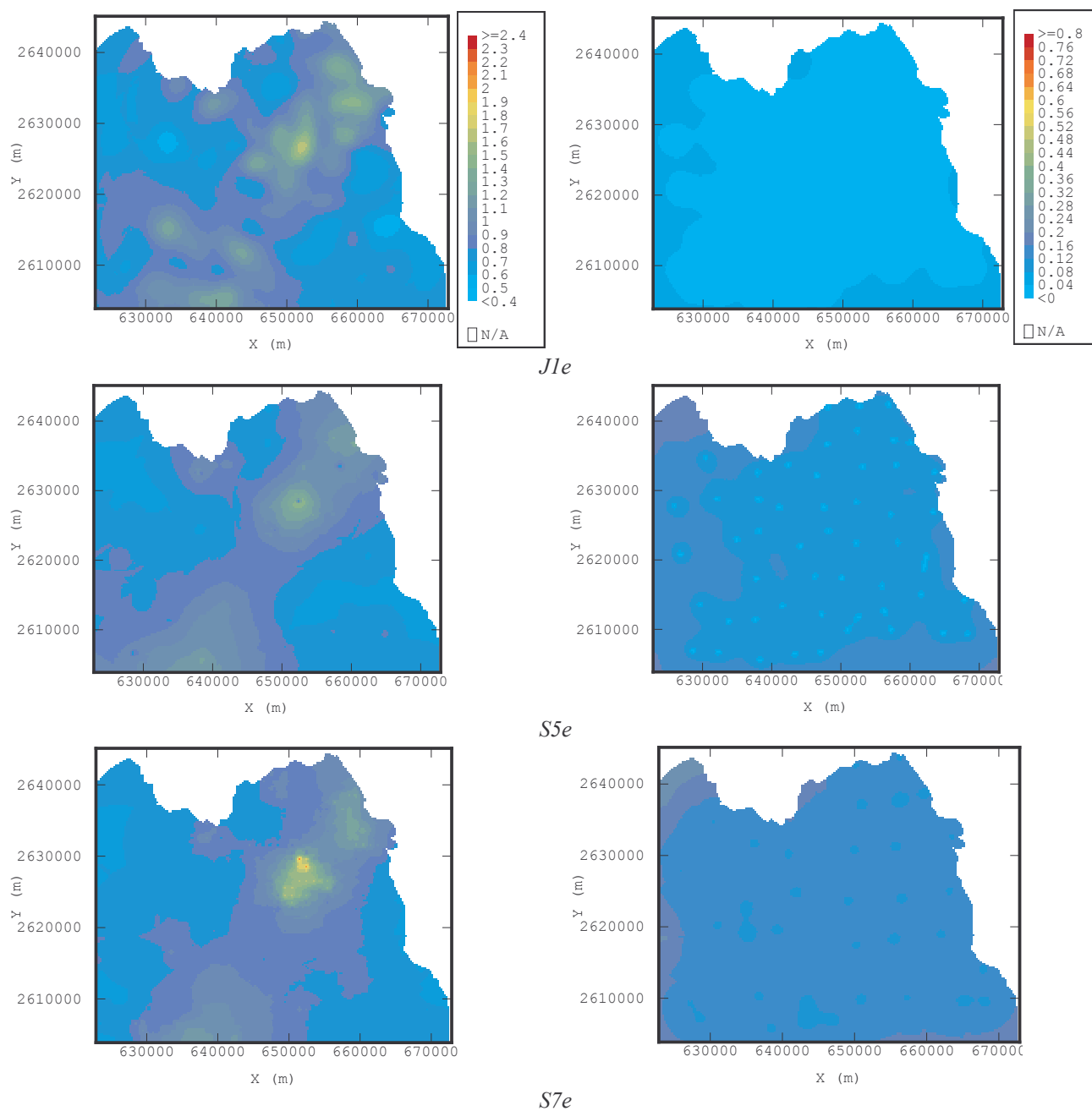


Figure 90 – Comparaison des cartes (gauche) d'estimation et (droite) d'écart-types établies par krigeage ponctuel colocalisé sur la grille de référence pour les différentes sélections

On évalue la robustesse du krigeage en déterminant les concentrations estimées pour une sélection donnée en utilisant non pas le modèle défini pour cette sélection, mais le modèle de référence. La sensibilité au variogramme est fortement marquée à partir de la sélection S7e.

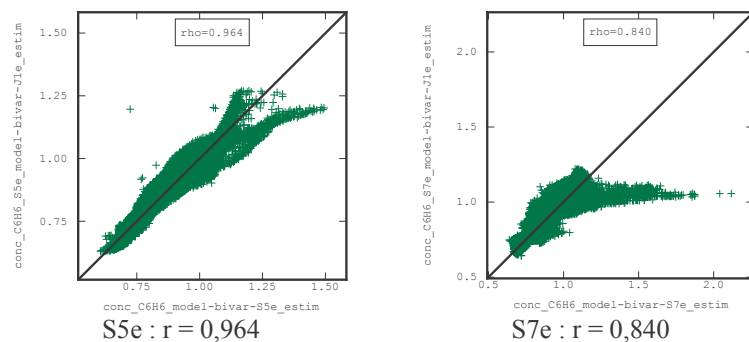


Figure 91 – Comparaison entre les estimations obtenues pour chaque sélection avec le modèle de référence (ordonnée) et avec le modèle de la sélection (abscisse)

Finalement, on calcule les différences absolues et relatives des cartes des différentes sélections avec la carte de référence.

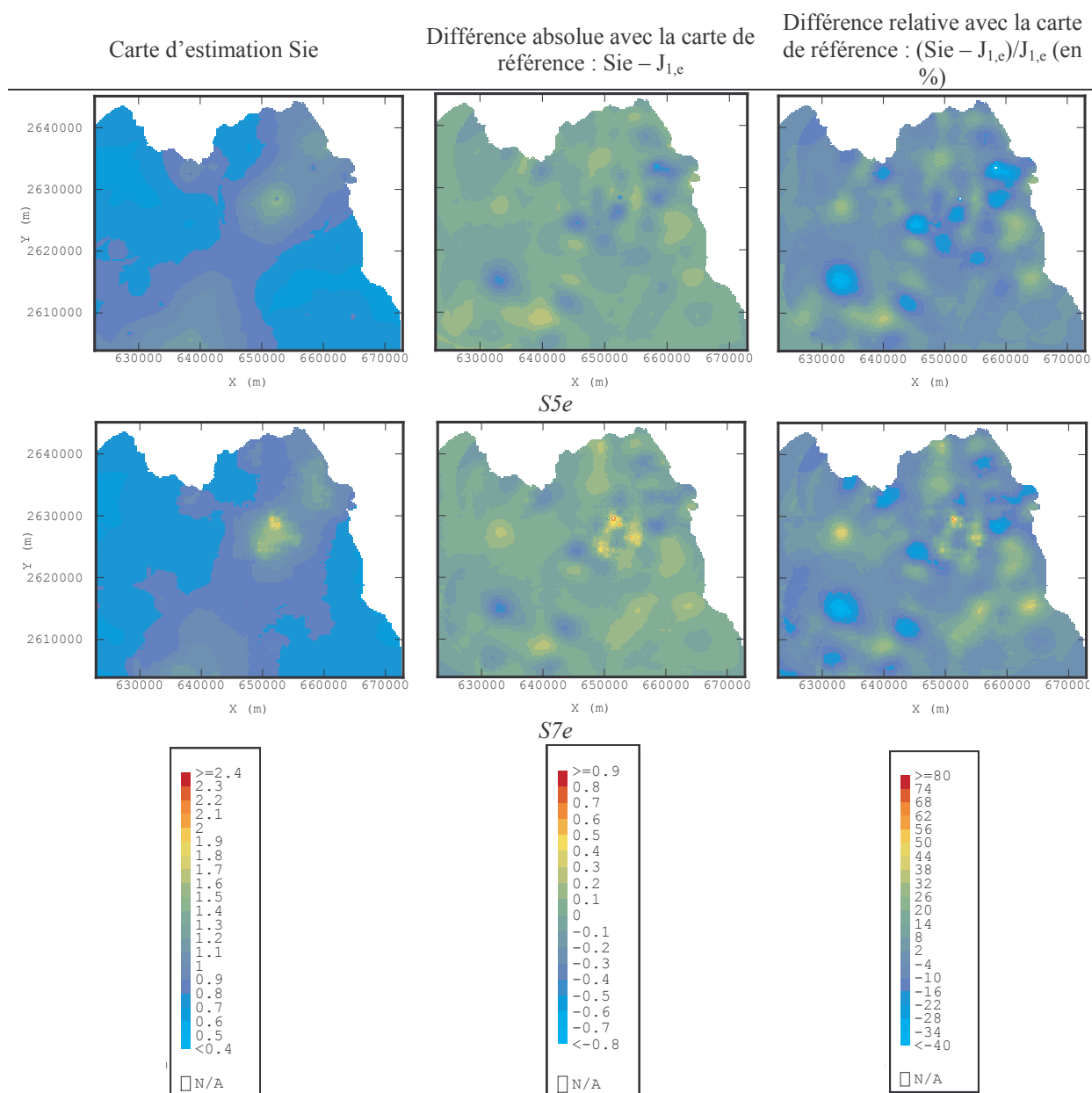


Figure 92 – Comparaison entre les cartes d'estimation obtenues avec le modèle de référence et avec les différentes sélections

La sensibilité au modèle de variogramme observée pour la sélection S7e se confirme ; elle entraîne une forte surestimation des concentrations pour certains points de la sélection par rapport aux concentrations estimées pour ces mêmes points avec le modèle de référence. Quelle que soit la taille de la maille, on n'observe pas d'amélioration significative par rapport aux estimations réalisées sans variable auxiliaire.

17.2 Cas de l'hiver 2004

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.

Tableau 44 – Synthèse des résultats de validation pour l'estimation par cokrigage colocalisé

Valeurs	J _{1,h}	S3h	S5h	S7h
Nombre de données	208	104	55	34
Minimum ($\mu\text{g m}^{-3}$)	0,75	0,75	0,75	0,75
Maximum ($\mu\text{g m}^{-3}$)	1,92	1,77	1,90	1,59
Moyenne ($\mu\text{g m}^{-3}$)	1,18	1,09	1,09	1,08
Ecart-type ($\mu\text{g m}^{-3}$)	0,073	0,042	0,051	0,041
Variance ($\mu\text{g m}^{-3}$)	0,270	0,205	0,226	0,202

Validation croisée					
Erreur					
	Moyenne	-0,00610	-0,00781	0,01418	0,01128
	Variance	0,03066	0,02335	0,02742	0,02222
Erreur standardisée					
	Moyenne	-0,03072	-0,04719	0,08513	0,09073
	Variance	0,82268	0,94940	0,96355	1,24974
Coefficient de corrélation $r =$		0,762	0,606	0,563	0,396

Validation absolue					
Erreur					
	Moyenne	-	0,14	0,13	0,07
	Variance	-	0,000	0,001	0,000
Coefficient de corrélation $r =$		-	0,592	0,585	0,624

La figure ci-après représente les corrélations entre les concentrations estimées sur la grille de référence à partir du jeu de données et du modèle de référence, et des différentes sélections et modèles associés. On observe une dégradation significative de la corrélation pour la sélection S5h ($r = 0,750$), comparé au maillage régulier sans variable auxiliaire ($r = 0,905$). L'effet est moins marqué pour la sélection S7h ($r = 0,716$ à $0,697$).

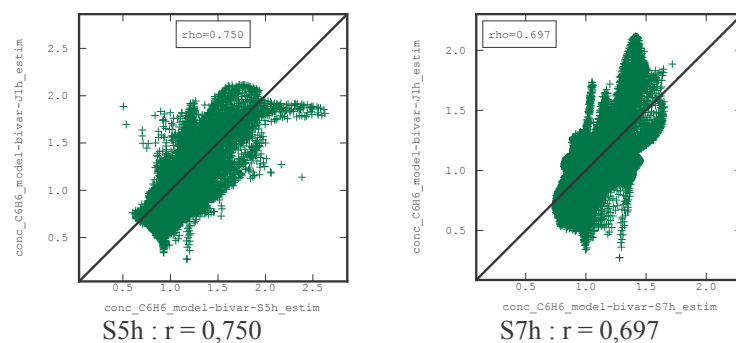


Figure 93 – Comparaison entre l'estimation de référence et les estimations obtenues à partir des différentes sélections

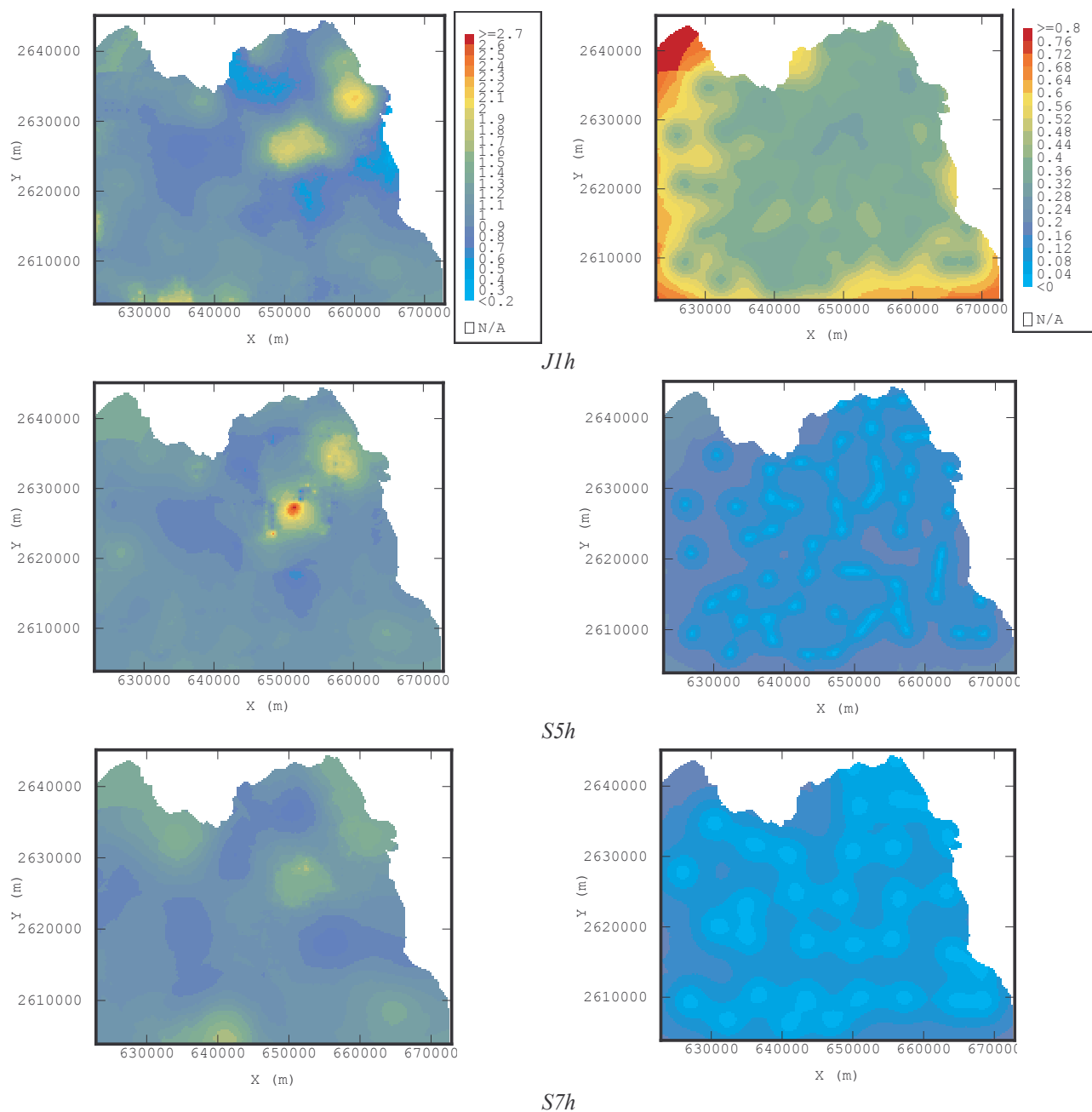


Figure 94 – Comparaison des cartes (gauche) d'estimation et (droite) d'écarts-types établies par krigeage ponctuel colocalisé sur la grille de référence pour les différentes sélections

On évalue la robustesse du krigeage en déterminant les concentrations estimées pour une sélection donnée en utilisant non pas le modèle défini pour cette sélection, mais le modèle de référence. La sensibilité au variogramme est marquée à partir de la sélection S7h.

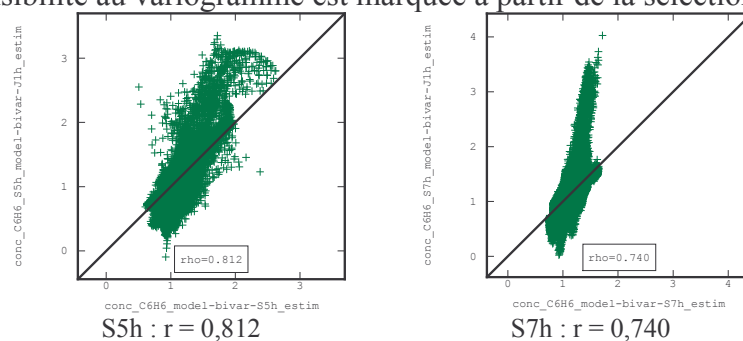
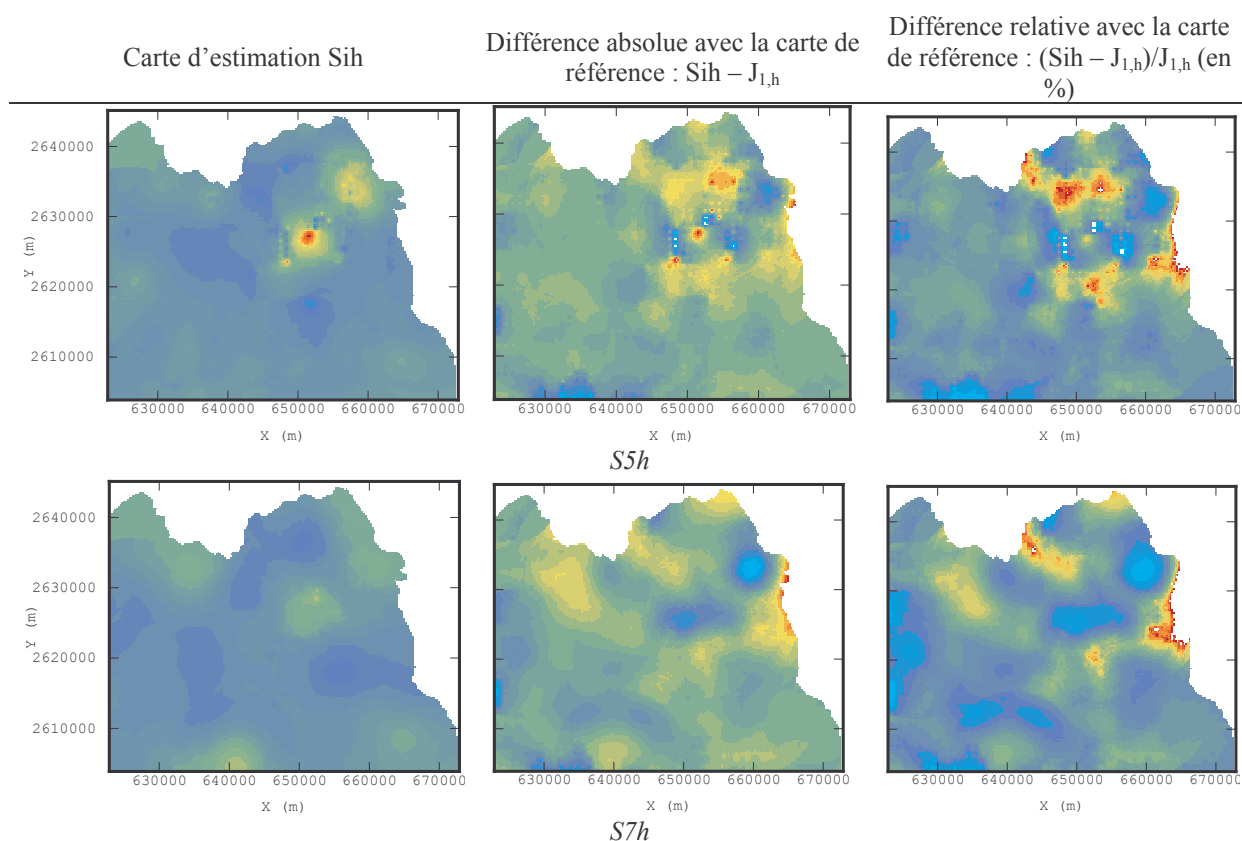


Figure 95 – Comparaison entre les estimations obtenues pour chaque sélection avec le modèle de référence (ordonnée) et avec le modèle de la sélection (abscisse)

Finalement, on calcule les différences absolues et relatives des cartes des différentes sélections avec la carte de référence.



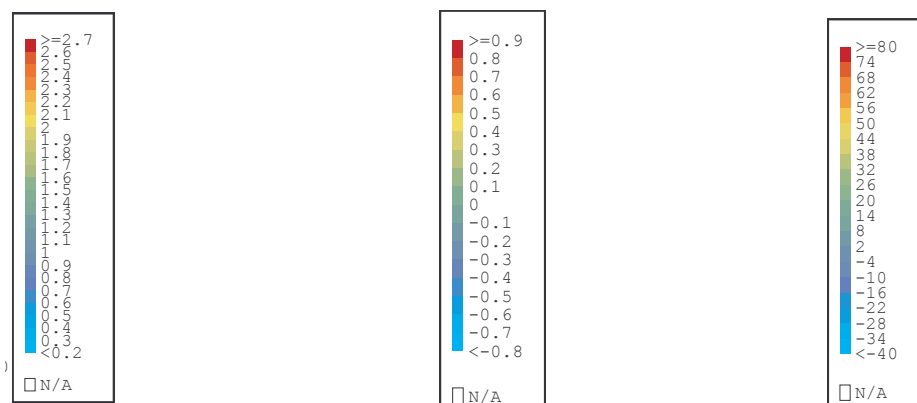


Figure 96 – Comparaison entre les cartes d'estimation obtenues avec le modèle de référence et avec les différentes sélections

La figure ci-dessous montre une nette dégradation des estimations avec des effets de bord importants à l'ouest du domaine (surestimation) et une sous-estimation systématique des centres urbains.

Par rapport au maillage régulier sans variable auxiliaire, on note de plus une augmentation des écarts, et une dégradation de la qualité des cartes : ainsi, la sélection S5h qui respectait le critère de qualité de 30% d'écart maximum par rapport au jeu de données de référence présente dans ce krigeage avec variable auxiliaire des différences relatives dépassant les 50%.

PARTIE III. 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 benzène à l'échelle d'une agglomération. Dans l'impossibilité de disposer de plusieurs jeux de données conséquents pour des villes de 10 000 à 100 000 habitants, nous avons également étudié le cas de l'arrondissement lillois.

Les données de deux campagnes estivale et hivernale fournies par Arema-LM ont été traitées séparément. A l'aide de l'outil géostatistique, des cartes de référence pour chaque saison ont pu être établies. Nous avons ensuite procédé à la dégradation du maillage en sélectionnant les points les plus proches des centres de mailles présentant des dimensions de plus en plus grandes.

Les estimations des concentrations en benzène effectuées sur ce maillage régulier nous ont permis de déterminer un maillage régulier optimal de 5 km de côté pour les deux saisons, soit environ 55 points répartis régulièrement sur le domaine d'étude. La densité d'échantillonnage est donc d'environ 1 point pour 30 km² à l'échelle du domaine.

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

Nous avons utilisé le cokrigage colocalisé pour prendre en compte le cofacteur d'émissions (lié aux émissions de NO_x pour la période estivale et aux émissions de NO_x et COV pour la période hivernale) dans l'estimation des concentrations de benzène sur l'arrondissement lillois.

Les estimations effectuées à partir des mêmes maillages réguliers en tenant compte des variables auxiliaires ne nous ont pas permis d'améliorer les estimations des concentrations de benzène. Nous avons au contraire observé une dégradation de la qualité d'estimation.

Bien qu'il soit difficile d'avancer une explication satisfaisante, on peut toutefois noter que les modèles de variogramme déterminés pour l'étude sans variable auxiliaire sont beaucoup plus robustes que ceux déterminés avec variable auxiliaire, et que ceci peut expliquer – du moins en partie – les différences moindres entre les estimations réalisées à partir des sélections et à partir du jeu de données de référence.

PARTIE IV. Références

- [1] AREMA Lille Métropole, *Cartographie de la pollution urbaine sur l'arrondissement de Lille par tubes à diffusion passive – été 2003 – hiver 2004*, Rapport d'études

Annexe 12 – Urbain : Troyes, Orléans, Bourges

Le nombre de points dans les jeux de données suivants est insuffisant pour dégrader l'échantillonnage. Nous nous limitons à l'étude de la structure spatiale.

TROYES – Benzène, 1998.

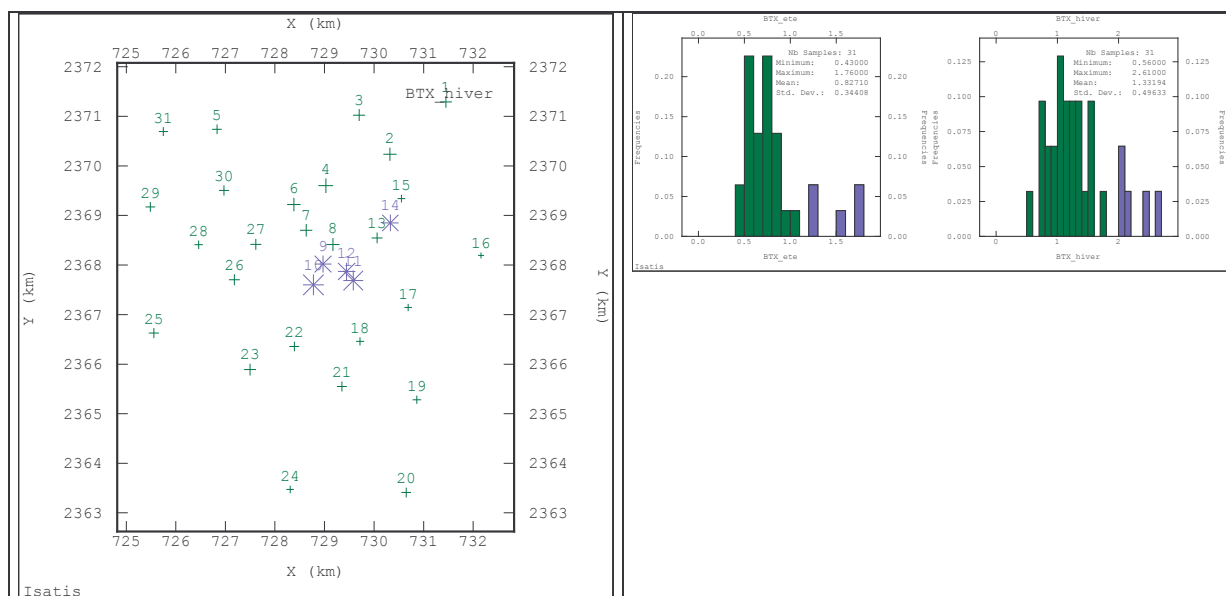


Figure 13 – Implantation des points d'échantillonnage et histogrammes des concentrations estivales et hivernales de benzène.

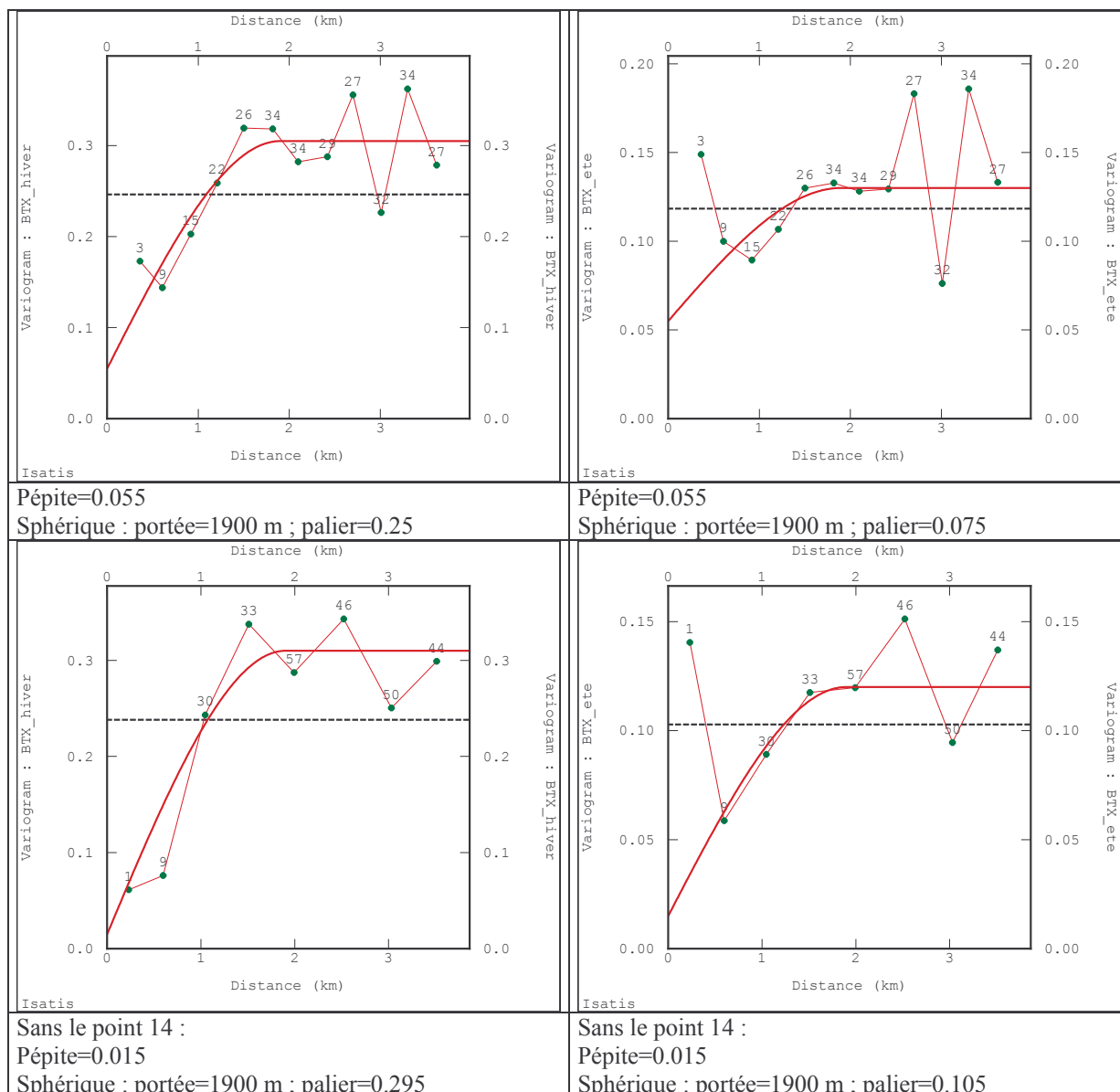


Figure 14 – Variogramme des concentrations de benzène. A gauche : hiver ; à droite : été. Ligne du haut : avec tous les points ; ligne du bas : sans le point 14, qui accroît sensiblement la variabilité à courte distance.

Dans les deux cas, en été comme en hiver, la portée se situe entre 1,5 km et 2 km. Les modèles ajustés sont également représentés. Le choix des structures repose sur des tests de validation croisée.

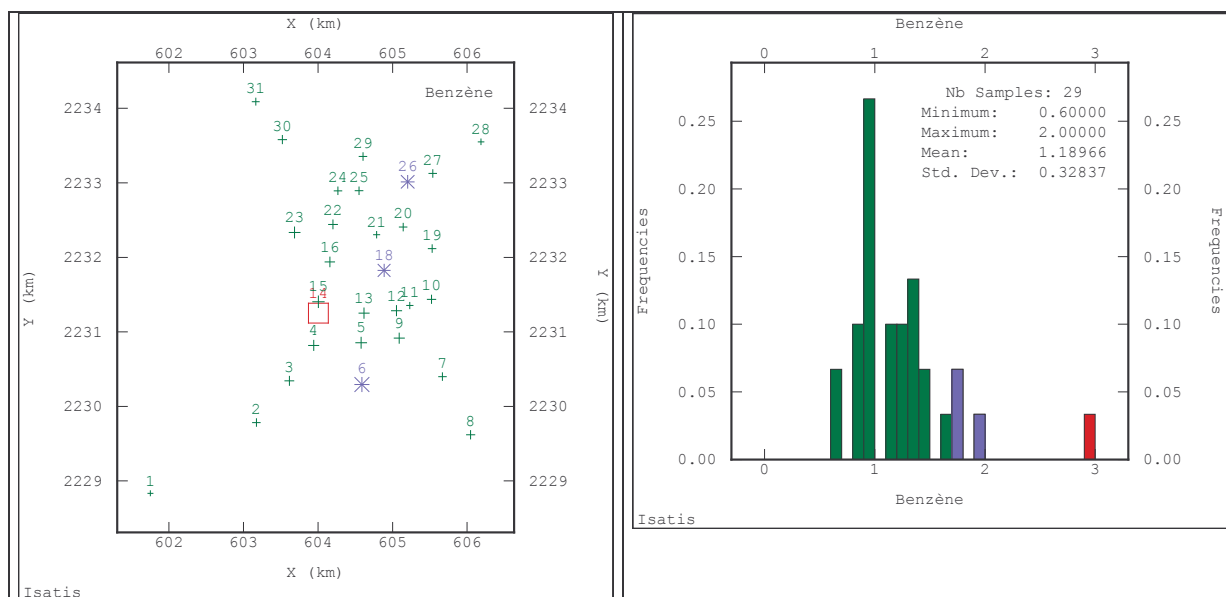


Figure 15 – Implantation des points d'échantillonnage et histogramme des concentrations de benzène.

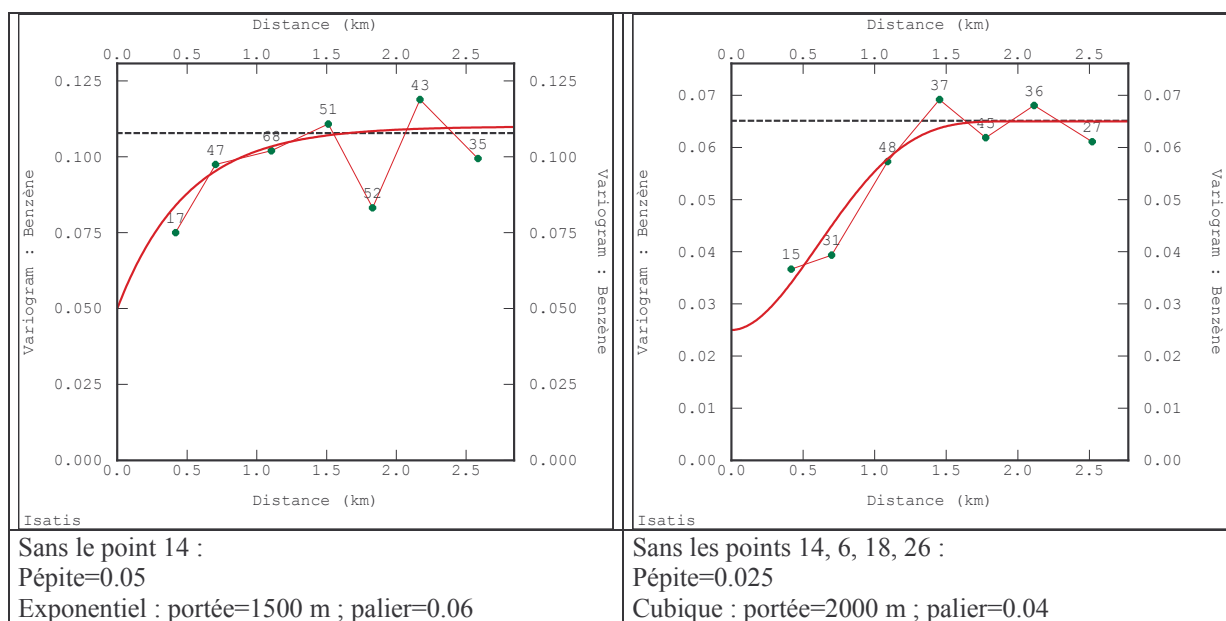


Figure 16 – Variogramme des concentrations de benzène. A gauche : sans le point 14, qui se démarque de la distribution et accroît sensiblement la variabilité à courte distance ; à droite : sans les points de plus forte concentration situés sur les grands axes (14, 6, 18, 26). La portée se situe entre 1,5 km et 2 km.

Les modèles ajustés sont également représentés. Le choix des structures repose sur des tests de validation croisée.

ORLEANS – Benzène, octobre-novembre 2001.

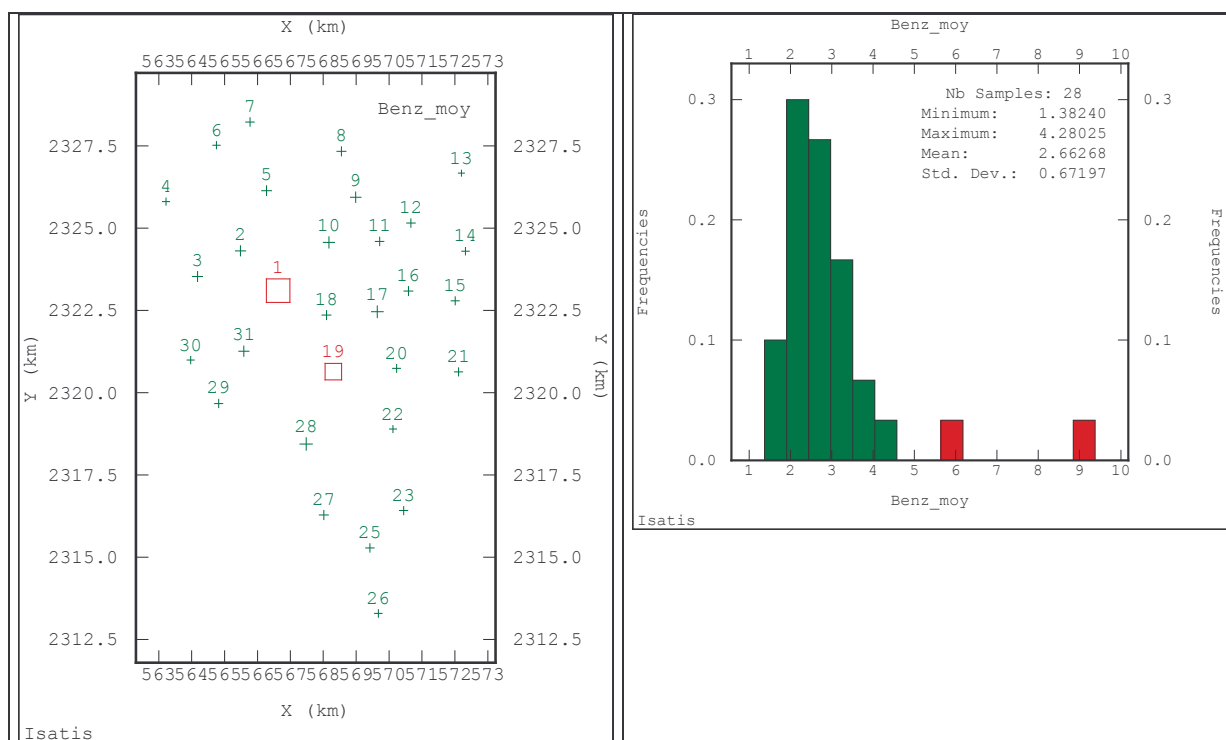


Figure 17 – Implantation des points d'échantillonnage et histogramme des concentrations de benzène. En bleu : points de concentration plus élevée, se distinguant sur l'histogramme.

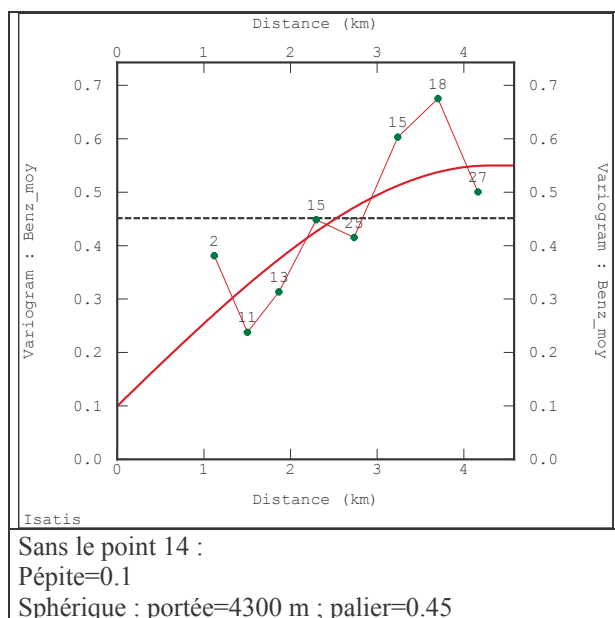


Figure 18 – Variogramme des concentrations de benzène, sans les points 1 et 19 qui masquent la structure spatiale. La portée se situe vers 4 km. La maille d'échantillonnage ne permet pas de déterminer s'il existe une structure de plus courte portée.

Le modèle ajusté est également représenté. Le choix des structures repose sur des tests de validation croisée.