

ECHANTILLONNAGE ET CARTOGRAPHIE DE O₃

Annexe 7 – Région : AQUITAINE

Annexe 8 – Région : RHÔNE-ALPES

Annexe 9 – Région : AUVERGNE

Annexe 7 – Région : AQUITAINE

TABLE DES MATIÈRES

▪ DESCRIPTION DU CAS	3
▪ Présentation du jeu de données initial.....	3
▪ Présentation des données auxiliaires.....	3
▪ Méthodologie de sélection des données	3
▪ ETUDE DU JEU DE DONNÉES INITIAL ET ÉLABORATION DE LA CARTE DE RÉFÉRENCE	4
▪ Histogramme.....	4
▪ Etude du variogramme expérimental.....	5
▪ Correlation avec l'altitude	7
▪ Modélisation du variogramme et validation croisée.....	7
▪ Estimation sur toute la région.....	8
▪ ETUDE DE LA SÉLECTION DE DONNÉES (MAILLE DE 30KM)	9
▪ Histogramme.....	9
▪ Etude du variogramme expérimental.....	9
▪ Modélisation du variogramme et validation croisée.....	10
▪ Vérification sur le jeu de données complémentaire	11
▪ Krigage sur la grille d'estimation	12
▪ RÉFÉRENCES	13

▪ DESCRIPTION DU CAS

▪ Présentation du jeu de données initial

Le jeu de données étudié a été fourni par l'association agréée de surveillance AIRAQ. Il est issu d'une campagne de mesure par échantillonnage passif conduite du 3 au 26 juin 2003 sur l'ensemble de la région Aquitaine (AIRAQ, 2003).

Le domaine d'étude s'étend sur environ 300km du nord au sud et 250km d'est en ouest ; il couvre une superficie de 42000 km². L'échantillonnage se compose de **108 tubes de fond** répartis selon un maillage de **25km** de résolution.

21 tubes ont été ajoutés à ces 108 échantillonneurs de façon à former 10 triplets et un doublon.

32 tubes au total, dont 8 triplets, sont situés au niveau de stations fixes.

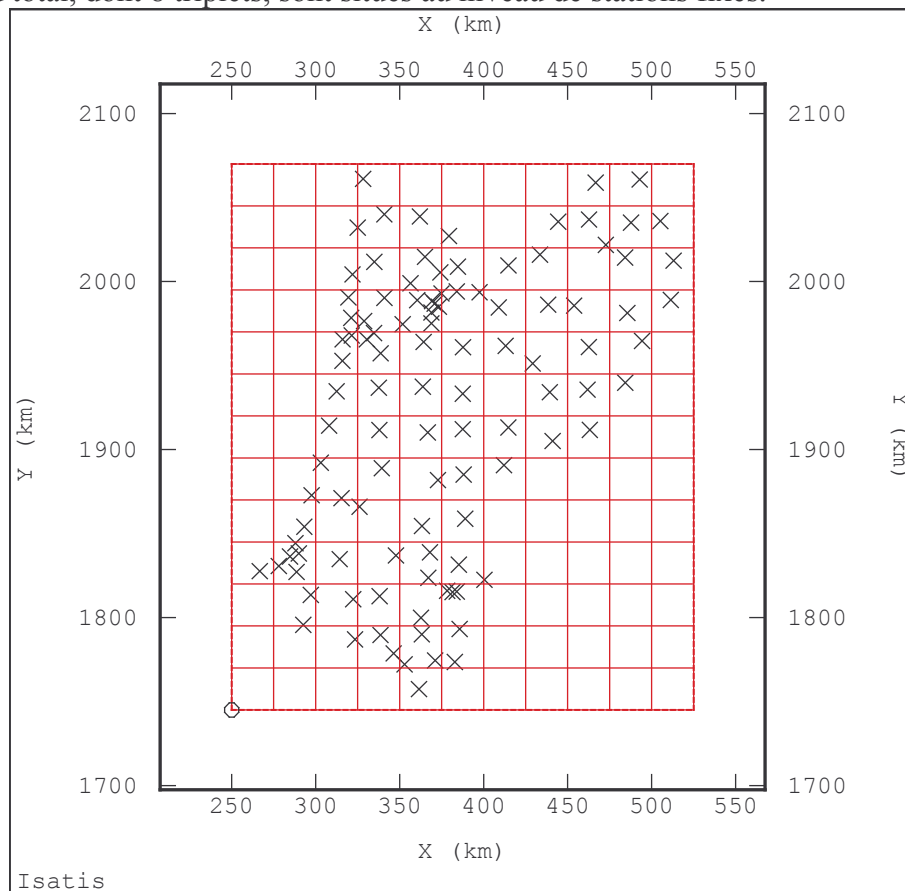


Figure 1 - Carte de répartition des 108 points de mesure

Les écarts relatifs entre les mesures de triplets, très faibles (entre 0 et 3%), et les écarts relatifs entre tubes et stations, faibles eux aussi (généralement inférieurs à 10%) indiquent une bonne qualité de mesure.

▪ Présentation des données auxiliaires

Une grille kilométrique d'altitude a été extraite par l'EMD d'une base nationale de données topographiques.

▪ Méthodologie de sélection des données

Un jeu de données réduit a été prélevé dans le jeu de données complet en calant sur le domaine d'étude un maillage régulier de résolution 30 km. Pour chaque maille, le site de

mesure le plus proche du centre a été sélectionné. Des points ont été ajoutés si besoin était, afin que les principales villes restent représentées dans l'échantillonnage. D'autre part, certains resserrements du maillage initial ont été conservés afin de pouvoir évaluer le variogramme aux plus courtes distances. Le jeu de données ainsi constitué compte 68 points de mesure (Figure 2).

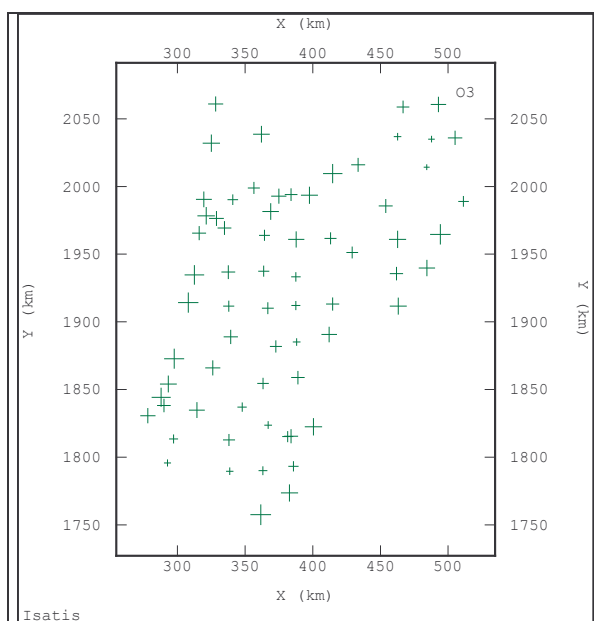


Figure 2 – Sélection des données selon une maille d'échantillonnage de 30km de côté.

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

Cette partie a pour objet d'élaborer la carte de référence, c'est-à-dire la carte de répartition des concentrations d'ozone établie avec le maximum de points de mesure. Par la suite, on comparera à cette carte les représentations obtenues à partir d'un nombre réduit de points.

▪ Histogramme

La distribution des concentrations s'approche une loi normale (Figure 3). Deux points de plus forte concentration s'en détachent : ils s'agit de mesures effectuées dans les Pyrénées. La concentration d'ozone prend également des valeurs assez élevées le long du littoral.

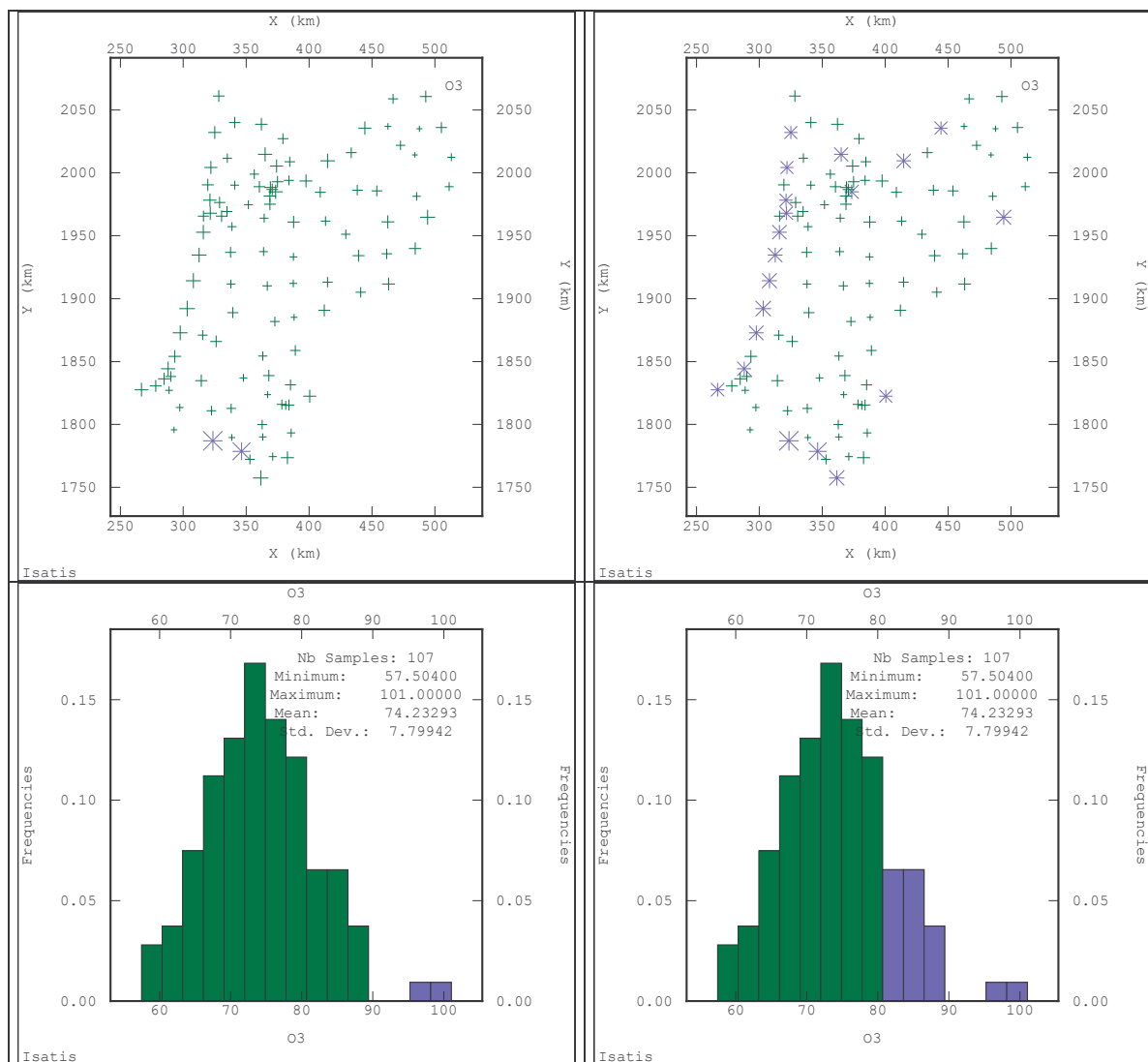


Figure 3 – Histogramme des concentrations et localisation des plus fortes valeurs (signalées en bleu).

• Etude du variogramme expérimental

Les deux points de concentration maximale situés dans les Pyrénées accroissent notablement la variabilité des concentrations (voir la nuée variographique en Figure 4). Vu qu'ils représentent un phénomène très local par rapport au domaine, ils sont éliminés du calcul du variogramme. Quelques points se distinguent également sur la nuée variographique. Supprimer les sites de mesure qui en sont responsables n'a pas d'effet sensible sur le variogramme expérimental. Ces sites sont conservés dans l'analyse, à l'exception du site 106, localisé à plus de 1400m d'altitude (par cohérence avec l'élimination des deux autres points de montagne 101 et 103)

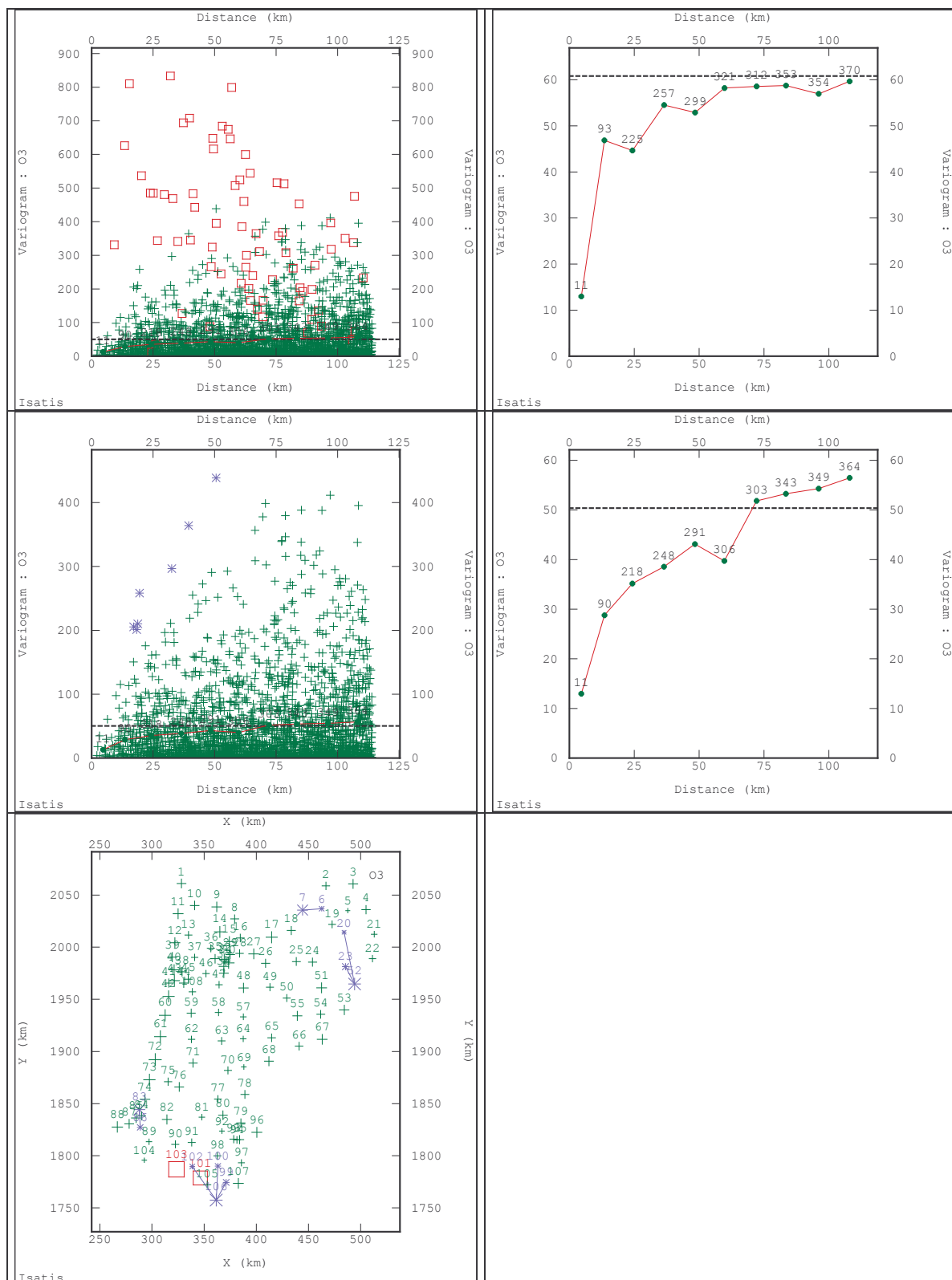


Figure 4 – Nuée variographique et variogramme expérimental. Ligne du haut : avec les points 101 et 103. Ligne du bas : sans ces points. Il y a correspondance entre les points rouges et bleus des nuées et les points rouges et bleus de la carte.

L'échelle d'étude est réduite aux cent premiers kilomètres, ce qui, selon la direction, correspond approximativement à la moitié ou au tiers du champ, conformément à la pratique. Notons qu'aux grandes distances le variogramme expérimental ne montre pas de dérive. On ne met pas en évidence d'anisotropie.

▪ Correlation avec l'altitude

Au sud du domaine, dans les Pyrénées, la concentration d'ozone augmente avec l'altitude, évolution fréquemment observée. Mais comme l'indique aussi la Figure 3, elle est relativement élevée le long du littoral, qui est d'altitude nulle. Ces niveaux de concentration peuvent s'expliquer par les brises de terre et de mer (AIRAQ, 2005).

A cause de phénomènes variés qui régissent la formation et la dispersion de l'ozone en Aquitaine, aucune relation linéaire n'a pu être établie entre la concentration d'ozone et l'altitude (ou sa transformée). **Cette variable n'est donc pas utilisée.**

▪ Modélisation du variogramme et validation croisée

Un modèle isotrope composé d'un effet de pépite et de deux structures sphériques est ajusté sur le variogramme expérimental (Figure 5).

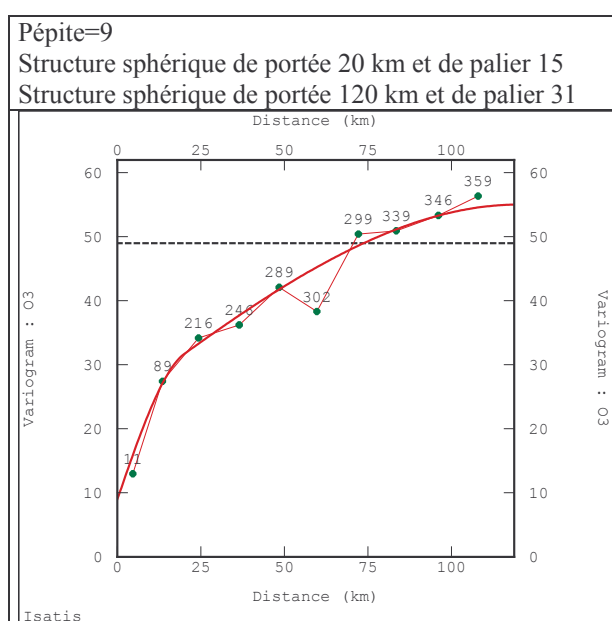


Figure 5– Toutes données (sauf points de montagne). Variogramme expérimental et modèle ajusté

Le choix des structures et des paramètres variographiques découle d'ajustements successifs testés par validation croisée. Un voisinage unique a été employé.

Nous présentons ci-après, pour le modèle retenu, quelques statistiques de validation croisée.

Tableau 1 – Modèle de référence - Statistiques des valeurs estimées et des erreurs de validation croisée

	N	P50	P75	Max	Moyenne	Variance	Ecart-type
Valeurs mesurées d'ozone - sans les 3 points de montagne	104	73.97	78.71	88.86	72.94	99.57	9.98
Valeurs mesurées d'ozone - avec les points de montagne	107	74.09	79.36	101.00	73.55	110.83	10.53
Valeurs estimées d'ozone - sans les 3 points de montagne en entrée	104	73.96	76.47	81.39	72.92	65.59	8.10
Valeurs estimées d'ozone - avec les points de montagne en entrée	107	74.52	76.92	83.64	73.52	63.72	7.98

	N	P50	P75	Max	Moyenne	Variance	Ecart-type
Erreur – sans les points de montagne en entrée	104	-0.60	5.05	10.68	-0.02	33.37	5.78
Erreur – avec les points de montagne en entrée	107	-0.09	5.13	21.07	-0.03	57.36	7.57
Erreur relative – sans les points de montagne en entrée	104	6.55	9.60	18.56	6.67	20.21	4.50
Erreur relative– avec les points de montagne en entrée	107	6.69	10.27	34.68	7.70	41.97	6.48

L'introduction des points 101 et 103 en entrée du krigeage dégrade l'estimation en quelques points de la partie sud du domaine (surestimation). Mais exclure ces deux mesures conduit à masquer les fortes valeurs observées en altitude.

Ainsi, dans cette zone localement soumise à des gradients de concentration élevés, on peut s'attendre à une plus grande incertitude d'estimation. (NB : cette incertitude ne sera pas retranscrite par l'écart-type de krigeage puisque celui-ci ne tient pas compte de la variabilité locale des données.)

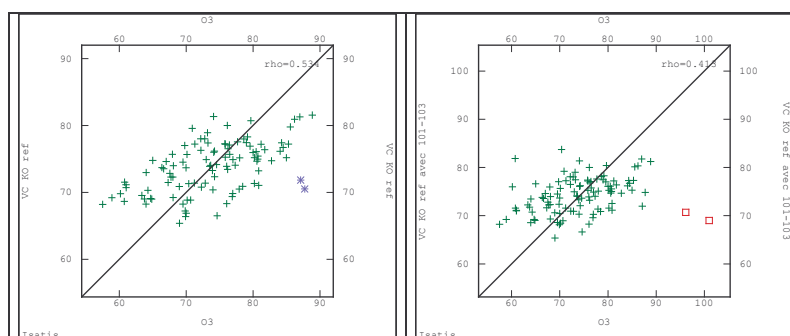
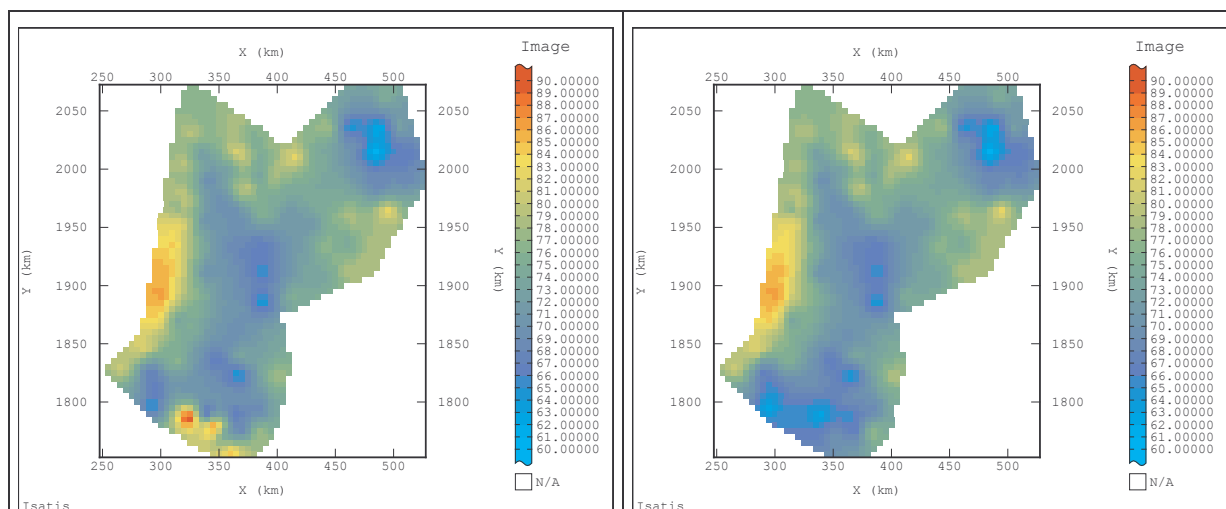


Figure 6 – Comparaison entre les valeurs mesurées et estimées par validation croisée sans (figure de gauche) et avec (figure de droite) les points 101 et 103 en entrée du krigeage

▪ Estimation sur toute la région

La maille d'estimation a 5km de côté. Comme dans la validation croisée, un voisinage d'estimation unique est adopté.

La suppression des points de montagne en entrée du krigeage abaisse les concentrations au sud du domaine mais n'a pas d'impact notable sur le reste de la cartographie (Figure 7).



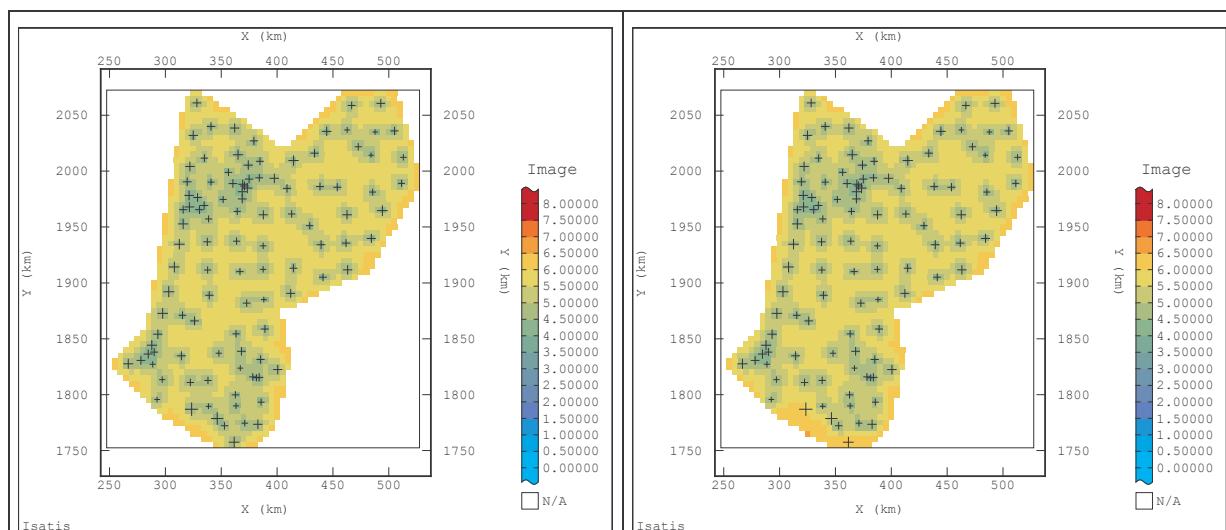


Figure 7 – Estimation par krigeage ponctuel à partir de toutes les données sur une grille de pas égal à 5km. A gauche : les points de montagne (101, 103, 106) sont utilisés en entrée du krigeage ; à droite : les points de montagne ne sont pas pris en compte. Ligne du haut : concentrations d'ozone estimées $[\mu\text{g}/\text{m}^3]$; ligne du bas : écarts-types de krigeage.

▪ ETUDE DE LA SELECTION DE DONNEES (MAILLE DE 30KM)

▪ Histogramme

Remarque : ce nouveau jeu de données ne contient pas les points 101 et 103.

Si l'histogramme a un profil plus irrégulier, la réduction de l'échantillonnage modifie assez peu les principales statistiques (sites de montagne exclus).

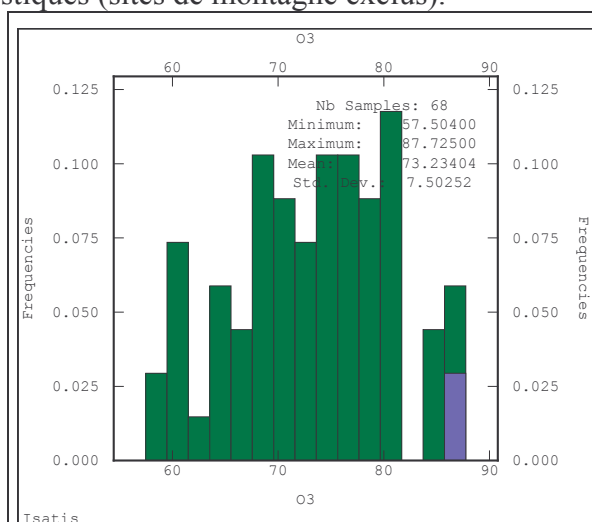


Figure 8 – Histogramme des concentrations pour la sélection S30km. En bleu : les points 52 et 106

▪ Etude du variogramme expérimental

Trois couples de points se détachent de la nuée variographique (Figure 9). Le point de montagne 106, qui seul accroît sensiblement la variabilité aux courtes et moyennes distances, est éliminé du calcul du variogramme.

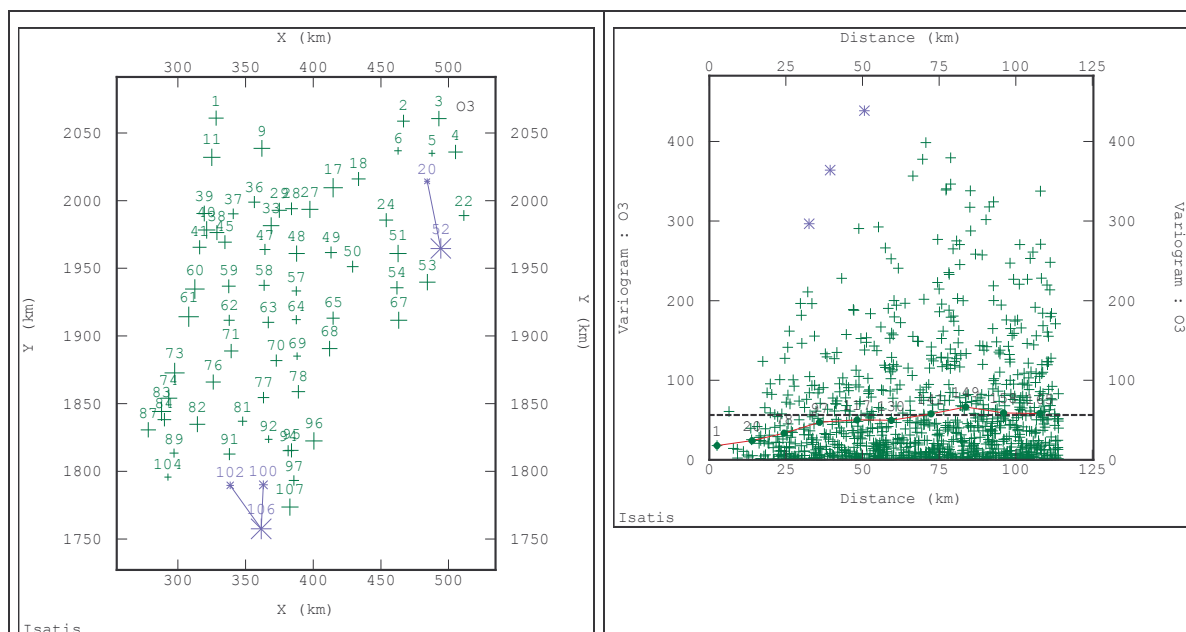


Figure 9 – Nuée variographique et variogramme expérimental. Il y a correspondance entre les points bleus de la nuée et les points bleus de la carte.

• Modélisation du variogramme et validation croisée

Le modèle ajusté (Figure 10) est soumis au test de validation croisée, avec ou sans le point 106 en entrée du krigeage (Tableau 2, Figure 11). Aux points de la sélection, on ne note pas de dégradation de l'estimation par rapport à la validation croisée précédente.

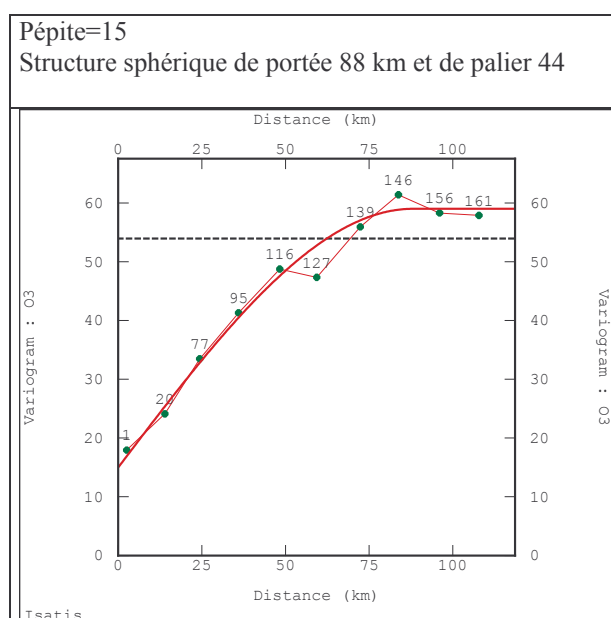


Figure 10– Sélection 30km (sauf points de montagne). Variogramme expérimental et modèle ajusté

Tableau 2 – Modèle associé à la sélection S30km - Statistiques des valeurs estimées et des erreurs de validation croisée

	N	P50	P75	Max	Moyenne	Variance	Ecart-type
Valeurs mesurées d'ozone - sans le point 106	66	73.67	78.54	87.12	71.95	131.28	11.46
Valeurs mesurées d'ozone - avec le point 106	67	73.80	78.76	87.73	72.18	132.96	11.53
Valeurs estimées d'ozone - sans le point 106 en entrée	66	73.78	75.52	81.74	71.83	94.74	9.73
Valeurs estimées d'ozone - avec le point 106 en entrée	67	73.97	75.55	81.85	72.04	93.45	9.67
Erreur – sans le point 106 en entrée	66	0.02	5.15	10.35	-0.11	33.89	5.82
Erreur – avec le point 106 en entrée	67	-0.15	5.00	10.35	-0.13	35.85	5.99
Erreur relative – sans le point 106 en entrée	66	6.86	9.65	16.97	6.94	18.01	4.24
Erreur relative – avec le point 106 en entrée	67	7.10	10.14	16.97	7.07	19.08	4.37

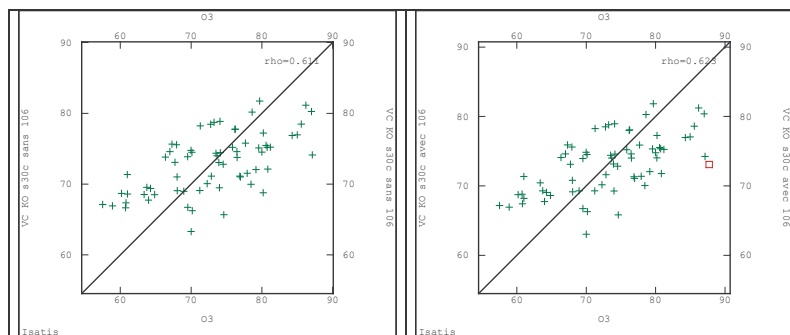


Figure 11 – Comparaison entre les valeurs mesurées et estimées par validation croisée sans (figure de gauche) et avec (figure de droite) le point 106 (point rouge) en entrée du krigeage

• Vérification sur le jeu de données complémentaire

Les concentrations d'ozone sont estimées aux 39 points de mesure qui n'appartiennent pas à la sélection puis elles sont comparées aux observations (Tableau 3, Figure 12). Le nuage de corrélation est assez dispersé, avec une sous-estimation des valeurs les plus élevées (ce qui était déjà le cas en validation croisée, pour le jeu de données complet). A l'exception des points 101 et 103, l'écart avec la mesure reste cependant limité.

Tableau 3 – Modèle associé à la sélection S30km - Statistiques des valeurs estimées et des erreurs de validation absolue

	N	P50	P75	Max	Moyenne	Variance	Ecart-type
Valeurs mesurées d'ozone	39	75.78	80.59	101.00	74.09	206.39	14.37
Valeurs estimées d'ozone - sans le point 106 en entrée	39	74.37	75.85	83.63	71.40	153.31	12.38
Valeurs estimées d'ozone - avec le point 106 en entrée	39	74.77	75.91	83.68	71.76	152.84	12.36
Erreur – sans le point 106 en entrée	39	-2.67	3.66	9.71	-2.68	81.61	9.03
Erreur – avec le point 106 en entrée	39	-2.77	3.75	10.58	-2.32	78.36	8.85
Erreur relative – sans le point 106 en entrée	39	6.20	9.97	36.72	7.92	49.97	7.07

	N	P50	P75	Max	Moyenne	Variance	Ecart-type
Erreur relative– avec le point 106 en entrée	39	6.21	9.79	35.61	8.01	45.58	6.75
Erreur relative – sans le point 106 en entrée, sans compter 101 et 103	37	6.11	9.48	14.28	6.54	13.41	3.66
Erreur relative– avec le point 106 en entrée, sans compter 101 et 103	37	6.03	9.45	16.27	6.76	15.04	3.88

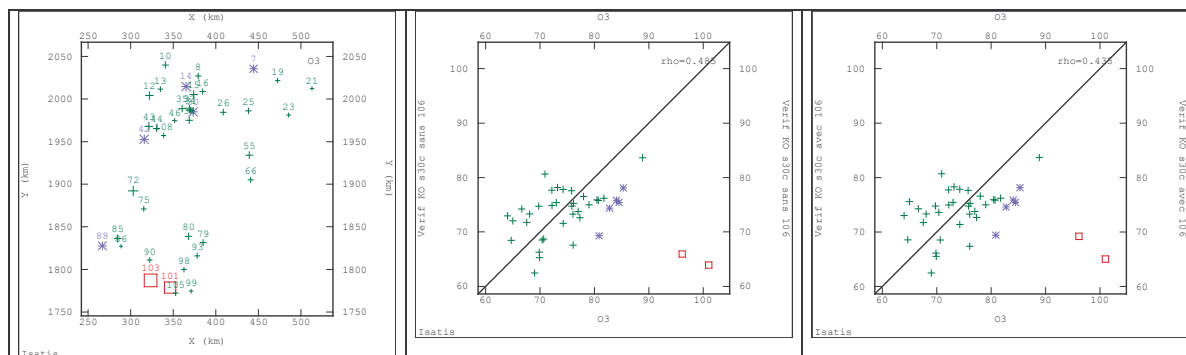
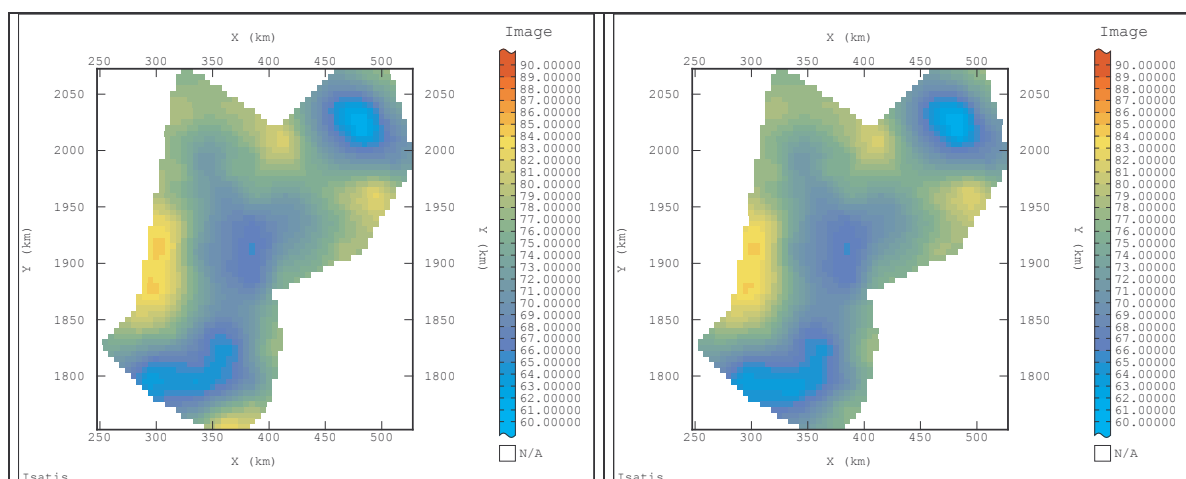


Figure 12 – Comparaison entre les valeurs mesurées et estimées sur le complémentaire de la sélection S30km sans (nuage de gauche) et avec (nuage de droite) le point 106 en entrée du krigage

• Krigeage sur la grille d'estimation

La maille d'estimation de 5km de côté est conservée. Les données et le modèle de variogramme utilisés sont ceux de la sélection. Le point 106 est soit réintroduit au moment du krigage soit ignoré.



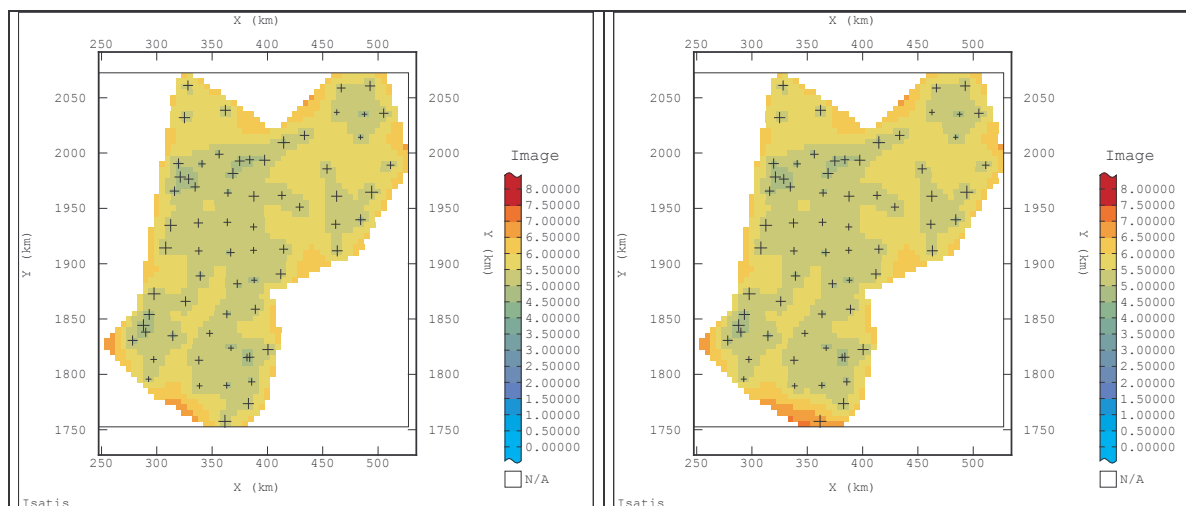


Figure 13 – Estimation par krigeage ponctuel à partir de la sélection S30km sur une grille de pas égal à 5km. A gauche : le point de montagne 106 est utilisé en entrée du krigeage ; à droite : il n'est pas pris en compte. Ligne du haut : concentrations d'ozone estimées [$\mu\text{g}/\text{m}^3$] ; ligne du bas : écarts-types de krigeage.

Compte tenu de l'incertitude des cartographies au sud du domaine, l'effet de la dégradation du maillage est évalué sur les estimations élaborées sans les données de montagne.

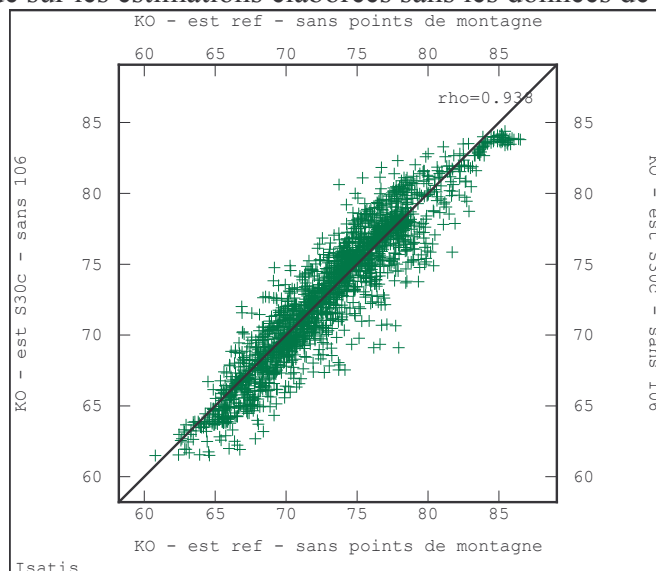


Figure 14 – Comparaison entre les estimations par krigeage ordinaire obtenues à partir de l'ensemble des points (sauf Pyrénées) et de la sélection S30km (sauf Pyrénées)

Les estimations sont en assez bon accord. L'écart relatif entre l'estimation dégradée et l'estimation de référence est de 1,8% en moyenne et de 11,3% au maximum. La différence de concentration entre les deux krigeages n'excède pas $9 \mu\text{g}/\text{m}^3$ en valeur absolue. Les concentrations obéissent globalement à la même répartition spatiale. La carte qui correspond à l'échantillonnage dégradé a un aspect plus lisse et accentue la zone de plus faible concentration au nord-est du domaine.

■ REFERENCES

AIRAQ, 2003. Ozone. Cartographie et bilan estival. Rapport n° ET/CA/CA/03.

Annexe 8 – Région : Rhône-Alpes

OZONE / Etude à l'échelle régionale : cas de la région Rhône-Alpes

PARTIE I. ECHANTILLONNAGE RÉGULIER SANS VARIABLE AUXILIAIRE 17

1	Description du cas	17
• 1.1	Présentation du jeu de données initial	17
• 1.2	Méthodologie de sélection des données	18
2	Etude du jeu de données initial et élaboration de la carte de référence	19
• 2.1	Etude du variogramme expérimental	19
• 2.2	Modélisation du variogramme et validation croisée	20
• 2.3	Etablissement de la carte de référence	21
3	Etude de la sélection de données S15 (maille de 15 × 15 km)	22
• 3.1	Etude du variogramme expérimental	23
• 3.2	Modélisation du variogramme et test de validation croisée	23
• 3.3	Test de validation absolue sur le complémentaire C15	24
• 3.4	Etablissement de la carte pour la sélection S15	25
4	Etude de la sélection de données S20 (maille de 20 × 20 km)	26
• 4.1	Etude du variogramme expérimental	27
• 4.2	Modélisation du variogramme et test de validation croisée	27
• 4.3	Test de validation absolue sur le complémentaire C20	28
• 4.4	Etablissement de la carte pour la sélection S20	29
5	Etude de la sélection de données S30 (maille de 30 × 30 km)	29
• 5.1	Etude du variogramme expérimental	30
• 5.2	Modélisation du variogramme et test de validation croisée	31
• 5.3	Test de validation absolue sur le complémentaire C30	32
• 5.4	Etablissement de la carte pour la sélection S30	32
6	Etude de la sélection de données S45 (maille de 45 × 45 km)	33
• 6.1	Etude du variogramme expérimental	34
• 6.2	Modélisation du variogramme et test de validation croisée	34
• 6.3	Test de validation absolue sur le complémentaire C45	35
• 6.4	Etablissement de la carte pour la sélection S45	36

7	Comparaison des estimations et discussion	37
---	---	----

PARTIE II. ECHANTILLONNAGE RÉGULIER AVEC VARIABLE AUXILIAIRE .43

1	Description de la variable auxiliaire : altitude	43
2	Calcul du cofacteur	44
3	Etude du krigeage avec dérive externe	45
• 3.1	Calcul de la dérive externe	46
• 3.2	Ajustement du modèle de covariance généralisé	48
• 3.3	Test de validation absolue sur le complémentaire C15	50
4	Etude du krigeage colocalisé sur la sélection S15	50
• 4.1	Etude des variogrammes expérimentaux	51
• 4.2	Modélisation des variogrammes et validation croisée	51
• 4.3	Test de validation absolue sur le complémentaire C15	53
5	Etude du jeu de données initial et élaboration de la carte de référence	54
• 5.1	Etude des variogrammes expérimentaux	55
• 5.2	Modélisation des variogrammes et validation croisée	55
• 5.3	Etablissement de la carte de référence	57
6	Etude de la sélection de données S20 (maille de 20 × 20 km)	58
• 6.1	Etude des variogrammes expérimentaux	58
• 6.2	Modélisation des variogrammes et validation croisée	59
• 6.3	Test de validation absolue sur le complémentaire C20	61
• 6.4	Etablissement de la carte pour la sélection S20	62
7	Etude de la sélection de données S30 (maille de 30 × 30 km)	62
• 7.1	Etude des variogrammes expérimentaux	62
• 7.2	Modélisation des variogrammes et validation croisée	63
• 7.3	Test de validation absolue sur le complémentaire C30	64
• 7.4	Etablissement de la carte pour la sélection S30	65
8	Etude de la sélection de données S45 (maille de 45 × 45 km)	65
• 8.1	Etude des variogrammes expérimentaux	65
• 8.2	Modélisation des variogrammes et validation croisée	66
• 8.3	Test de validation absolue sur le complémentaire C45	67
• 8.4	Etablissement de la carte pour la sélection S45	68

9	Comparaison des estimations et discussion	68
PARTIE III.	CONCLUSIONS ET PERSPECTIVES	73
PARTIE IV.	RÉFÉRENCES	74

▪ DESCRIPTION DU CAS

▪ Présentation du jeu de données initial

Les mesures ont été réalisées par les 6 associations agréées de surveillance de la qualité de l'air en Rhône-Alpes (Air APS, AMPASEL, ASCOPARG, ASQUADRA, COPARLY et SUP'air) entre avril et septembre 2002 dans le cadre de 6 campagnes d'une semaine utilisant des capteurs à diffusion passive. Ces campagnes avaient pour objectif de permettre de représenter les niveaux d'ozone de la région en période estivale.

Une étude géostatistique des données obtenues a été effectuée par les associations [1] en s'appuyant sur plusieurs variables auxiliaires : altitude, données d'émissions et de population, etc.

Le territoire observé a une superficie de 43 698 km². La stratégie de répartition des sites d'échantillonnage était basée sur une étude préliminaire réalisée sur le département de l'Isère en 2001 [2], et comprenait : un maillage principal de 15 × 15 km, un maillage de proximité de 5 × 5 km, un maillage complémentaire de 10 × 10 km et un maillage grande distance de 30 × 30 km.

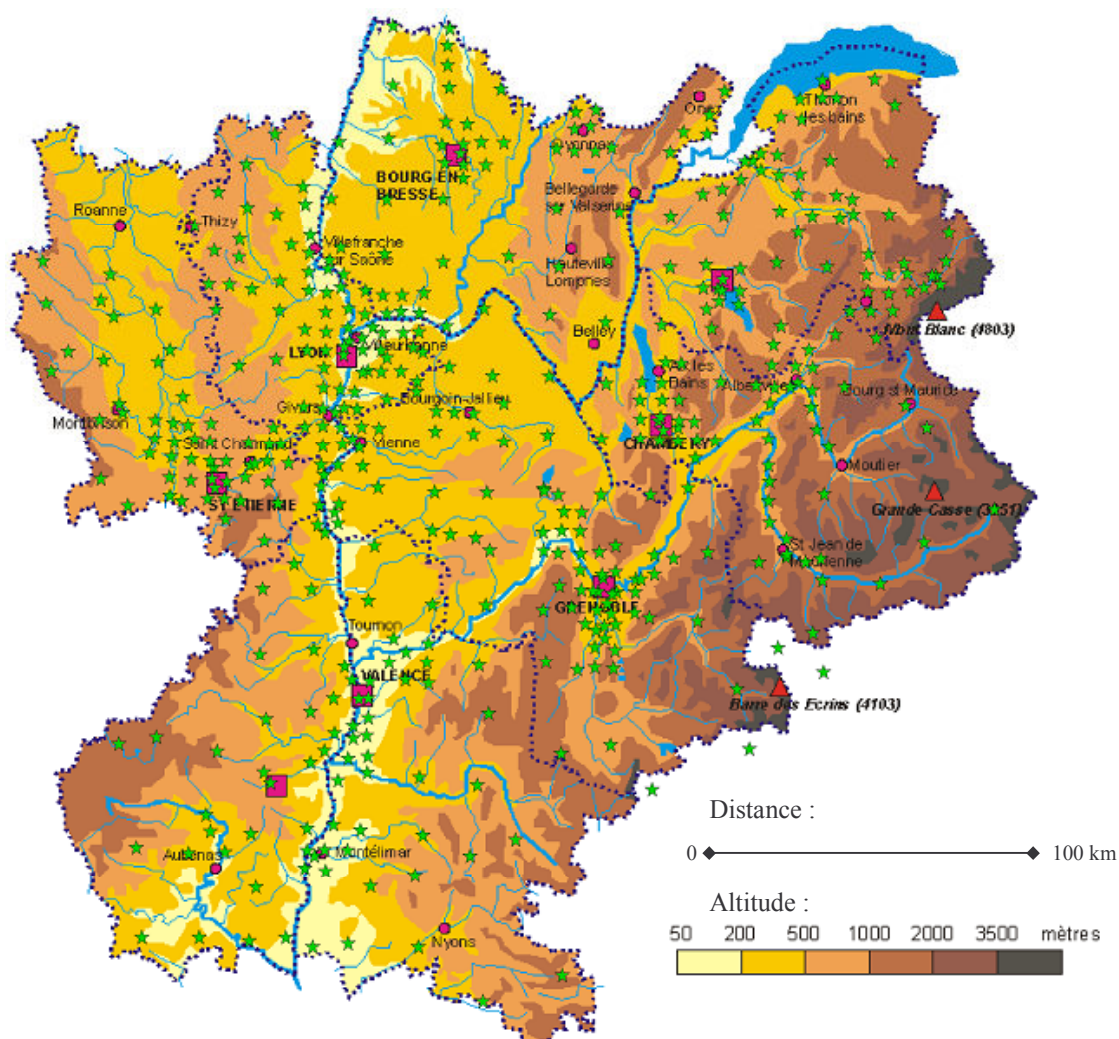


Figure 15 – Zone couverte par l'étude et implantation des sites de mesure

▪ Méthodologie de sélection des données

La variabilité spatiale des concentrations peut être liée à l'altitude (valeurs généralement plus fortes pour les altitudes élevées) et à la présence d'agglomérations (valeurs faibles dues à la présence de NO_x principalement). De plus, l'histogramme des concentrations d'ozone présente au moins 2 populations distinctes (Figure 16).

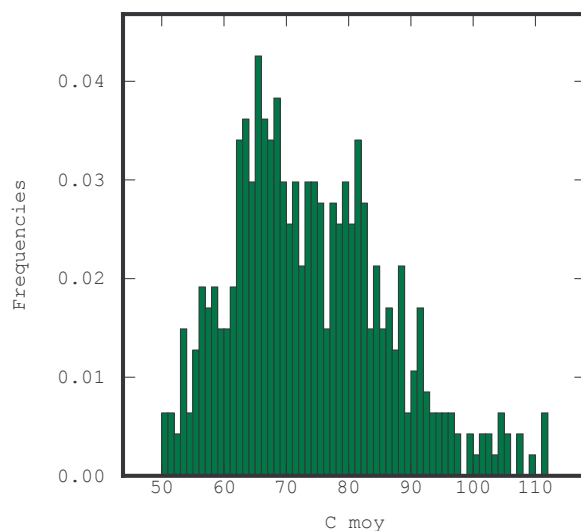


Figure 16 – Histogramme des concentrations moyennes d'ozone pour les 470 sites de mesure

Nous avons utilisé la valeur moyenne de concentration d'ozone obtenue sur les 6 campagnes, et avons repéré les points atypiques (valeurs supérieures à $100 \mu\text{g m}^{-3}$). Sur les 470 données initiales (J_0), 17 ont été ainsi écartées du jeu de données de référence pour constituer le jeu de données J_1 qui sera utilisé pour définir le modèle de référence. La totalité des données J_0 sera en revanche utilisée pour la définition de la carte de krigeage de référence. La figure ci-après présente la répartition de ces deux jeux de données.

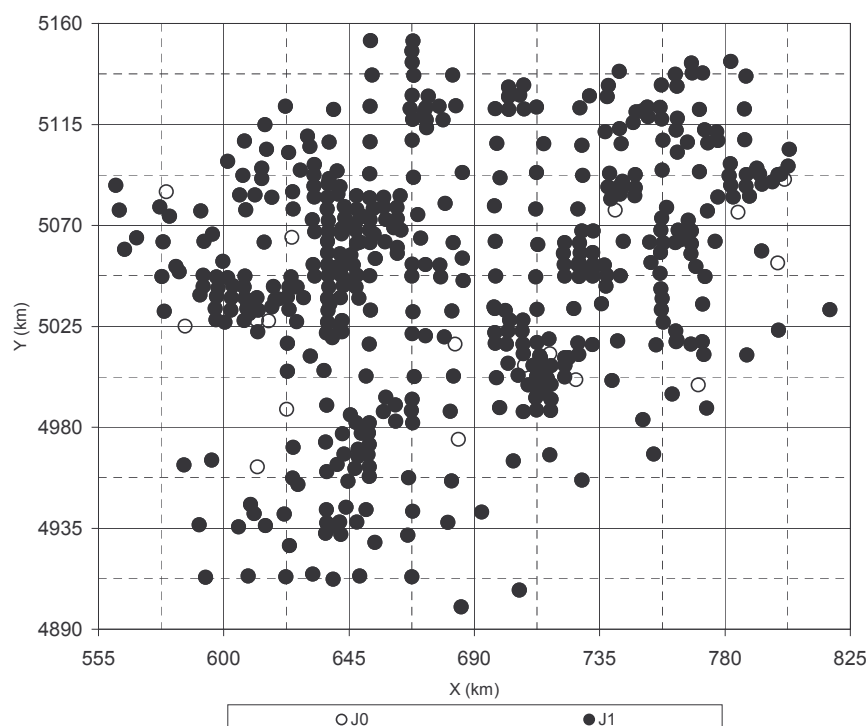


Figure 17 - Répartition des jeux de données J_0 et J_1

Les données ont été sélectionnées selon quatre maillages réguliers à partir de J_1 :

- mailles carrées de 15 km de côté (sélection S15)
- mailles carrées de 20 km de côté (sélection S20)
- mailles carrées de 30 km de côté (sélection S30)
- mailles carrées de 45 km de côté (sélection S45)

A chaque sélection correspond un jeu de données complémentaires noté C15, C20, C30 et C45, 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 le Tableau 1.

Tableau 4 – Caractéristiques statistiques des jeux de données étudiés

	J_0	J_1	S15	C15	S20	C20	S30	C30	S45	C45
Nombre de données	470	453	176	277	115	338	60	393	31	422
Minimum	50,43	50,43	50,75	50,43	53,72	50,43	50,75	50,43	58,40	50,43
Percentile 25	64,88	64,26	66,57	63,05	67,32	63,28	69,67	63,12	68,13	64,21
Médiane	72,48	71,63	74,70	69,15	75,28	70,53	78,49	70,56	76,91	71,30
Percentile 75	81,82	80,72	82,23	78,74	85,50	79,95	83,63	79,99	84,12	80,39
Max	111,69	99,83	92,91	99,83	97,95	99,83	99,83	99,19	96,02	99,83
Etendue	61,26	49,40	42,16	49,40	44,23	49,40	49,08	48,76	37,62	49,40
Moyenne	73,9	72,7	74,4	70,6	76,0	71,5	77,1	72,0	77,3	72,3
Ecart-type	12,41	10,96	9,90	10,44	11,16	10,66	11,55	10,72	10,54	10,92
Variance	153,9	120,1	98,0	109,1	124,6	113,7	133,3	114,9	111,1	119,3
CV%	16,80	15,08	13,30	14,79	14,68	14,91	14,98	14,89	13,64	15,10

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

▪ Etude du variogramme expérimental

Le variogramme retenu est omnidirectionnel avec 16 pas de 8 km chacun. Aucune donnée supplémentaire n'a été écartée.

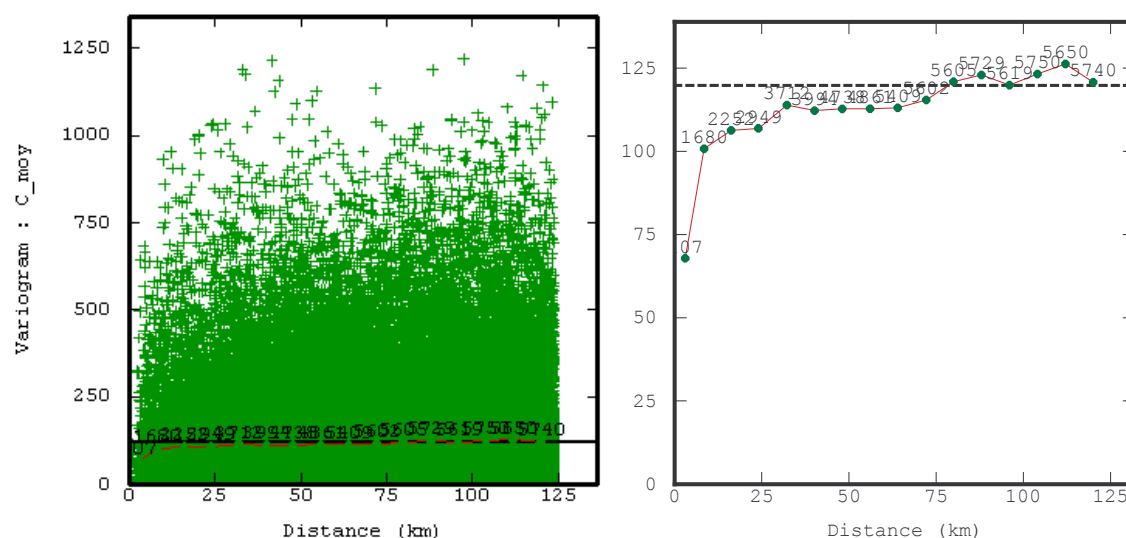


Figure 18 - Variogramme expérimental (16 pas ; 8 km) pour les 453 sites de mesure

- **Modélisation du variogramme et validation croisée**

Plusieurs modèles ont été testés pour obtenir une validation croisée de bonne qualité. Toutefois, quelles que soient les structures utilisées, le coefficient de corrélation de la validation croisée reste faible ($r < 0,5$), même en sélectionnant séparément les populations. Il a donc été décidé de privilégier le modèle utilisé par l'étude géostatistique des campagnes en Rhône-Alpes [1], composé d'une structure pépitique et de 2 structures exponentielles à courte et longue portée, pour l'étude de la dégradation de la carte de référence et du variogramme en fonction de la taille de la maille d'échantillonnage.

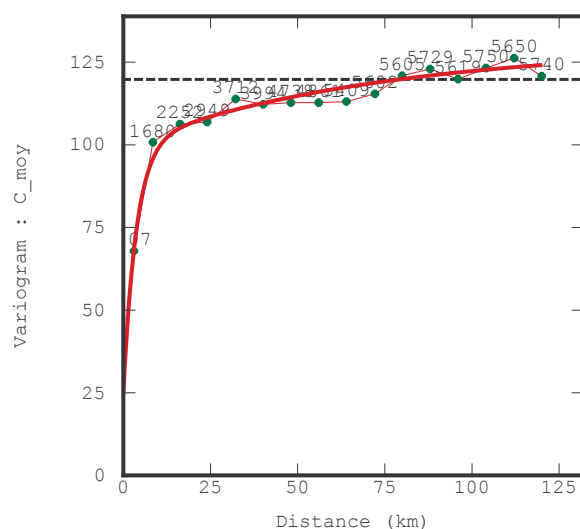


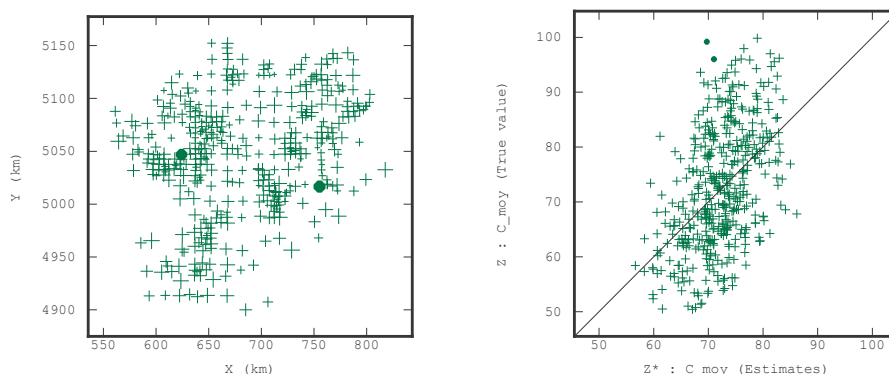
Figure 19 – Variogramme simulé : pépite (25) + exp (11 km ; 75) + exp (220 km ; 30)

En raison du grand nombre de points de mesure, on a considéré un voisinage circulaire de 70 km (1 seul secteur, 35 points). Ce voisinage sera celui utilisé par la suite lors de l'étude de dégradation du maillage.

Tableau 5 – Statistiques sur l’erreur d’estimation, obtenues par validation croisée sur le jeu de données J₁

J_1	Moyenne	Variance
Erreur	-0,452	93,672
Erreur standardisée	-0.033	0.978

2 valeurs sur 453 non robustes.



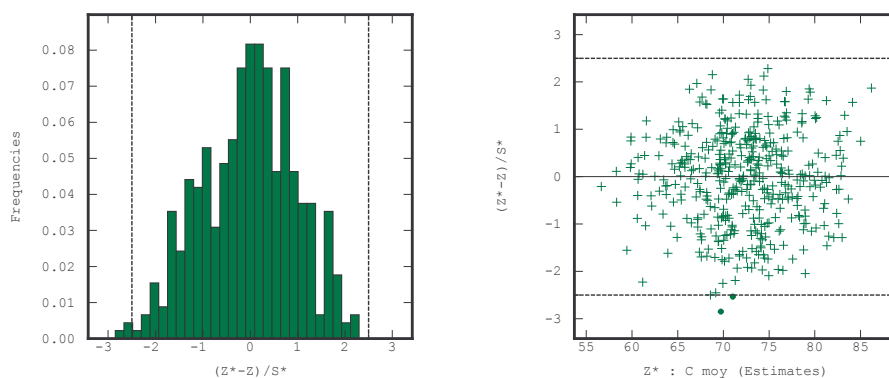


Figure 20 – Résultats de la validation croisée

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. Cependant la corrélation entre données mesurées et données estimées est plutôt faible ($r = 0,468$).

▪ Etablissement de la carte de référence

Pour établir la carte de référence, les concentrations en ozone ont été estimées aux nœuds d'une grille régulière de 2 km de côté. Les caractéristiques de cette maille d'estimation sont les suivantes :

$X_{\min} = 550 \text{ km}$	$X_{\max} = 830 \text{ km}$	$\Delta X = 280 \text{ km}$	141 nœuds
$Y_{\min} = 4880 \text{ km}$	$Y_{\max} = 5160 \text{ km}$	$\Delta Y = 280 \text{ km}$	141 nœuds

On a considéré le même voisinage que celui utilisé pour la validation croisée, et utilisé le jeu de données J_0 .

Tableau 6 – Statistiques sur l'estimation sur une grille régulière de 2 km de côté, par krigeage ordinaire ponctuel

J_0	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	470	11 337	11 337
Minimum	50,43	55,37	6,63
Percentile 25	64,88	72,22	9,50
Médiane	72,48	76,90	10,10
Percentile 75	81,82	81,07	10,36
Maximum	111,69	101,18	11,33
Etendue	61,26	45,81	4,70
Moyenne	73,9	76,63	9,87
Ecart-type	12,41	6,11	0,75
Variance	153,9	37,33	0,56
CV%	16,80	7,97	7,58

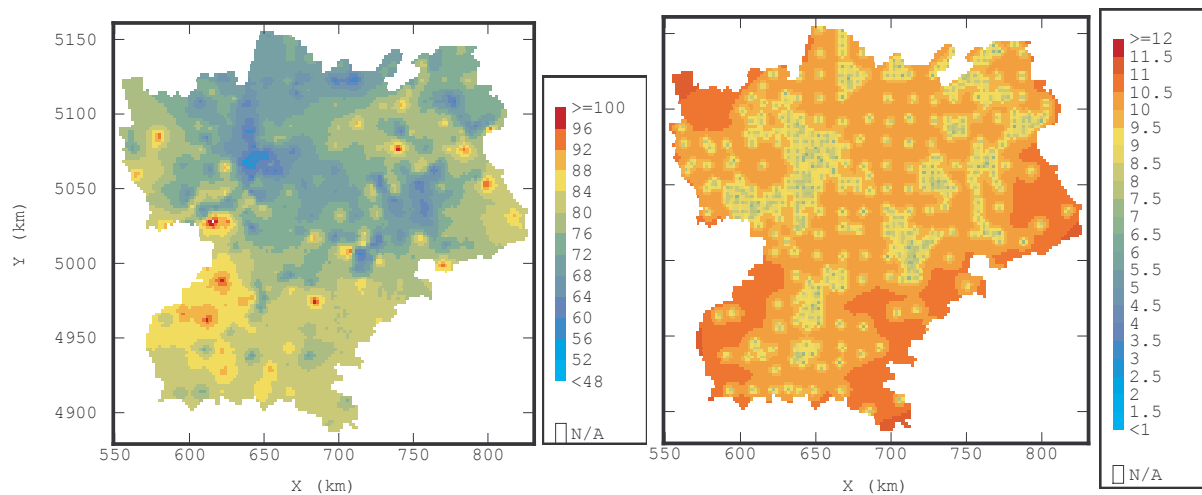


Figure 21 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données J_0 , avec le modèle de référence et (droite) écarts-types de krigeage

De manière générale, on voit bien apparaître sur la carte de krigeage, les zones de concentrations faibles (agglomérations), moyennes (plaines) et plus fortes (montagne). Cependant, on observe des continuités spatiales entre les vallées là où elles ne devraient pas apparaître : la prise en compte de variables auxiliaires (que nous utiliserons dans la 2^{ème} partie de cette étude) permettrait vraisemblablement d'améliorer la cartographie. De plus, à cause de la forte variabilité à petite distance, l'écart-type de krigeage augmente rapidement dès qu'on s'éloigne du point de mesure.

A partir de ce jeu de référence, l'impact de la dégradation du maillage a été étudié sur des maillages carrés réguliers.

▪ ETUDE DE LA SELECTION DE DONNEES S15 (MAILLE DE 15 × 15 KM)

La sélection S15 contient 176 données correspondant à 1 point de mesure par maille carrée de 15 km de côté.

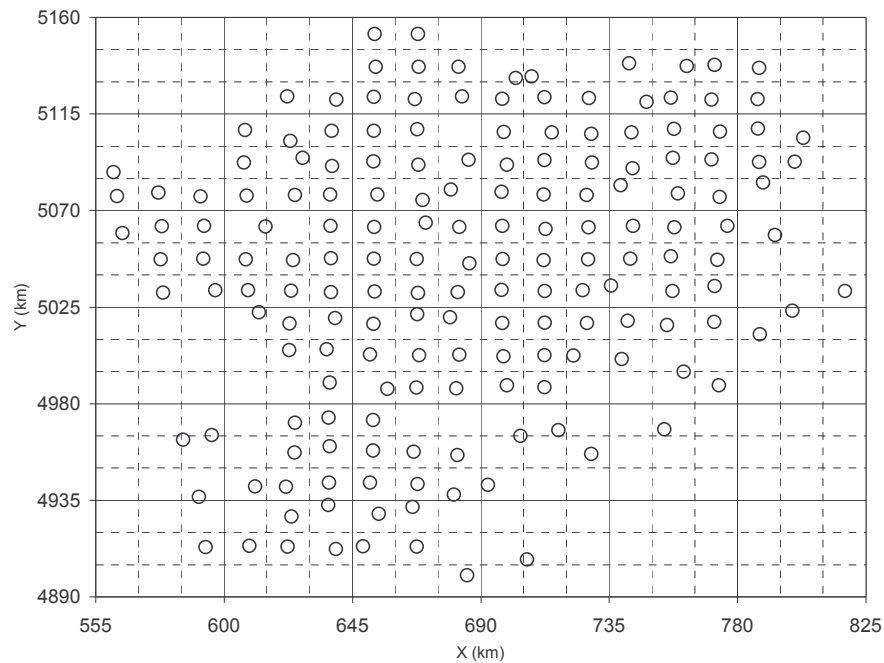


Figure 22 - Répartition du jeu de données S15

• Etude du variogramme expérimental

Le variogramme choisi est omnidirectionnel et comprend 16 pas de 8 km chacun. Aucun point n'a été éliminé.

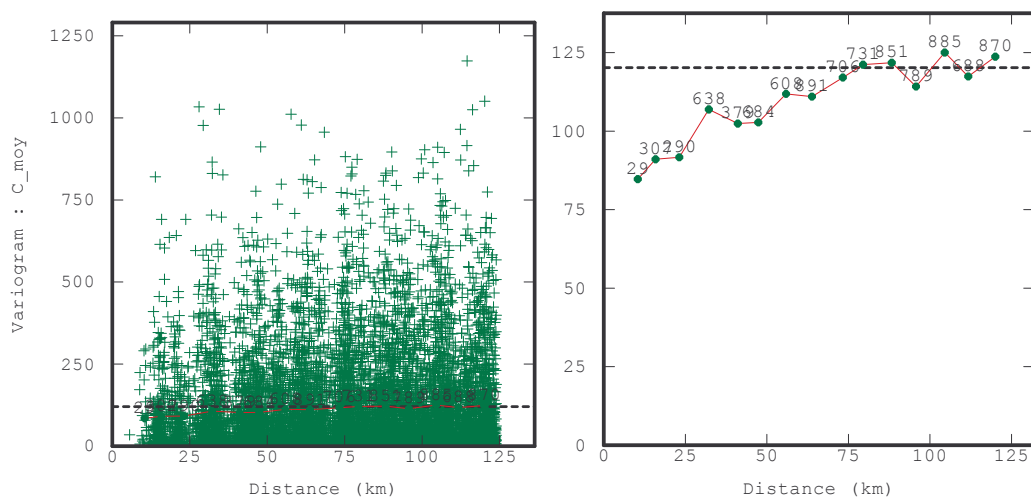


Figure 23 - Variogramme expérimental (16 pas ; 8 km) pour les 176 sites de mesure

• Modélisation du variogramme et test de validation croisée

Un modèle omnidirectionnel est ajusté au variogramme expérimental, puis soumis au test de validation croisée.

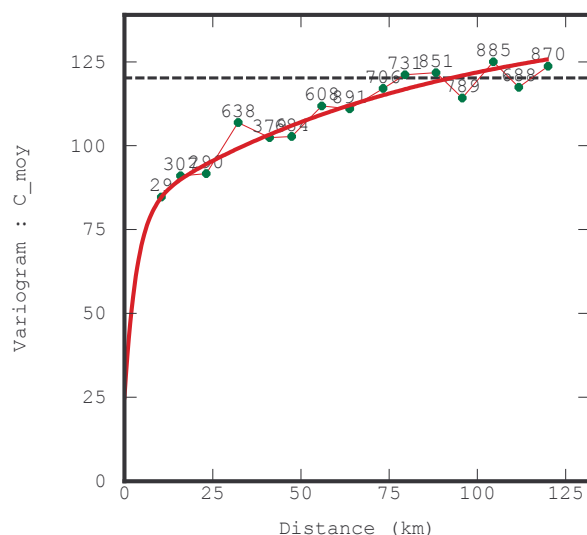


Figure 24 – Variogramme simulé : pépite (25) + exp (10 km ; 55) + exp (270 km ; 70)

Tableau 7 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur le jeu de données S15

S15	Moyenne	Variance
Erreur	-0,0208	101,03
Erreur standardisée	-0,0023	1,037

1 valeur non robuste sur 176 données

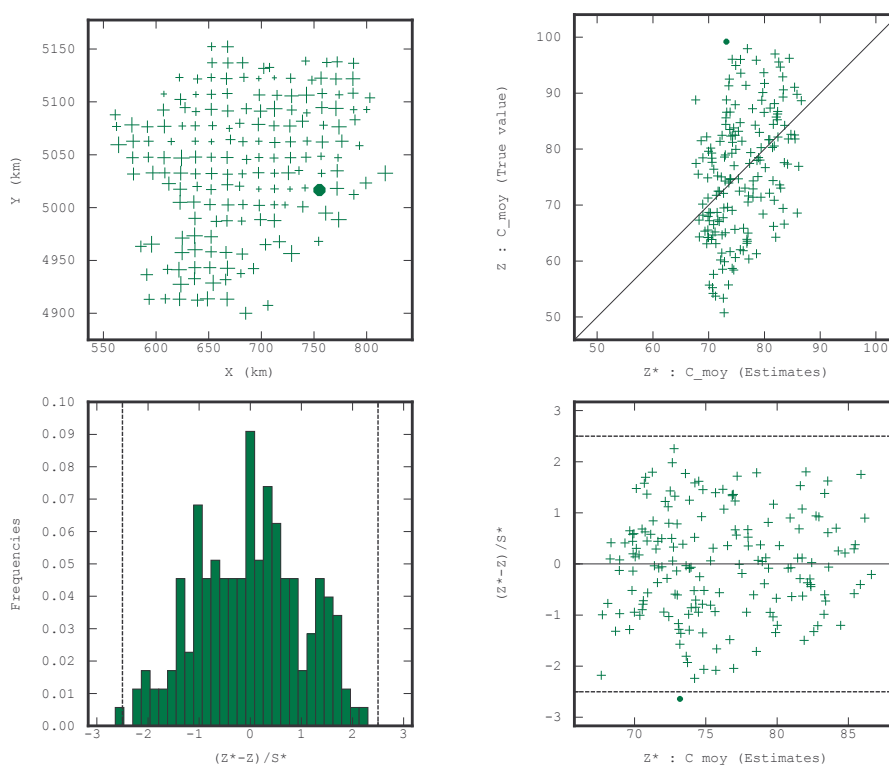


Figure 25 – Résultats de la validation croisée

La corrélation entre valeurs mesurées et estimées s'est légèrement dégradée par rapport au jeu de données de référence ($r = 0,402$).

• Test de validation absolue sur le complémentaire C15

Les concentrations d'ozone sont estimées aux points n'ayant pas été sélectionnés dans le jeu de données S15, puis comparées aux valeurs réelles.

Tableau 8 – Résultats de la validation absolue sur le complémentaire C15 pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S15

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S15	
C15	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	277	277	277
Minimum	50,43	64,76	7,65
Percentile 25	63,05	69,51	9,39
Médiane	69,15	73,30	9,49
Percentile 75	78,74	77,88	9,57
Maximum	99,83	87,14	10,02
Etendue	49,40	22,38	2,37
Moyenne	70,6	74,05	9,46
Ecart-type	10,44	5,30	0,22
Variance	109,1	28,05	0,05
CV%	14,79	7,15	2,35
Corrélation mesures/estimations		$r = 0,318$	

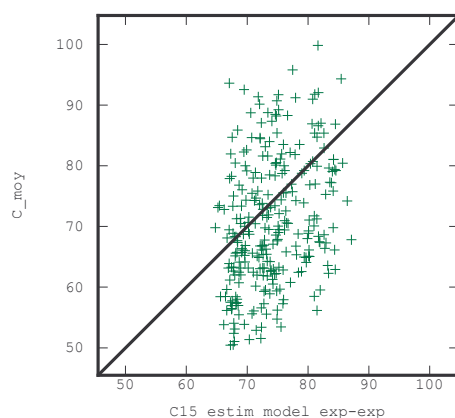


Figure 26 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C15

- **Etablissement de la carte pour la sélection S15**

La carte de krigeage correspondante a été établie.

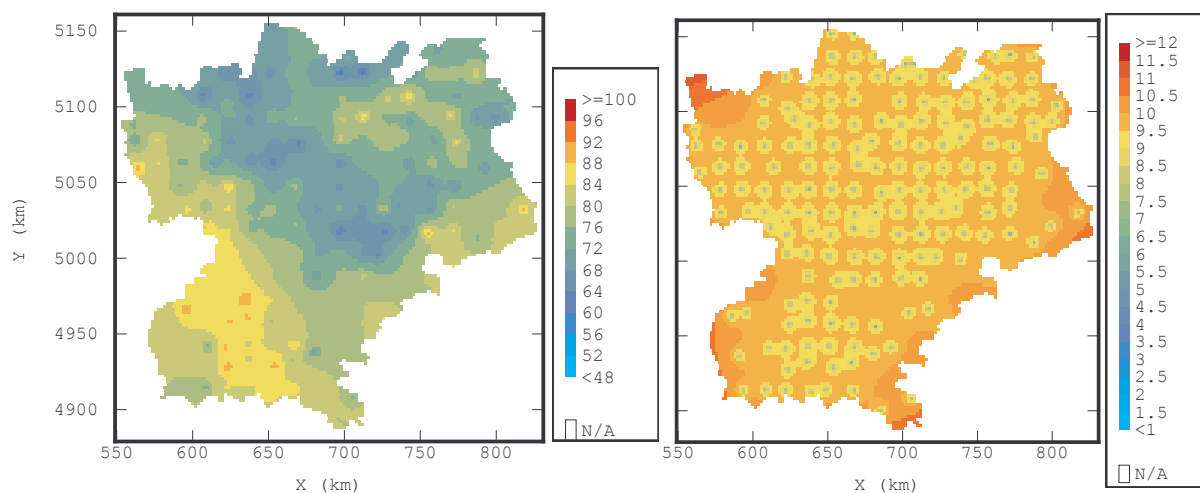


Figure 27 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S15, avec le modèle de la sélection et (droite) écarts-types de krigeage

Avec ce maillage de 15×15 km, calé sur le maillage principal adopté par l'étude de 2002, on voit diminuer l'écart-type de krigeage puisque l'ensemble du territoire est correctement couvert. Les concentrations estimées ont subi un lissage visible.

▪ ETUDE DE LA SELECTION DE DONNEES S20 (MAILLE DE 20×20 KM)

La sélection S20 contient 115 données correspondant à 1 point de mesure par maille carrée de 20 km de côté.

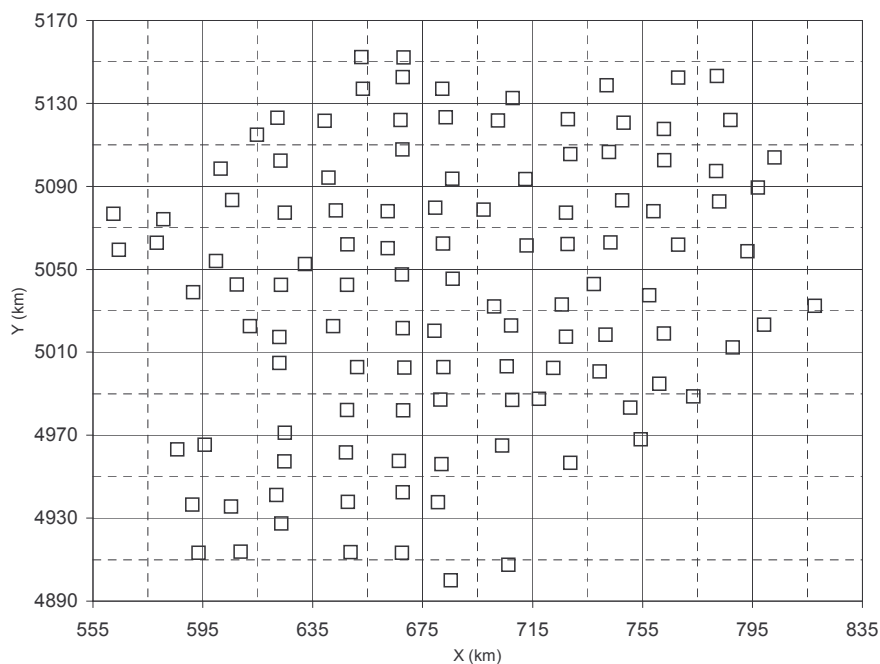


Figure 28 - Répartition du jeu de données S20

- **Etude du variogramme expérimental**

Le variogramme choisi est omnidirectionnel et comprend 15 pas de 9 km chacun. Aucun point n'a été éliminé.

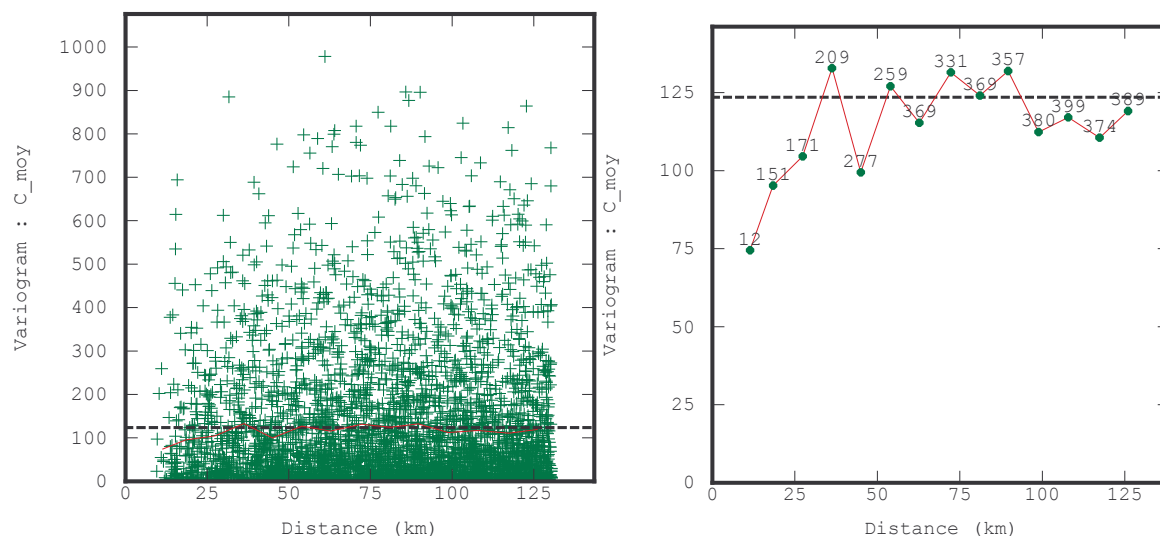


Figure 29 - Variogramme expérimental (15 pas ; 9 km) pour les 115 sites de mesure

- **Modélisation du variogramme et test de validation croisée**

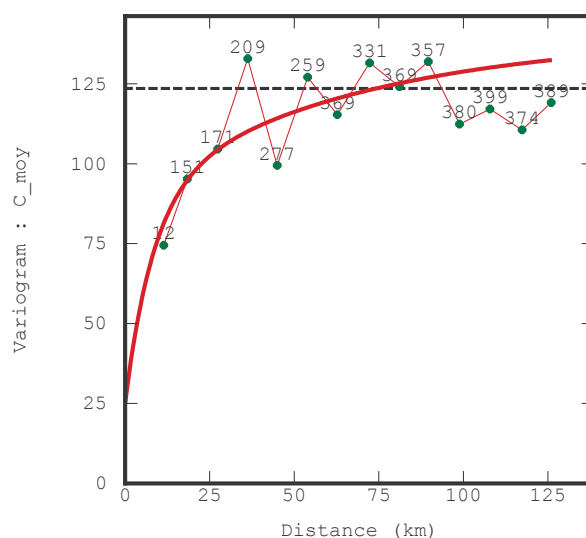


Figure 30 – Variogramme simulé : pépite (25) + exp (25 km ; 65) + exp (200 km ; 50)

Tableau 9 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur le jeu de données S20

S20	Moyenne	Variance
Erreur	-0,045	107,91
Erreur standardisée	-0,006	1,020

Toutes les données sont robustes.

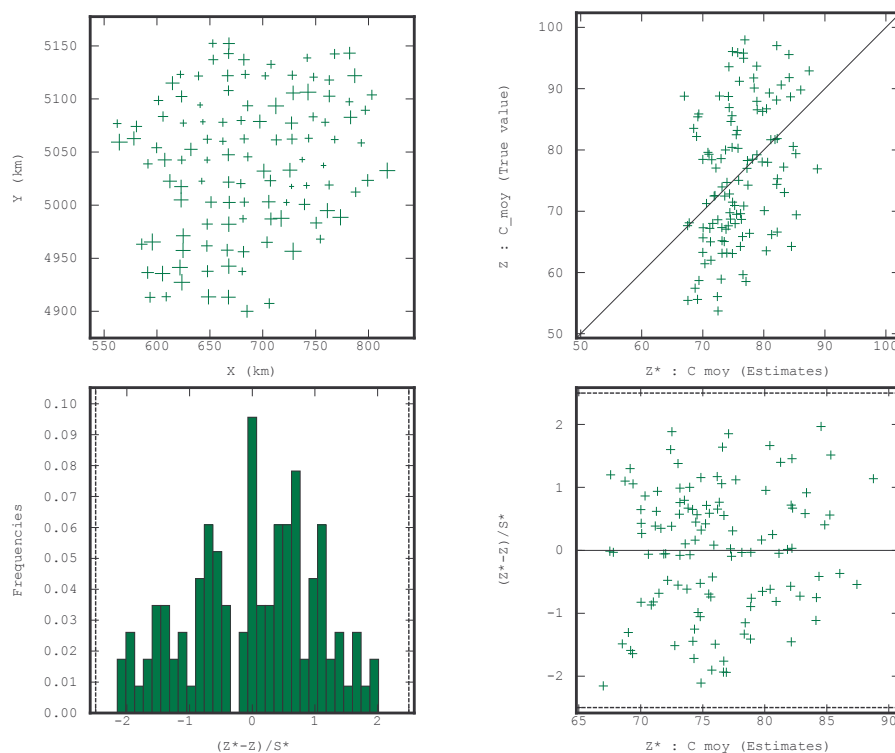


Figure 31 – Résultats de la validation croisée

La corrélation entre valeurs mesurées et estimées est de $r = 0,364$.

• Test de validation absolue sur le complémentaire C20

Les concentrations d’ozone sont estimées aux points n’ayant pas été sélectionnés dans le jeu de données S20, puis comparées aux valeurs réelles.

Tableau 10 – Résultats de la validation absolue sur le complémentaire C20 pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S20

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S20	
C20	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d’estimation ($\mu\text{g m}^{-3}$)
Nombre de données	338	338	338
Minimum	50,43	61,69	7,33
Percentile 25	63,28	68,66	9,15
Médiane	70,53	73,07	9,42
Percentile 75	79,95	78,07	9,74
Maximum	99,83	89,47	10,65
Etendue	49,40	27,78	3,32
Moyenne	71,5	73,39	9,43
Ecart-type	10,66	5,87	0,43
Variance	113,7	34,45	0,19
CV%	14,91	8,00	4,58
Corrélation mesures/estimations		$r = 0,324$	

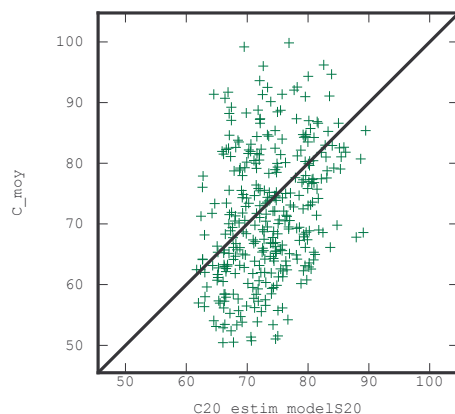


Figure 32 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C20

- **Etablissement de la carte pour la sélection S20**

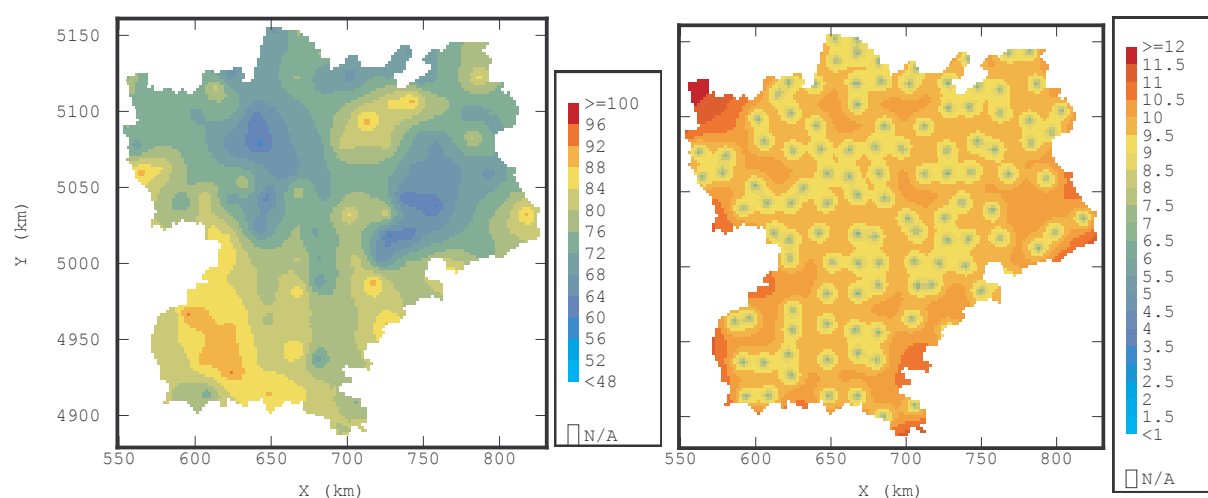


Figure 33 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S20, avec le modèle de la sélection et (droite) écarts-types de krigeage

- **ETUDE DE LA SELECTION DE DONNEES S30 (MAILLE DE 30 x 30 KM)**

La sélection S30 contient 60 données correspondant à 1 point de mesure par maille carrée de 30 km de côté.

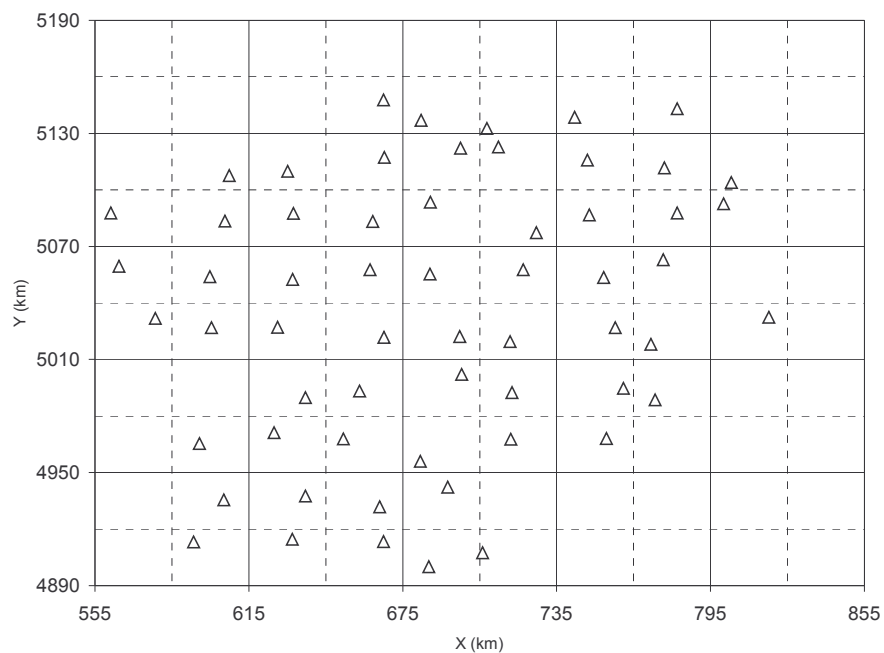


Figure 34 - Répartition du jeu de données S30

• Etude du variogramme expérimental

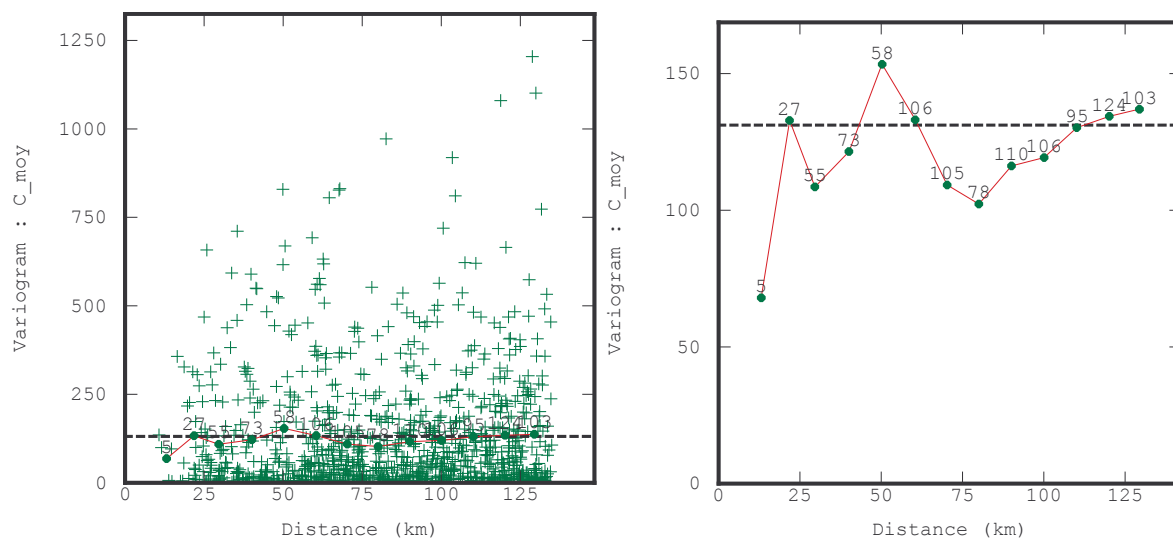


Figure 35 - Variogramme expérimental (14 pas ; 10 km) pour les 60 sites de mesure

- Modélisation du variogramme et test de validation croisée

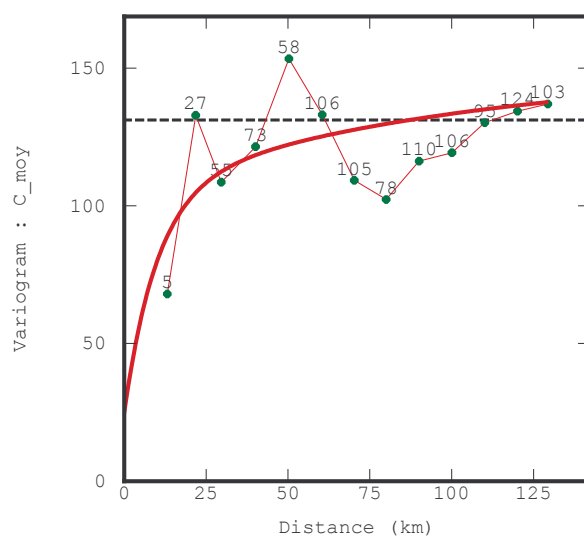


Figure 36 – Variogramme simulé : pépite (25) + exp (30 km ; 80) + exp (300 km ; 45)

Tableau 11 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur le jeu de données S30

S30	Moyenne	Variance
Erreur	-0,041	132,66
Erreur standardisée	0,0006	1,081

Toutes les données sont robustes.

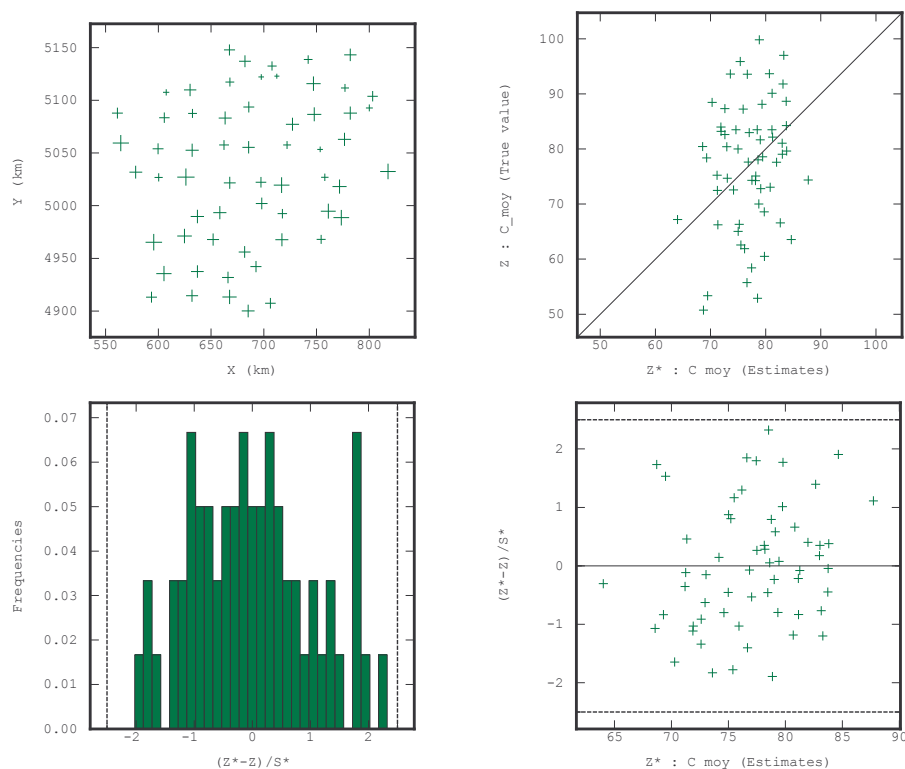


Figure 37 – Résultats de la validation croisée

La corrélation entre valeurs estimées et valeurs réelles s'est nettement dégradée ($r = 0,196$).

- **Test de validation absolue sur le complémentaire C30**

Tableau 12 – Résultats de la validation absolue sur le complémentaire C30 pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S30

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S30	
C30	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	393	393	393
Minimum	50,43	58,58	7,49
Percentile 25	63,12	73,37	9,93
Médiane	70,56	76,43	10,35
Percentile 75	79,99	80,03	10,60
Maximum	99,19	88,05	11,27
Etendue	48,76	29,47	3,78
Moyenne	72,0	76,49	10,22
Ecart-type	10,72	4,86	0,53
Variance	114,9	23,62	0,28
CV%	14,89	6,35	5,19
Corrélation mesures/estimations		$r = 0,231$	

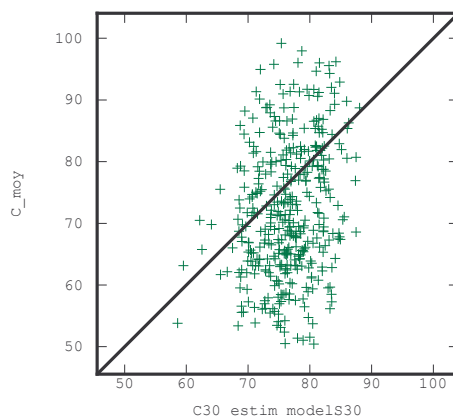


Figure 38 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C30

- **Etablissement de la carte pour la sélection S30**

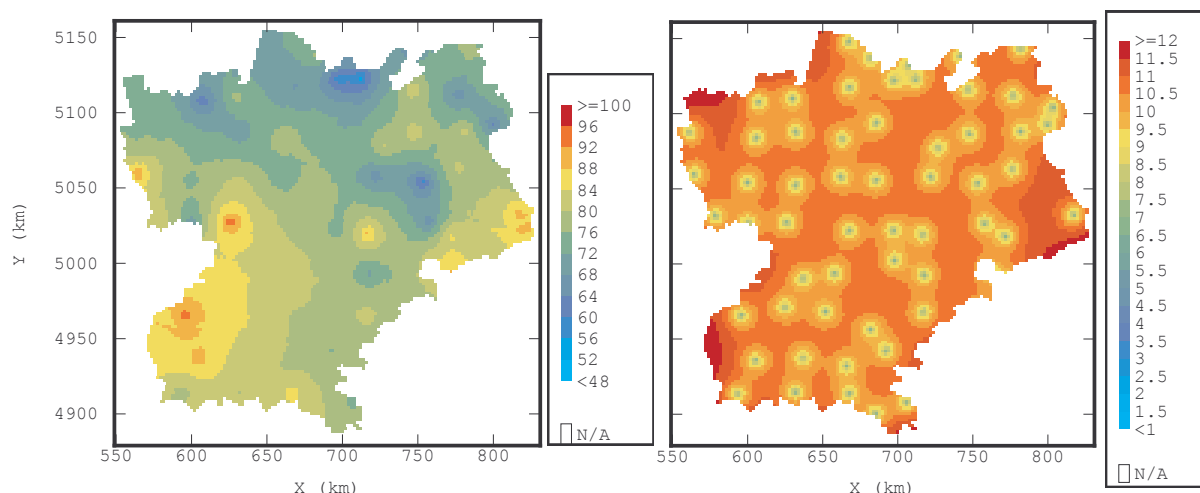


Figure 39 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S30, avec le modèle de la sélection et (droite) écarts-types de krigeage

▪ ETUDE DE LA SELECTION DE DONNEES S45 (MAILLE DE 45 × 45 KM)

La sélection S45 contient 31 données correspondant à 1 point de mesure par maille carrée de 45 km de côté.

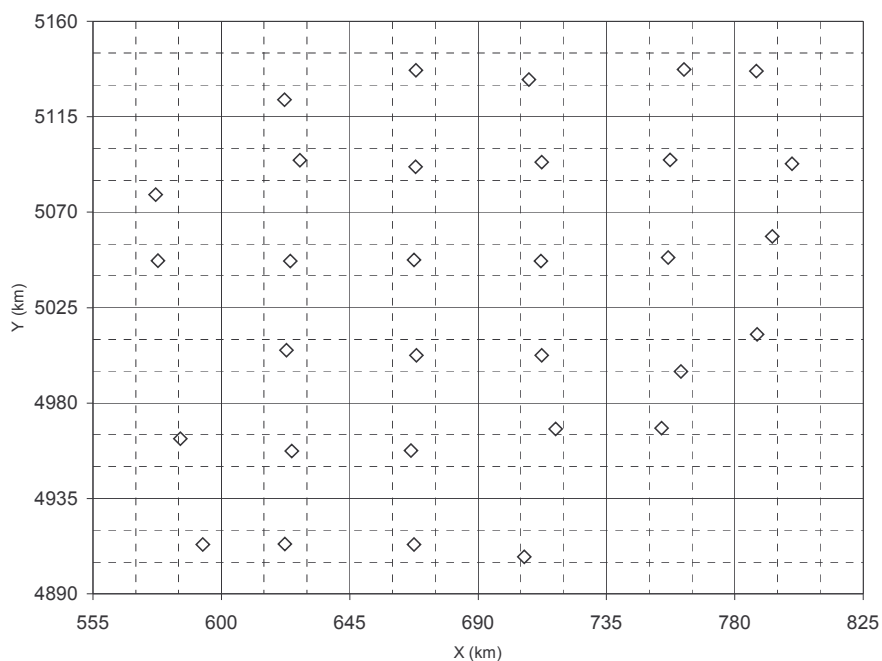


Figure 40 - Répartition du jeu de données S45

- Etude du variogramme expérimental

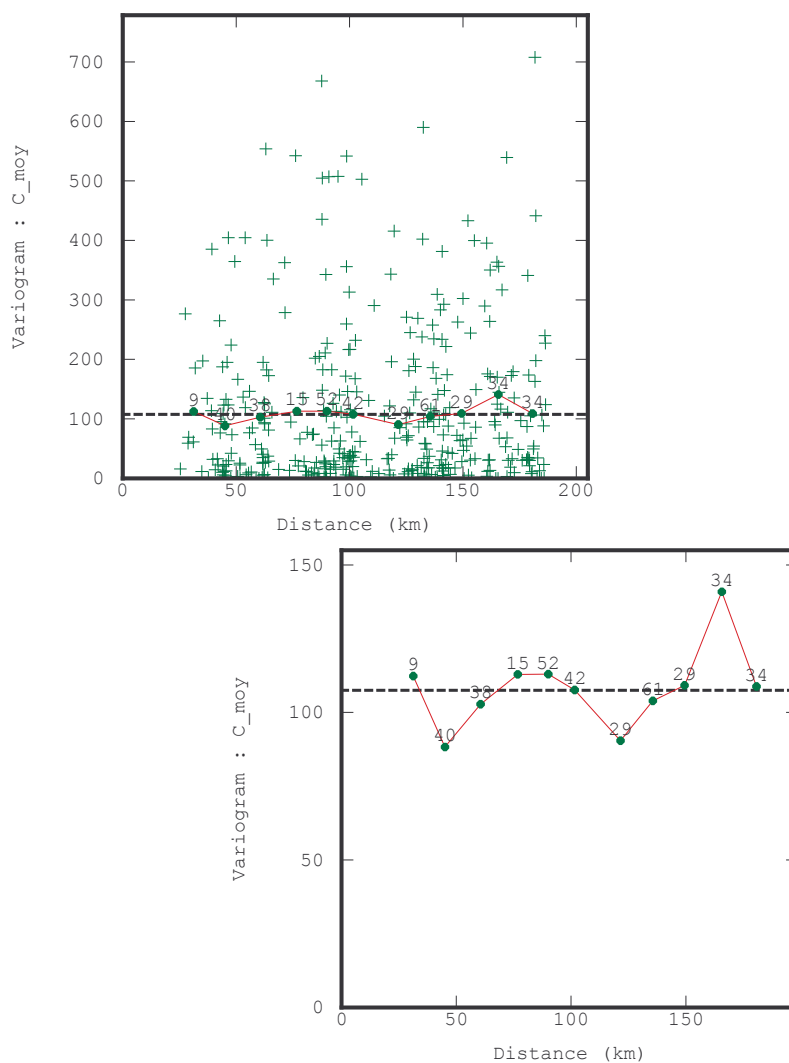


Figure 41 - Variogramme expérimental (13 pas ; 15 km) pour les 31 sites de mesure

- Modélisation du variogramme et test de validation croisée

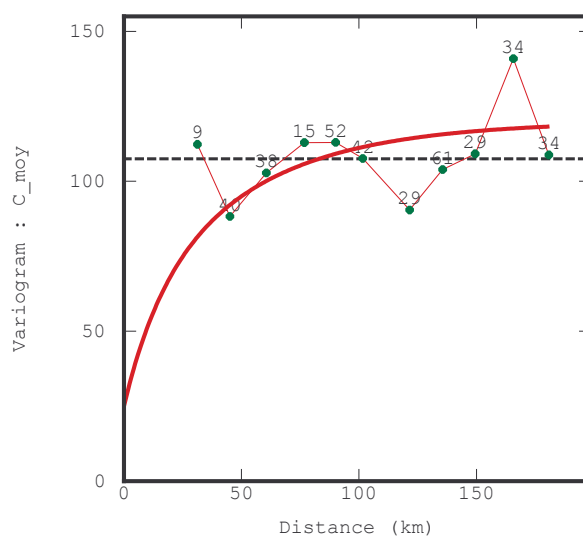


Figure 42 – Variogramme simulé : pépite (25) + exp (45 km ; 30) + exp (150 km ; 65)

Tableau 13 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur le jeu de données S45

S45	Moyenne	Variance
Erreur	0,6495	106,31
Erreur standardisée	0,0522	1,023

Toutes les données sont robustes.

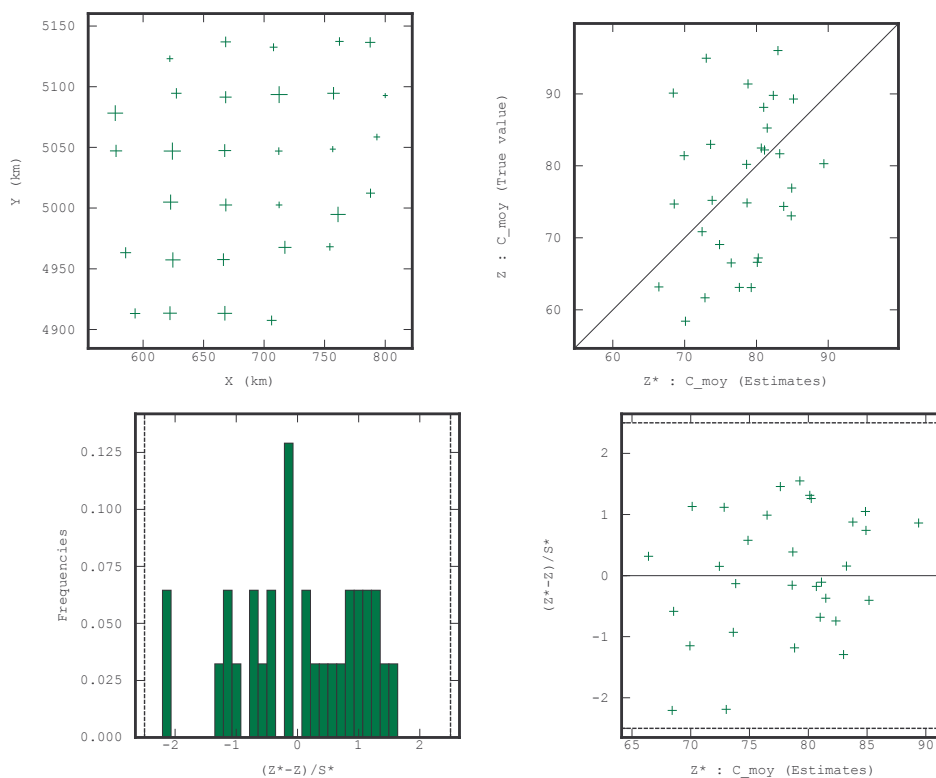


Figure 43 – Résultats de la validation croisée

La corrélation entre concentrations estimées et mesurées est de $r = 0,285$.

• Test de validation absolue sur le complémentaire C45

Tableau 14 – Résultats de la validation absolue sur le complémentaire C45 pour le krigeage ordinaire ponctuel, avec le modèle de la sélection S45

		Validation croisée par krigeage ordinaire ponctuel selon le modèle de la sélection S45	
C45	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	422	422	422
Minimum	50,43	61,54	6,98
Percentile 25	64,21	72,20	9,00
Médiane	71,30	78,45	9,23
Percentile 75	80,39	84,24	9,36
Maximum	99,83	91,40	10,55
Etendue	49,40	29,86	3,57
Moyenne	72,3	78,17	9,11
Ecart-type	10,92	6,84	0,43
Variance	119,3	46,79	0,18
CV%	15,10	8,75	4,68
Corrélation mesures/estimations		$r = 0,131$	

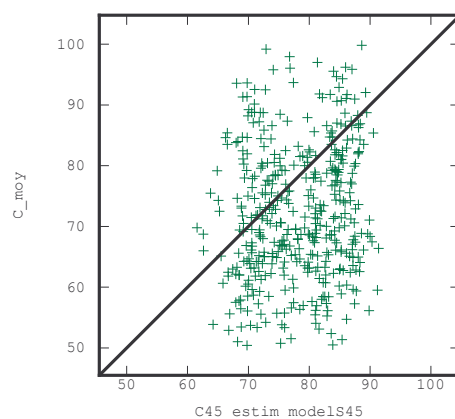


Figure 44 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C45

- **Etablissement de la carte pour la sélection S45**

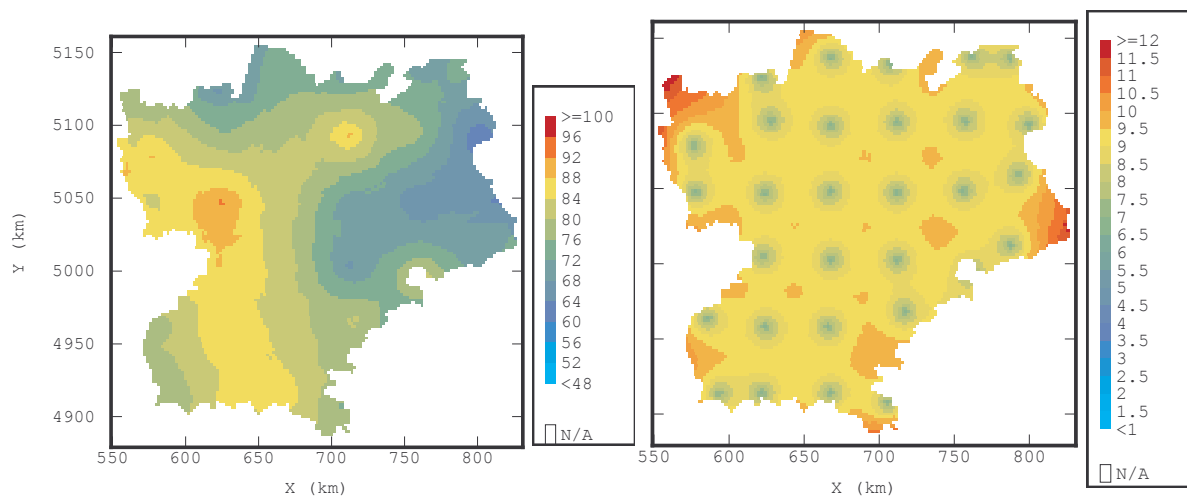


Figure 45 – (gauche) Estimation par krigeage ponctuel, à partir du jeu de données S45, avec le modèle de la sélection et (droite) écarts-types de krigeage

■ COMPARAISON DES ESTIMATIONS ET DISCUSSION

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 12).

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 15 – Synthèse des résultats de validation

Valeurs	J ₁	S15	S20	S30	S45
Nombre de données	453	176	115	60	31
Minimum ($\mu\text{g m}^{-3}$)	50,43	50,75	53,72	50,75	58,40
Max ($\mu\text{g m}^{-3}$)	99,83	92,91	97,95	99,83	96,02
Moyenne ($\mu\text{g m}^{-3}$)	72,7	74,4	76,0	77,1	77,3
Ecart-type ($\mu\text{g m}^{-3}$)	10,96	9,90	11,16	11,55	10,54
Variance ($\mu\text{g m}^{-3}$)	120,1	98,0	124,6	133,3	111,1

Validation croisée						
Erreur						
	Moyenne	-0,452	-0,021	-0,045	-0,041	0,650
	Variance	93,67	101,03	107,91	132,66	106,31
Erreur standardisée						
	Moyenne	-0,033	-0,0023	-0,006	0,0006	0,0522
	Variance	0,978	1,037	1,020	1,081	1,023
Coefficient de corrélation $r =$		0,468	0,402	0,364	0,196	0,285

Validation absolue						
Erreur						
	Moyenne	-	9,46	9,43	10,22	9,11
	Variance	-	0,05	0,19	0,28	0,18
Coefficient de corrélation $r =$		-	0,318	0,324	0,231	0,131

Les différents variogrammes expérimentaux peuvent également être comparés.

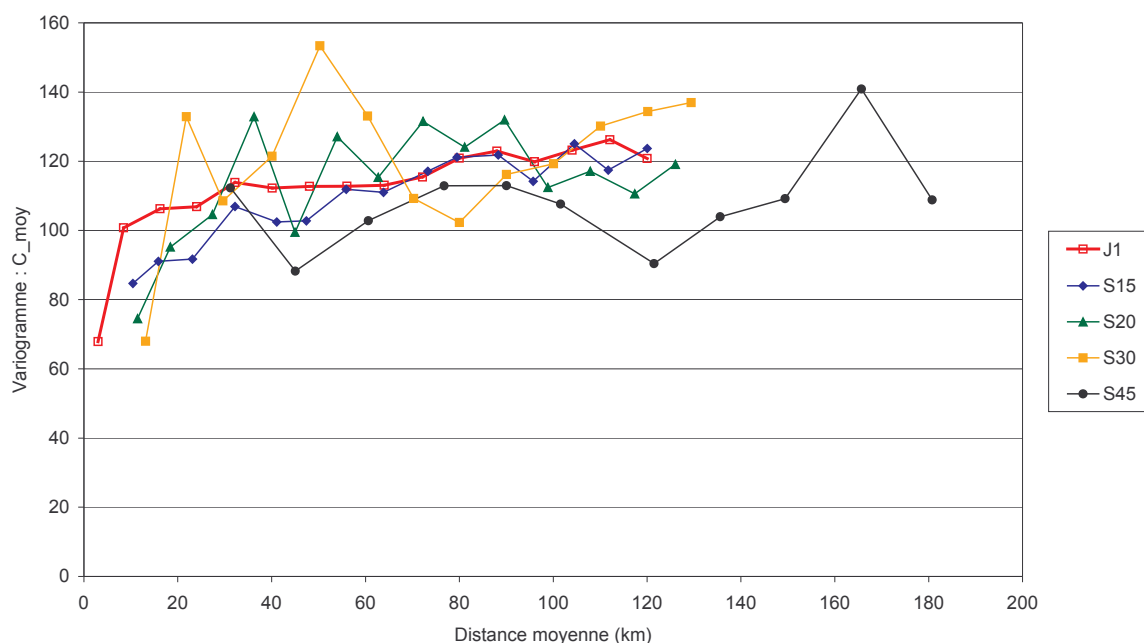


Figure 46 – Comparaison des variogrammes expérimentaux

La structure du variogramme varie avec le nombre de points. On observe un net changement de structure à partir de S30, mais dès S20 la structure du variogramme est plus irrégulière et la portée de la première exponentielle augmente plus nettement.

On peut d'ailleurs remarquer que jusqu'à la sélection S20, comme le montrent les figures ci-dessous, les corrélations entre valeurs estimées de la carte de référence et valeurs estimées des différentes sélections sont bonnes et que les grandes zones définies par la carte de référence sont globalement correctement reproduites.

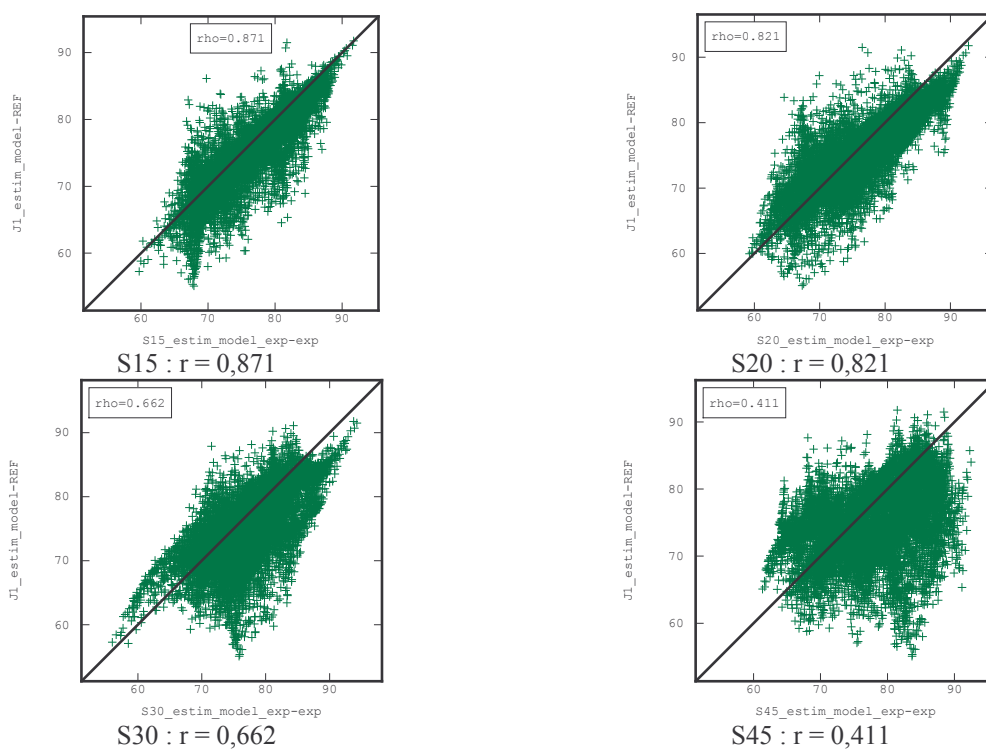


Figure 47 – Comparaison entre l'estimation de référence et les estimations obtenues à partir des différentes sélections

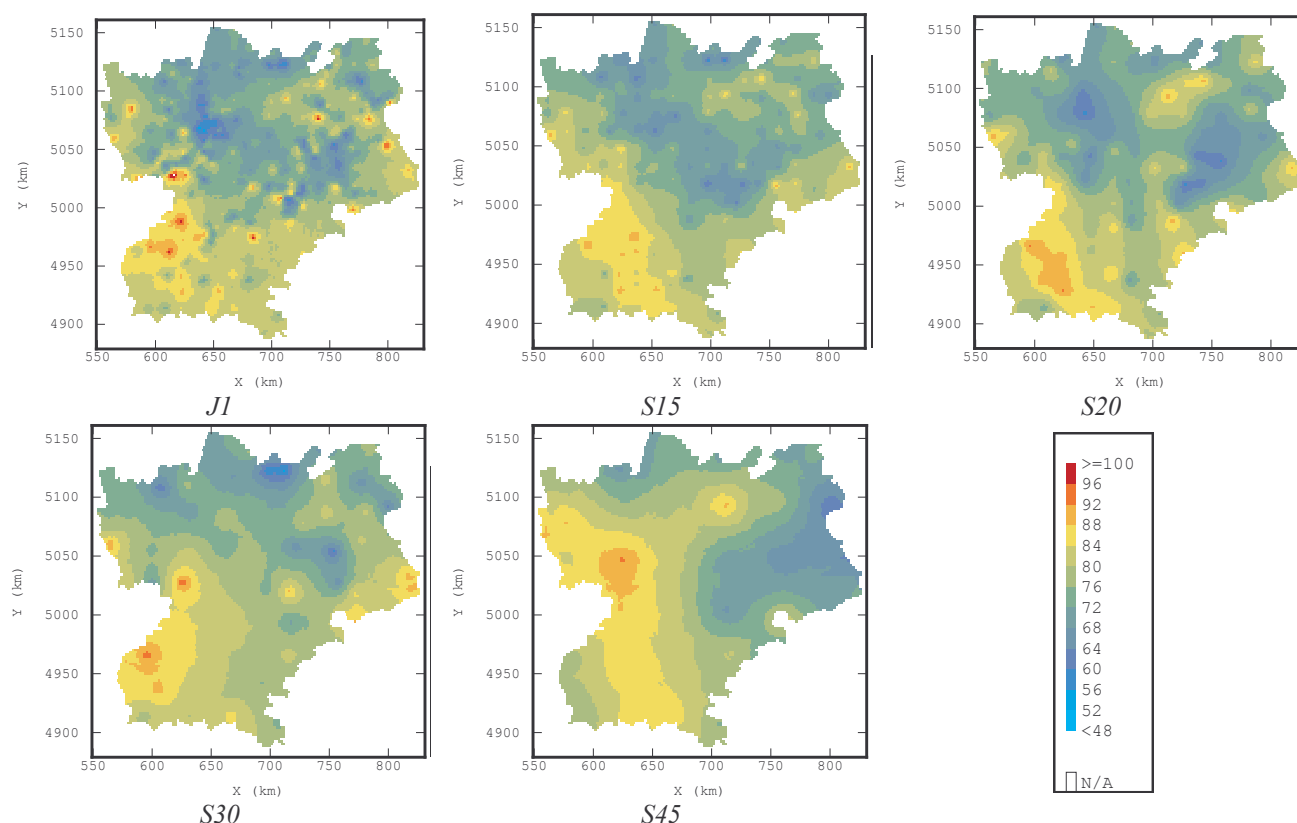


Figure 48 – Comparaison des cartes d'estimation établies par krigeage ponctuel sur la grille de référence pour les différentes sélections

La comparaison des écarts-types de krigeage est également intéressante: les cartes obtenues pour les sélections S30 et S45 confirment un résultat notable dans l'étude précédente traitant du NO_2 particulièrement en zone urbaine [3], à savoir qu'une estimation réalisée avec des données peu nombreuses mais homogènes spatialement conduit à des erreurs de krigeage moindres qu'avec un nombre plus important de données réparties de façon plus hétérogène sur la zone d'étude, comme c'est le cas pour la sélection S30.

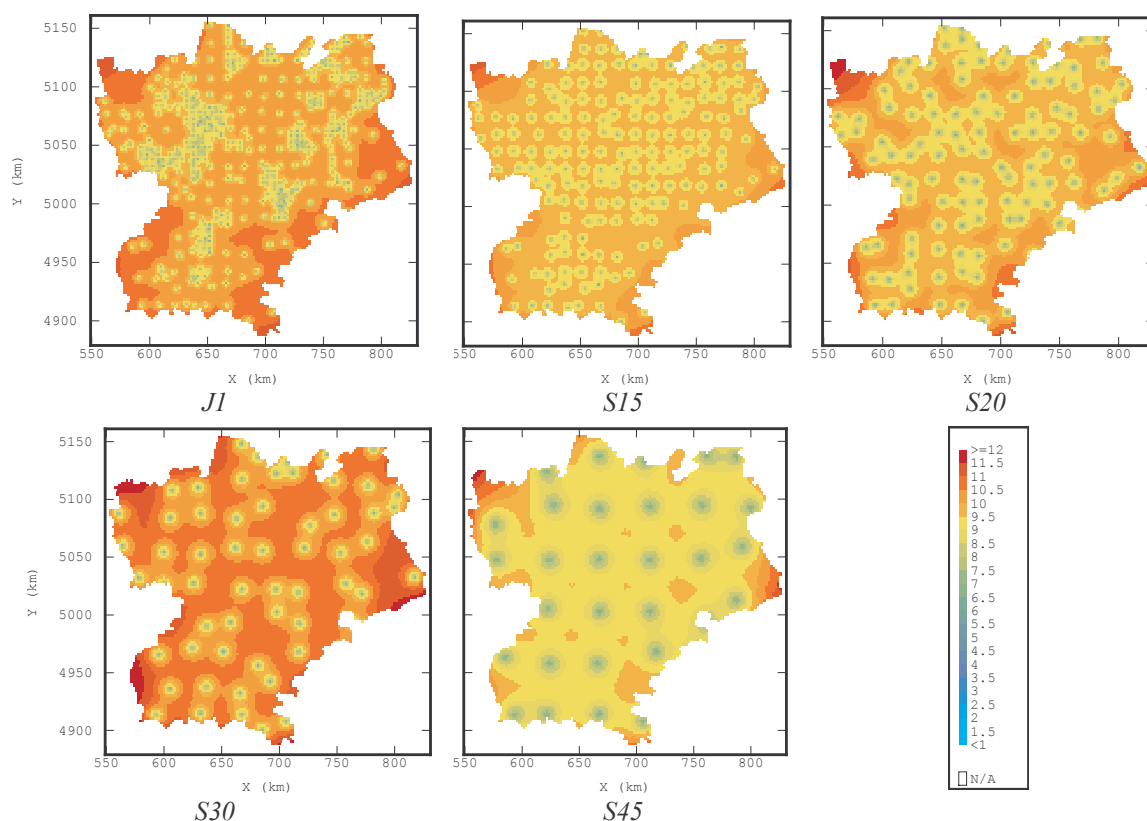


Figure 49 – Comparaison des cartes d'écarts-types de krigeage établies par krigeage ponctuel sur la grille de référence pour les différentes sélections

On peut évaluer la robustesse du krigeage en estimant les concentrations d'ozone sur la grille à partir des points de chaque sélection mais en utilisant, non pas le modèle de variogramme ajusté pour cette sélection, mais le modèle de référence défini à partir du jeu de données J_1 . On observe une sensibilité importante au variogramme à partir de la sélection S20. Cette dégradation est vraisemblablement liée (au moins partiellement) à la structure spatiale du phénomène, et notamment à sa corrélation avec l'altitude.

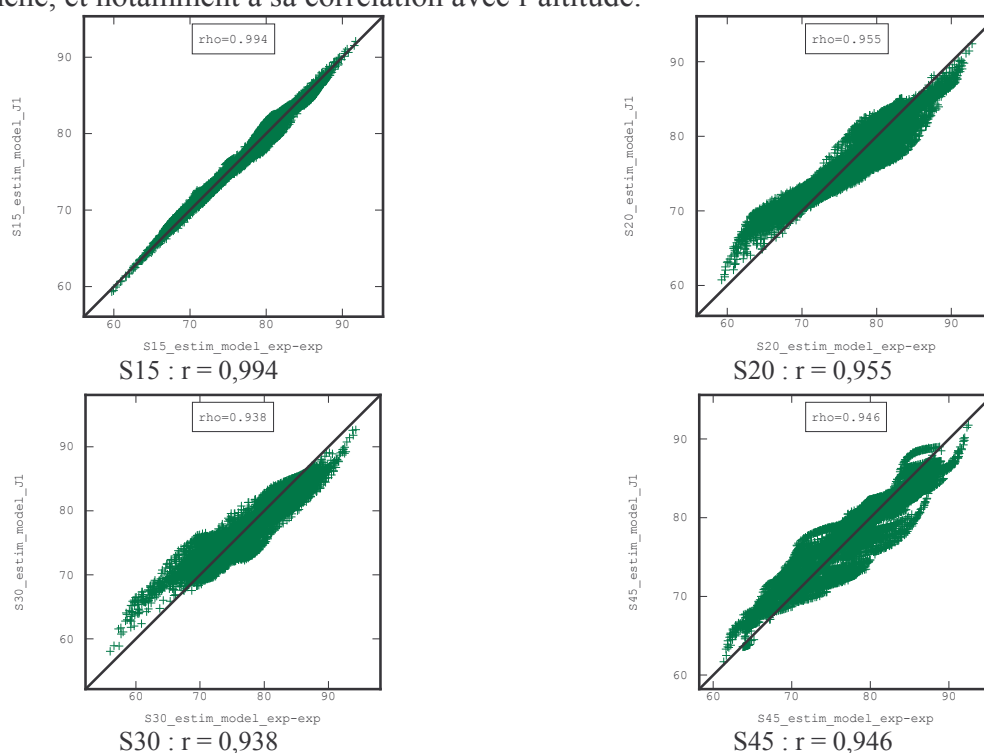
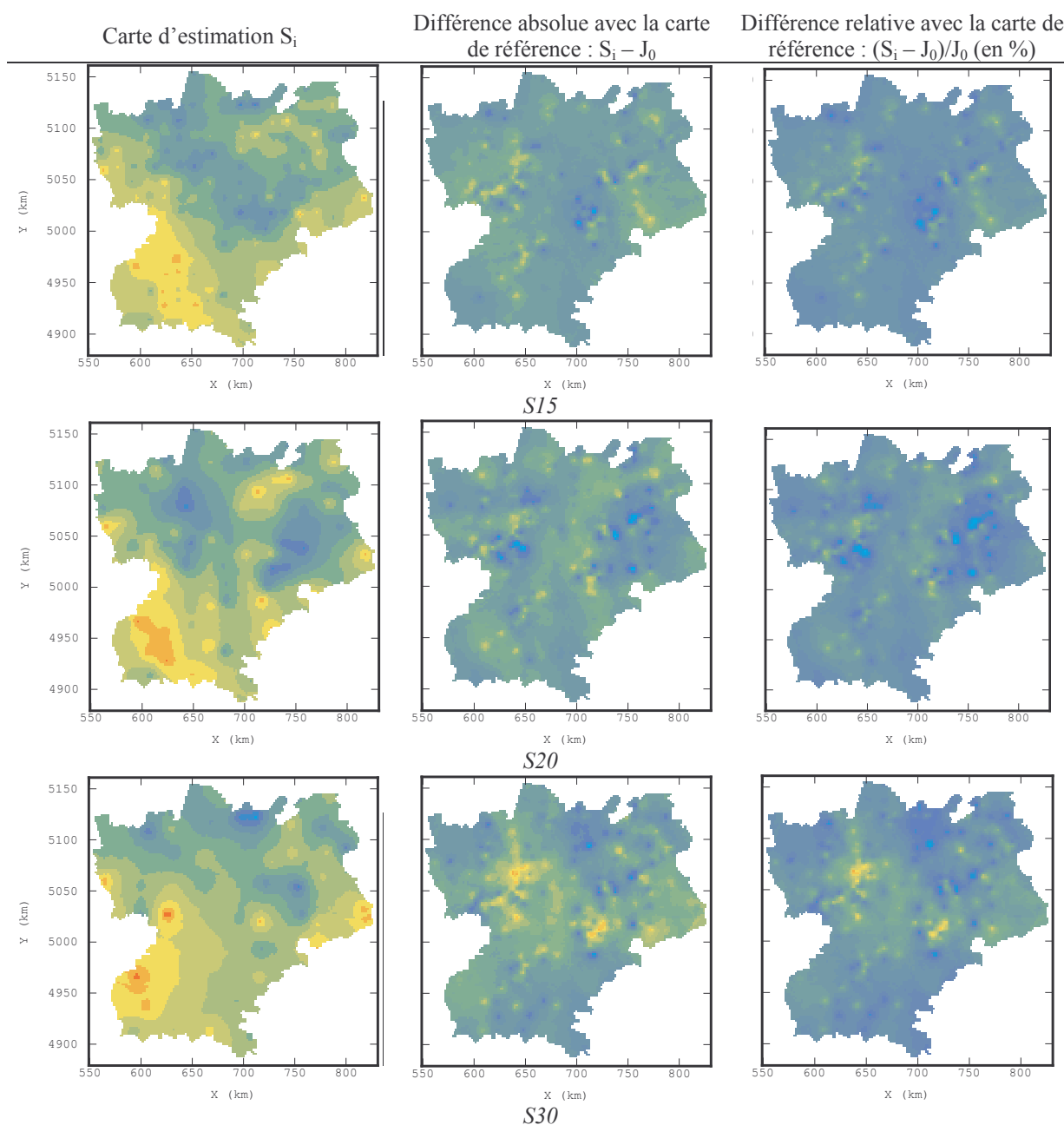


Figure 50 – Comparaison entre les estimations obtenues pour chaque sélection avec le modèle de référence (ordonnée) et avec le modèle de la sélection (abscisse)

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.



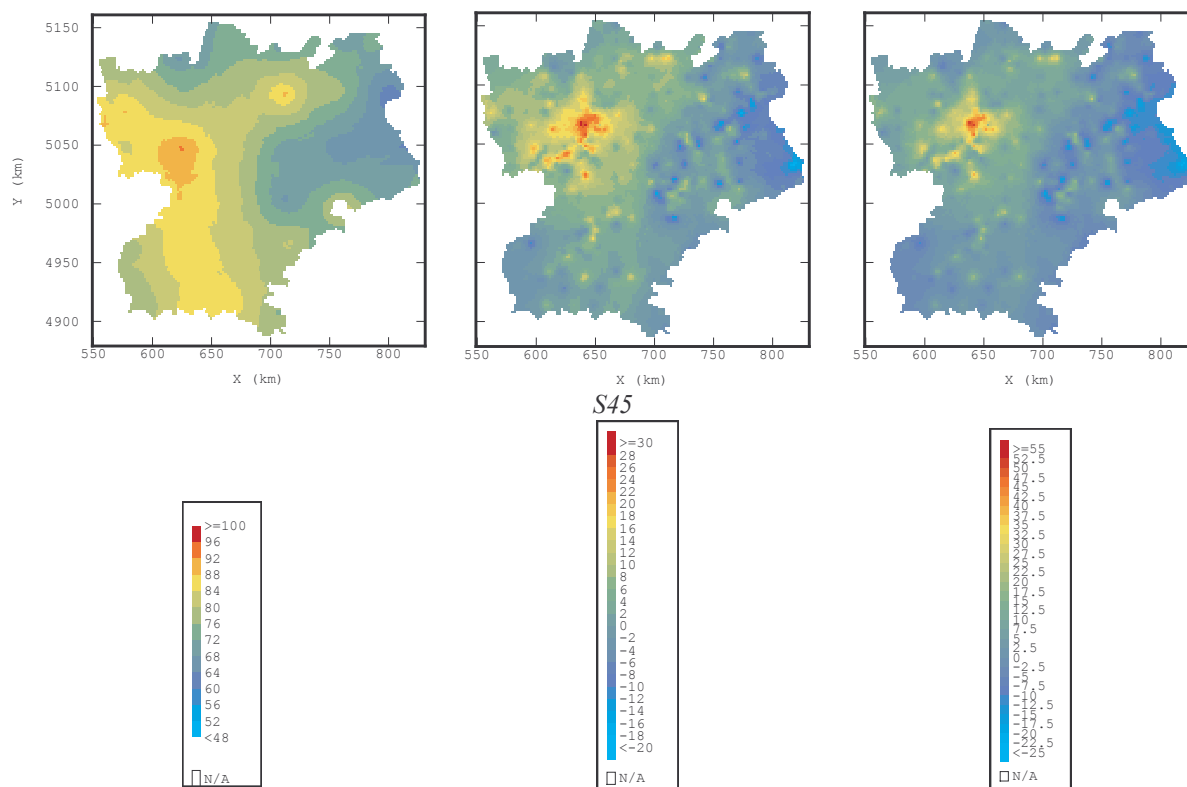


Figure 51 – Comparaison entre les cartes d'estimation obtenues avec le modèle de référence et avec les différentes sélections (de bas en haut : S15, S20, S30, S45)

En valeur absolue, l'écart avec la carte de référence peut atteindre $30 \mu\text{g m}^{-3}$ pour les sélections S30 et S45. En valeur relative, il atteint presque 55% pour les zones de surestimation des concentrations d'ozone (centre-ville de Lyon), et 25% pour les zones de sous-estimation (sites d'altitude). On observe donc un net effet de lissage : sous-estimation en zones éloignées de toute pollution par les oxydes d'azote, surestimation dans les zones influencées par la pollution de proximité (centres-villes, axes routiers).

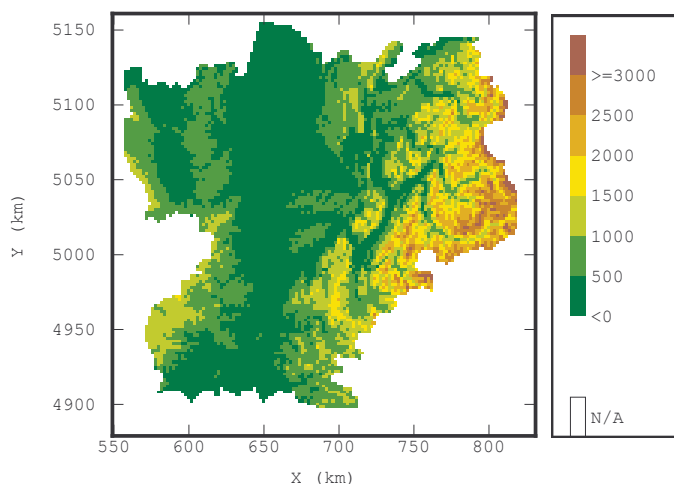
PARTIE II. Echantillonnage régulier avec variable auxiliaire

En vue d'améliorer la qualité de l'estimation, on va conserver les mailles régulières d'échantillonnage pour déterminer les jeux de données sélectionnés et leurs complémentaires, et on va intégrer une 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 connaissance pour l'estimation.

1 DESCRIPTION DE LA VARIABLE AUXILIAIRE : ALTITUDE

Pour cette étude, nous disposons d'une seule variable auxiliaire, l'altitude, connue sur des mailles de 1 km de côté. On conserve dans cette partie la grille de référence de 2 km de côté utilisée dans la partie I.

L'altitude joue un rôle important dans la modélisation de la distribution spatiale des données : on a notamment vu dans la partie précédente que les différences de concentration entre les vallées et les sommets étaient mal représentées. La figure 38 rappelle le relief de la zone, caractérisé par le couloir rhodanien s'étalant du Nord au Sud, et qui sépare les premiers contreforts du Massif Central à l'ouest des Alpes à l'Est.



Les concentrations en ozone augmentent globalement avec l'altitude dans les premiers kilomètres de la troposphère pour des sites situés loin de toutes sources de pollution [4]. On peut cependant s'attendre à ce que le facteur altitude ne soit pas pertinent sur toute la zone : en effet, si l'ozone va présenter en général des concentrations plus élevées en altitude (moins de puits d'ozone, transport de masses d'air, ...), des paramètres anthropiques vont également influencer sur les mesures des sites de proximité, surtout pour les basses altitudes où la densité de population est plus dense et les sources de pollution anthropique plus nombreuses.

■ CALCUL DU COFACTEUR

On cherche à obtenir par des transformations mathématiques appliquées à la variable auxiliaire (de manière empirique), une nouvelle variable présentant la meilleure corrélation linéaire possible avec la concentration en polluant. Selon les résultats obtenus, on choisira d'appliquer la méthode de cokrigage colocalisé (méthode stationnaire) ou la méthode de krigage avec dérive externe (méthode non stationnaire).

La figure 39 présente le diagramme de corrélation entre l'altitude et la concentration en ozone pour le jeu de données J_0 .

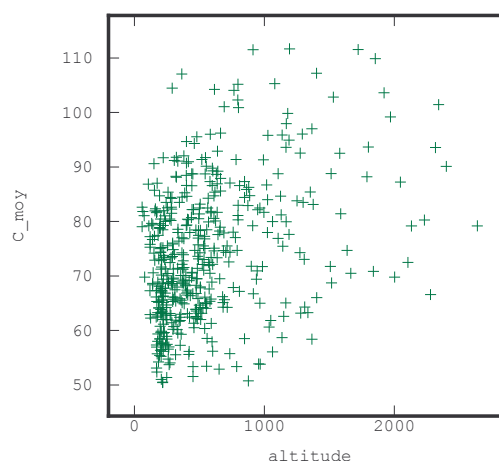


Figure 53 – Diagramme de corrélation altitude / concentration en ozone pour J_0 , de corrélation $r = 0,363$

Nous voyons que lorsque l'altitude est élevée, la concentration en ozone a globalement tendance à être plus importante. Mais une altitude faible ou moyenne n'implique pas nécessairement que la concentration en ozone soit faible. L'interprétation de la relation entre les deux variables est donc plus complexe qu'une simple relation linéaire ; le profil est plutôt logarithmique. On calcule donc une variable auxiliaire en tous points de mesure du jeu de données J_0 , égale au logarithme décimal de la variable {altitude + 1}. La constante « +1 » permet d'éviter une erreur mathématique en cas d'altitude nulle.

On obtient le diagramme de corrélation suivant.

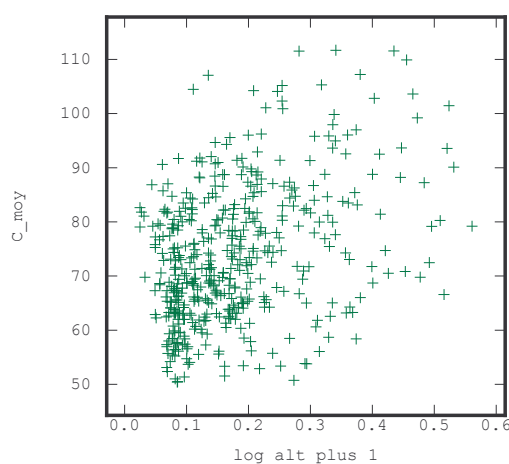


Figure 54 – Diagramme de corrélation entre la valeur auxiliaire $\log(\text{alt}+1)$ et la concentration en ozone pour J_0 , de corrélation $r = 0,373$

Pour tester les deux méthodes, nous allons utiliser le jeu de données S15, afin de disposer du complémentaire C15 pour la validation absolue.

▪ ETUDE DU KRIGEAGE AVEC DERIVE EXTERNE

Dans ce premier cas, nous allons travailler avec une dérive externe. Cela consiste à considérer que la variable auxiliaire (ici l'altitude) représente la tendance générale de la concentration en NO_2 , dont la moyenne n'est pas constante (méthode non stationnaire).

Le nuage de corrélation présenté figure 40 respecte bien la tendance globale des variations de la concentration en ozone. On fait donc l'hypothèse que même si globalement la corrélation n'est pas strictement linéaire (mais qu'il y a simplement existence d'une tendance), les concentrations d'ozone sont localement liées à la variable auxiliaire d'altitude par une fonction linéaire.

A titre de vérification, nous séparons les valeurs de la variable auxiliaire en 12 classes égales, et pour chacune de ces classes, nous calculons la concentration moyenne en ozone. On vérifie sur la figure 41 que cette tendance est effectivement globalement linéaire.

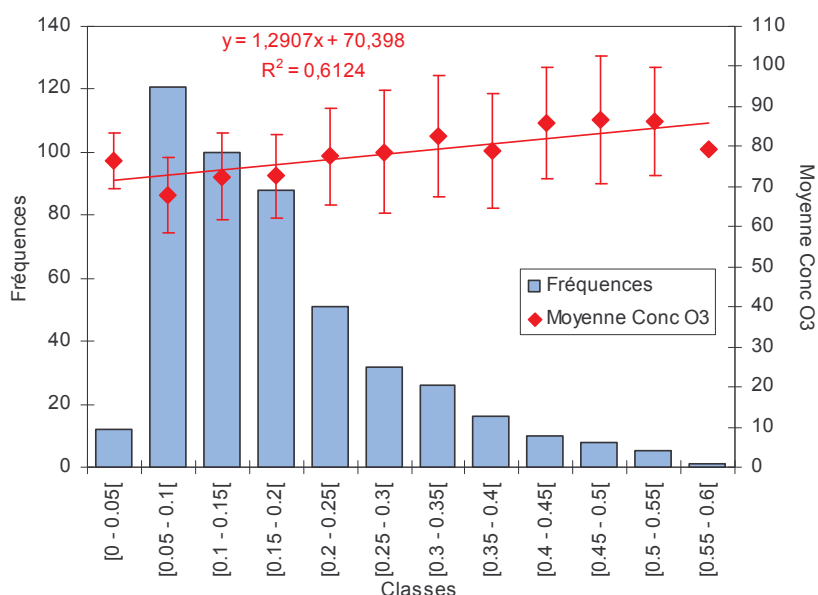


Figure 55 – Répartition par classes de la variable auxiliaire et moyenne des concentrations d'ozone pour chacune des classes

L'ajustement d'un modèle non-stationnaire se fait en deux étapes : calcul de la dérive et ajustement du modèle de covariance généralisée correspondant.

• Calcul de la dérive externe

Tout d'abord, on va considérer un voisinage unique sur tout le domaine de manière à établir la structure variographique des résidus. En effet, la corrélation entre la variable auxiliaire d'altitude $\{\log(\text{alt}+1)\}$ et la concentration d'ozone nous permet de calculer les résidus correspondant à chaque point du jeu de données :

$$R_i = (C_moy)_i - (a \times \{\log(\text{alt}+1)\}_i + b)$$

où R_i , $(C_moy)_i$ et $\{\log(\text{alt}+1)\}_i$ sont les valeurs du résidu, de la concentration d'ozone et de la variable auxiliaire d'altitude au point i , respectivement ;

a et b sont les coefficients de la régression linéaire entre la variable auxiliaire d'altitude et la concentration d'ozone.

On calcule la valeur d'altitude en tout point du jeu de données S15 (172 points déterminés), puis on établit le variogramme des résidus sur le jeu de données S15 et on modélise ce variogramme.

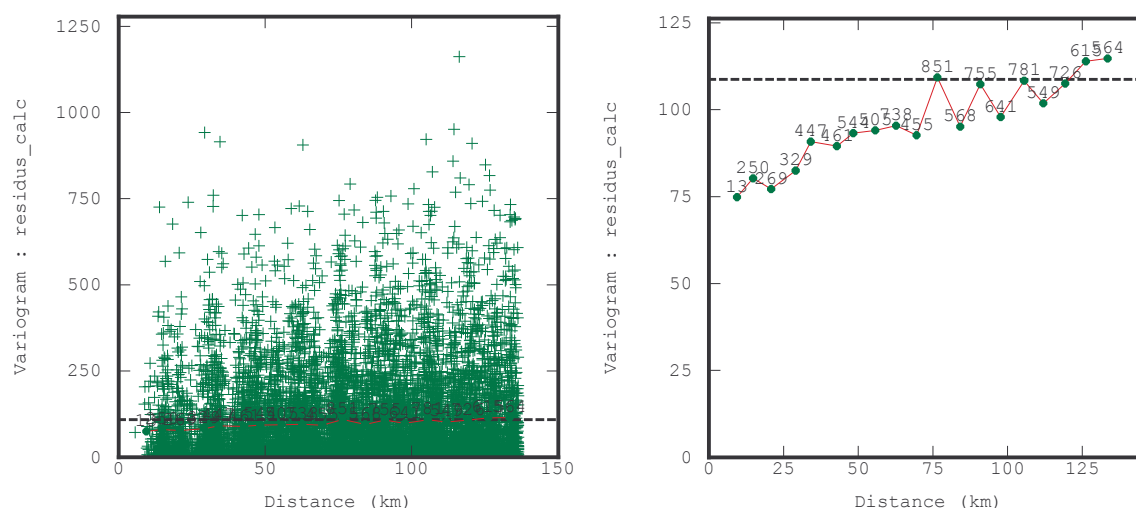


Figure 56 – Variogramme des résidus calculés sur le jeu de données S15 (172 sites ; 20 pas ; 7 km)

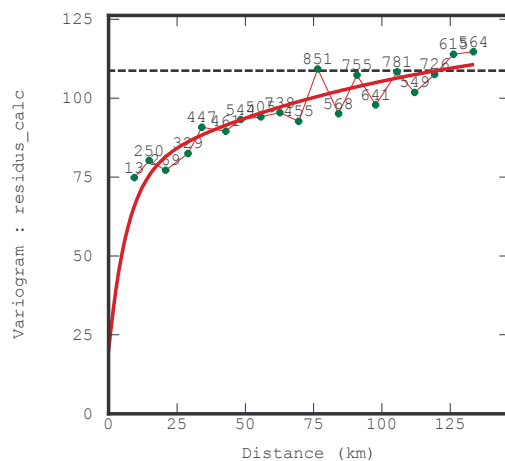


Figure 57 – Variogramme simulé : pépité (20) + exp (20 km ; 55) + exp (320 km ; 50)

On teste le modèle de variogramme des résidus par validation croisée en considérant un voisinage unique.

Tableau 16 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les résidus calculés

R_S15 (172 points)	Moyenne	Variance
Erreur	0,01921	85,09738
Erreur standardisée	0,00105	1,02483

1 donnée sur 172 non robuste.

R_S15 (171 points)	Moyenne	Variance
Erreur	0,16208	82,08454
Erreur standardisée	0,01654	0,98952

Toutes les données sont robustes.

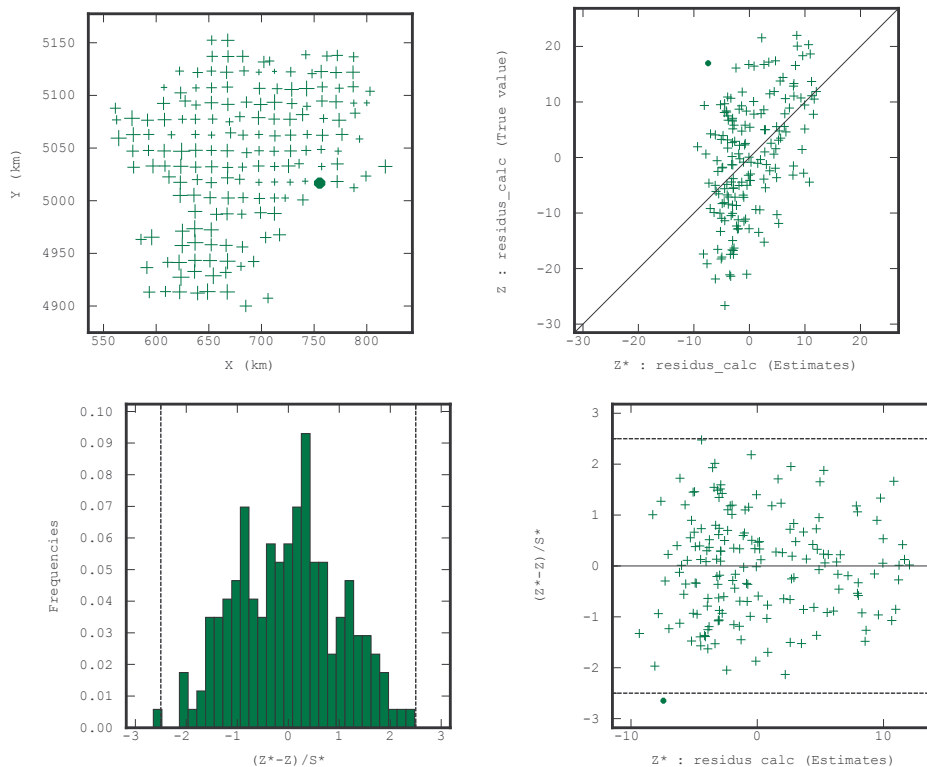


Figure 58 – Résultats de la validation croisée

On va maintenant chercher la structure du modèle de dérive externe pour un voisinage glissant de 70 km, en se basant sur le modèle de variogramme des résidus obtenu avec le voisinage unique.

A l'aide du logiciel, on détermine que la dérive externe optimale est une fonction linéaire de la variable auxiliaire d'altitude :

$$\text{dérive} = u \times \{\log(\text{alt}+1)\} + v$$

où u et v varient selon le point considéré du domaine.

▪ Ajustement du modèle de covariance généralisé

L'ajustement se fait automatiquement dans le logiciel à partir des structures élémentaires choisies d'après l'étude pour le voisinage unique, à savoir un effet pépité et deux structures exponentielles de courte (20 km) et longue (320 km) portée que l'on fera varier. On ne peut choisir que des structures isotropes dont seul le palier ou la pente est ajusté lors de l'optimisation.

Le modèle de covariance généralisée le mieux adapté est une structure exponentielle de portée 320 km sans effet pépitique.

On teste ce modèle de covariance généralisée par validation croisée sur les concentrations d'ozone en utilisant le modèle de dérive externe linéaire et pour un voisinage de 70 km. Les résultats sont présentés dans le tableau et la figure ci-après.

Tableau 17 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les concentrations

S15 (172 points)	Moyenne	Variance
Erreur	0,06369	89,01799
Erreur standardisée	0,00463	2,30866

18 données sur 172 non robustes.

S15 (154 points)	Moyenne	Variance
Erreur	-0,38404	56,00244
Erreur standardisée	-0,05695	1,43922

Toutes les données sont robustes.

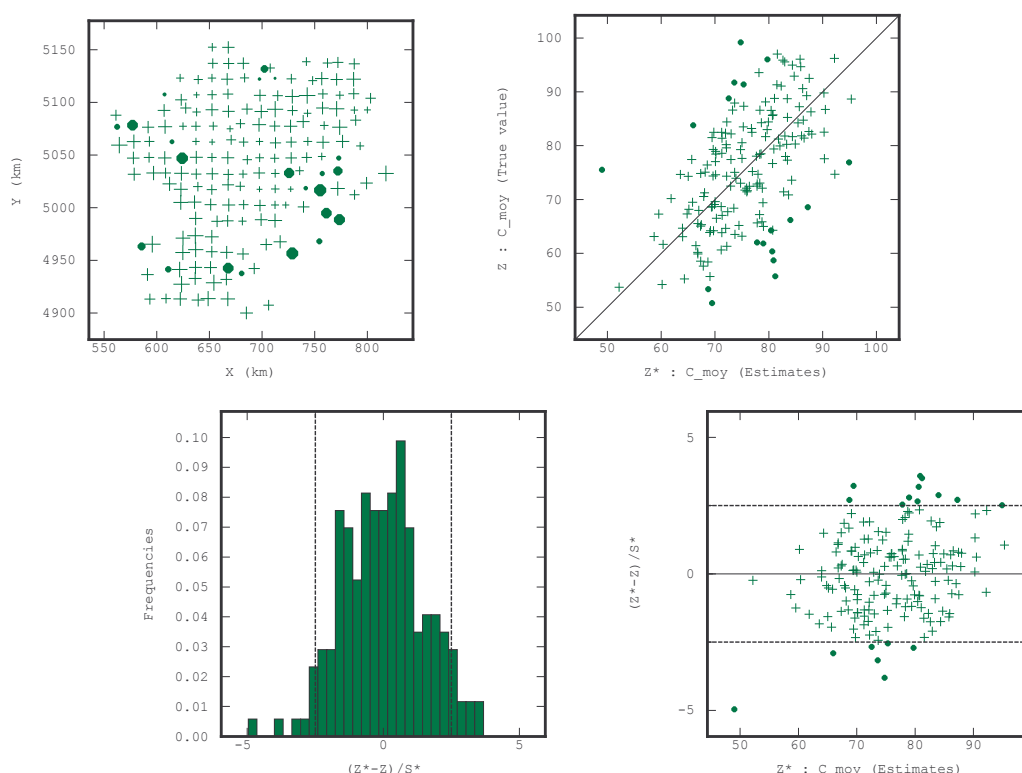


Figure 59 – Résultats de la validation croisée

La validation croisée apparaît donc comme peu satisfaisante pour valider le modèle de dérive externe défini par le logiciel. Néanmoins cela reste cohérent avec la faible corrélation établie entre la concentration d'ozone et la variable auxiliaire d'altitude.

- **Test de validation absolue sur le complémentaire C15**

Tableau 18 – Résultats de la validation absolue sur le complémentaire C15 pour le krigeage ponctuel avec dérive externe

		Validation croisée par krigeage ponctuel avec dérive externe selon le modèle de la sélection S15	
C15	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	277	276	276
Minimum	50,43	51,61	2,13
Percentile 25	63,05	65,06	4,44
Médiane	69,15	72,44	4,71
Percentile 75	78,74	78,42	4,95
Maximum	99,83	98,17	9,27
Etendue	49,40	46,56	7,14
Moyenne	70,6	71,99	4,70
Ecart-type	10,44	9,47	0,56
Variance	109,1	89,76	0,32
CV%	14,8	13,2	12,0
Corrélation mesures/estimations		$r = 0,419$	

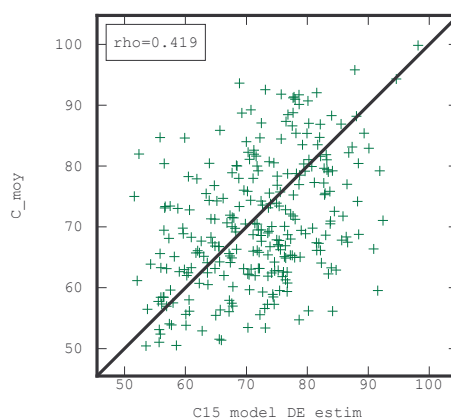


Figure 60 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C15

- **ETUDE DU KRIGEAGE COLOCALISE SUR LA SELECTION S15**

En utilisant le krigeage colocalisé, on fait l'hypothèse que les deux variables (concentration d'ozone et variable auxiliaire d'altitude) sont liées par un modèle variographique. Ce type de krigeage est plus flexible puisqu'il permet de définir une corrélation variant avec la distance. Il est en revanche très sensible au type de structure définissant la relation entre les deux variables.

On va établir le modèle bivariable à partir du jeu de données S15 et des valeurs de variable auxiliaire définies aux mêmes coordonnées, de manière à pouvoir tester le modèle sur le complémentaire C15.

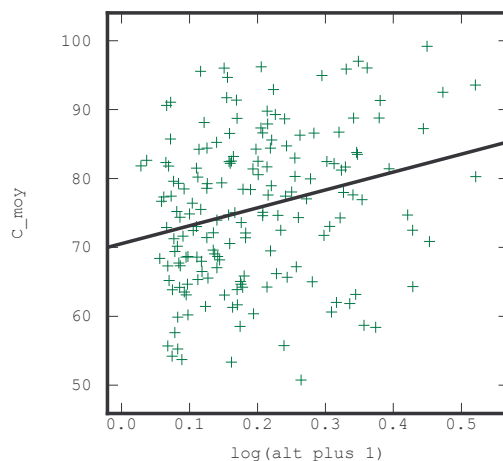


Figure 61 – Diagramme de corrélation entre la valeur auxiliaire $\log(\text{alt}+1)$ et la concentration en ozone pour S15, de corrélation $r = 0,260$

▪ Etude des variogrammes expérimentaux

On calcule pour toutes les directions les deux variogrammes simples et le variogramme croisé avec 20 pas de 7 km.

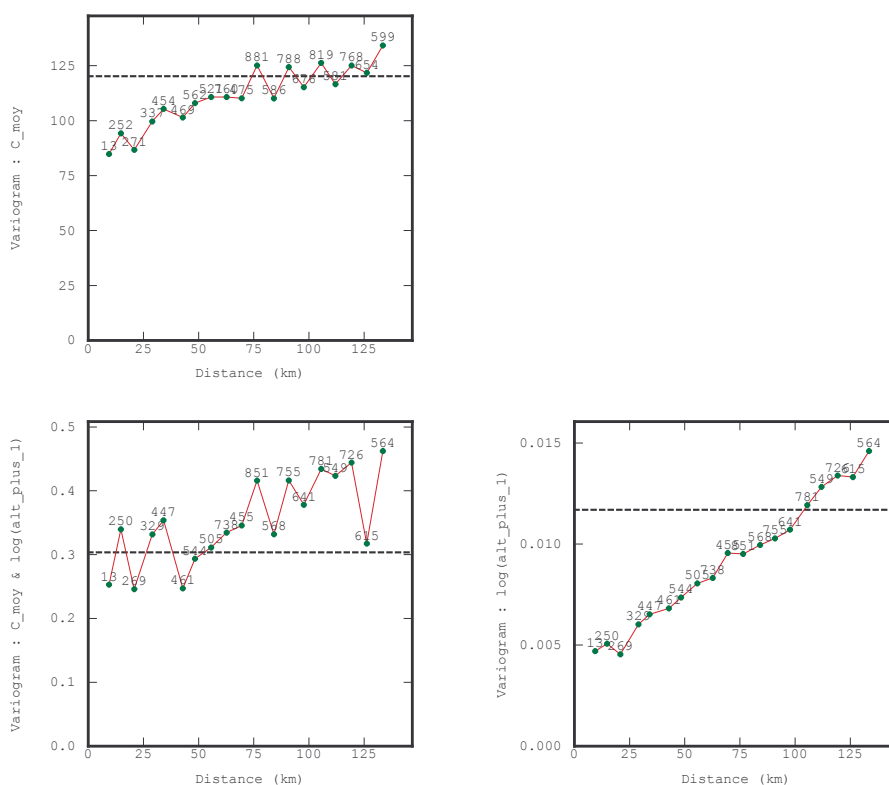


Figure 62 – Variogrammes expérimentaux simples et croisé (176 sites ; 20 pas ; 7 km)

▪ Modélisation des variogrammes et validation croisée

A partir de plusieurs structures simples, le logiciel définit les structures permettant de simuler les différents variogrammes. Le modèle est finalement constitué d'un effet pépite, d'une structure exponentielle de portée 35 km et d'une structure linéaire de portée 35 km.

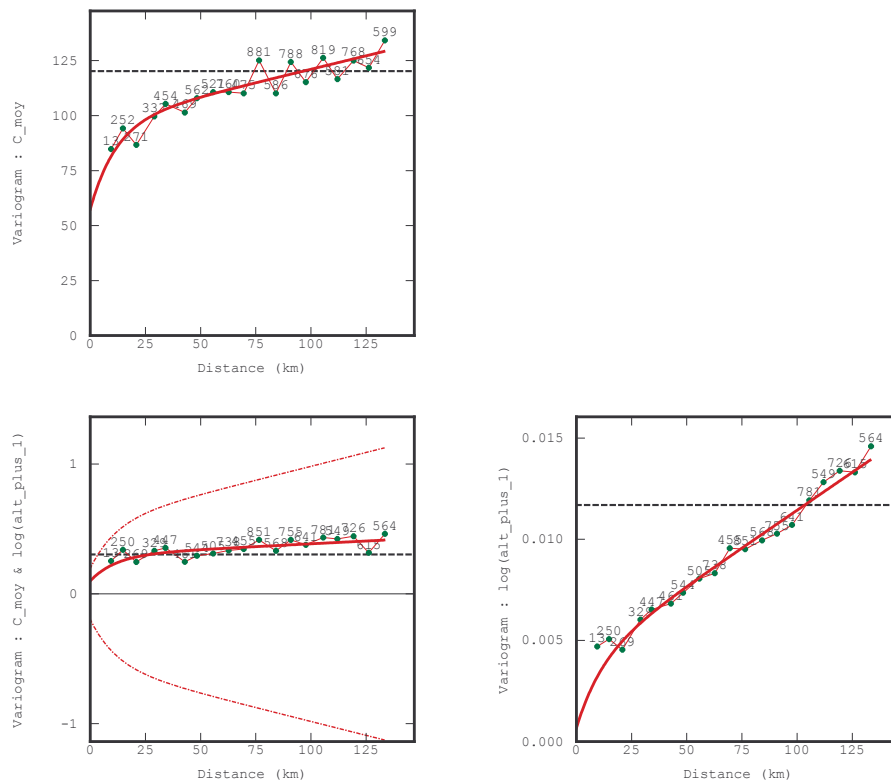


Figure 63 – Variogrammes simulés simples et croisé

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 70 km.

Tableau 19 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les concentrations

S15 (176 points)	Moyenne	Variance
Erreur	-0,07708	82,13215
Erreur standardisée	-0,00888	1,00489

1 donnée sur 176 non robuste.

S15 (175 points)	Moyenne	Variance
Erreur	-0,21862	79,07532
Erreur standardisée	-0,02492	0,96534

Toutes les données sont robustes.

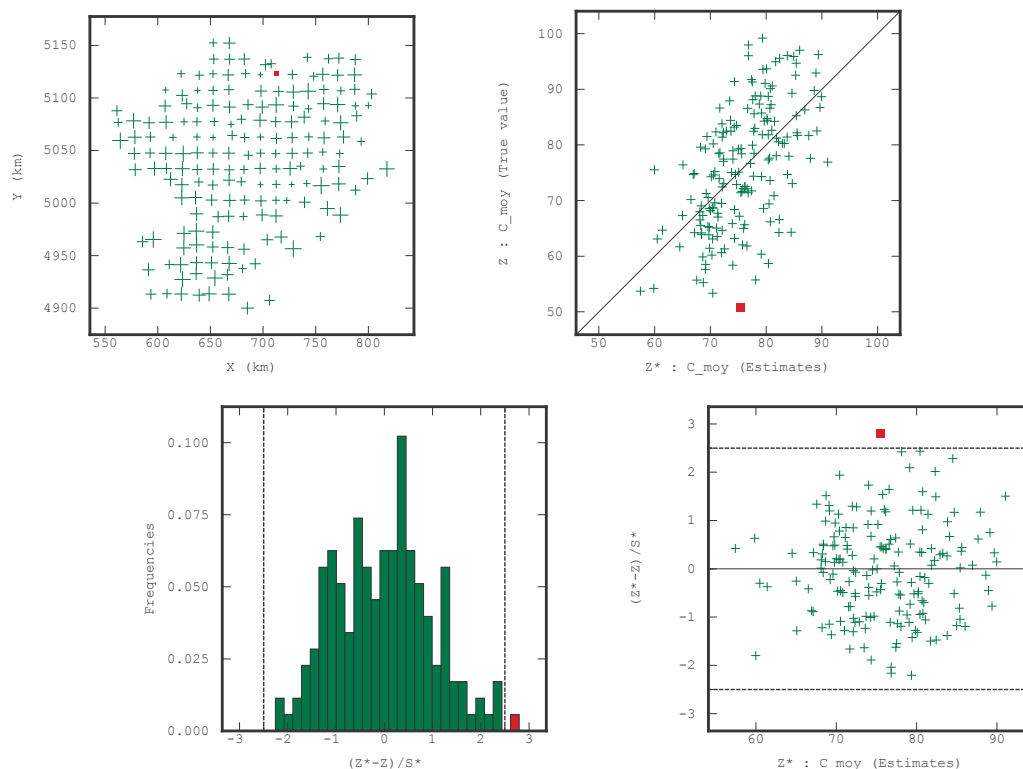


Figure 64 – Résultats de la validation croisée

La corrélation entre données mesurées et données estimées reste relativement faible ($r = 0,564$) mais s'améliore par rapport à la validation croisée pour le même jeu de données sans variable auxiliaire ($r = 0,402$).

▪ Test de validation absolue sur le complémentaire C15

Tableau 20 – Résultats de la validation absolue sur le complémentaire C15 pour le krigeage ponctuel colocalisé

		Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S15	
C15	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	277	276	276
Minimum	50,43	54,97	8,03
Percentile 25	63,05	67,21	8,48
Médiane	69,15	71,51	8,55
Percentile 75	78,74	77,84	8,59
Maximum	99,83	93,97	9,35
Etendue	49,40	39,00	1,32
Moyenne	70,6	72,21	8,54
Ecart-type	10,44	6,78	0,13
Variance	109,1	46,02	0,02
CV%	14,8	9,4	1,5
Corrélation mesures/estimations		$r = 0,477$	

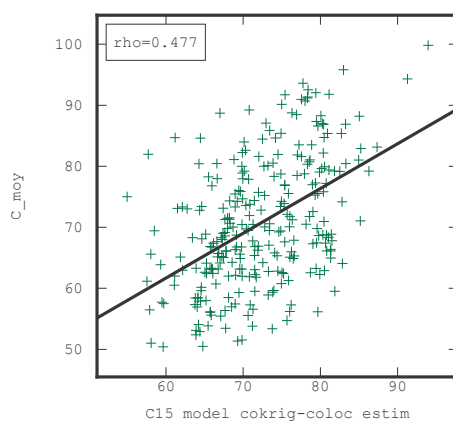


Figure 65 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C15

La corrélation entre valeurs estimées et mesurées s’est améliorée par rapport au krigeage ordinaire ponctuel sans variable auxiliaire, pour lequel le coefficient de corrélation s’établissait à 0,318.

Compte tenu des résultats pour le krigeage par dérive externe et le cokrigeage colocalisé obtenus sur le jeu de données S15 et testés sur le complémentaire C15, on choisit d’utiliser le cokrigeage colocalisé sur le jeu de données de référence J₁ (excluant les points atypiques) pour la définition de la carte de référence.

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

La corrélation entre la variable auxiliaire d’altitude et la concentration en ozone est présentée figure suivante pour le jeu de données J₁.

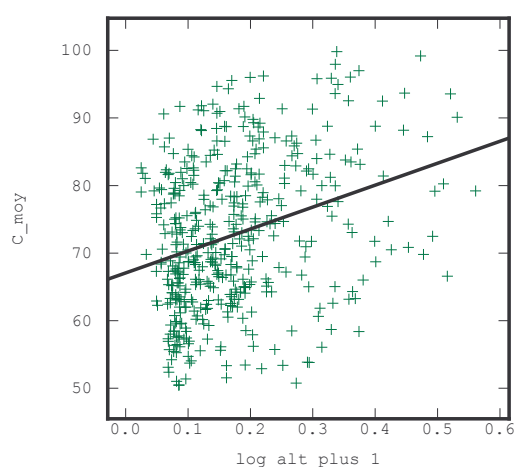


Figure 66 – Diagramme de corrélation entre la valeur auxiliaire $\log(\text{alt}+1)$ et la concentration en ozone J₁, de corrélation $r = 0,304$

▪ Etude des variogrammes expérimentaux

On calcule pour toutes les directions les deux variogrammes simples et le variogramme croisé avec 16 pas de 8 km.

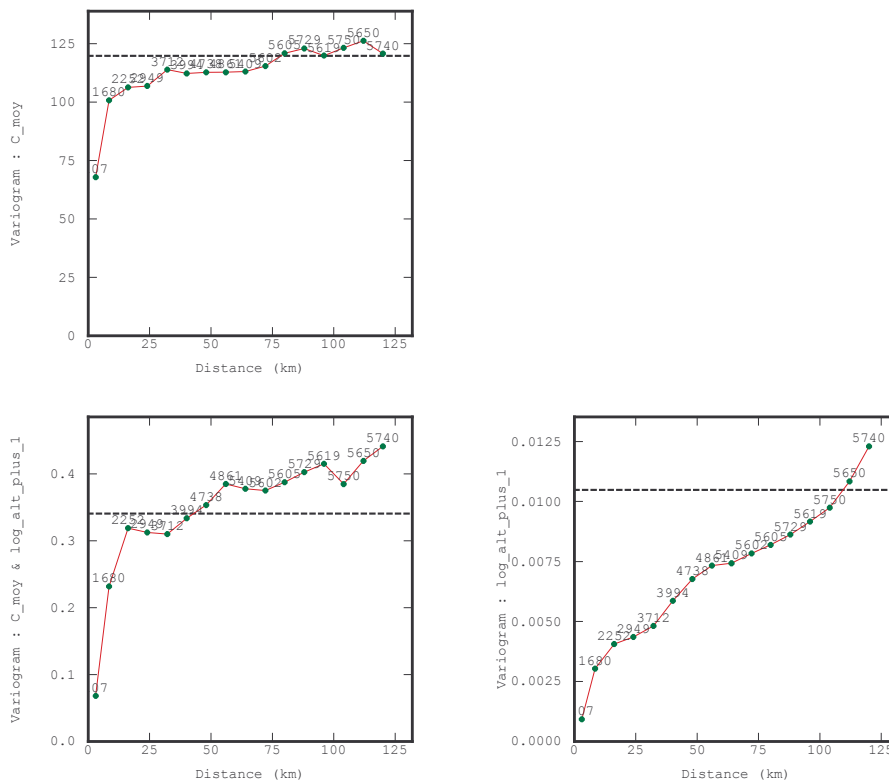


Figure 67 – Variogrammes expérimentaux simples et croisé (453 sites ; 20 pas ; 6 km)

▪ Modélisation des variogrammes et validation croisée

A partir de plusieurs structures simples, le logiciel définit les structures permettant de simuler les différents variogrammes. Le modèle est finalement constitué d'un effet pépite, d'une structure exponentielle de portée 15 km et d'une structure linéaire de portée 35 km.

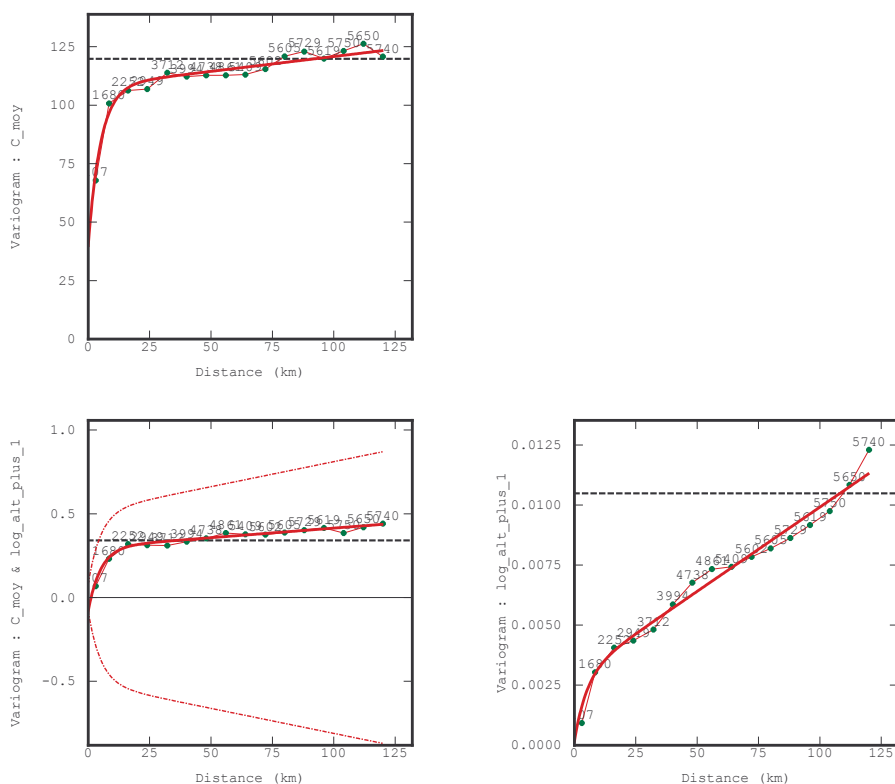


Figure 68 – Variogrammes simulés simples et croisé

Number of variables = 2
 - Variable 1 : C_moy
 - Variable 2 : log_alt_plus_1
 Number of basic structures = 3

S1 : Exponential - Scale = 15.00km

Variance-Covariance matrix :	Decomposition into factors	Decomposition into eigen vectors
Variable 1 Variable 2	(normalized eigen vectors) :	(whose variance is eigen values) :
Variable 1 68.1082 0.3800	Variable 1 Variable 2	Variable 1 Variable 2 Eigen Val. Var. Perc.
Variable 2 0.3800 0.0027	Factor 1 8.2528 0.0460	Factor 1 1.0000 0.0056 68.1103 100.00
	Factor 2 -0.0001 0.0247	Factor 2 -0.0056 1.0000 0.0006 0.00

S2 : Nugget effect

Variance-Covariance matrix :	Decomposition into factors	Decomposition into eigen vectors
Variable 1 Variable 2	(normalized eigen vectors) :	(whose variance is eigen values) :
Variable 1 40.0004 -0.0800	Variable 1 Variable 2	Variable 1 Variable 2 Eigen Val. Var. Perc.
Variable 2 -0.0800 0.0002	Factor 1 6.3246 -0.0126	Factor 1 1.0000 -0.0020 40.0005 100.00
	Factor 2 0.0000 0.0022	Factor 2 0.0020 1.0000 0.0000 0.00

S3 : Order-1 G.C. - Scale = 35.00km

Variance-Covariance matrix :	Decomposition into factors	Decomposition into eigen vectors
Variable 1 Variable 2	(normalized eigen vectors) :	(whose variance is eigen values) :
Variable 1 4.4456 0.0400	Variable 1 Variable 2	Variable 1 Variable 2 Eigen Val. Var. Perc.
Variable 2 0.0400 0.0025	Factor 1 2.1085 0.0190	Factor 1 1.0000 0.0090 4.4460 99.95
	Factor 2 -0.0004 0.0458	Factor 2 -0.0090 1.0000 0.0021 0.05

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 70 km.

Tableau 21 – Statistiques sur l'erreur d'estimation, obtenues par validation croisée sur les concentrations

J_1 (453 points)	Moyenne	Variance
Erreur	-0,24618	81,46404
Erreur standardisée	-0,02360	1,00180

6 données sur 453 non robustes.

J_1 (447 points)	Moyenne	Variance
Erreur	-0,15068	74,81654
Erreur standardisée	-0,01258	0,91729

Toutes les données sont robustes.

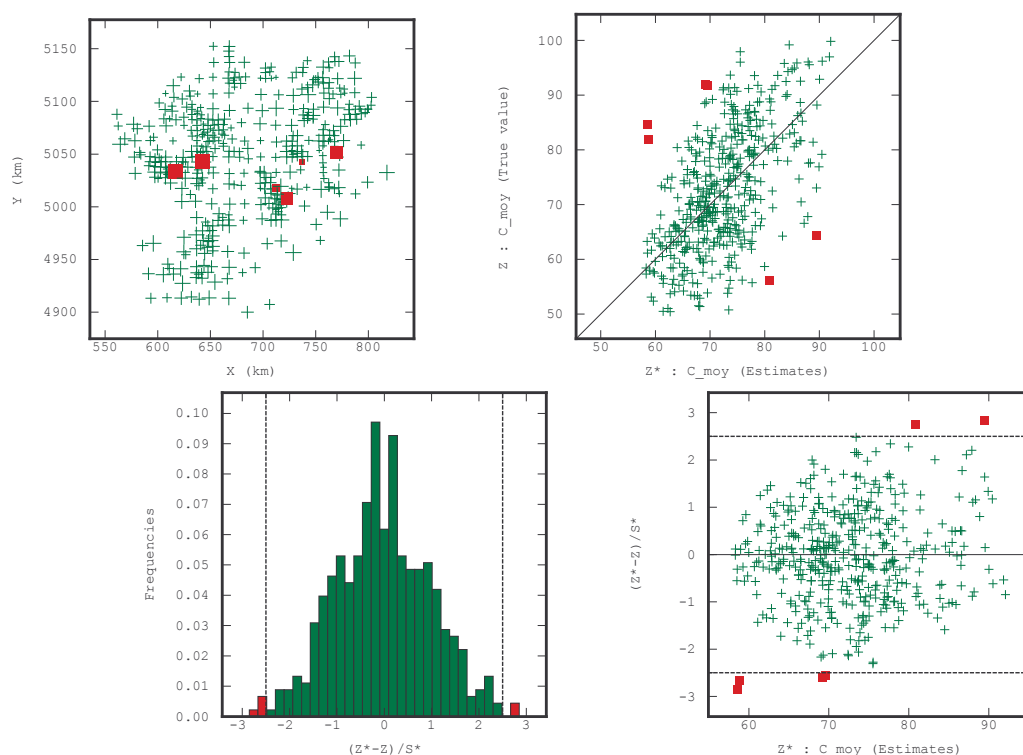


Figure 69 – Résultats de la validation croisée

La corrélation entre données mesurées et données estimées reste relativement faible ($r = 0,571$) mais s'améliore par rapport à la validation croisée pour le même jeu de données sans variable auxiliaire ($r = 0,468$).

▪ Etablissement de la carte de référence

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 $\{\log(\text{alt}+1)\}$ et du modèle bivariable, en considérant un voisinage de 70 km.

Tableau 22 – Statistiques sur l'estimation sur une grille régulière de 2 km de côté, par krigeage ponctuel colocalisé

J_0	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	470	11240	11240
Minimum	50,43	56,76	3,47
Percentile 25	64,88	71,87	8,88
Médiane	72,48	78,15	9,12
Percentile 75	81,82	84,51	9,26
Maximum	111,69	100,96	10,17
Etendue	61,26	44,20	6,70
Moyenne	73,9	78,40	9,05
Ecart-type	12,41	8,24	0,39
Variance	153,9	67,96	0,16
CV%	16,8	10,5	4,4

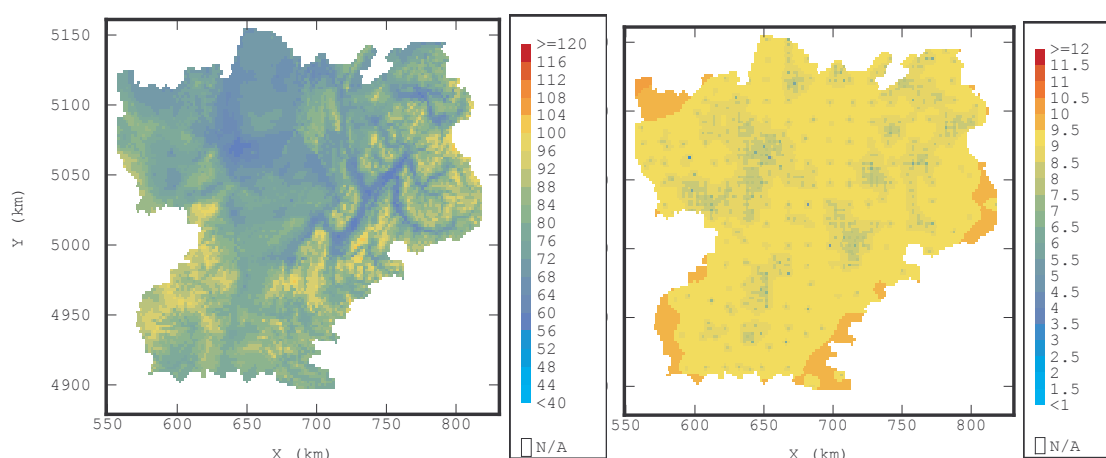


Figure 70 - (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données J_0 , et (droite) écarts-types de krigeage

Dans la partie précédente, nous avons pu observer que les estimations n'étaient pas acceptables pour les sélections S30 et S45 (écarts relatifs supérieurs à 30% par rapport à la carte de référence). Dans cette deuxième partie, nous allons donc déterminer si la présence d'une variable auxiliaire telle que l'altitude permet de réduire cet écart entre les estimations pour les sélections à partir de S20.

- **ETUDE DE LA SELECTION DE DONNEES S20 (MAILLE DE 20 × 20 KM)**
- **Etude des variogrammes expérimentaux**

On calcule pour toutes les directions les deux variogrammes simples et le variogramme croisé avec 15 pas de 9 km.

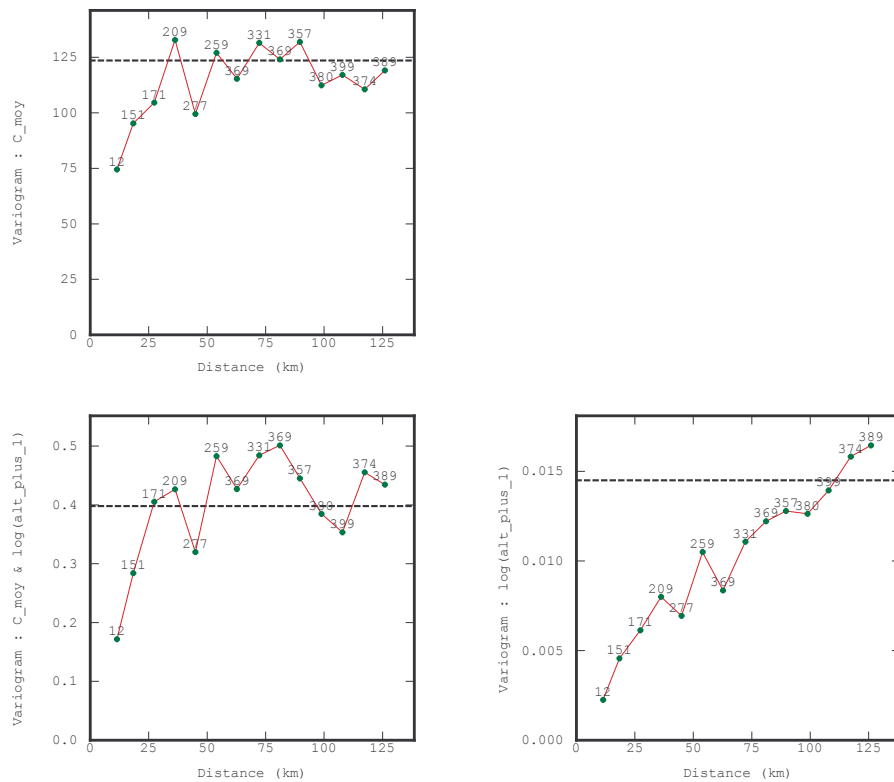
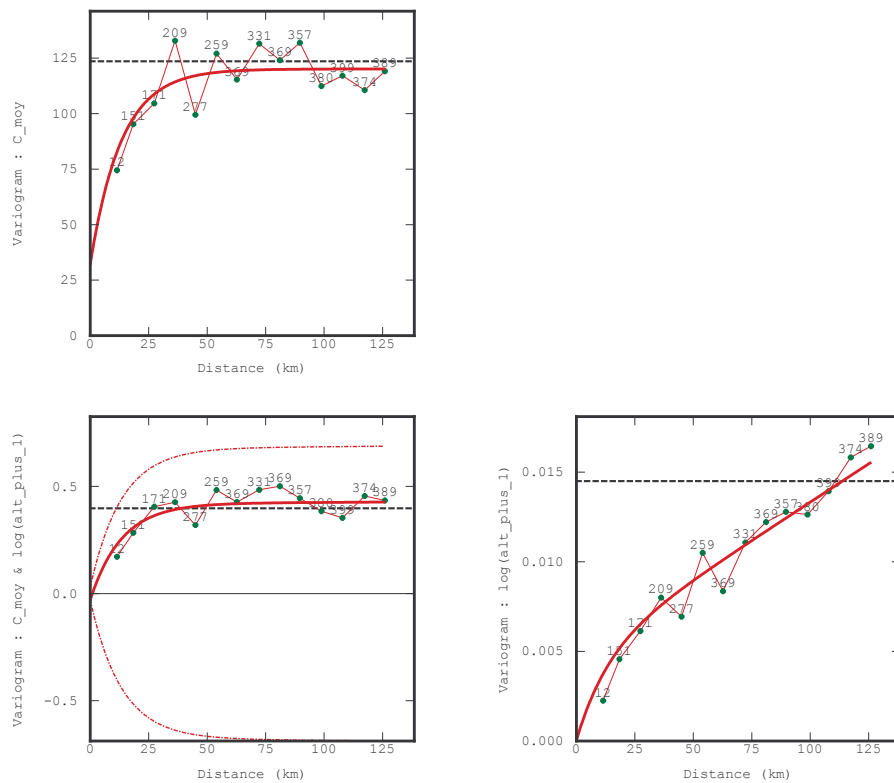


Figure 71 – Variogrammes expérimentaux simples et croisé (115 sites ; 15 pas ; 9 km)

▪ Modélisation des variogrammes et validation croisée



**Figure 72 – Variogrammes simulés simples et croisé :
pépité (-0,03) + exp (40 km ; 0,45) + lin (35 km ; 0,002)**

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 70 km.

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

S20 (115 points)	Moyenne	Variance
Erreur	-0,11401	88,83417
Erreur standardisée	-0,01471	0,99474

Toutes les données sont robustes.

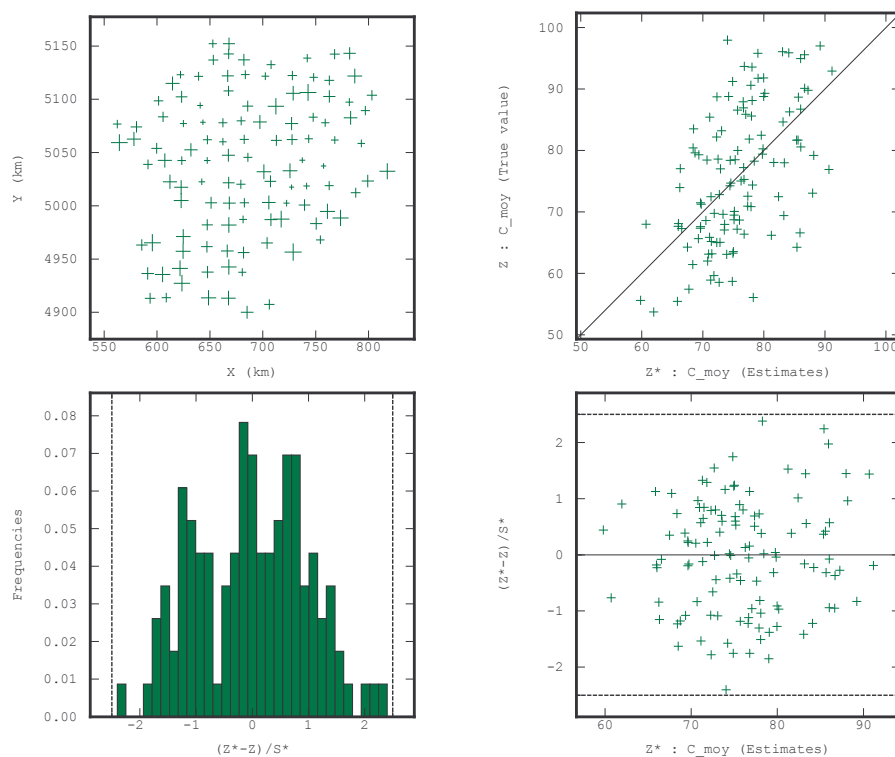


Figure 73 – Résultats de la validation croisée

La corrélation obtenue entre concentrations mesurées et estimées présente un coefficient de 0,533.

- **Test de validation absolue sur le complémentaire C20**

Tableau 24 – Résultats de la validation absolue sur le complémentaire C20 pour le krigeage ponctuel colocalisé

		Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S20	
C20	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	338	338	338
Minimum	50,43	56,05	7,67
Percentile 25	63,28	67,01	8,57
Médiane	70,53	71,02	8,73
Percentile 75	79,95	76,47	8,93
Maximum	99,83	88,37	9,75
Etendue	49,40	32,32	2,08
Moyenne	71,5	71,69	8,75
Ecart-type	10,66	6,47	0,28
Variance	113,7	41,80	0,08
CV%	14,9	9,0	3,2
Corrélation mesures/estimations		$r = 0,480$	

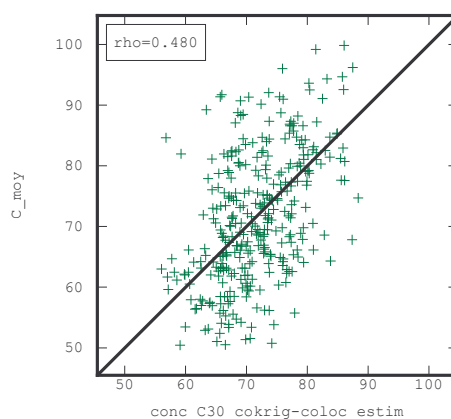


Figure 74 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C20

- **Etablissement de la carte pour la sélection S20**

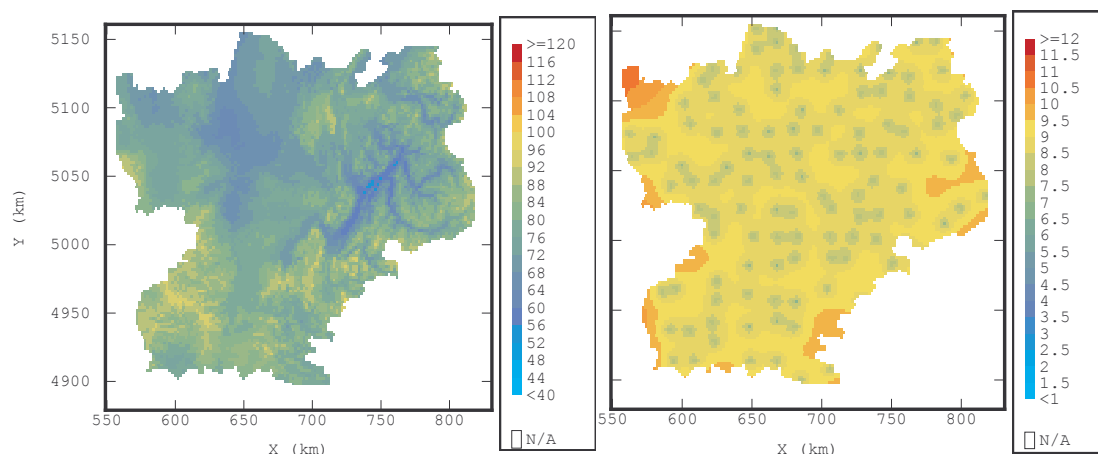


Figure 75 – (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données S20, et (droite) écarts-types de krigeage

- **ETUDE DE LA SELECTION DE DONNEES S30 (MAILLE DE 30 × 30 KM)**

- **Etude des variogrammes expérimentaux**

On calcule pour toutes les directions les deux variogrammes simples et le variogramme croisé avec 15 pas de 10 km.

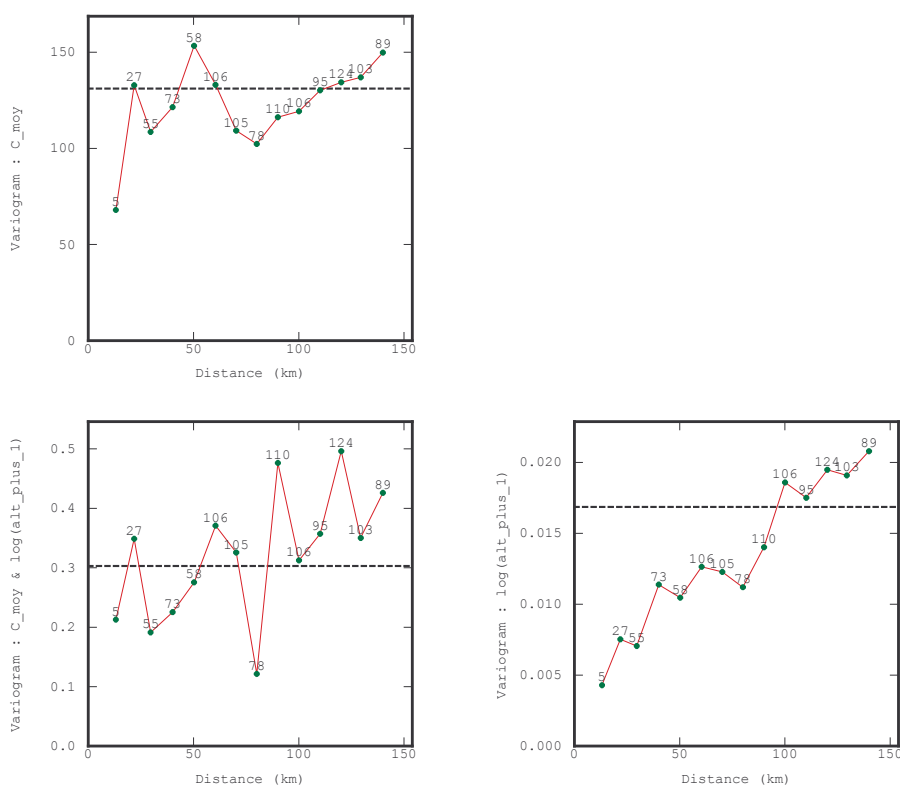
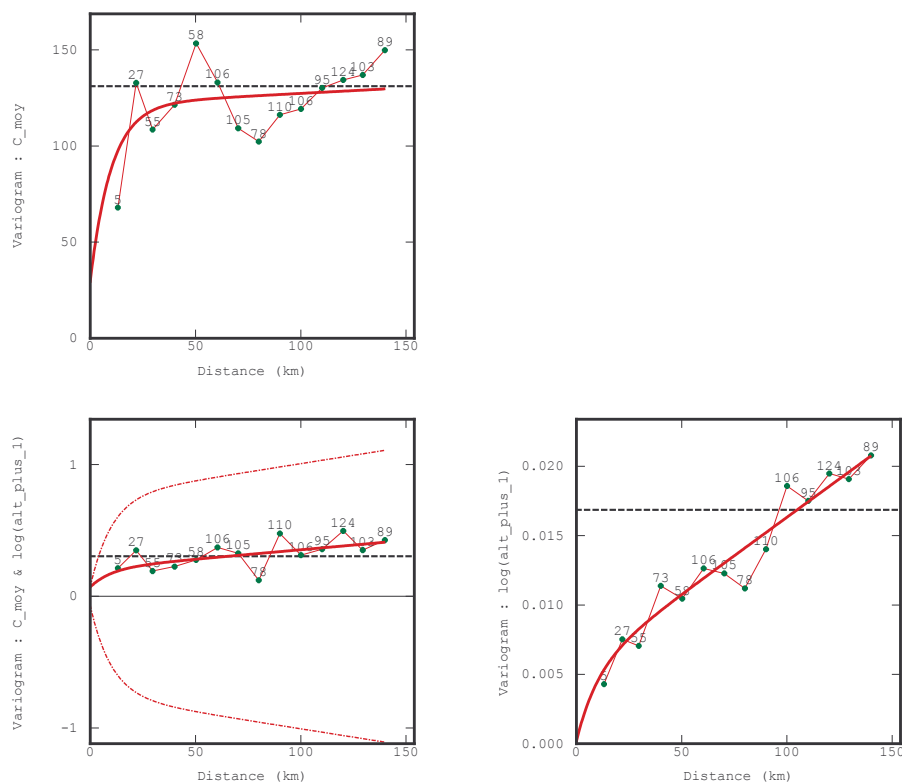


Figure 76 – Variogrammes expérimentaux simples et croisé (60 sites ; 15 pas ; 10 km)

- **Modélisation des variogrammes et validation croisée**



**Figure 77 – Variogrammes simulés simples et croisé :
pépite (0,07) + exp (30 km ; 0,14) + lin (35 km ; 0,05)**

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 70 km.

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

S30 (60 points)	Moyenne	Variance
Erreur	0,08856	118,71078
Erreur standardisée	0,01117	0,98192

Toutes les données sont robustes.

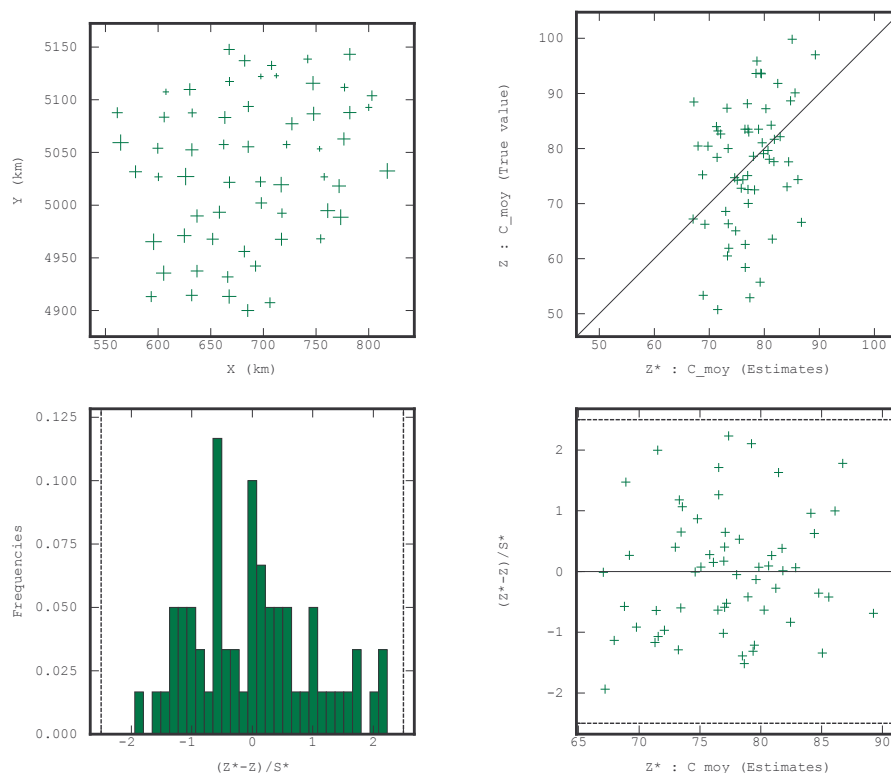


Figure 78 – Résultats de la validation croisée

La corrélation obtenue entre concentrations mesurées et estimées présente un coefficient de 0,332.

▪ Test de validation absolue sur le complémentaire C30

Tableau 26 – Résultats de la validation absolue sur le complémentaire C30 pour le krigeage ponctuel colocalisé

Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S30			
C30	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d'estimation ($\mu\text{g m}^{-3}$)
Nombre de données	393	393	393
Minimum	50,43	59,17	7,06
Percentile 25	63,12	72,20	9,98
Médiane	70,56	75,22	10,39
Percentile 75	79,99	78,92	10,62
Maximum	99,19	88,31	11,18
Etendue	48,76	29,14	4,12
Moyenne	72,0	75,47	10,25
Ecart-type	10,72	4,88	0,52
Variance	114,9	23,84	0,27
CV%	14,9	6,5	5,1
Corrélation mesures/estimations	$r = 0,433$		

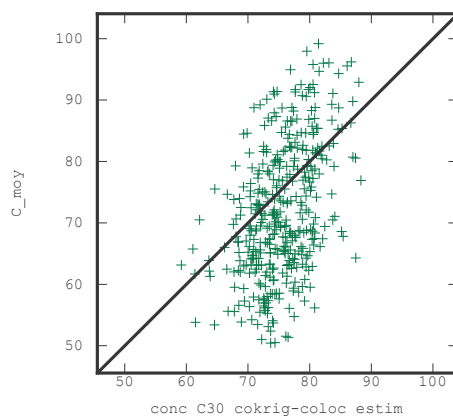


Figure 79 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C30

- **Etablissement de la carte pour la sélection S30**

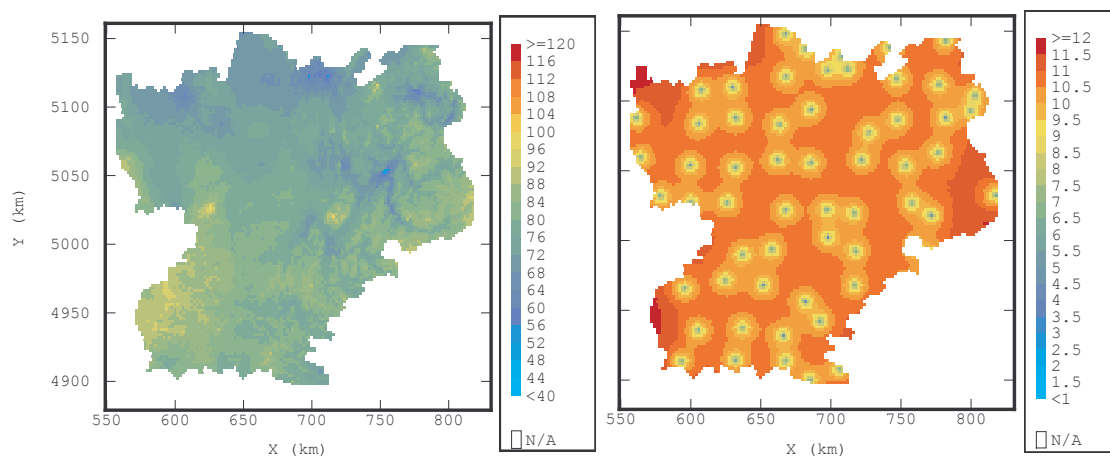


Figure 80 – (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données S30, et (droite) écarts-types de krigeage

- **ETUDE DE LA SELECTION DE DONNEES S45 (MAILLE DE 45 x 45 KM)**

- **Etude des variogrammes expérimentaux**

On calcule pour toutes les directions les deux variogrammes simples et le variogramme croisé avec 13 pas de 15 km.

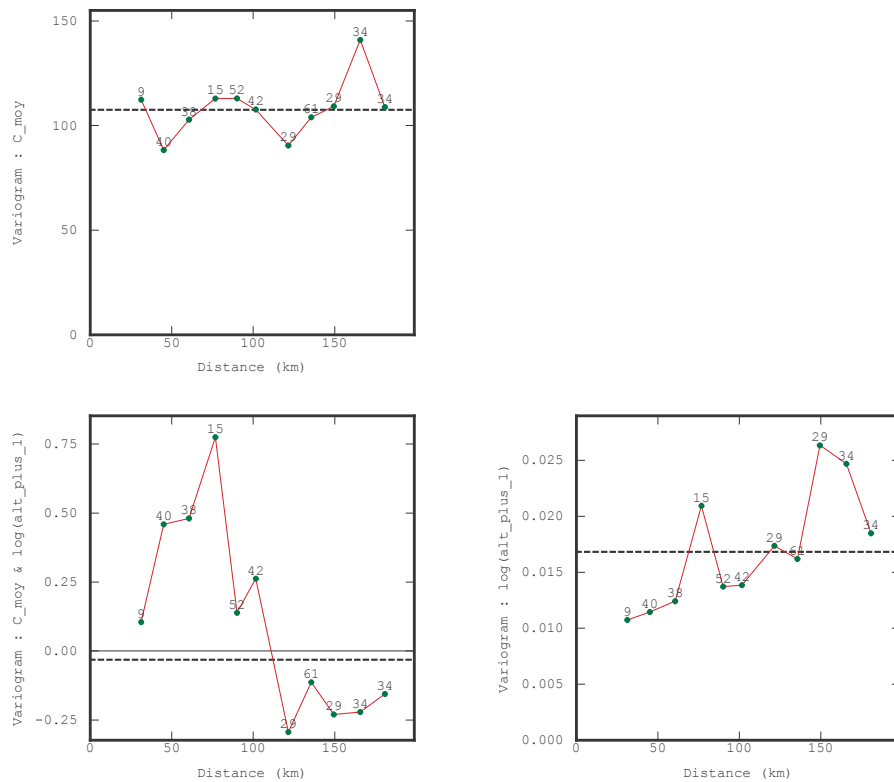
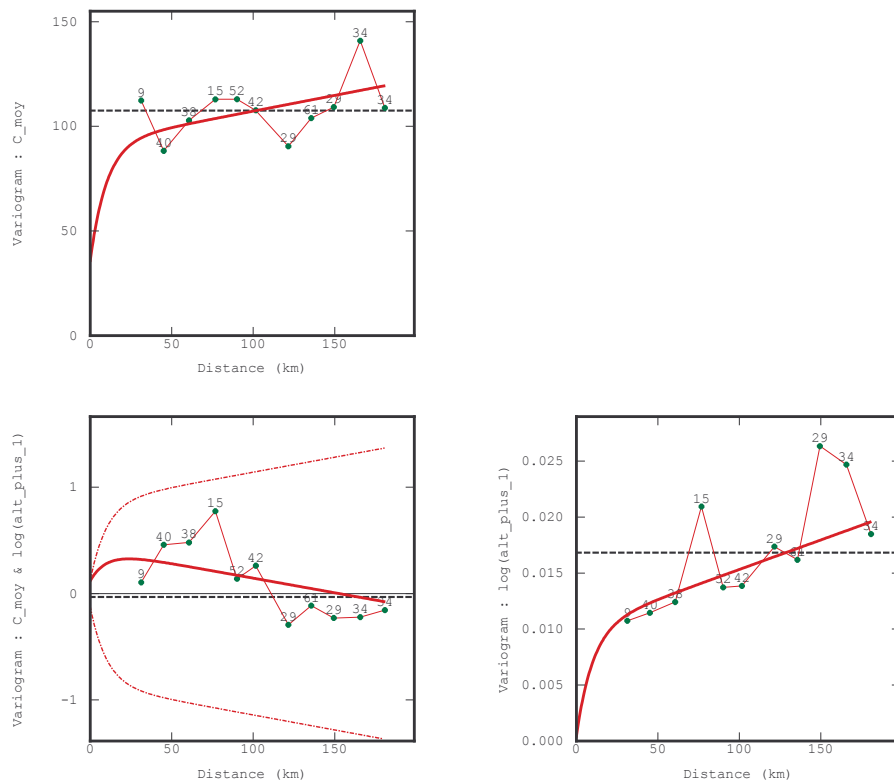


Figure 81 – Variogrammes expérimentaux simples et croisé (31 sites ; 13 pas ; 15 km)

▪ Modélisation des variogrammes et validation croisée



**Figure 82 – Variogrammes simulés simples et croisé :
pépité (0,12) + exp (30 km ; 0,30) + lin (120 km ; -0,33)**

On teste ces résultats par validation croisée en utilisant le modèle bivariable et un voisinage de 70 km.

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

S45 (31 points)	Moyenne	Variance
Erreur	0,72575	98,06942
Erreur standardisée	0,05808	0,92470

Toutes les données sont robustes.

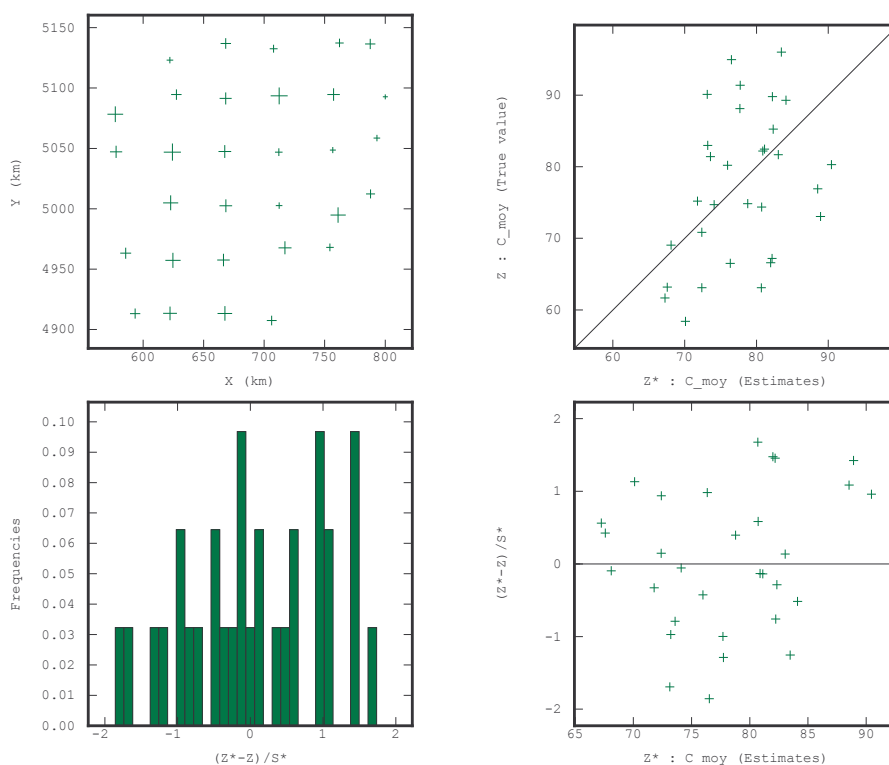


Figure 83 – Résultats de la validation croisée

La corrélation obtenue entre concentrations mesurées et estimées présente un coefficient de 0,369.

• Test de validation absolue sur le complémentaire C45

Tableau 28 – Résultats de la validation absolue sur le complémentaire C45 pour le krigeage ponctuel colocalisé

		Validation croisée par krigeage ponctuel colocalisé selon le modèle de la sélection S45	
C45	Mesures ($\mu\text{g m}^{-3}$)	Estimations ($\mu\text{g m}^{-3}$)	Erreur d’estimation ($\mu\text{g m}^{-3}$)
Nombre de données	422	422	422
Minimum	50,43	61,54	6,70
Percentile 25	64,21	71,54	9,43
Médiane	71,30	77,90	9,58
Percentile 75	80,39	81,07	9,66
Maximum	99,83	91,44	10,59
Etendue	49,40	29,90	3,89
Moyenne	72,3	76,73	9,49
Ecart-type	10,92	6,06	0,36
Variance	119,3	36,67	0,13

CV%	15,1	7,9	3,8
Corrélation mesures/estimations	r = 0,230		

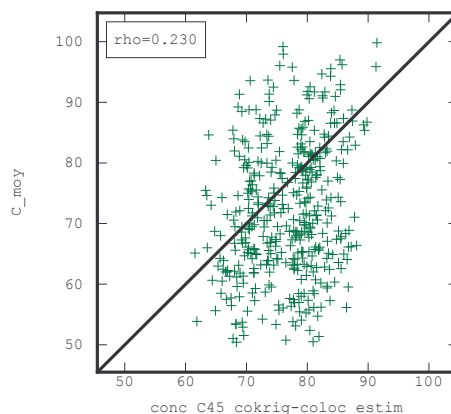


Figure 84 – Nuage de corrélations entre les concentrations estimées et mesurées aux points complémentaires C20

▪ Etablissement de la carte pour la sélection S45

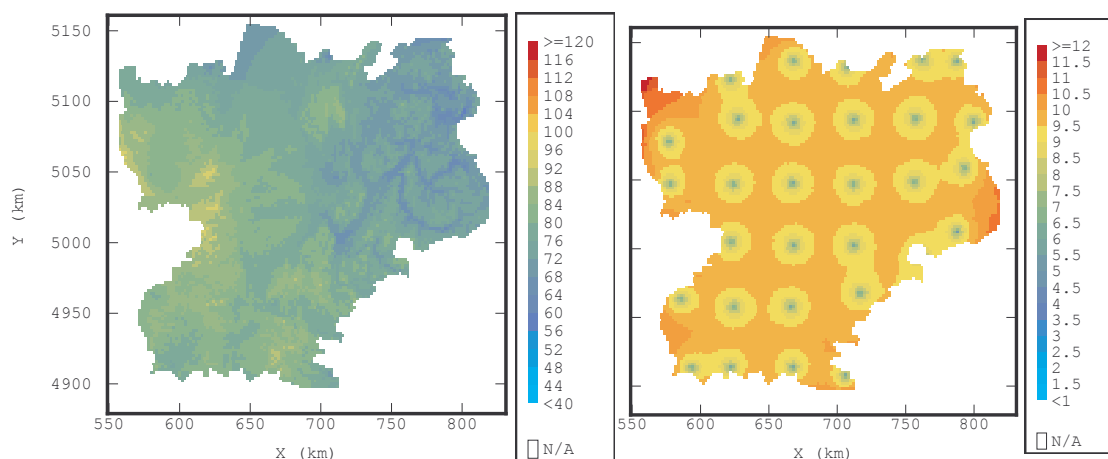


Figure 85 – (gauche) Estimation par krigeage ponctuel colocalisé, à partir du jeu de données S45, et (droite) écarts-types de krigeage

▪ COMPARAISON DES ESTIMATIONS ET DISCUSSION

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

Valeurs	J ₁	S15	S20	S30	S45
Nombre de données	453	176	115	60	31
Minimum (µg m ⁻³)	50,43	50,75	53,72	50,75	58,40
Maximum (µg m ⁻³)	99,83	92,91	97,95	99,83	96,02
Moyenne (µg m ⁻³)	72,7	74,4	76,0	77,1	77,3
Ecart-type (µg m ⁻³)	10,96	9,90	11,16	11,55	10,54
Variance (µg m ⁻³)	120,1	98,0	124,6	133,3	111,1

Validation croisée						
Erreur						
Moyenne	-0,24618	-0,07708	-0,11401	0,08856	0,72575	
Variance	81,46404	82,13215	88,83417	118,71078	98,06942	
Erreur standardisée						
Moyenne	-0,02360	-0,00888	-0,01471	0,01117	0,05808	
Variance	1,00180	1,00489	0,99474	0,98192	0,92470	
Coefficient de corrélation r =	0,571	0,564	0,533	0,332	0,369	

Validation absolue						
Erreur						
Moyenne	-	8,54	8,75	10,25	9,49	
Variance	-	0,02	0,08	0,27	0,13	
Coefficient de corrélation r =	-	0,477	0,480	0,433	0,230	

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 amélioration significative de la corrélation pour la sélection S30 ($r = 0,806$), comparé au maillage régulier sans variable auxiliaire ($r = 0,662$).

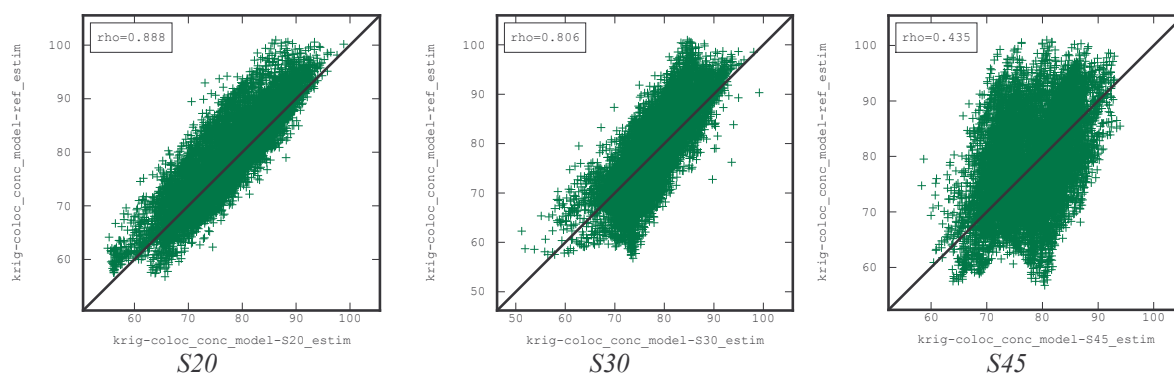
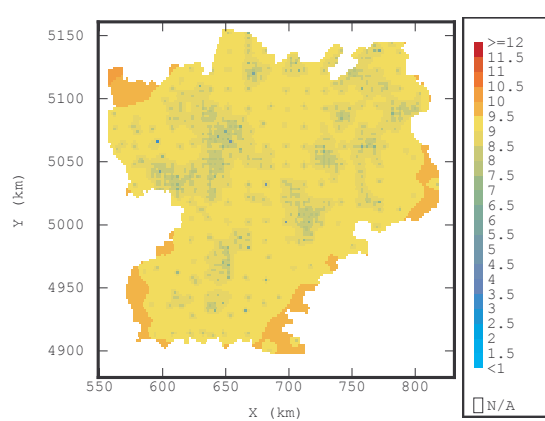
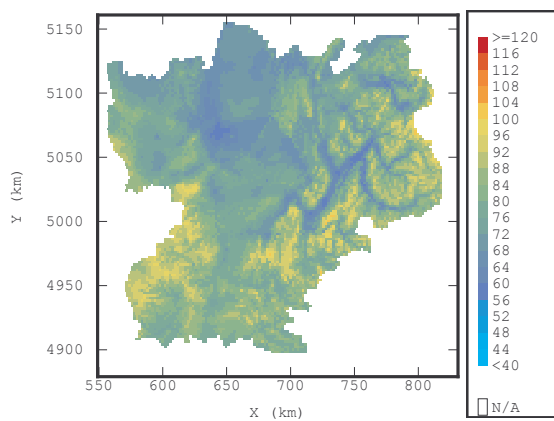
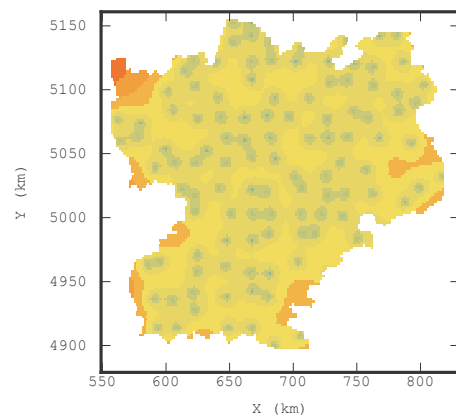
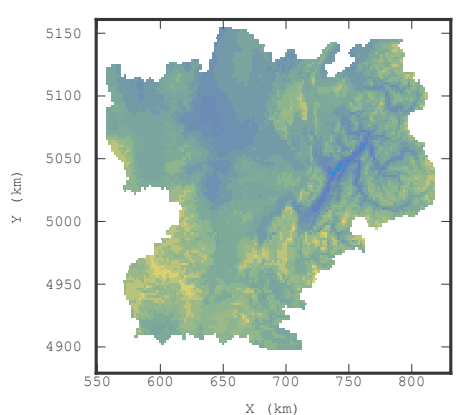


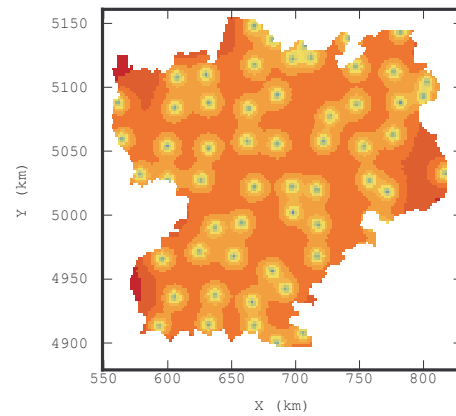
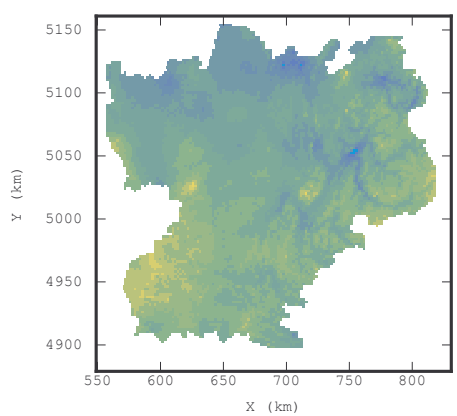
Figure 86 – Comparaison entre l'estimation de référence et les estimations obtenues à partir des différentes sélections



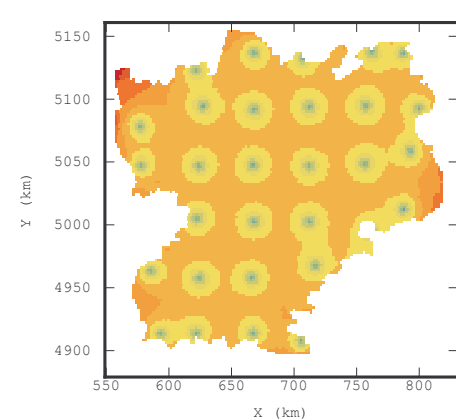
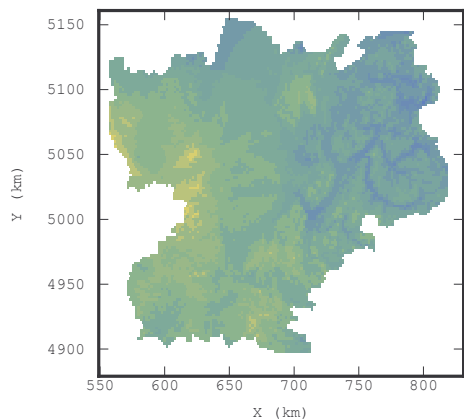
J1



S20



S30



S45

Figure 87 – 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 plus marquée à partir de la sélection S30.

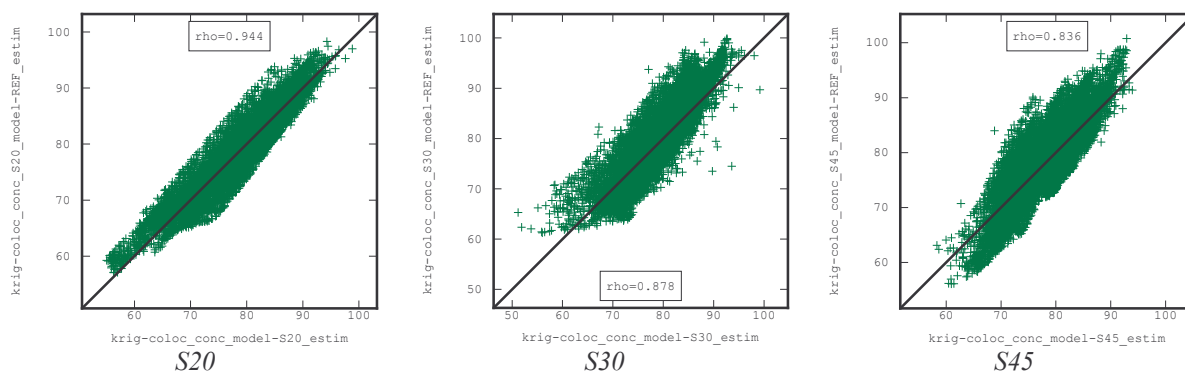
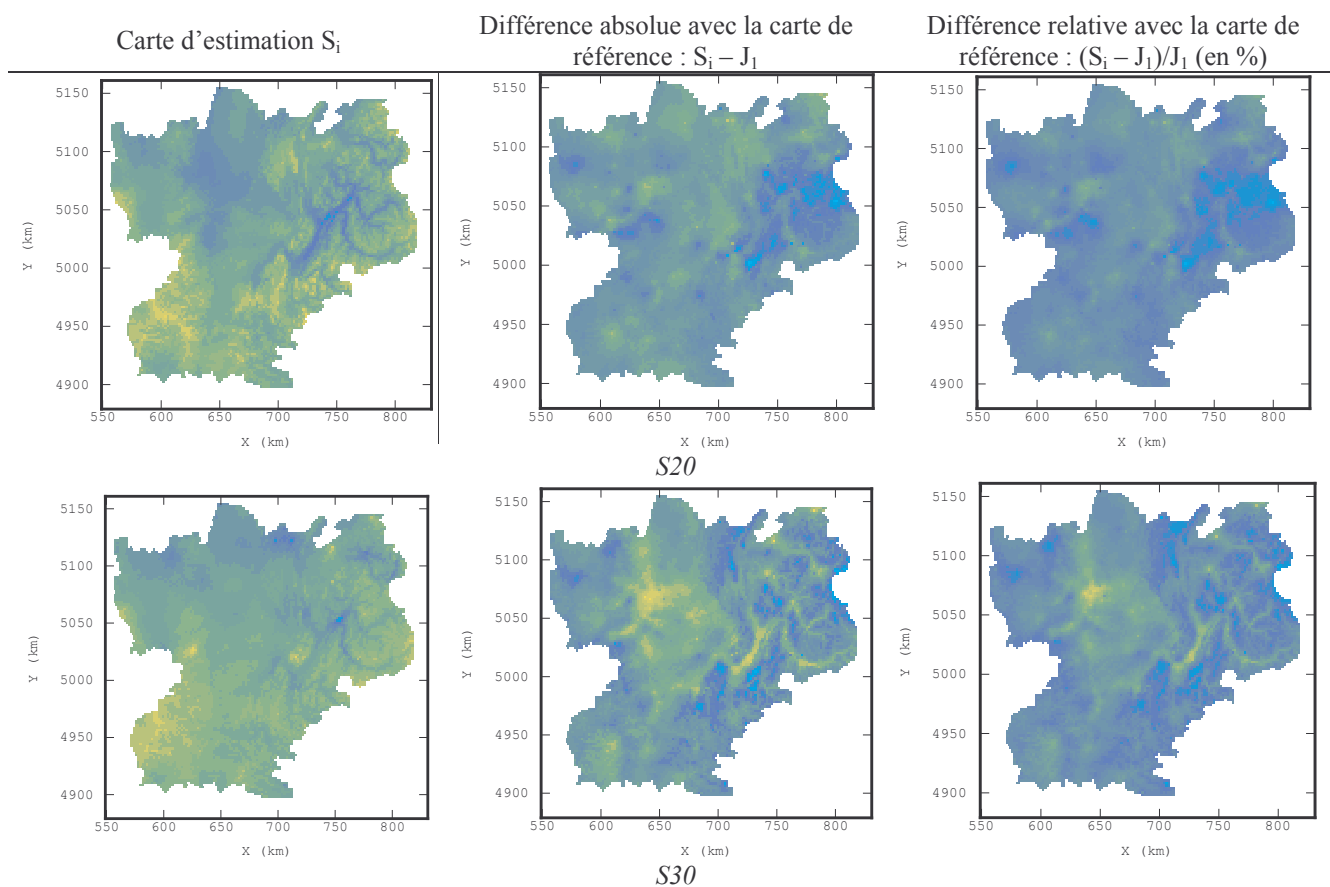


Figure 88 – Comparaison entre les estimations obtenues pour chaque sélection avec le modèle de référence (ordonnée) et avec le modèle de la sélection (abscisse)

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



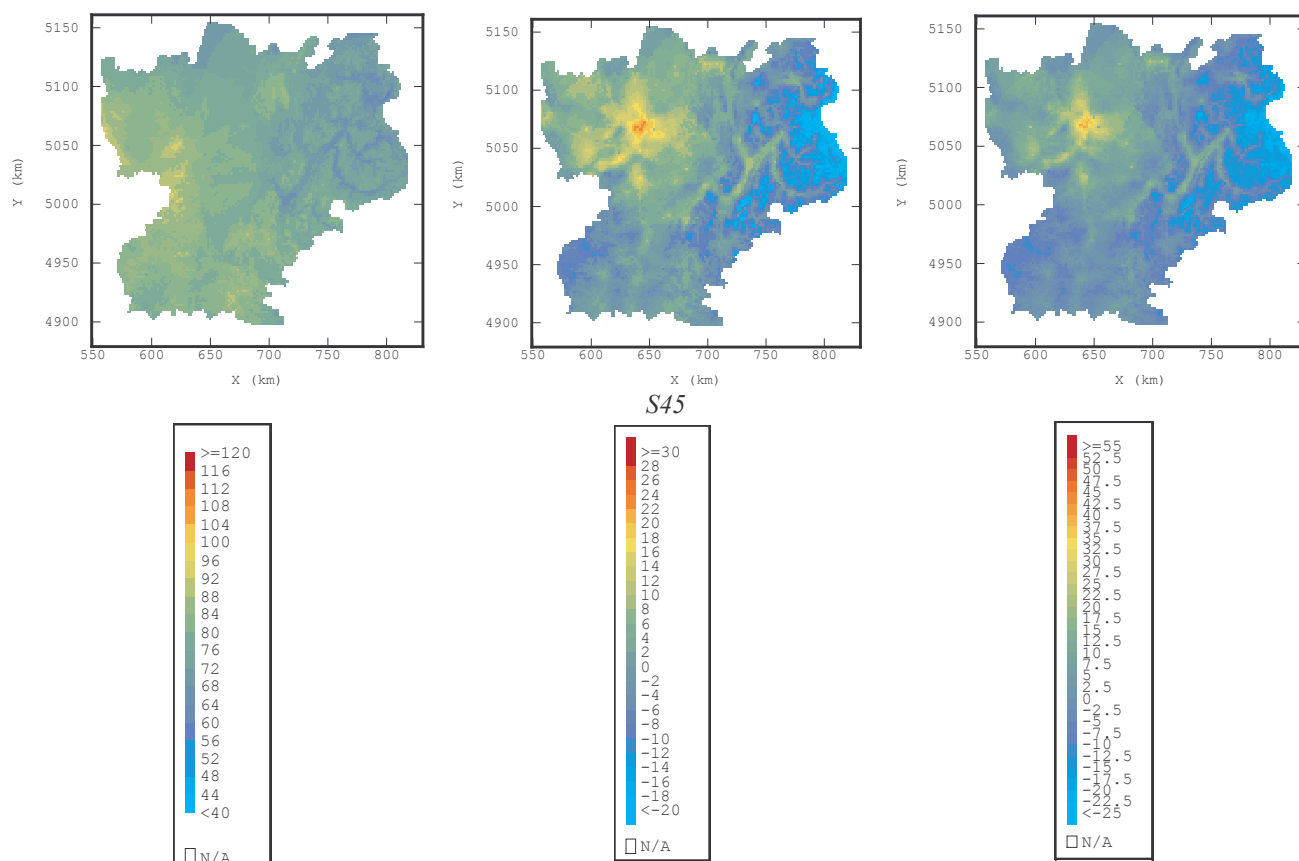


Figure 89 – Comparaison entre les cartes d’estimation obtenues avec le modèle de référence et avec les différentes sélections (de bas en haut : S20, S30, S45)

En valeur absolue, la différence reste inférieure à $20 \mu\text{g m}^{-3}$ pour les sélections S20 et S30 ; elle atteint $27 \mu\text{g m}^{-3}$ pour la sélection S45. En valeur relative, l’écart avec la carte de référence reste inférieur à 30% pour les sélections S20 et S30, mais dépasse 40% pour la sélection S45. Le lissage des concentrations reste important : sous-estimation en altitude et surestimation en agglomération, lyonnaise particulièrement.

Par rapport au maillage régulier sans variable auxiliaire, on note toutefois une réduction des écarts, en particulier pour les reliefs marqués des vallées, et une amélioration de la qualité des cartes : ainsi, la sélection S30 respecte dans ce cas le critère de qualité des 30% d’écart maximum par rapport au jeu de données de référence.

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 de l'ozone à l'échelle régionale en prenant le cas de la région Rhône-Alpes.

Dans un premier temps, les données de campagnes d'échantillonnage par tubes passifs, fournies par Atmo-Rhône-Alpes, ont permis d'établir une carte de référence et d'étudier la dégradation des estimations de concentrations d'ozone en utilisant un maillage régulier.

On a pu observer que la structure des variogrammes expérimentaux se dégrade nettement lorsque le nombre de points diminue, notamment à partir du maillage 30×30 km (60 points).

Ce travail a également montré qu'une maille de 20 km maximum de côté était nécessaire pour vérifier les critères de qualité, soit 115 points, ce qui, compte tenu de la superficie du domaine, correspond à une densité d'échantillonnage de 1 point pour 380 km².

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

Nous avons utilisé le cokrigage colocalisé pour prendre en compte le cofacteur d'altitude dans l'estimation des concentrations d'ozone. La prise en compte de cette variable auxiliaire a permis d'améliorer de manière significative les cartes d'estimation déjà satisfaisantes avec le krigage ordinaire en réduisant les écarts avec les cartes de référence. Ainsi, on a pu montrer qu'une maille de 30 km de côté (60 points) pouvait satisfaire dans ces conditions les critères de qualité d'estimation des concentrations en ozone, soit une densité d'échantillonnage de 1 point pour 730 km².

L'altitude n'apparaît pas cependant comme une variable auxiliaire suffisamment pertinente pour estimer correctement les concentrations d'ozone sur la région Rhône-Alpes, en particulier parce que les concentrations d'ozone à altitude faible peuvent varier dans une large gamme en raison du mode de formation de ce polluant.

Il nous paraîtrait judicieux d'explorer d'autres variables auxiliaires potentielles dans les travaux prévus pour 2007, notamment les données issues de la base Corin Land Cover gérée par l'IFEN et dont une classification opportune restera à définir à partir des données brutes, et les données de population, qui devraient toutes les deux refléter de manière plus significative le comportement spatial de l'ozone. Nous pourrions alors étudier la dégradation d'un maillage stratifié en fonction de ces nouvelles variables.

PARTIE IV. Références

- [1] Atmo-Rhône-Alpes (2003), *Cartographie de l’ozone estival sur la région Rhône-Alpes – Eté 2002*.
- [2] ASCOPARG (2001), *Mise en place d’une méthodologie pour la cartographie de l’ozone à l’échelle du département de l’Isère*, E. CHAXEL.
- [3] LCSQA-EMD (2005), *Adaptation des plans d’échantillonnage spatial aux objectifs de campagne*, A. WROBLEWSKI et L. MALHERBE.
- [4] Atmospheric Environment 34, 5191 (2000), *A climatology of regional background ozone at different elevations in Switzerland (1992-1998)*, S. BRÖNNIMANN et al.

Annexe 9 – Région : Auvergne

▪	DESCRIPTION DU CAS	76
▪	Présentation des données.....	76
▪	Histogramme.....	77
▪	Corrélation avec l'altitude	78
▪	Etude du variogramme expérimental (moyenne de l'été).....	79
▪	Modélisation du variogramme et validation croisée.....	80
▪	Commentaires sur la validation croisée et la qualité de l'estimation	82
▪	RÉFÉRENCES	83

Etude d'une région : surveillance de l'ozone en Auvergne

■ DESCRIPTION DU CAS

NB : à cause du nombre de sites assez restreint, aucune réduction de l'échantillonnage n'a été pratiquée. Cette analyse se limite à dégager les caractéristiques spatiales du jeu de données et à en tirer des informations utiles pour l'élaboration de plans d'échantillonnage.

■ Présentation des données

• *DONNEES DE CONCENTRATION*

Le jeu de données étudié a été fourni par l'association agréée de surveillance de la qualité de l'air ATMO Auvergne. Il est issu du réseau de mesure fixe de l'Auvergne et des départements limitrophes (LIMAIR, LIG'AIR, ATMOSF'AIR Bourgogne Centre Nord, ATMOSF'AIR Bourgogne Sud, AMPASEL, ASQUADRA) et de sites complémentaires temporaires (moyens mobiles).

Il compte 48 points dont 19 sont de type rural, 8 de type périurbain, 20 de type urbain et 1 de type industriel.

Il couvre la période allant du 1^{er} juin 2005 au 30 septembre 2005. Les données journalières sont exploitées ; on en fait la moyenne mois par mois et sur l'été puisqu'il s'agit de formuler des recommandations d'échantillonnage dans un but de cartographie moyenne (saisonnière par exemple).

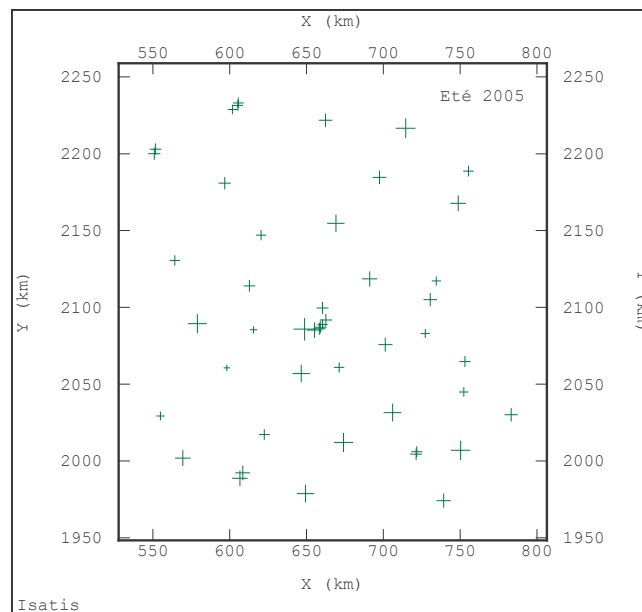


Figure 90 - Carte de répartition des points de mesure

• *VARIABLES AUXILIAIRES*

L'**altitude** de la plupart des sites de mesure est indiquée dans la base de données d'ATMO Auvergne.

D'autre part, une grille kilométrique d'altitude a été extraite par l'EMD d'une base nationale de données topographiques.

LCSQA

**Ecole des Mines de Douai - Département Chimie et Environnement –
INERIS, DRC-07-74775-03939A**

• Histogramme

Remarques préliminaires :

Le site n°30 (MO_CHAT), dont le taux de fonctionnement est de 25% seulement sur la période, est éliminé.

Les sites n°5 (BULLY), n°18 (JABRUN) et 26 (CHATELET) ont fonctionné moins de 25% du temps en juin, juin et septembre respectivement, mais plus de 75% du temps en l'été. Dans les histogrammes de juin et septembre, les valeurs de concentration enregistrées par ces stations ont approximativement la même position que dans les histogrammes des autres mois. Afin de maintenir un effectif suffisant de points de mesure, ces trois sites sont conservés dans l'analyse de la moyenne estivale.

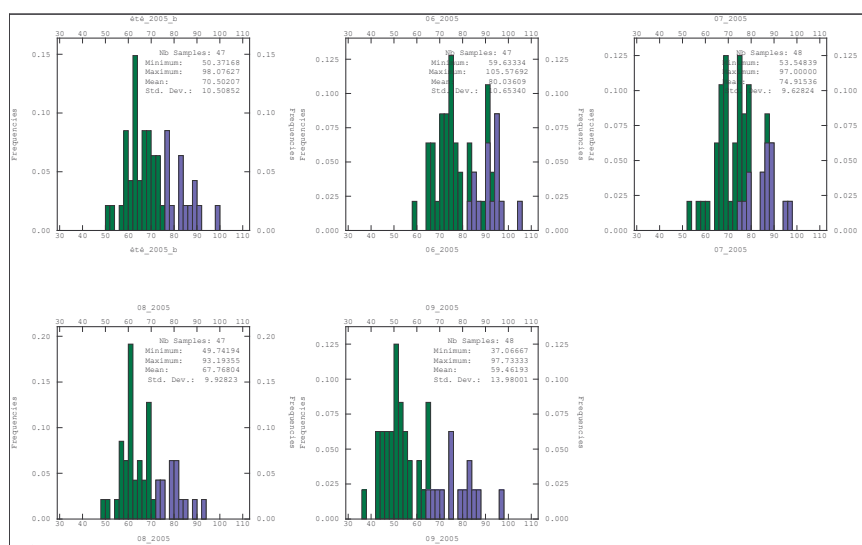


Figure 91 – Histogrammes des concentrations, tous sites confondus. De gauche à droite et de haut en bas : moyennes sur l'été, juin, juillet, août et septembre 2005. En bleu : plus fortes valeurs estivales (en bleu).

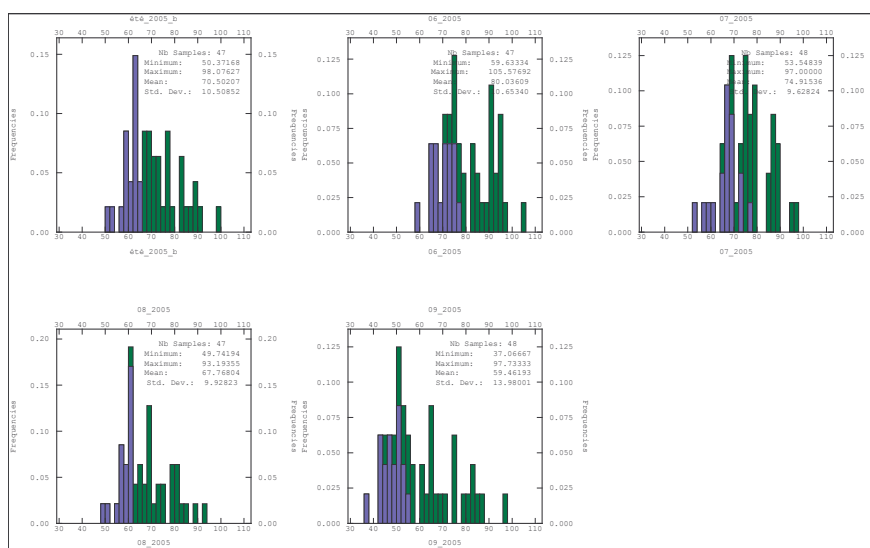


Figure 92 – Histogrammes des concentrations, sites ruraux et périurbains. De gauche à droite et de haut en bas : moyennes sur l'été, juin, juillet, août et septembre 2005. En bleu : plus faibles valeurs estivales (en bleu).

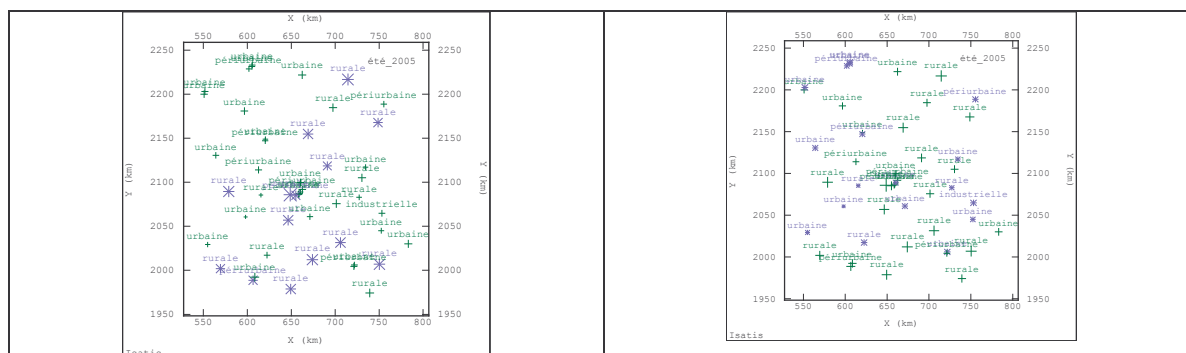


Figure 93 – Localisation des plus fortes (carte de gauche) et des plus faibles (carte de droite) concentrations. Il y a correspondance avec les histogrammes.

Les sites associés aux valeurs les plus élevées (Figure 91, Figure 93) sont pour la plupart de type rural et, pour quelques uns, de type périurbain. Les sites associés aux plus faibles valeurs (Figure 92, Figure 93) sont plus mélangés : de type rural, périurbain, urbain, industriel. Les histogrammes saisonniers ont été tracés par typologie (Figure 94). Les mesures rurales et périurbaines s'étendent sur une plus large gamme de concentration.

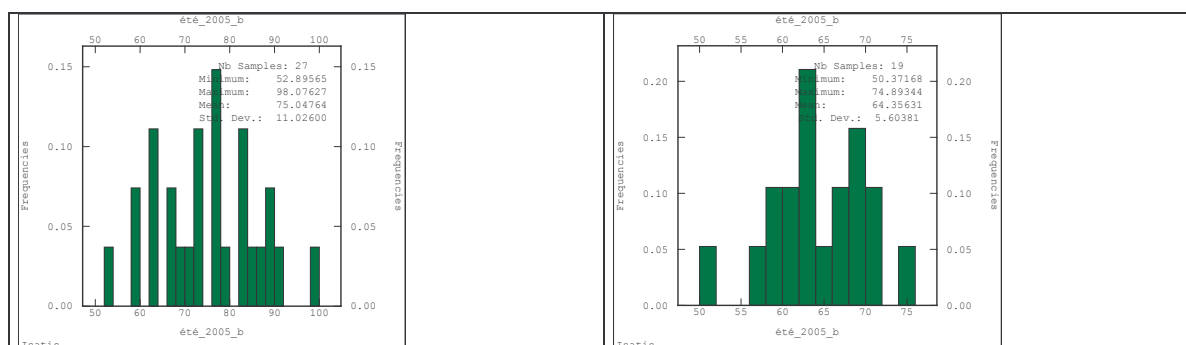


Figure 94 – Histogrammes des concentrations rurales et périurbaines (figure de gauche) et urbaines (figure de droite)

• Corrélation avec l'altitude

Remarque : pour sept des sites, l'altitude est inconnue mais elle est estimée par interpolation à partir de la grille kilométrique.

Aux points où la donnée d'altitude existe, une comparaison entre les valeurs mesurées et interpolées montre, dans l'ensemble, une excellente concordance. Ponctuellement cependant, des écarts peuvent apparaître du fait de gradients d'altitude localement élevés. Afin de ne pas introduire de biais, les sept sites mentionnés ci-dessus ne sont pas considérés dans l'étude de la corrélation.

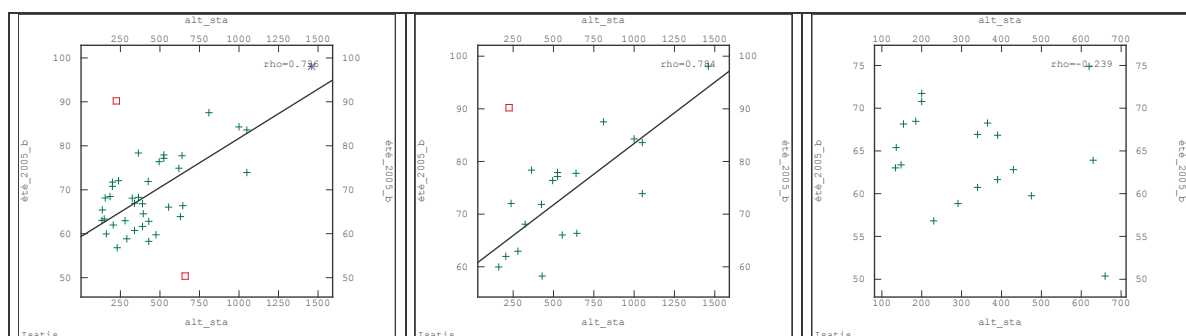


Figure 95 – Nuages de corrélation entre concentration d'ozone et altitude. A gauche : tous sites confondus. Au milieu : sites ruraux et périurbains. A droite : sites urbains.

Deux points s'écartent notablement de la bissectrice. Il s'agit des stations HONORE (rurale) et USSEL (urbaine). Ces deux sites mis à part, la concentration apparaît positivement corrélée à l'altitude. Comme il a été constaté (LIMAIR, 2006) et ainsi que le confirme une analyse par typologie, cette corrélation est plutôt le fait des stations rurales et périurbaines. Sur l'ensemble des stations urbaines, la corrélation est nulle, plusieurs nuages semblant se différencier.

La relation entre la concentration moyenne estivale et l'altitude est établie par régression linéaire sur les sites périurbains et ruraux, après retrait de la station HONORE. Le site Puy-de-Dôme (en bleu) tire le nuage de corrélation vers le haut mais ne perturbe pas sensiblement les coefficients de la régression linéaire (il augmente légèrement le poids accordé à l'altitude). On décide ici de le conserver. La relation s'écrit :

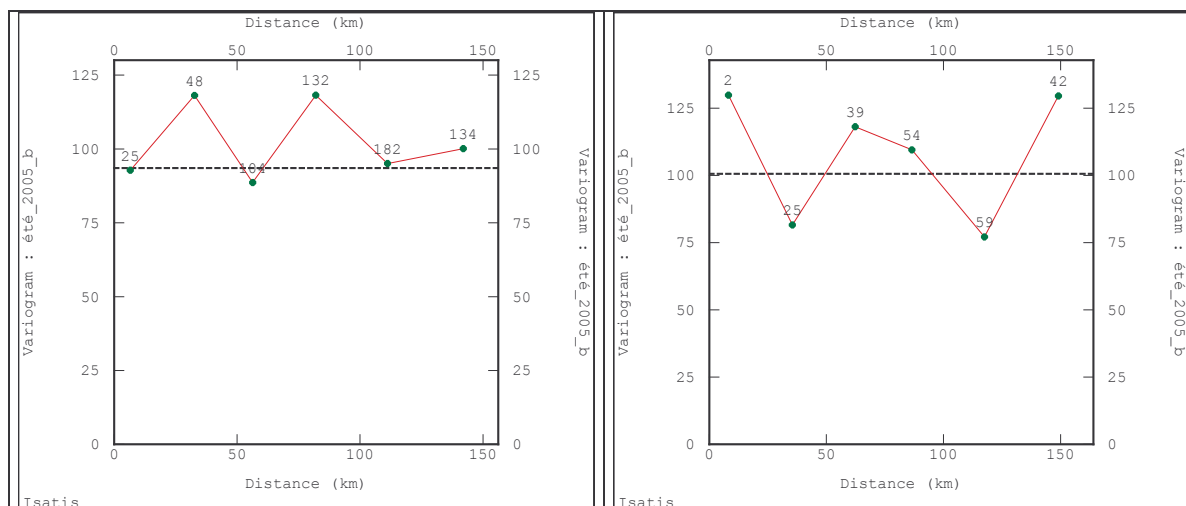
$$Z_{O_3} = 0,023 \cdot \text{altitude} + 60,1$$

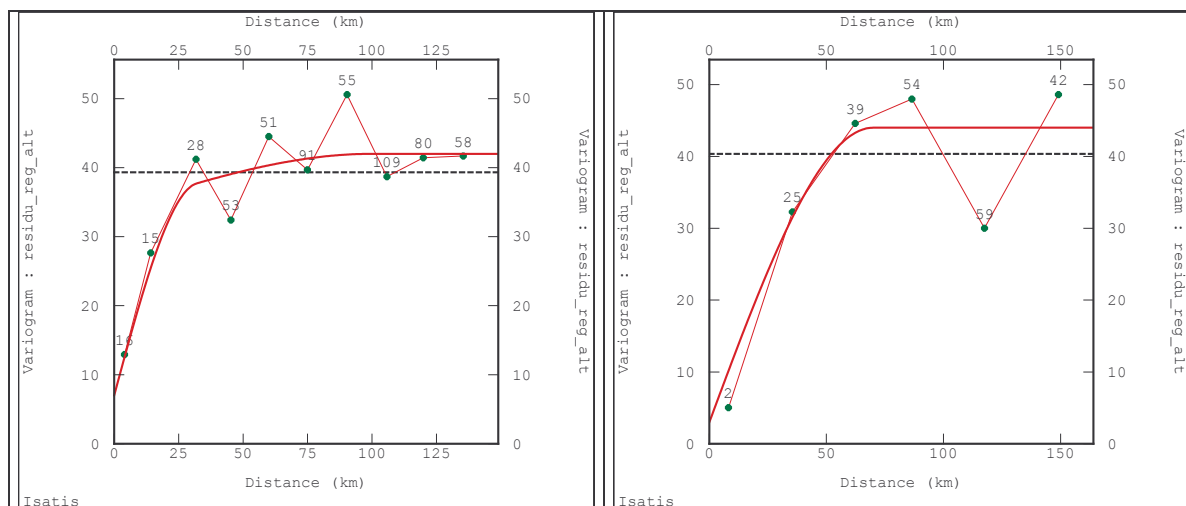
Elle est ensuite appliquée à la totalité des données (toutes typologies) en utilisant, le cas échéant, l'altitude interpolée.

• Etude du variogramme expérimental (moyenne de l'été)

Quel que soit l'ensemble de points considéré (toutes typologies ou sites ruraux-périurbains), le variogramme expérimental de la concentration d'ozone a une allure pépitique alors que la pollution d'ozone est connue comme un phénomène de grande portée (voir par exemple AIRMARAIX, 2004). Ce résultat peut s'expliquer par la variabilité liée aux gradients d'altitude, qui masque la structure spatiale de la concentration. Cette interprétation se confirme par l'examen du variogramme des résidus de la régression. Aucune anisotropie n'est mise en évidence.

Remarque : les stations rurales HONORE, VERNEU et FALGOUX et la station urbaine USSEL sont retirées du calcul des variogrammes. En effet, les résidus expérimentaux calculés en ces points s'écartent de l'histogramme des résidus et destructurent les variogrammes (Figure 4). Notons que pour Verneu et Falgoux, la donnée d'altitude est une valeur interpolée et que dans la commune contenant Falgoux, les variations topographiques sont importantes.





**Figure 96 – Ligne du haut : variogramme expérimental des concentration d’ozone.
Ligne du bas : variogramme expérimental et modélisé des résidus de la régression, sans les quatre sites listés ci-dessus.**

Colonne de gauche : tous sites confondus ; colonne de droite : sites ruraux et périurbains

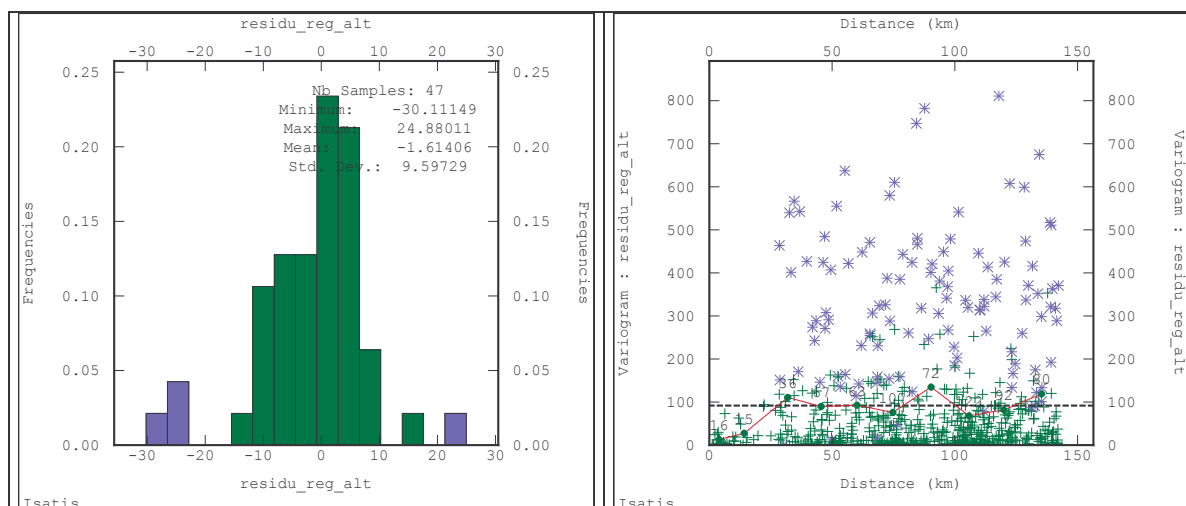


Figure 97 – Histogramme des résidus de la régression (tous sites) et nuée variographique. En bleu : stations HONORE, VERNEU, FALGOUX, USSEL.

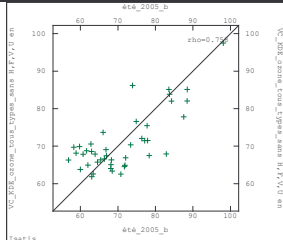
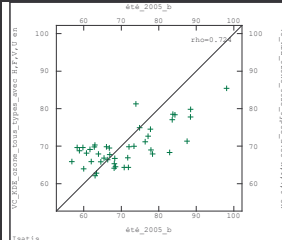
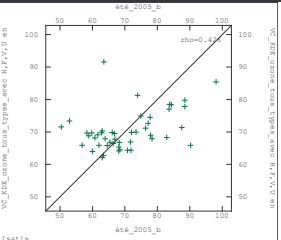
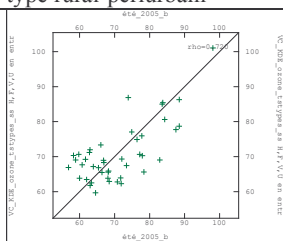
• Modélisation du variogramme et validation croisée

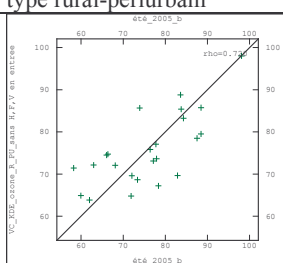
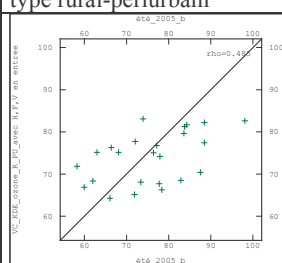
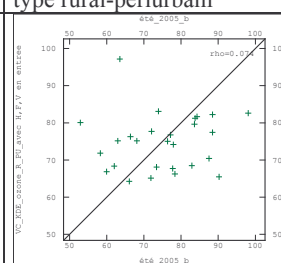
Un modèle est ajusté sur le variogramme expérimental des résidus construit soit avec tous les types de sites, soit avec les sites ruraux et périurbains. Dans le premier cas, il se compose d’un effet de pépite et de deux structures sphériques de portées 32 km et 100 km ; dans le second cas, il se limite à un effet de pépite et à une structure sphérique de portée intermédiaire (70 km).

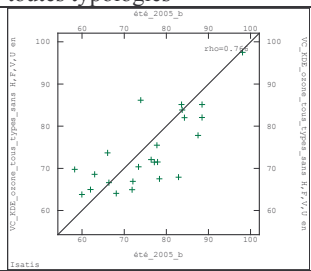
Le choix des structures et des paramètres variographiques découle d’ajustements successifs testés par validation croisée.

Les concentrations sont estimées par krigeage en dérive externe, en voisinage unique, avec l’altitude comme dérive. Le modèle de variogramme est l’un ou l’autre des deux modèles précédemment définis (en spécifiant qu’il s’agit d’un modèle à dérive - *Drift part* dans Isatis -).

Les deux tableaux ci-après présentent quelques résultats de validation croisée.

Données d'entrée	Tous sites sauf HONORE, VERNEU, FALGOUX, USSEL	Tous sites avec HONORE, VERNEU, FALGOUX, USSEL	Tous sites avec HONORE, VERNEU, FALGOUX, USSEL
Points pris en compte dans le tracé du nuage de corrélation et le calcul des statistiques	Tous sites sauf HONORE, VERNEU, FALGOUX, USSEL	Tous sites sauf HONORE, VERNEU, FALGOUX, USSEL	Tous sites avec HONORE, VERNEU, FALGOUX, USSEL
Modèle de variogramme	Variogramme des résidus toutes typologies	Variogramme des résidus toutes typologies	Variogramme des résidus toutes typologies
Nuage de corrélation			
Moyenne et variance de l'erreur/de l'erreur relative	m=-0,15g/m ³ ; var=39,95 m=7,43 % ; var=29,66	m=-1,21g/m ³ ; var=47,55 m=8,11 % ; var=27,65	m=-0,14 µg/m ³ ; var=92,7 m=10,65 % ; var=98,64
Modèle de variogramme	Variogramme des résidus de type rural-périurbain		
Nuage de corrélation			
Moyenne et variance de l'erreur/de l'erreur relative	m=-1,68g/m ³ ; var=51,43 m=8,19 % ; var=33,66		

Données d'entrée	Sites ruraux-périurbains HONORE, VERNEU, FALGOUX	Sites ruraux-périurbains avec HONORE, VERNEU, FALGOUX	Sites ruraux-périurbains avec HONORE, VERNEU, FALGOUX
Points pris en compte dans le tracé du nuage de corrélation et le calcul des statistiques	Sites ruraux-périurbains sauf HONORE, VERNEU, FALGOUX	Sites ruraux-périurbains sauf HONORE, VERNEU, FALGOUX	Sites ruraux-périurbains avec HONORE, VERNEU, FALGOUX
Modèle de variogramme	Variogramme des résidus de type rural-périurbain	Variogramme des résidus de type rural-périurbain	Variogramme des résidus de type rural-périurbain
Nuage de corrélation			
Moyenne et variance de l'erreur/de l'erreur relative	m=-0,07 µg/m ³ ; var=49,78 m=7,95 % ; var=34,30	m=-1,84 µg/m ³ ; var=78,09 m=10,33 % ; var=37,52	m=-0,30 µg/m ³ ; var=164,3 m=14,06 % ; var=160,0

Données d'entrée	Tous sites sauf HONORE, VERNEU, FALGOUX, USSEL
Points pris en compte dans le tracé du nuage de corrélation et le calcul des statistiques	Sites ruraux-périurbains sauf HONORE, VERNEU, FALGOUX
Modèle de variogramme	Variogramme des résidus toutes typologies
Nuage de corrélation	
Moyenne et variance de l'erreur/de l'erreur relative	$m = -1,69 \mu\text{g}/\text{m}^3$; $\text{var} = 42,61$ $m = 7,45 \%$; $\text{var} = 28,72$

• Commentaires sur la validation croisée et la qualité de l'estimation

L'introduction dans le krigeage des sites HONORE, VERNEU, FALGOUX et USSEL dégrade légèrement l'estimation aux autres stations. En ces quatre points, quel que soit le modèle, la concentration est nettement surestimée (VERNEU, FALGOUX et USSEL) ou sous estimée (HONORE).

Le modèle construit toutes typologies confondues se révèle le plus efficace pour une estimation sur la totalité des sites.

Aux stations rurales et périurbaines, la qualité de l'estimation est statistiquement peu différente selon qu'on utilise les données et le modèle de type rural-périurbain ou les données et le modèle relatifs à toutes les typologies.

Ainsi :

- avec une bonne répartition spatiale des stations de mesure (la distance moyenne entre sites ruraux/périurbains est voisine de 30 km, valeur légèrement inférieure à la portée de la première structure, à quoi s'ajoutent des resserrements près des agglomérations) ;
- avec des données secondaires soigneusement choisies (l'altitude se révèle ici indispensable pour mettre au jour une structure spatiale) ;

le krigeage fournit une estimation assez satisfaisante des moyennes estivales observées.

En quelques points particuliers cependant, la concentration n'est pas correctement estimée ; d'autre part, les nuages de corrélation présentent une certaine dispersion autour de la bissectrice. C'est pourquoi il pourrait être utile de rechercher des variables auxiliaires supplémentaires afin d'accroître la précision de l'estimation. Ce travail sort du cadre actuel de l'étude.

▪ REFERENCES

AIRMARAIX, avec la participation de Géovariances, 2004. Evaluation des outils de cartographie de la pollution par l’ozone dans les Bouches-du-Rhône, le Var et le Vaucluse. Rapport disponible sur le site www.airmaraix.com.

LIMAIR, 2006. Etude ozone – été 2005. Région Limousin. Rapport référencé ETD/2005/13, disponible sur le site www.limair.asso.fr.