

Estimation du dioxyde d'azote routier dans la vallée de la Thur

Serge Séguret, février 2003

Résumé de l'étude

Nous cherchons à cartographier le dioxyde d'azote (NO_2) de part et d'autre d'une route nationale pour les saisons été et hiver. Sont fournis: en 55 points de mesure (tubes), les moyennes saisonnières et par quinzaine; en 3 stations, des mesures quotidiennes ; sur une grille de 1km, des variables auxiliaires (émissions du cadastre, variables botaniques et autres telles l'occupation des sols et la population). L'altitude est donnée sur une grille de 250m.

L'Analyse en Composantes Principales ne permet de dégager que "x", la distance à la route des tubes, et l'altitude, cette dernière étant fortement corrélée à la distance le long de la route (notée "y", ordonnée curviligne). En conséquence, la variable est régionalisée dans l'espace (x,y). Le problème ainsi posé, l'ensemble des tubes dont on peut exploiter les mesures se réduit à 39 éléments. Tous ces points sont détaillés dans le **Chapitre I, "Problématique"**.

La variable $\text{NO}_2(x,y)$ présente une forte non-stationnarité qui est double : en moyenne, et en variance. Obtenir l'estimation procède alors de quatre étapes : calcul d'une dérive, d'un écart-type, modélisation des variogrammes, estimation. Chaque étape fait l'objet d'un chapitre :

Chapitre II – Modélisation de la dérive $m(x,y)$. La non-stationnarité de $\text{NO}_2(x,y)$ est examinée le long de la route et transversalement, dans le but de donner à $m(x,y)$ une forme analytique permettant d'exprimer la dérive partout. Pour que l'analyse en x soit cohérente avec celle en y, les mêmes données doivent être utilisées. Ceci oblige, dans un premier temps, à établir un modèle de dérive sur l'ensemble T1 des 26 tubes sur transect dont les distances à la route sont 1 ou 2m, 50m, 200 ou 400m. Les coefficients de la dérive $m(x,y)$ sont ensuite estimés sur l'ensemble T2 des 39 tubes exploitables.

Chapitre III – Modélisation de l'écart-type $\sigma(x,y)$. Des résidus sont construits en recentrant $\text{NO}_2(x,y)$ autour de $m(x,y)$. L'examen de leur variance indique qu'elle n'est pas stationnaire et l'on construit un modèle $\sigma^2(x,y)$, fonction non linéaire de la dérive $m(x,y)$.

Chapitre IV – Variographie des résidus normés. On norme par $\sigma(x,y)$ les résidus centrés autour de la moyenne $m(x,y)$ et l'on calcule pour chaque saison les variogrammes expérimentaux de ces quantités. Les variogrammes sont ajustés de deux manières : une approche minimale qui conduit à des modèles purement pépitiqes, une autre où l'on aboutit à un modèle exponentiel et pépitiqes en été, purement sphérique en hiver.

Chapitre V – Estimations. En cas de variogramme purement pépitiqes, l'estimateur du NO_2 est la dérive $m(x,y)$ pour une variance d'estimation donnée par $\sigma^2(x,y)$. Cette première estimation est comparée à celle déduite du krigeage simple des résidus normés (variogrammes exponentiels ou sphériques). Les mécanismes du krigeage, et son importance en cas de variogrammes robustes, sont ainsi montrés. En **Annexe**, deux autres types de krigeage sont proposés, qui diffèrent par le degré de confiance accordé à la dérive $m(x,y)$:

Krigeage Universel (KU) : une fonction affine de la dérive est réestimée

Krigeage Universel plus général (KUG) : les coefficients des fonctions de base sur lesquelles repose la dérive sont réestimés.

Dans la conclusion, **chapitre VI**, certains points sont discutés :

- L'habitat et les points isolés
- L'altitude
- Les rond-points
- Les portions de route courbes
- Le retour à l'espace géographique
- La stationnarité temporelle

SOMMAIRE

I – Problématique	5
0 Résumé du chapitre	5
1 NO ₂ moyen, variable d'intérêt	5
2 Données auxiliaires	6
3 Objectif de l'étude	7
4 Sélection et regroupement des tubes	7
5 Repère de travail	9
II – Modélisation de la dérive $m(x,y)$	11
0 Résumé du chapitre	11
1 Symétrie axiale	11
2 Construction de la dérive	12
2-1 Comportement en y, ordonnée curviligne	12
2-2 Comportement en x, distance à la route	12
2-3 Expression de $m(x,y)$	15
2-4 Synthèse : $m(x,y)$ sur T1	18
2-5 Remarques	19
3 $m(x,y)$ sur T2	20
4 Hiver	21
4-1 Construction	21
4-2 synthèse : $m(x,y)$ sur T1	22
4-3 $m(x,y)$ sur T2	22
5 Histogrammes des résidus	23
6 Discussion	23
6-1 Été/Hiver	23
6-2 Dérive nulle	23
III – Modélisation de l'écart–type $\sigma(x,y)$	25
0 Résumé du chapitre	25
1 Remarque préliminaire	25
2 Par classes de distances à la route	25
3 Par les résidus	27
3-1 Remarque préliminaire	27
3-2 Modélisation	28
4 Discussion	30

4-1 Autres modèles	30
4-2 Variance nulle	31
IV – Variographie des résidus normés	32
0 Résumé du chapitre	32
1 Remarques préliminaires	32
1-1 Séparer les parties gauche et droite ?	32
1-2 Traverser la route ?	32
1-3 Tolérance angulaire	32
1-4 Statut des données en bordure de route	34
2 Variogrammes expérimentaux – Été	34
2-1 En h_x : traversant	34
2-2 En h_x : gauche, droit, moyen, traversant	35
2-3 En h_y : gauche, droit, moyen, traversant	35
3 Variogrammes expérimentaux – Hiver	36
3-1 En h_x : traversant	36
3-2 En h_x : gauche, droit, moyen, traversant	36
3-3 En h_y : gauche, droit, moyen, traversant	36
4 Ajustements	37
4-1 Été : par un effet de pépité	37
4-2 Été : par une structure non purement pépitique	37
4-3 Hiver : par un effet de pépité	38
4-4 Hiver : par une structure non pépitique	38
V – Estimations	39
0 Résumé du chapitre	39
1 Rappels	39
2 Equations du krigeage simple (KS) des résidus	39
3 Cas pépitique	40
3-1 Été	41
3-2 Hiver	42
4 Cas non purement pépitique	43
4-1 Été	43
4-2 Hiver	45
VI – Conclusions	48
1 Synthèse	48
1-1 Objectif	48
1-2 Les étapes	48
1-3 Portée de l'étude	48
2 Discussion	49

2-1 Habitat et points isolés	49
2-2 Altitude	49
2-3 Rond-points	50
2-4 Portions de route courbes	50
2-5 Retour à l'espace géographique	51
2-6 Stationnarité temporelle	51
Annexe – Modèles possibles de krigeage	53
1 KU : Krigeage Universel du NO ₂	53
2 KUG : Krigeage Universel plus Général du NO ₂	54
Index des figures et tableau	55
Index des formules	60

Chapitre I – Problématique

0 Résumé du chapitre

Les données sont d'une part le dioxyde d'azote saisonnier, dont on cherche à réaliser la cartographie spatiale en été et en hiver, et d'autre part une vingtaine de variables auxiliaires connues sur des grilles. L'implantation des 55 capteurs (tubes passifs) est choisie de sorte qu'ils ne mesurent que le dioxyde d'azote émis par la route nationale RN66 de la vallée de la Thur.

L'analyse en composantes principales montre que les seules variables potentiellement explicatives sont la distance du capteur à la route et l'altitude, ce qui conduit à déplacer le problème d'estimation dans un nouvel espace où l'abscisse du point de mesure est la distance à la route et son ordonnée la position le long de la route.

Les points de mesure sont regroupés en sous-ensembles, selon qu'ils appartiennent ou non à des transects. Sur les 55 points de mesures, 10 sont abandonnés parce qu'ils ne peuvent pas être régionalisés dans le nouvel espace ou que la distance à la route ne mesure pas la distance à la source de l'émission étudiée, 4 parce qu'ils sont implantés dans une zone où l'habitat n'est pas représentatif de l'ensemble des mesures, et enfin 2 parce que les mesures qu'ils fournissent brouillent à eux seuls les statistiques déduites de la majorité. Au final, l'étude s'appuie sur 39 points de mesure.

1 NO₂ moyen, variable d'intérêt

L'ASPA (Association pour la Surveillance et l'Etude de la Pollution Atmosphérique en Alsace) a réalisé une campagne de mesures de dioxyde d'azote (NO₂) pour mesurer l'impact de la RN66 dans la vallée de la Thur.

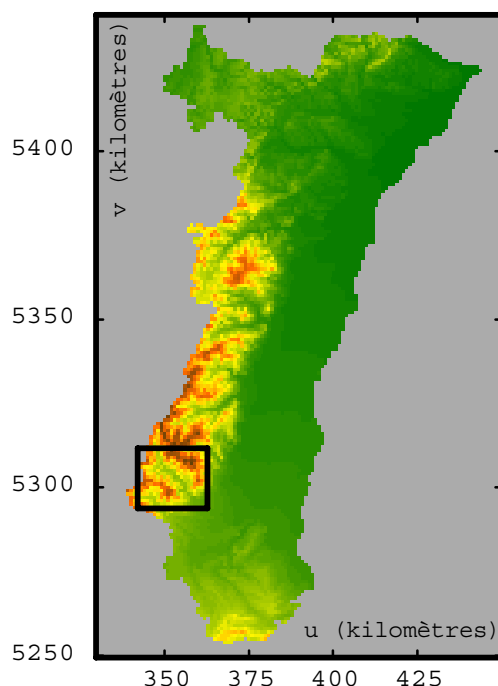


Figure 1 – Carte du relief de l'Alsace. Dans le rectangle, la zone d'étude.

Installés de part et d'autre de la route, 55 tubes passifs ont mesuré quotidiennement les concentrations de NO₂ sur deux périodes : été (jours 1 à 85) et hiver (jours 171 à 255). L'unité de mesure est le $\mu\text{g}/\text{m}^3$. Les données fournies en chaque tube sont une moyenne par quinzaine (6 par saison) et une moyenne saisonnière (été, hiver). 8 tubes sont isolés, les autres appartiennent à 8 transects. Par ailleurs, 2 stations et un camion fournissent des mesures quotidiennes sur des périodes comparables.

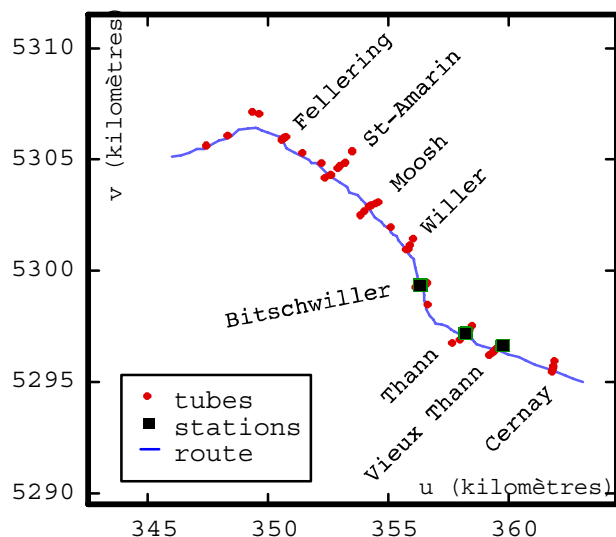


Figure 2 – 55 Tubes (mesures saisonnières et par quinzaine) et 3 stations (mesures quotidiennes)

2 Données auxiliaires

Sont fournies également la grille au pas de 1 km du cadastre des émissions (NO_x, benzène, COVNM, Particules, SO₂, CO et BAR), celle de la population, type de surface bâtie, végétation et autres, et enfin la grille au pas de 250 m de l'altitude. Après migration aux tubes des données auxiliaires (opération qui pourrait faire l'objet d'une étude à part entière, mais que nous ne développons pas ici pour rester concis), une ACP est conduite. L'axe F1 (46% de la variance totale) révèle un effet de taille opposant les variables émissions, démographiques et industrielles, fortement corrélées, aux variables d'intérêt plus l'altitude et la distance à la route. Les variables "botaniques" (conifères, vignes...) n'apportent rien qui ne soit connu et utile ici (conifères et feuillus en altitude, vignes, cultures et prairies proches des sources de NO₂). L'axe F2 (16 % de la variance) oppose le NO₂ à un regroupement altitude, distance_route, regroupement qui n'est qu'un effet de projection (l'axe F3, 12% de variance, les sépare).

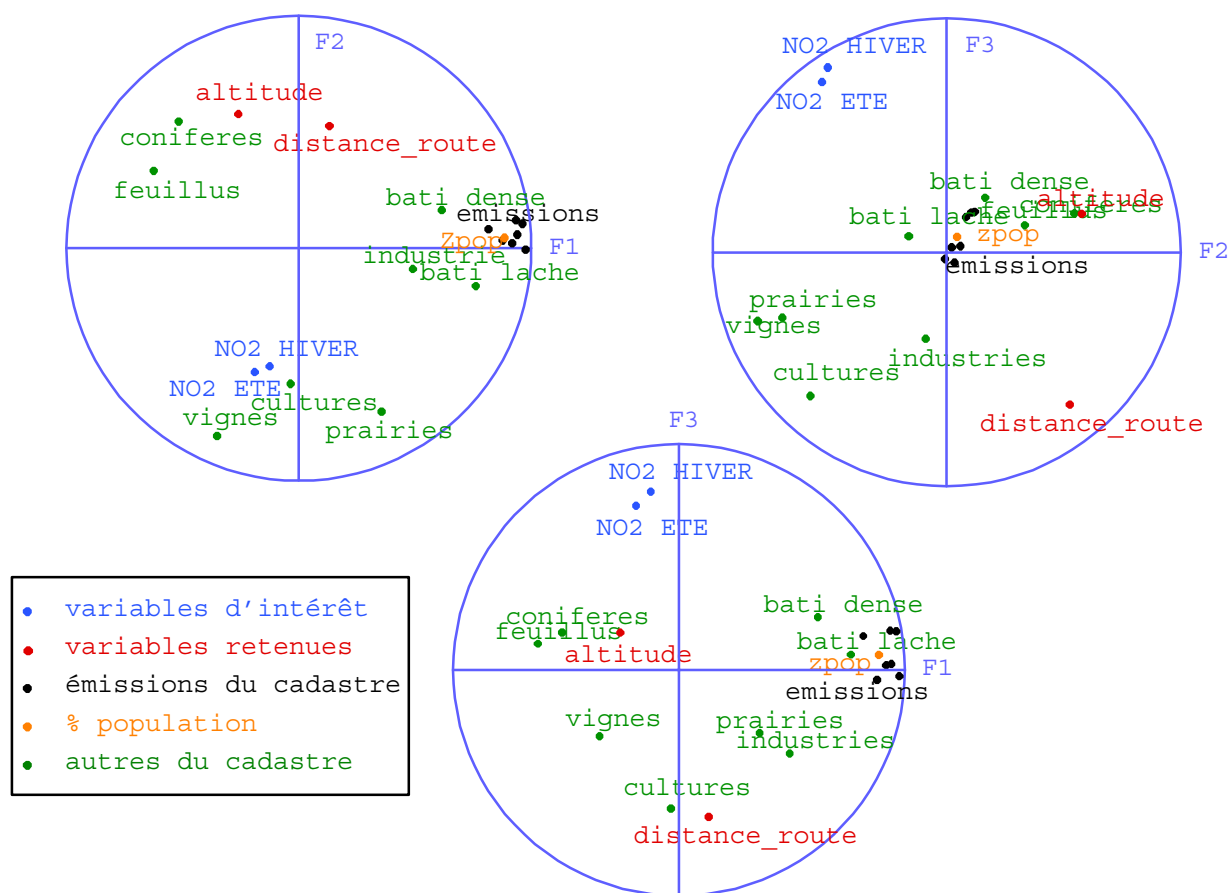


Figure 3 – Projection des variables dans les plans factoriels

3 Objectif de l'étude

Initialement, nous devons examiner s'il était possible de s'appuyer sur les variables auxiliaires pour cartographier la pollution de fond du NO₂ dans la vallée. L'ACP montre que les seules variables potentiellement explicatives sont la position des tubes par rapport à la route et l'altitude, prouvant de ce fait (et à ce niveau de l'analyse) que la campagne de mesure a bien satisfait son objectif : ne mesurer que les émissions routières. En concertation avec l'INERIS, il a été décidé d'aller dans le sens de la campagne de mesures en essayant de cartographier finement le NO₂ saisonnier moyen autour de la route. Compte tenu de la complexité de ce problème et du faible nombre de stations, seules les moyennes des tubes sont utilisées et la problématique tube/station n'est pas abordée.

4 Sélection et regroupement des tubes

Ignorant si la distance à la route se réfère à une position particulière (centre de la route ?) ou concerne chaque bord de la route, et faute d'informations à temps, 9 tubes sont écartés parce qu'ils ne disposent pas de cette variable ou qu'elle ne mesure pas la distance à la source de l'émission étudiée. Il s'agit de :

8VT : donné à 1m alors que graphiquement, sa position est à plus de 500 m

1W, 3W, 4W, 5W, 5, 6 : distance à la route non informée au moment de l'étude.

7SA : moyenne de l'été non disponible (4 quinzaines sur 6 sont informées)

3T : influencé par une autre route

Nous avons tenté de calculer les valeurs absentes de distance_route en utilisant la route donnée par le SIG. Nous nous sommes aperçu que cette route ne passe pas à proximité des points situés à moins de 2 m de la route. Il faut déplacer la route SIG d'environ 150m dans la direction NE 70 degrés pour obtenir quelque chose d'approximativement cohérent.

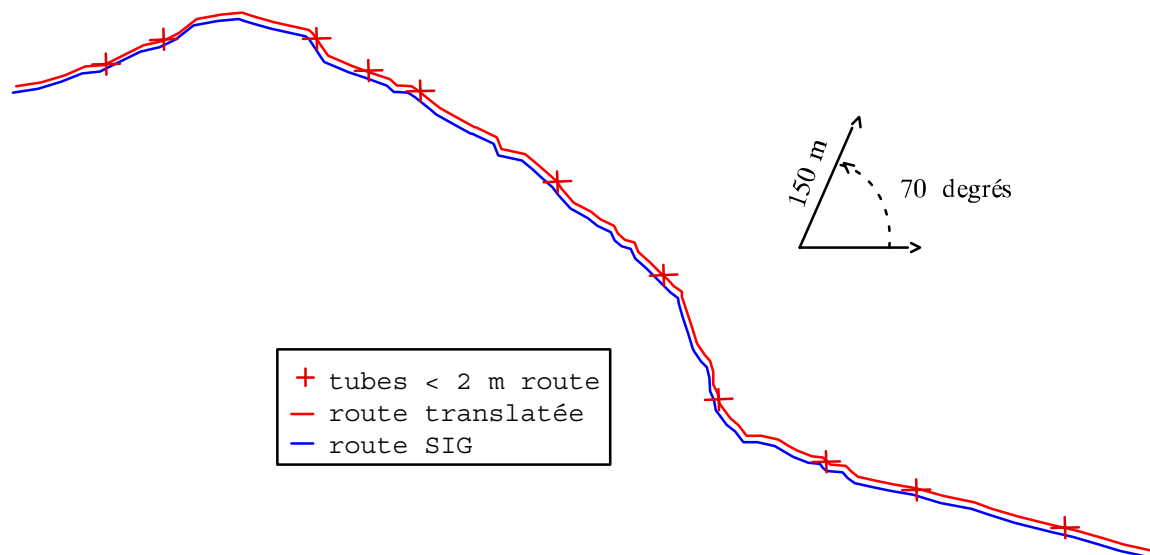


Figure 4 – Décalage entre route SIG et position des tubes

Si, à l'échelle de l'Alsace ou de la vallée, cette incohérence est négligeable, elle devient cruciale dans une étude où les distances à la route sont de l'ordre de 50 à 400 m. Nous avons en conséquence renoncé à calculer les distances manquantes.

7 autres tubes sont écartés pour les raisons suivantes :

Bitschwiller : ce transect alterne des valeurs extrêmes sans rapport avec la distance à la route qui est donnée. Ainsi, le tube 4B, donné pour 50m à la route, mesure une concentration en été égale à 16.82, valeur supérieure à celle du tube 3B (14.12) donné pour être à 1 m de la route. Explication : 4B est situé à proximité d'un rond-point. Sur le même transect, le tube 5B, donné à 200m de la route, présente une concentration de 63.24 en été, la valeur la plus importante de l'ensemble du jeu de données. Explication : ce tube serait *sur* la route. Mais dans ce cas, où se trouve 3B donné à 1m de la route ? N'y aurait-il que 1 m entre 3B et 5B ? La route se diviserait-elle en tronçons ? Quand aux autres valeurs du transect, elles sont anormalement basses. Explication du laboratoire : "ces tubes sont dans un milieu rural masqué par un habitat dense de type rue". Freiné par les surfaces bâties, le NO_2 ne s'y diffuse pas d'une manière comparable aux autres tubes. Compte tenu du caractère totalement atypique de ce transect, nous avons choisi de ne pas l'utiliser.

4SA : ce tube nous a posé des problèmes tout au long de l'étude. Donné pour 100m à la route, sa valeur est en été 32.46 et en hiver 38.62, alors que 50 m plus loin perpendiculairement à la route, elle chute à 10.35 (22.60 en hiver) pour un transect présentant des valeurs très basses dans l'ensemble. Le problème est que nous n'avons pas de tube proche de la route pour ce transect, et donc aucun moyen de comparaison permettant de conclure que la distance à la route n'est pas représentative. Son influence s'avérait considérable dans l'établissement d'un modèle de moyenne, et totalement (in)déterminante dans l'étude des variances et variogrammes (sa seule introduction multipliait les variances par 4). Par prudence, nous avons choisi de l'écartier et l'étude a été reconduite sans lui.

Au total, l'ensemble des tubes "exploitables" se réduit à 39. Ne sachant pas si les tubes situés à proximité immédiate de la route sont à gauche ou à droite de cette dernière, la route est dans la suite considérée de largeur nulle, ce qui n'est pas sans conséquence sur la modélisation.

Par la suite, nous examinons les mesures le long de la route et transversalement, ce qui oblige à regrouper les tubes en 2 ensembles :

1) Les 26 tubes des transects Cernay, Vieux-Thann, Thann, Moosh, Fellinging pour lesquels les distances à la route sont 1m, 2m, 50m, 200m ou 400m.

T1 = {1C, 2C, 3C, 4C, 7VT, 6VT, 5VT, 4VT, 3VT, 2VT, 7T, 6T, 5T, 4T, 2T, 1T, 7M, 6M, 5M, 4M, 3M, 2M, 1M, 2, 1, 4}

2) Les 39 tubes exploitables

$T2 = T1 \cup \{5C, 1VT, 8T, 2W, 1SA, 2SA, 3SA, 5SA, 6SA, 8SA, 3, 1U, 2U\}$

5 Repère de travail

La route est la source des émissions, la distance à la route est une variable prépondérante, il est naturel de se placer dans un nouveau repère (x,y) défini comme suit : l'abscisse "x" est la distance à la route exprimée négativement à gauche de la route et positivement à la droite d'un observateur sur la route qui fait face à la vallée. L'ordonnée curviligne "y" est la distance parcourue sur la route, l'origine étant le point milieu du transect Cernay.

Cette nouvelle régionalisation de la variable NO₂ a nécessité certaines approximations. Tout d'abord, les transects utilisés sont considérés perpendiculaires à la route, ce qui est acceptable pour la plupart des transects sauf à l'évidence Willer, lequel n'est pas informé en distance_route. Cette approximation équivaut à considérer que les données d'un transect ont la même ordonnée curviligne, en conséquence de quoi il est possible d'analyser séparément les comportements le long de la route et transversalement. Pour l'ordonnée curviligne "y" et faute de mieux, nous avons décomposé la route SIG translattée en segments et cumulé leurs longueurs.

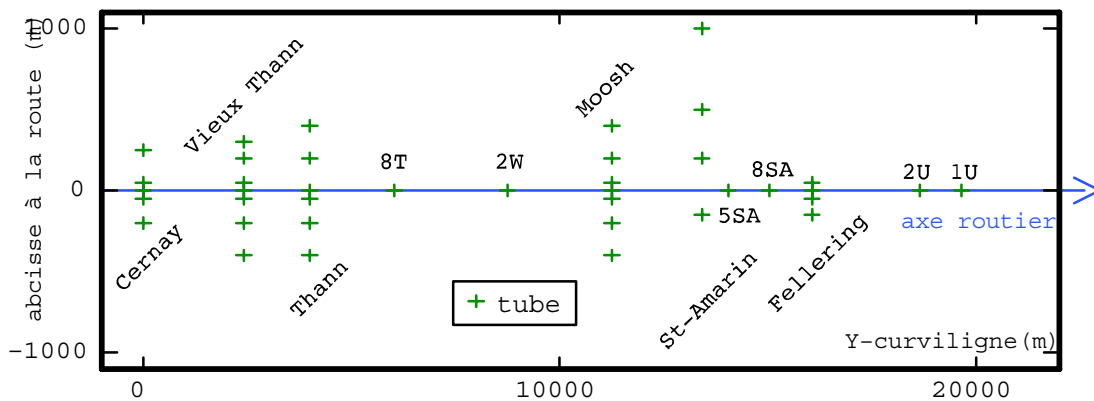


Figure 5 – Les 39 tubes régionalisés dans le repère de travail

Si l'ordonnée curviligne s'impose immédiatement comme la deuxième dimension du problème, dimension qui se doit d'être perpendiculaire à la première pour préserver le caractère cartésien de la régionalisation, elle s'impose aussi par ses corrélations avec la variable altitude, laquelle s'était dégagée de l'ACP comme variable potentiellement explicative. Son coefficient de corrélation est 0.84 pour un nuage relativement bien structuré, à l'exception du tube 2T.

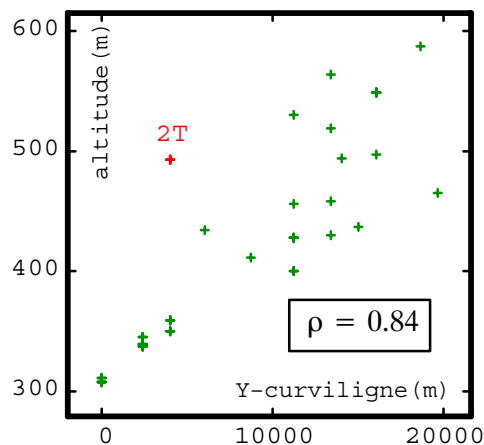


Figure 6 – Nuage de corrélation entre la distance sur la route et l'altitude migrée aux tubes à partir d'une grille de pas 250m

Ce nuage est établi sur les tubes de l'ensemble T2. La dispersion de l'altitude augmente avec "y". Noter que le relief devient plus accidenté quand on s'enfonce dans la vallée.

En passant aux coordonnées (x,y), la potentialité que possède l'altitude d'expliquer le NO₂ moyen est implicitement prise en compte par "y" et l'étude se réduit à l'analyse du NO₂ saisonnier dans cette nouvelle régionalisation. Les ACP qui suivent résument et justifient cette approche.

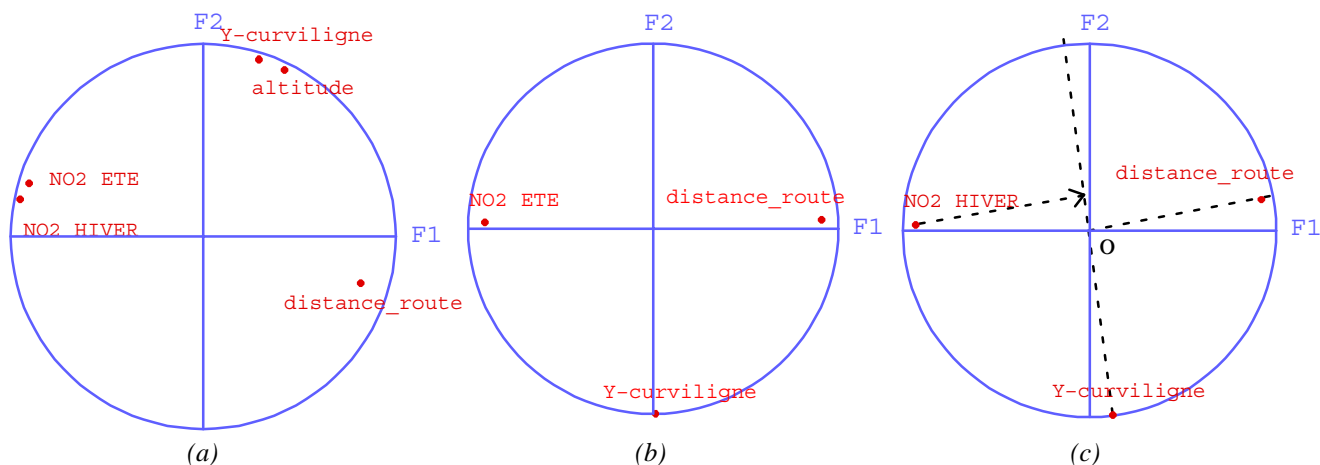


Figure 7 – (a) Projection de toutes les variables retenues dans le plan factoriel
 (b) Dioxyde_été dans le plan factoriel des coordonnées de travail
 (c) Dioxyde_hiver dans le plan factoriel des coordonnées de travail

En (a), on retrouve la proximité de "y" avec l'altitude, et l'éloignement de ces deux variables par rapport au NO₂ saisonnier d'une part, et la distance à la route d'autre part. Les deux variables NO₂ se regroupent dans leur opposition à la distance à la route. En (b), on constate que l'axe F2 est confondu avec "y", et l'axe F1 quasiment confondu avec les deux autres variables. En régionalisant le NO₂ dans le repère (x,y), on fait passer la distance_route d'un statut de variable aléatoire à celui d'une simple coordonnée. F1 et F2 étant portés par ces coordonnées, et F1 étant par construction non corrélé (orthogonal) à F2, on en déduit que 1) aucune information statistique n'est perdue dans ce changement de repère, et 2) aucun biais n'est introduit. En (b) toujours, on voit que la projection du NO₂ sur l'axe F2 est quasiment nulle, alors qu'elle se situe en opposition directe de x. On peut en conséquence s'attendre à une relation fonctionnelle marquée perpendiculairement à la route, et quasi inexistante le long de la route. En (c), et puisque x et y ne sont pas exactement sur F1 et F2, c'est le repère (ox,oy) qu'il faut considérer. Il est quasiment orthogonal, comme en été, avec les mêmes conclusions. Par contre, la projection du NO₂ sur l'axe oy donne une corrélation plus forte en hiver qu'en été. La relation fonctionnelle le long de la route sera en conséquence plus marquée en hiver qu'en été.

Nous serons amené à considérer des moyennes (spatiales) de valeurs de NO₂ qui sont elle-mêmes des moyennes temporelles. Pour ne pas alourdir les notations, la variable d'intérêt s'appelle simplement NO₂(x,y), tout en gardant à l'esprit que chaque valeur a un poids de 6 échantillons (les quinzaines).

Chapitre II – Modélisation de la dérive $m(x,y)$

0 Résumé du chapitre

Nous construisons un modèle de dérive $m(x,y)$ qui rend compte de la non-stationnarité de la variable $NO_2(x,y)$ le long de la route et transversalement. On procède en deux étapes : construction du modèle, puis ajustement.

Construction. On examine les évolutions, à $x=x_i$ fixé, de $NO_2(x_i, y)$, puis à $y=y_j$ fixé, celles de $NO_2(x,y)$. Les évolutions longitudinales conduisent à une fonction linéaire de l'ordonnée curviligne y , les évolutions transversales à une fonction bi-exponentielle de la distance à la route x . On en déduit la forme analytique de $m(x,y)$ et l'on vérifie qu'elle est cohérente avec les observations. La nécessité de cette vérification aura imposé, dès le début, de construire le modèle sur l'ensemble T1 des tubes sur transect.

Ajustement. Les coefficients de la fonction $m(x,y)$ sont dans un premier temps ajustés par moindres-carrés sur T1, puis sur l'ensemble T2 de tous les tubes exploitables. La comparaison des résultats permet de mettre en avant le problème de 6 tubes, dits isolés, dont la moyenne est nettement plus basse que celle des autres tubes situés à la même distance de la route.

Ces opérations sont faites en été puis en hiver. On conclut par une discussion sur les différences constatées entre l'été et l'hiver, et sur l'opportunité ou non d'avoir une dérive qui s'annule à une certaine distance de la route.

1 Symétrie axiale

Perpendiculairement à la route, les moyennes de NO_2 , calculées par classe de distance à la route, évoluent d'une manière sensiblement symétrique, en hiver comme en été.

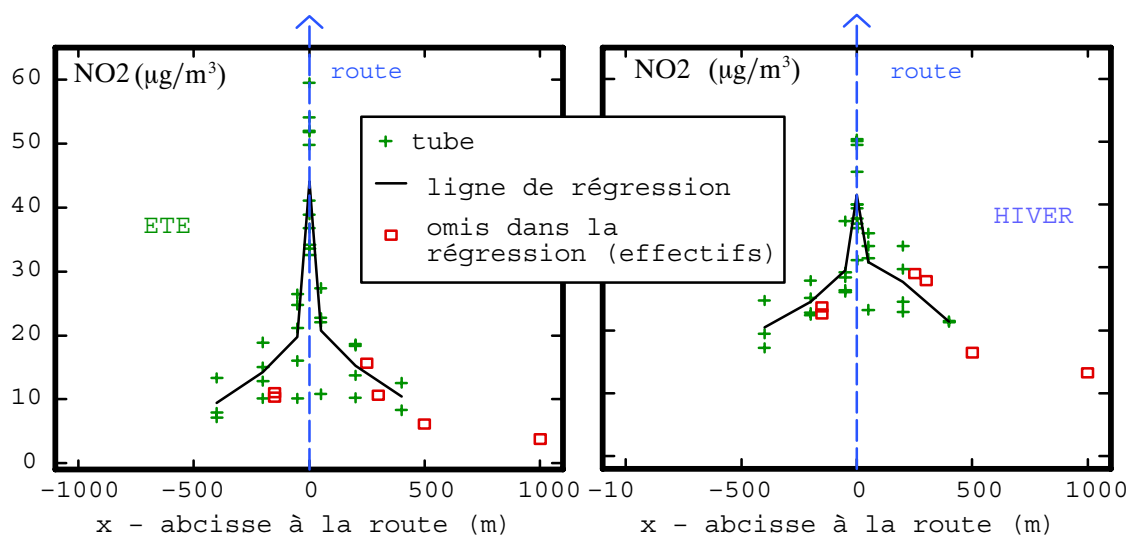


Figure 8 –Projection des valeurs dans un plan perpendiculaire à la route

Cette symétrie est utilisée pour augmenter les effectifs statistiques. Dans la suite, on ne différencie pas ce qui est à droite de la route de ce qui est à gauche, et l'on utilise la "distance à la route", valeur absolue de l'abscisse. L'abscisse à la route ne réapparaît que dans la phase de mise en oeuvre des modèles (chapitre V – Estimations). Pour éviter toute confusion, les formules utilisent l'expression " $|x|$ ".

On désigne par $\{x_i\}$ l'ensemble des distances à la route disponibles et $\{y_j\}$ celui des abscisses curvilignes. Mis à part les tubes 5C, 1VT, 6SA, 1SA et 3, les autres tubes appartiennent aux classes de distance 1.23m, 50m, 200m et 400m, les tubes distants à 2m ayant été associés à ceux de 1m pour obtenir un plus grand effectif statistique, ce qui explique le chiffre 1.23. Nous distinguerons l'ensemble $\{x_i, \text{tous tubes exploitables}\}$ de $\{x_i, i=1,4\}$ où seules les 4 classes de distance citées sont utilisées. Et de même pour $\{y_j, \text{tous tubes exploitables}\}$ à distinguer de $\{y_j, j=1,5\}$ où sont considérées les seules ordonnées curvilignes des transects Cernay, Vieux-Thann, Thann, Moosh et Fellingring. Avec ces notations, l'ensemble T1 est l'ensemble des tubes de coordonnées (x_i, y_j) avec $\{x_i, i=1,4\}$ et $\{y_j, j=1,5\}$.

On commence par l'été.

2 Construction de la dérive

On examine $NO_2(x,y)$ en y puis en x pour identifier des fonctions susceptibles de représenter les tendances. On se limite pour l'instant à l'ensemble T1 des 26 tubes sur transect pour être sûr d'utiliser la même information dans les deux axes.

2-1 Comportement en y , ordonnée curviligne

L'évolution de $NO_2(x,y)$ le long de la route pour $x \in \{x_i, i=1,4\}$ et $y \in \{y_j, j=1,5\}$ peut être vue comme linéaire décroissante quand on s'enfonce dans la vallée. En conséquence, on regarde si le modèle

$$m(x_i, y) = a(x_i) y + b(x_i) \quad (1)$$

convient et l'on effectue des régressions linéaires pour chaque x_i .

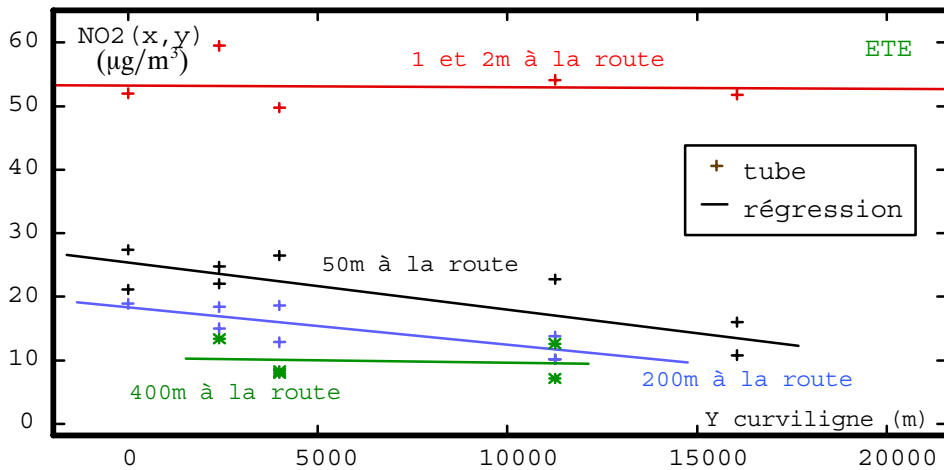


Figure 9 – Régressions linéaires par classe de distance à la route

Les échantillons sont peu nombreux, mais il faut bien s'en contenter. Les coefficients $a(x_i)$ et $b(x_i)$ sont

ÉTÉ	$a(x_i)$	$b(x_i)$
$x_1 = 1 \text{ m}$	$-1.04 \cdot 10^{-4}$	54.14
$x_2 = 50 \text{ m}$	$-7.44 \cdot 10^{-4}$	28.28
$x_3 = 200 \text{ m}$	$-5.87 \cdot 10^{-4}$	18.32
$x_4 = 400 \text{ m}$	$-0.75 \cdot 10^{-4}$	10.36

Tableau 1 – Pentes $a(x)$ et constantes $b(x)$ des régressions linéaires par classe de distance à la route

2-2 Comportement en x , distance à la route

Pour chaque transect, $NO_2(x,y)$ décroît fortement quand on s'éloigne de la route.

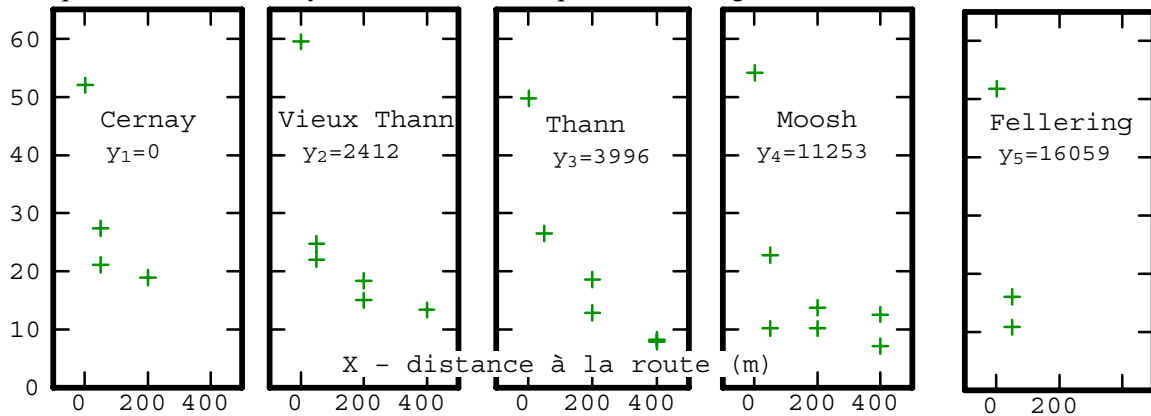


Figure 10 – Valeurs du dioxyde_été aux transects

Il faut donner une forme analytique à cette décroissance. Pour choisir quelles fonctions utiliser, on s'appuie sur un comportement transversal moyen. Pour chaque $x \in \{x_i, i=1,4\}$, on peut être tenté de calculer la moyenne de $\text{NO}_2(x_i, y_j)$ prise sur l'ensemble des y_j associés, mais les moyennes ainsi obtenues ne sont pas comparables entre elles car elles ne sont pas calculées chaque fois avec les mêmes effectifs (pour certains transects, certains x_i ne sont pas représentés de part et d'autre de la route, voire pas du tout). Comme le montre la figure suivante, chacune de ces moyennes associée à x_i doit l'être également à une ordonnée curviligne égale à la moyenne des y_j impliqués dans le calcul. Comme cette moyenne dépend de x_i , on adopte la notation

$$\bar{y}_i = \overline{y(x_i)} = \frac{1}{n(x_i)} \sum_{j=1}^{n(x_i)} y_j$$

avec $n(x_i)$ = nombre d'échantillons situés à la distance x_i de la route.

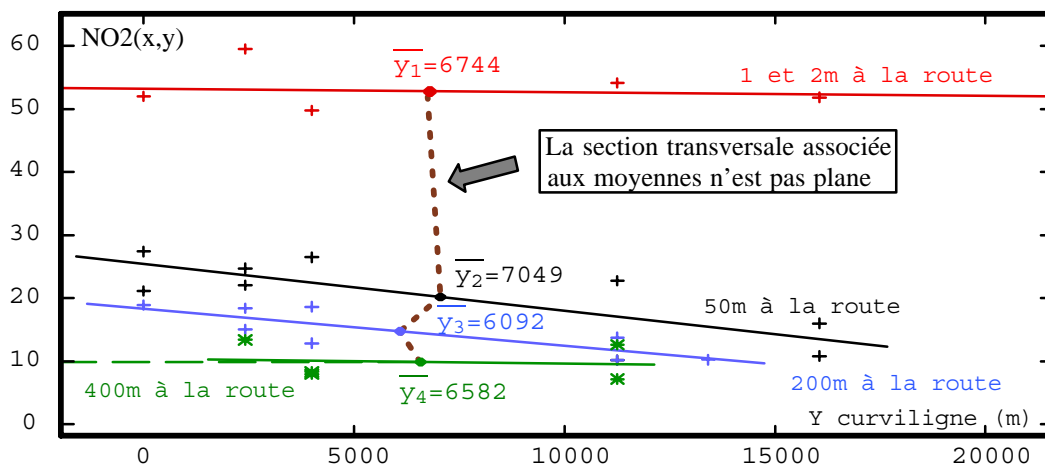


Figure 11 – Régressions linéaires par classe de distance à la route (été). Les moyennes du dioxyde par classe de distance à la route sont associées aux moyennes des ordonnées curvilignes

Sur la figure, la ligne brisée représente l'échantillonnage transversal associé à ces moyennes. Comme ce n'est pas une droite verticale, une correction est nécessaire, qui permettra de comparer ces moyennes entre elles. À défaut de mieux, nous avons décidé de nous fier aux régressions en recalculant chacune des moyennes selon la formule (1) et pour $y=7049$. Dans ce cas d'étude, les corrections sont minimales car les pentes sont faibles, mais elles peuvent être importantes dans d'autres cas. À 1m, la valeur passe de 53.44 à 53.41, à 50m, la valeur de la moyenne ne change pas, pour 200m, elle passe de 14.74 à 14.18, et pour 400m de 9.87 à 9.83.

Pour trouver une fonction de x qui ajuste au mieux le comportement transversal moyen, on opère par moindres-carrés. Du fait de la correction appliquée sur les moyennes, cette minimisation doit se faire sur les moyennes corrigées pondérées par le nombre d'échantillons utilisés dans chaque moyenne, et non directement sur les échantillons eux-même.

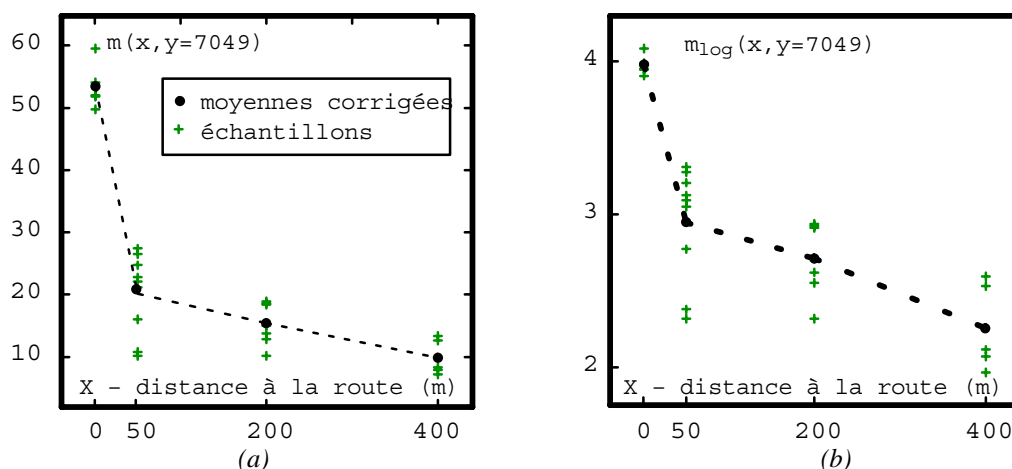


Figure 12 – Moyennes du dioxyde par classe de distance à la route (été). (a) Moyenne corrigées pour se ramener à l'ordonnée curviligne 7049m. (b) Moyennes des logarithmes du dioxyde

Sauf à utiliser des degrés élevés, une fonction polynomiale ne peut expliquer une telle décroissance. Après divers essais, nous nous sommes arrêté sur les fonctions de type $e^{-\frac{|x|}{p}}$. Si la largeur de la route était connue, de même que la position à gauche ou à droite des tubes donnés à 1 et 2m de la route, nous aurions probablement utilisé des fonctions gaussiennes : leur tangente horizontale en $x=0$ donne une “bulle” au dessus de la route (bulle qui pourrait être calibrée sur la largeur de la route) alors que les exponentielles sont en “pointe” (route de largeur nulle). Une seule exponentielle suffit-elle ? La figure 12–(b), qui représente les moyennes de $\log[\text{NO}_2(x,y)]$, montre que non : les moyennes des logarithmes ne sont pas alignées. La décroissance entre 1m et 50m est telle qu’une première fonction est indispensable pour ajuster le passage de 1m à 50m, et il en faut une autre qui ajuste les derniers points.

L’optimalité conduit à l’ensemble $\{ e^{-\frac{|x|}{18}}, e^{-\frac{|x|}{550}} \}$ pour un modèle qui devient

$$m(x,y_j) = c_1(y_j) e^{-\frac{|x|}{18}} + c_2(y_j) e^{-\frac{|x|}{550}} \quad (2)$$

L’influence de la première exponentielle s’estompe à partir de 50m et celle de la deuxième exponentielle au delà de 1600m. La figure 13 montre l’ajustement obtenu avec $c_1(7049)=35.79$ et $c_2(7049)=20$. Ces coefficients ne serviront pas ultérieurement car à ce stade, on se borne à constater que deux fonctions de type $e^{-\frac{|x|}{p}}$, avec $p=18$ et $p=550$, permettent de rendre compte du comportement transversal moyen.

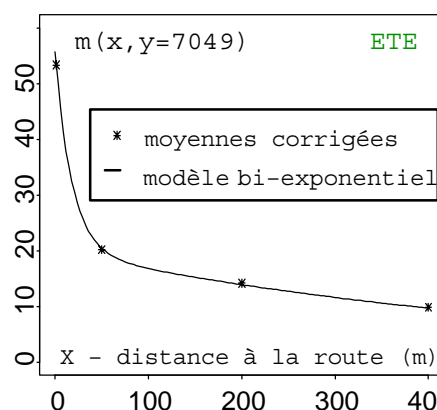


Figure 13 – Ajustement, des moyennes corrigées, par le modèle bi-exponentiel

Les coefficients multiplicatifs $c_1(y_j)$ et $c_2(y_j)$ de l’expression (2) sont ajustés par moindres-carrés sur chacun des 5 transects de l’ensemble T1.

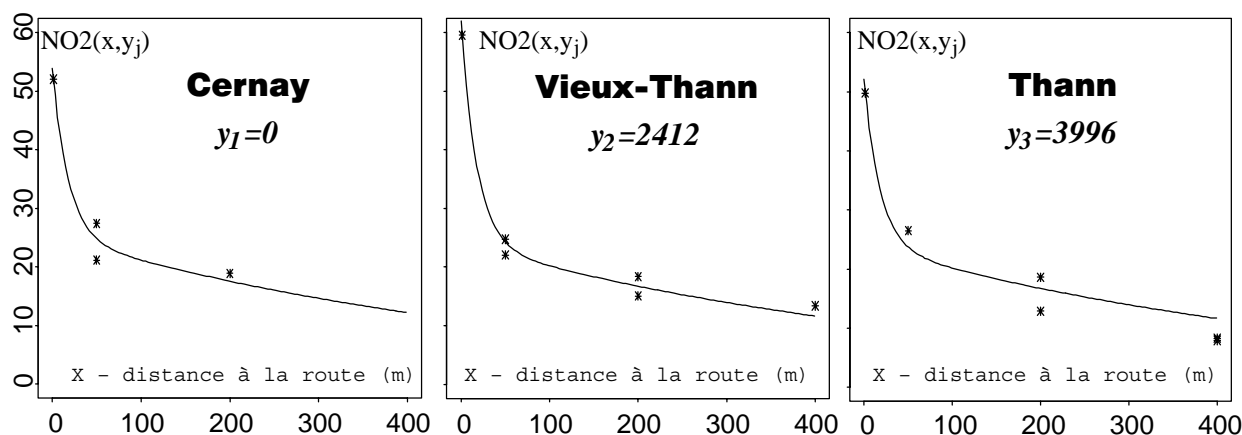


Figure 14-1 – Régressions bi-exponentielles à ordonnée curviligne fixée (été)

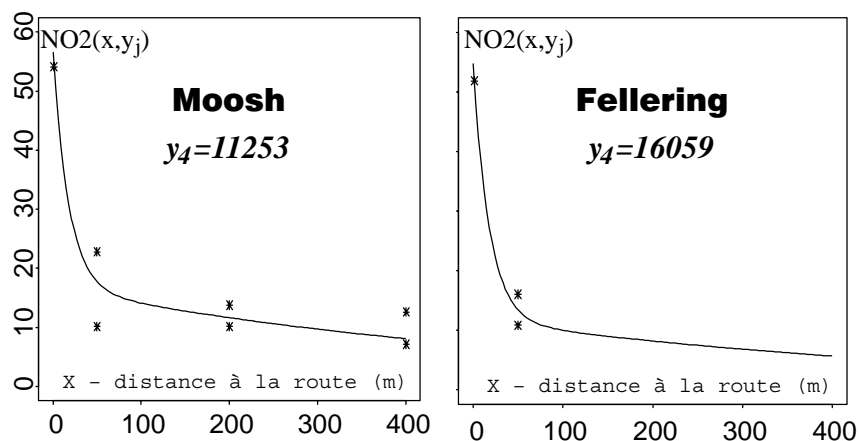


Figure 14-2 – Régressions bi-exponentielles à ordonnée curviligne fixée (été)

	$c_1(y_j)$	$c_2(y_j)$
Cernay ($y_1=0$)	28.46	25.24
Vieux-Thann ($y_2=2412$)	37.85	24.07
Thann ($y_3=3996$)	28.00	24.07
Moosh ($y_4=11253$)	39.89	16.73
Fellingering ($y_5=16059$)	42.92	11.76

Tableau 2 – Coefficients obtenus pour chaque ordonnée curviligne (été).
 $c_1()$ concerne l'exponentielle de paramètre 18, $c_2()$ celle de paramètre 550

2-3 Expression de $m(x,y)$

- Le comportement du NO_2 le long de la route conduit au modèle linéaire

$$m(x_i,y) = a(x_i) y + b(x_i) \quad (1\text{-rappel})$$

et nous avons calculé les coefficients $a(x_i)$ et $b(x_i)$ pour $x_i = 1.23, 50, 200$ et 400m , les seules distances ayant un effectif statistique suffisant.

- Le comportement du NO_2 perpendiculairement à la route a conduit au modèle bi-exponentiel

$$m(x,y_j) = c_1(y_j) e^{-\frac{|x|}{18}} + c_2(y_j) e^{-\frac{|x|}{550}} \quad (2\text{-rappel})$$

et nous avons calculé les coefficients $c_1(y_j)$ et $c_2(y_j)$ pour les ordonnées curvilignes y_j associées aux 5 transects de l'ensemble T1.

Les expressions (1) et (2) représentent des sections en x et en y d'une surface $m(x,y)$ dont il faut trouver l'expression analytique.

Examinons l'évolution, le long de la route, des coefficients $c_1(y_j)$ et $c_2(y_j)$. (on aurait pu tout aussi bien commencer par l'évolution en x des coefficients $a(x_i)$ et $b(x_i)$)

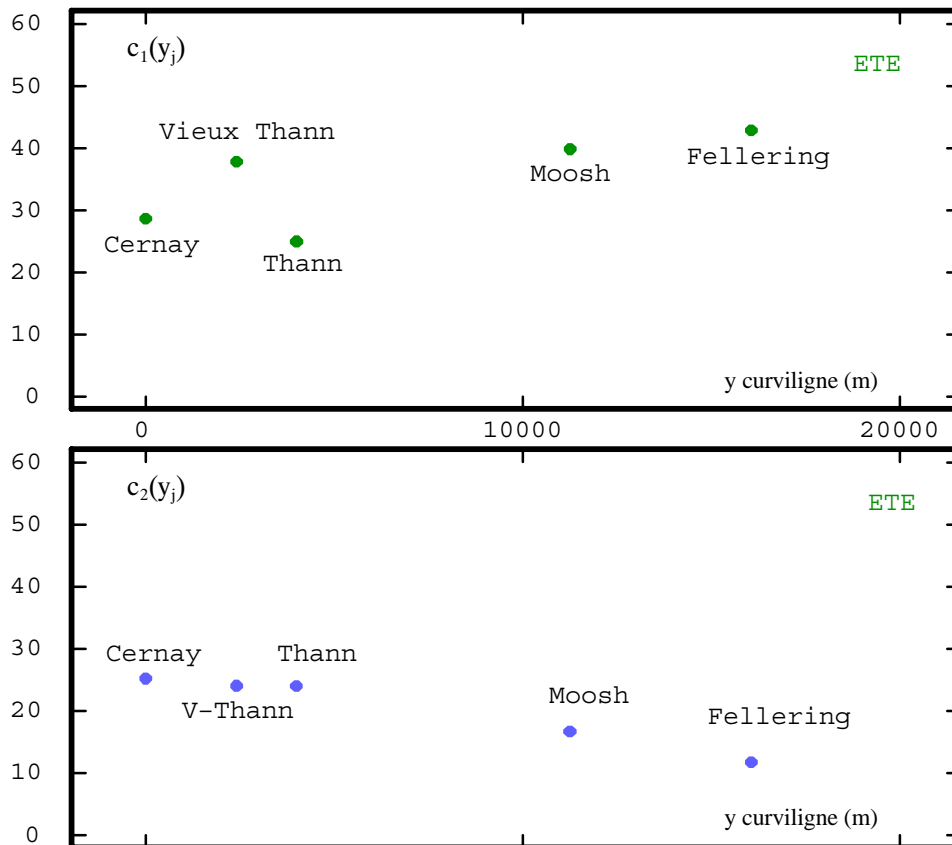


Figure 15 – Coefficients obtenus pour chaque ordonnée curviligne. $c_1()$ concerne l'exponentielle de paramètre 18, $c_2()$ celle de paramètre 550

L'alignement des points $(y_j, c_2(y_j))$ est marqué ; il l'est moins pour les points $(y_j, c_1(y_j))$. Nous choisissons la simplicité et posons

$$c_1(y) = c_1^1 y + c_1^2$$

$$c_2(y) = c_2^1 y + c_2^2$$

ce qui donnerait, pour l'expression de la dérive

$$m(x,y) = (c_1^1 e^{-\frac{|x|}{18}} + c_2^1 e^{-\frac{|x|}{550}}) y + (c_1^2 e^{-\frac{|x|}{18}} + c_2^2 e^{-\frac{|x|}{550}})$$

Ce modèle est-il cohérent avec les valeurs expérimentales $a(x_i)$ et $b(x_i)$ obtenues dans le paragraphe 2-1? Pour le vérifier, procédons en deux étapes :

- 1) Calculer les coefficients c_1^1 , c_1^2 , c_2^1 , et c_2^2 et vérifier que la régression linéaire convient
- 2) Pour $x \in \{x_i, i=1,4\}$, vérifier que

$$(c_1^1 e^{-\frac{|x|}{18}} + c_2^1 e^{-\frac{|x|}{550}}) \text{ interpole correctement } a(x) \text{ et}$$

$$(c_2^2 e^{-\frac{|x|}{18}} + c_2^2 e^{-\frac{|x|}{550}}) \text{ interpole correctement } b(x)$$

(identification avec (1) de l'expression de la dérive)

Étape 1 : Par régression linéaire sur les valeurs expérimentales $c_1(y_j)$ et $c_2(y_j)$, on obtient

c_1^1	c_1^2	c_2^1	c_2^2
0.000860	29.06	-0.000870	26.23

Tableau 3 – Coefficients obtenus par régression linéaire sur $c_1(y_j)$ et $c_2(y_j)$.

c_1^1 et c_1^2 sont la pente et la constante associées à $c_1(y)$

c_2^1 et c_2^2 sont la pente et la constante associées à $c_2(y)$

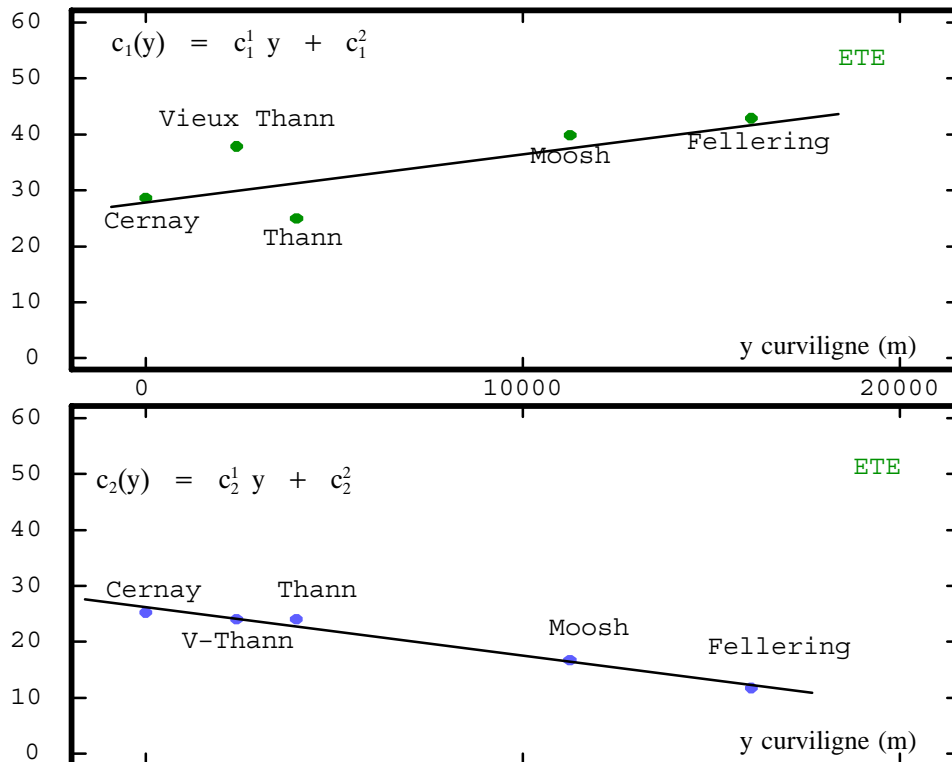


Figure 16 – Régressions linéaires sur $c_1(y_j)$ et $c_2(y_j)$.
 $c_1(y)$ concerne l'exponentielle de paramètre 18, $c_2(y)$ celle de paramètre 550

Étape 2 : on confronte la fonction $c_1^1 e^{-\frac{|x|}{18}} + c_1^2 e^{-\frac{|x|}{550}}$ aux valeurs expérimentales $a(x_i)$. De même pour $c_2^1 e^{-\frac{|x|}{18}} + c_2^2 e^{-\frac{|x|}{550}}$ avec les valeurs expérimentales $b(x_i)$.

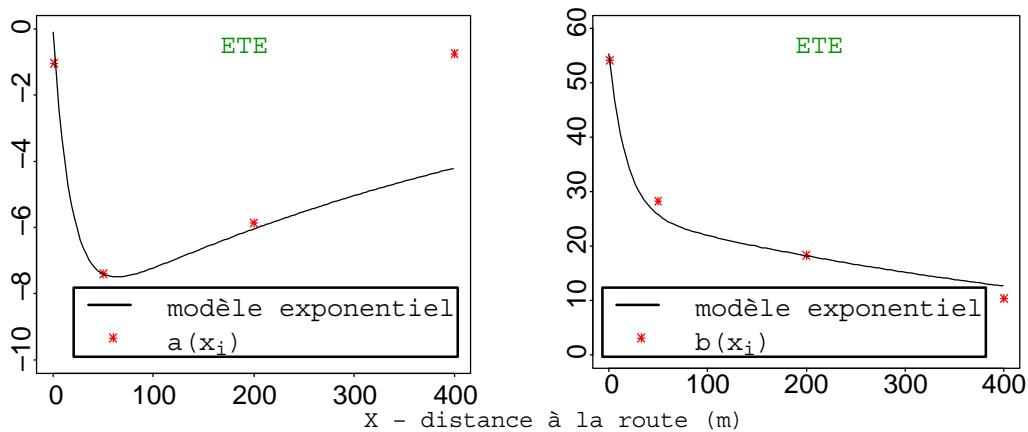


Figure 17 – Modèle bi-exponentiel issu des régressions transversales, confronté aux pentes $a(x_i)$ et constantes $b(x_i)$ des régressions linéaires longitudinales

La cohérence est bonne partout sauf pour a(400m). Cette pente est calculée sur les 3 transects Vieux–Thann, Thann, Moosh. Elle fait intervenir des ordonnées curvilignes $\in [2412\text{m}, 11253\text{m}]$ alors que les autres pentes sont établies sur $[0, 16059\text{m}]$. Calculée sur une longueur de 9 km, la pente a(400) n’est probablement pas généralisable aux 16 km couverts par les autres pentes.

2–4 Synthèse : $m(x,y)$ sur T1

Simplifions les notations en posant

$$m(x,y) = c_1 e^{-\frac{|x|}{18}} + c_2 e^{-\frac{|x|}{550}} + c_3 y e^{-\frac{|x|}{18}} + c_4 y e^{-\frac{|x|}{550}} \quad (3\text{--}été)$$

On vient de montrer que cette expression analytique de la dérive rend compte de ce qui est observé le long de la route et transversalement. On peut en conséquence calculer les coefficients c_i par **régression globale** sur les 26 tubes, en une seule opération. Mais à ce stade, il est peut-être utile de donner plus de flexibilité au modèle en lui ajoutant une constante c_0 :

$$m(x,y) = c_0 + c_1 e^{-\frac{|x|}{18}} + c_2 e^{-\frac{|x|}{550}} + c_3 y e^{-\frac{|x|}{18}} + c_4 y e^{-\frac{|x|}{550}} \quad (4\text{--}été)$$

L’intérêt de cette constante est qu’elle nous permet d’envisager une dérive qui ne s’annule pas nécessairement quand x tend vers l’infini. (ce point est discuté plus loin). Dans la phase d’identification des fonctions exponentielles (paragraphe 2–2), il n’était pas possible d’introduire cette constante : nous disposons de seulement 4 moyennes corrigées, et nous avons 4 paramètres à ajuster (2 coefficients multiplicatifs, 2 portées pour les exponentielles).

Le tableau suivant résume les coefficients obtenus quand on utilise (3) ou (4). Les résultats sont comparés à ceux obtenus en procédant par sections.

ETE/T1	c_0	c_1	c_2	c_3	c_4
(1) ajustement par sections		29.06	26.23	$8.60 \cdot 10^{-4}$	$-8.70 \cdot 10^{-4}$
(2) ajustement global sans c_0		29.98	25.83	$7.66 \cdot 10^{-4}$	$-8.34 \cdot 10^{-4}$
(3) ajustement global avec c_0	0.685	30.23	24.92	$7.63 \cdot 10^{-4}$	$-8.32 \cdot 10^{-4}$

Tableau 4 – Coefficients de la dérive selon trois approches. (1) Approche par sections transversale et longitudinale ayant permis de construire le modèle. (2) et (3) Coefficients obtenus en une seule régression sur les 26 tubes de T1

La valeur de c_0 étant faible, les autres coefficients diffèrent peu si on utilise ou pas cette constante dans l’approche globale. Comparativement à l’approche par sections, les variations des coefficients sont plus importantes (surtout pour c_3), mais demeurent mesurées. On peut trouver dans cette stabilité une justification *à fortiori* de la méthode utilisée pour construire la dérive : ce qui est optimal en voisinage glissant (approche par sections) ne l’est pas nécessairement en voisinage unique (approche globale).

2-5 Remarques

- On se réfère à l'expression $m(x,y) = c_1(y) e^{-\frac{|x|}{18}} + c_2(y) e^{-\frac{|x|}{550}}$. Sur les figures suivantes sont représentés en vert les coefficients c_1 et c_2 calculés sur les transects, et en rouge les coefficients calculés sur les moyennes corrigées (paragraphe 2-2, phase d'identification des fonctions à utiliser pour modéliser $m(x,y)$). Ces coefficients sont associés à l'ordonnée curviligne 7049 m choisie pour les corrections.

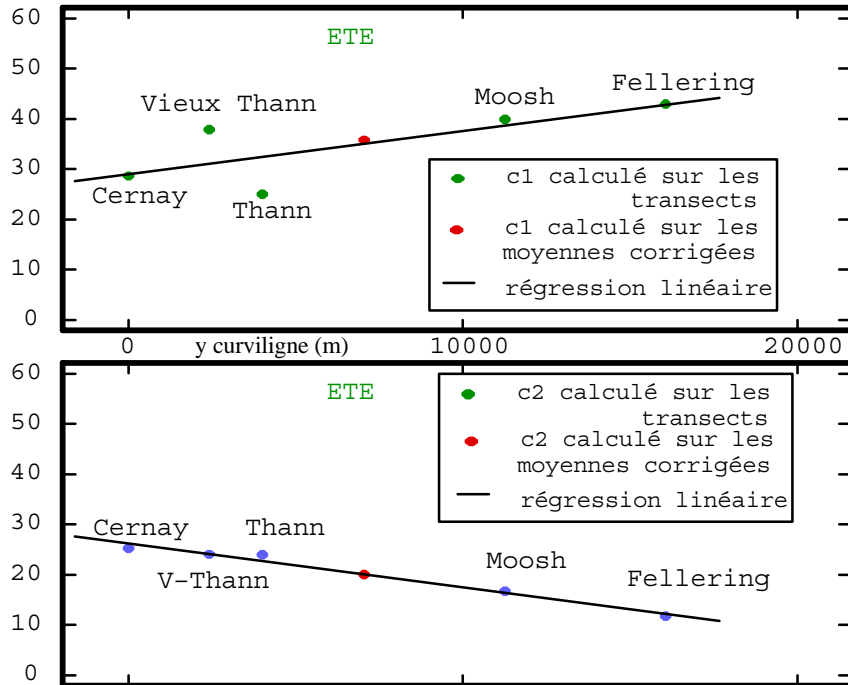


Figure 18 – Régressions linéaires sur $c_1(y)$ et $c_2(y)$. $c_1(y)$ concerne l'exponentielle de paramètre 18, $c_2(y)$ celle de paramètre 550. En rouge, les coefficients associés au modèle transversal moyen

Les points rouges s'inscrivent dans la parfaite continuité de la ligne de régression calculée sur les points verts. Ce constat est une autre manière d'exprimer la cohérence de la démarche adoptée pour construire $m(x,y)$. Si, au lieu de choisir de positionner les corrections des moyennes à $y=7049$ m, nous avions choisi de les positionner à $y=6744$ m, barycentre des transects utilisés, nous pourrions alors dire que les coefficients établis sur les moyennes sont les moyennes des coefficients calculés sur chaque transect.

- Sur la figure suivante sont représentées les évolutions, le long de la route, de deux rapports:

$$\frac{c_1(y_j)}{c_1(y_j) + c_2(y_j)}$$

qui représente l'influence de c_1 par rapport à c_1+c_2

$$\frac{c_2(y_j)}{c_1(y_j) + c_2(y_j)}$$

qui représente l'influence de c_2 par rapport à c_1+c_2 .

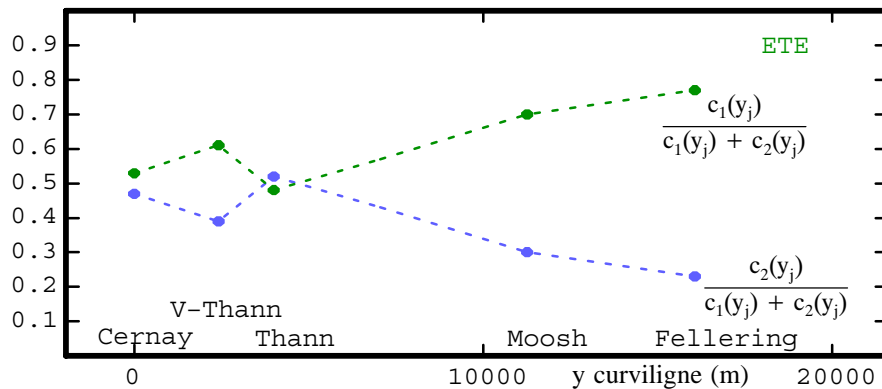


Figure 19 – Importances relatives de $c_1(y)$ et $c_2(y)$ par rapport à leur somme. $c_1(y)$ concerne l'exponentielle de paramètre 18, $c_2(y)$ celle de paramètre 550

Dans l'expression $m(x,y) = c_1(y_j) e^{-\frac{|x|}{18}} + c_2(y_j) e^{-\frac{|x|}{550}}$, le coefficient c_1 a en moyenne une importance qui augmente quand on s'enfonce dans la vallée, et celle de c_2 décroît. Cette remarque donne une idée des déformations de la dérive. En considérant que le comportement parallèlement à la route est linéaire, on donne aux modèles (3) et (4) le degré de liberté qui permet de rendre compte de ces déformations. C'est en cela, sans doute, que réside l'importance de la modélisation longitudinale linéaire.

3 m(x,y) sur T2

Établie sur les seules distances 1m, 50m, 200m et 400m et sur les seuls transects cités, $m(x,y)$ est-elle représentative du comportement moyen des 39 tubes exploitables ?

Non, essentiellement parce que T2 contient 6 tubes situés au bord de la route qui n'appartiennent à aucun transect. On a reconduit des régressions linéaires sur tous les tubes proches de la route.

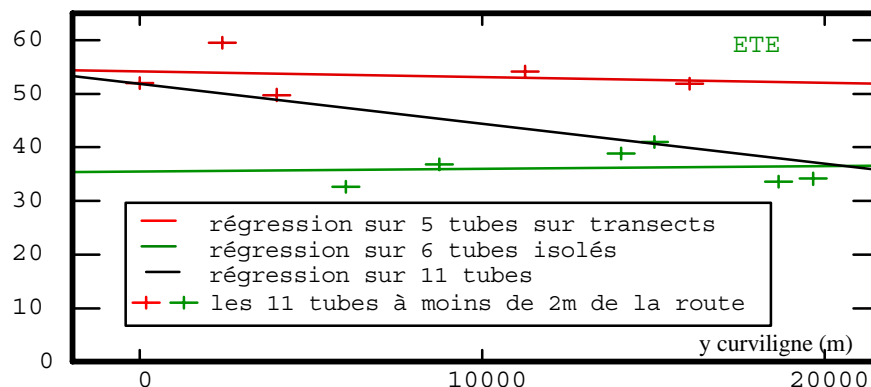


Figure 20 – Influence des tubes à moins de 2m de la route et isolés (non sur transect)

La moyenne des 6 tubes isolés est plus basse que celle des 5 tubes sur transect, les deux premiers tubes “isolés”(i.e. non sur transect) sont informés bien après les premiers tubes sur transect, les deux derniers tubes isolés sont informés bien après les derniers tubes sur transect. Cette configuration influence considérablement les régressions linéaires, comme le montre le tableau suivant.

ÉTÉ	$a(x_i)$	$b(x_i)$
5 tubes sur transect	$-1.04 \cdot 10^{-4}$	54.14
6 tubes isolés	$0.46 \cdot 10^{-4}$	35.53
les 11 tubes	$-7.44 \cdot 10^{-4}$	51.85

Tableau 5 – Pentas $a(x)$ et constantes $b(x)$ des régressions linéaires longitudinales sur les tubes situés à moins de 2m de la route

Deux chiffres ressortent de ces régressions : $-7.44 \cdot 10^{-4}$, la pente associée aux 11 tubes, et 35.53, niveau moyen des tubes isolés. Pourquoi est-il si bas ? Ces tubes sont-ils en zone non urbaine ? Et donc non soumis à des émissions non routières qui conditionneraient les autres tubes ? Ou faut-il voir, en regard de l'ordonnée curviligne, une alternance cyclique à modéliser, non linéairement, mais par une fonction périodique ? Ou encore, ces tubes ne montrent-ils pas que la variance des moyennes saisonnières de NO_2 sur la route est plus importante que ne le laisseraient supposer les transects ? Les autres tubes qui appartiennent à T2, mais pas à T1, sont 5C (250m de la route), 1VT(300m), 4 tubes du transect Saint-Amarin et 3 (150m). Les tests montrent qu'ils sont d'une influence négligeable sur les régressions.

On a reconduit le calcul des coefficients de la formule (4) par régression sur les 39 tubes de T2. Dans le tableau qui suit, les résultats sont comparés à ceux obtenus par régression globale sur les tubes de T1.

ETE	c_0	c_1	c_2	c_3	c_4
(1) sur T2	0.338	28.23	24.97	$0.66 \cdot 10^{-4}$	$-7.70 \cdot 10^{-4}$
(2) sur T1	0.685	30.23	24.92	$7.63 \cdot 10^{-4}$	$-8.32 \cdot 10^{-4}$

Tableau 6 – Coefficients de la dérive obtenus par régression globale sur les deux ensembles de tubes.
(1) Régression globale sur T2, ensemble des 39 tubes exploitables. (2) Régression globale sur T1, ensemble des 26 tubes sur transect (on rappelle que T2 contient T1)

La constante c_0 est négligeable, il se confirme donc que l'introduction de ce paramètre est d'une importance secondaire en été. Comparativement aux coefficients obtenus sur T1, c'est au niveau de c_3 que la différence est plus grande. Quand on utilise les valeurs de ces coefficients dans la formule (4) et pour $x=1$, on obtient l'équation $m(1,y) = 51.97 - 7.05 \cdot 10^{-4} y$, une formule pratiquement identique à celle obtenue par régression linéaire en y des 11 tubes ($51.85 - 7.4 \cdot 10^{-4} y$). Les différences sont essentiellement dues aux tubes isolés.

4 Hiver

Comme pour l'été, on recherche sur T1 les fonctions de forme qui conviennent puis on vérifie qu'une approche linéaire en y et bi-exponentielle en x est cohérente, suite à quoi la dérive est ajustée par moindres-carrés sur T2.

4-1 Construction

La fonction $c_1(y)e^{-\frac{|x|}{14}} + c_2(y)e^{-\frac{|x|}{940}}$ est retenue pour rendre compte d'une dérive qui, transversalement, s'étend plus loin en hiver qu'en été. (L'ajustement donne $c_1(7049)=17$ et $c_2(7049)=31.5$).

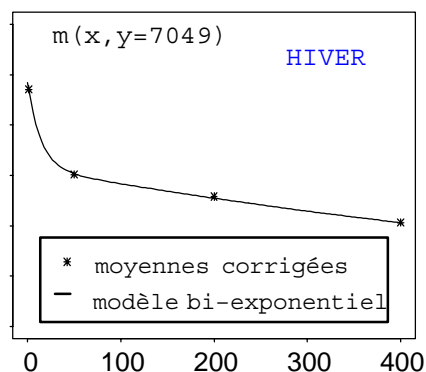


Figure 21 – Ajustement des moyennes corrigées par le modèle bi-exponentiel

Comme pour l'été, l'influence relative des coefficients de ces nouvelles exponentielles, ajustés sur chacun des 5 transects, évolue quand on s'enfonce dans la vallée

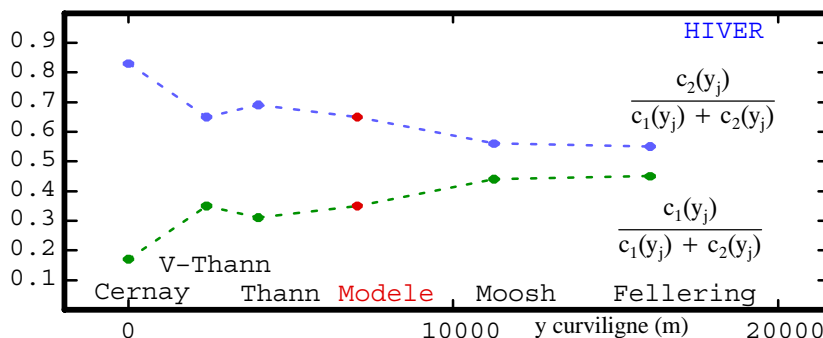


Figure 22 – Importances relatives de $c_1(y)$ et $c_2(y)$ par rapport à leur somme. $c_1(y)$ concerne l'exponentielle de paramètre 14, $c_2(y)$ celle de paramètre 940

La cohérence entre l'ensemble $\{a(x_i)$ et $b(x_i)\}$ des régressions linéaires parallèlement à la route, et l'ensemble $\{c_1(y_j), c_2(y_j)\}$ des régressions bi-exponentielles transversales est meilleure qu'en été, surtout pour $a(400)$.

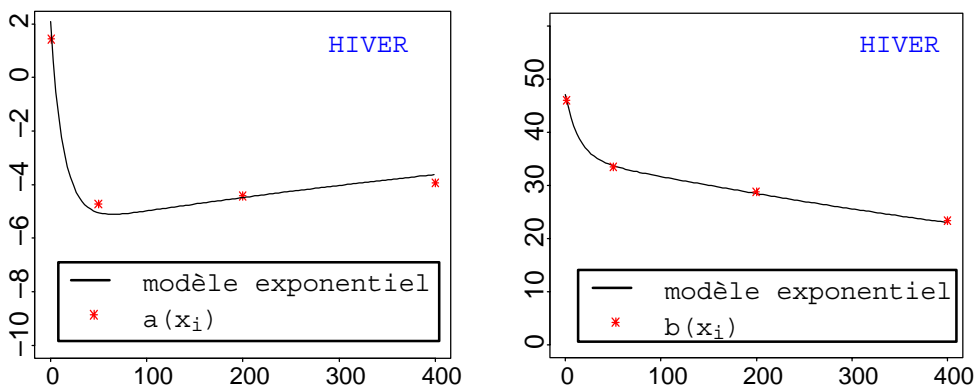


Figure 23 – Modèle bi-exponentiel issu des régressions transversales, confronté aux pentes $a(x)$ et constante $b(x)$ des régressions linéaires longitudinales

Obtenu par sections longitudinales et transversales, un premier modèle de dérive est établi sur T1

$$m(x,y) = c_1 e^{-\frac{|x|}{14}} + c_2 e^{-\frac{|x|}{940}} + c_3 y e^{-\frac{|x|}{14}} + c_4 y e^{-\frac{|x|}{940}} \quad (3\text{-hiver})$$

avec les coefficients suivants

c_1	c_2	c_3	c_4
11.89	35.23	$7.63 \cdot 10^{-4}$	$-5.55 \cdot 10^{-4}$

Tableau 7 – Coefficients obtenus par régression linéaire sur les valeurs $c_1(y_j)$ et $c_2(y_j)$ issues des régressions bi-exponentielles sur transect

4-2 synthèse : $m(x,y)$ sur T1

Le modèle (3) est ajusté globalement sur T1, ainsi que le modèle (4) suivant, qui diffère du précédent par l'introduction d'une constante.

$$m(x,y) = c_0 + c_1 e^{-\frac{|x|}{14}} + c_2 e^{-\frac{|x|}{940}} + c_3 y e^{-\frac{|x|}{14}} + c_4 y e^{-\frac{|x|}{940}} \quad (4\text{-hiver})$$

HIVER/T1	c_0	c_1	c_2	c_3	c_4
(1) ajustement par sections		11.89	35.23	$7.63 \cdot 10^{-4}$	$-5.55 \cdot 10^{-4}$
(2) ajustement global sans c_0		11.59	35.29	$7.35 \cdot 10^{-4}$	$-5.40 \cdot 10^{-4}$
(3) ajustement global avec c_0	0.95	11.80	34.15	$7.33 \cdot 10^{-4}$	$-5.39 \cdot 10^{-4}$

Tableau 8 – Coefficients de la dérive selon trois approches. (1) Approche par sections transversale et longitudinale ayant permis de construire le modèle. (2) et (3) Coefficients obtenus en une seule régression sur les 26 tubes de T1

Les valeurs diffèrent peu selon les approches. Noter que l'introduction de la constante n'apporte pas d'amélioration significative : la somme des carrés des résidus (RSS) passe de 265 (sans la constante) à 264 (avec la constante).

4-3 $m(x,y)$ sur T2

Le problème des 6 tubes isolés existe également en hiver

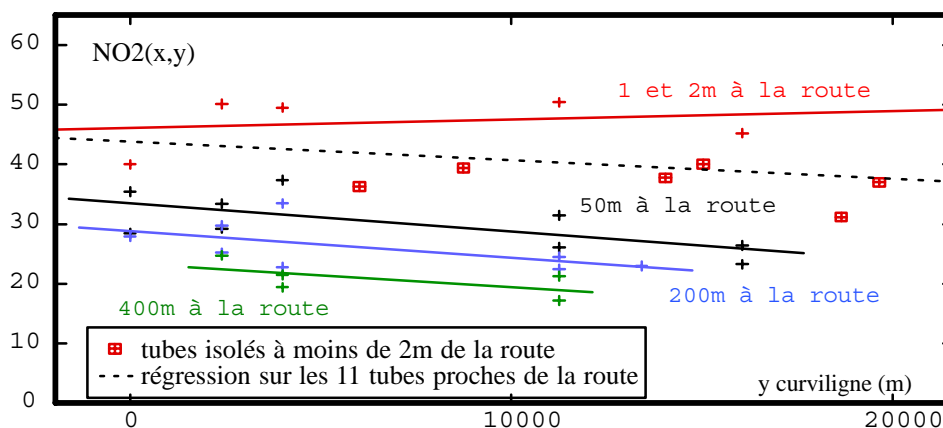


Figure 24 – Régressions linéaires par classe de distance à la route (Hiver). En pointillé, la ligne de régression calculée sur les 11 tubes proches de la route

De positive lorsqu'elle est établie sur les 5 tubes des transects, la droite de régression devient négative quand elle incorpore les 6 tubes isolés.

La régression globale du modèle (4) sur T2 donne les coefficients suivants.

HIVER	c_0	c_1	c_2	c_3	c_4
(1) sur T2	3.52	11.93	31.46	$1.78 \cdot 10^{-4}$	$-5.64 \cdot 10^{-4}$
(2) sur T1	0.95	11.80	34.15	$7.33 \cdot 10^{-4}$	$-5.39 \cdot 10^{-4}$

Tableau 9 – Coefficients de la dérive par régression globale sur les deux ensembles de tubes. (1) Régression globale sur T2, ensemble des 39 tubes exploitables. (2) Régression globale sur T1, ensemble des 26 tubes sur transect

Comparativement au modèle établi sur T1, le coefficient c_0 a pris de l'importance et l'influence des tubes isolés s'est essentiellement portée sur c_3 .

5 Histogrammes des résidus

À ce stade, pour juger de la qualité des modèles de dérive, il faut vérifier qu'ils assument convenablement leur rôle de "comportement moyen autour duquel les mesures sont censées fluctuer". Il faut donc examiner la moyenne des résidus

$$R(x,y) = NO_2(x,y) - m(x,y) \quad (5)$$

et leurs histogrammes. La notation $\overline{R(x,y)}$ désigne la moyenne des résidus.

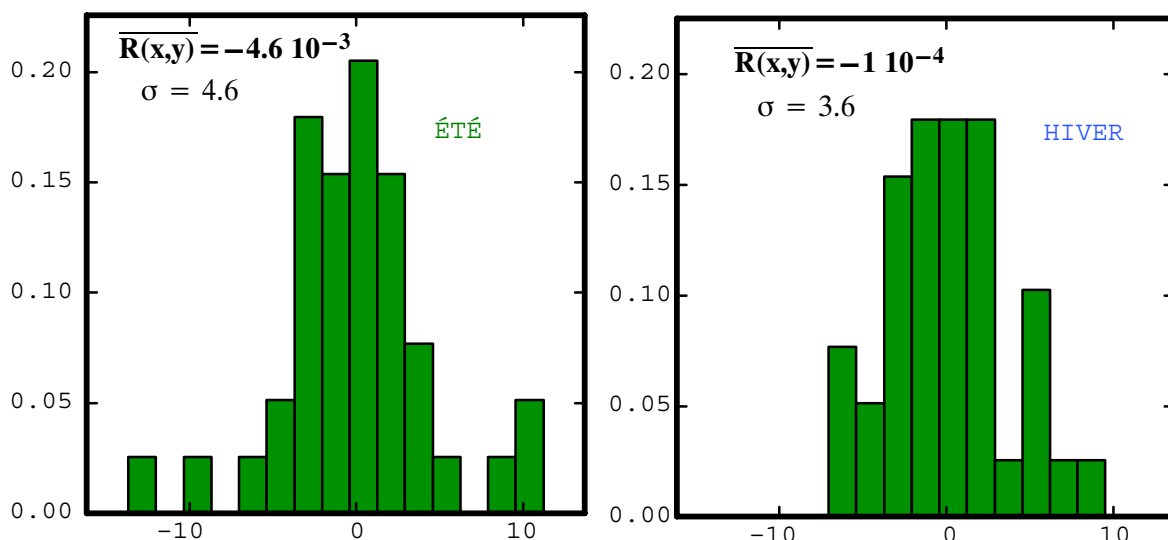


Figure 25 – Histogrammes des résidus du NO_2 centré sur la moyenne modélisée

Les résidus sont de moyenne nulle et leur symétrie autour de la moyenne acceptable. Les extrêmes négatifs sont essentiellement dus aux 6 tubes isolés proches de la route, les extrêmes positifs aux 5 tubes sur transect (et proches de la route). La dispersion est moins grande en hiver.

6 Discussion

6-1 Été/Hiver

La différence de comportement entre l'été et l'hiver s'explique par des conditions purement climatiques : "En été, le rayonnement solaire favorise les phénomènes d'advection, propices à la dispersion des polluants, et les brises de vallée. Les orages sont fréquents et les pluies plus abondantes en été"(citation ASPA).

6-2 Dérive nulle

Concernant les formules (4)

$$m(x,y) = c_0 + c_1 e^{-\frac{|x|}{18}} + c_2 e^{-\frac{|x|}{550}} + c_3 y e^{-\frac{|x|}{18}} + c_4 y e^{-\frac{|x|}{550}} \quad (4\text{-été-rappel})$$

$$m(x,y) = c_0 + c_1 e^{-\frac{|x|}{14}} + c_2 e^{-\frac{|x|}{940}} + c_3 y e^{-\frac{|x|}{14}} + c_4 y e^{-\frac{|x|}{940}} \quad (4\text{-hiver-rappel})$$

examinons les coefficients c_0 obtenus par régression globale en été et en hiver.

	ETE	HIVER
(1) Constante c_0 obtenue par régression globale sur T2	0.338	3.52
(2) Constante c_0 obtenue par régression globale sur T1	0.685	0.95

Tableau 10 – Constantes de la dérive en été et en hiver, sur deux ensembles de tubes. (1) Par régression globale sur T2, ensemble des 39 tubes exploitables. (2) Par régression globale sur T1, ensemble des 26 tubes sur transect

En été, la valeur négligeable obtenue par régression sur T2 implique que dans le modèle, la moyenne $m(x,y)$ s'annule en pratique à environ 1600m de la route. En hiver, cette moyenne s'établit sur 3.5 pour des distances supérieures à 3000m. Les concentrations ne pouvant être que positives, et leurs fluctuations centrées (dans le modèle) sur cette moyenne, ce choix d'une dérive nulle en été implique qu'à une certaine distance de la route, les concentrations et donc les fluctuations sont censées disparaître. Ce choix est-il judicieux ? N'aurait-il pas fallu prendre pour l'été un modèle de moyenne qui impose une constante positive plus importante ? Et dans ce cas, quelle constante ? On ne peut se fier aux valeurs obtenues par régression car elles ne sont conditionnées par aucune mesure : nous n'avons qu'un tube à 1000 m. Ces valeurs résultent de l'extrapolation de fonctions exponentielles ajustées sur des distances inférieures à 500 m et sont à interpréter avec précaution. (En hiver, selon que la régression globale est faite sur T1 ou T2, la valeur de c_0 passe du simple au quadruple).

Mais peut-être faut-il formuler la question ainsi : 1) Peut-on considérer que les tubes ne sont soumis qu'aux émissions routières ? 2) Dans ce cas, est-il judicieux de penser qu'à une certaine distance de la route, les émissions routières ne sont plus mesurables (cela peut sembler évident à certains) 3) Dans le cas où les tubes mesurent également d'autres émissions, n'est-il pas préférable de modéliser ce "bruit de fond" ?

À la question 1) nous répondons que nous n'avons pas constaté de corrélation avec les variables du cadastre, mais la résolution de la grille ne permettait de toutes manières pas une étude fine dans ce sens. Notre choix de dérive se justifie simplement par la volonté d'aller dans le sens de la campagne de mesures dont l'objectif était de mesurer l'influence de la route. Notre hypothèse est que la campagne est bien faite, et rien dans les tests statistiques ne nous indique le contraire.

Aux questions 2) et 3) nous répondons qu'il aurait fallu des mesures éloignées, mais nous n'avons qu'un tube à 1000m et rien au delà. Dans le doute et pour conclure, on ne peut que préconiser de limiter l'utilisation de ce modèle aux seules distances où les mesures sont bien connues, c'est à dire approximativement à moins de 500m de la route.

Chapitre III – Modélisation de l'écart-type $\sigma(x,y)$

0 Résumé du chapitre

La non-stationnarité concerne aussi la variance du NO_2 . Comme pour la dérive, il faut pouvoir calculer cette variance en tout point de l'espace et un modèle est requis. En première approche (paragraphe 2), un modèle d'écart-type basé sur $\text{NO}_2(x,y)$ est établi, mais il s'avère que cette approche est incorrecte. On opère ensuite sur la variable recentrée autour du modèle $m(x,y)$ établi précédemment (paragraphe 3). Ces résidus, égaux à $\text{NO}_2(x,y) - m(x,y)$, conduisent à un modèle $\sigma(x,y)$ qui s'exprime d'une manière simple en fonction de la dérive. Suit une discussion (paragraphe 4) où l'on envisage d'autres possibilités, notamment l'opportunité d'avoir une variance qui ne s'annule pas quand la dérive est nulle.

1 Remarque préliminaire

Les figures suivantes rappellent, pour l'été et l'hiver, le comportement transversal moyen calculé sur les tubes T2 informés aux distances 1 et 2m, 50m, 200 et 400m.

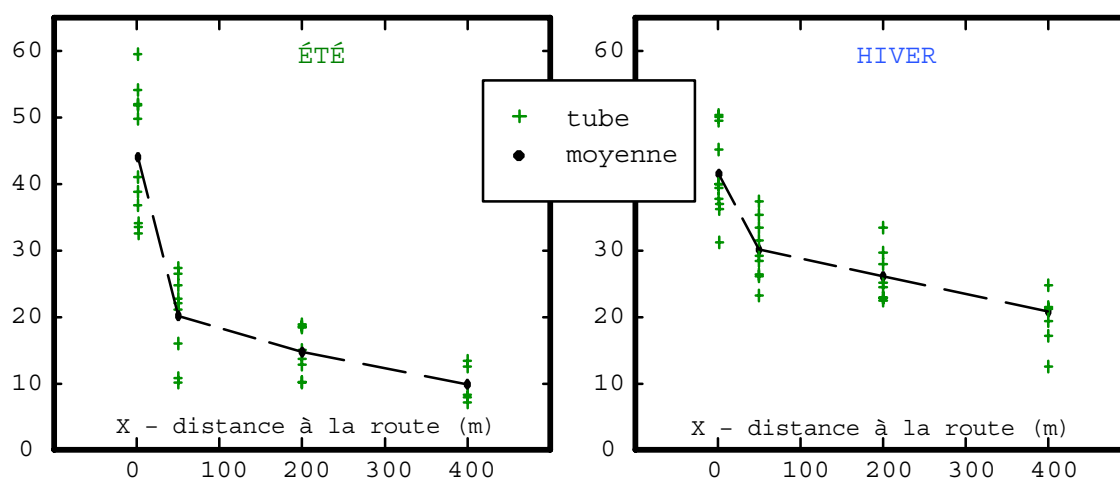


Figure 26 – Moyennes (non corrigées) du NO_2 par classe de distance à la route. Concerne les tubes de T2 dont les distances à la route sont 1 et 2m, 50m, 200 et 400m

On constate (surtout en été) que les variations autour de chaque moyenne décroissent avec cette dernière. Ce constat incite à rechercher une éventuelle relation entre moyenne et variance.

L'identification de $m(x,y)$ s'est faite sur l'ensemble T1 des 26 tubes sur transects car 1) nous souhaitons caractériser analytiquement les évolutions longitudinales et transversales et 2) obtenir un modèle bi-dimensionnel cohérent avec ces expressions analytiques. La nécessité d'utiliser la même information pour chaque axe examiné nous a en conséquence restreint à 5 transects. Ayant obtenu une expression analytique globale de $m(x,y)$, nous avons alors estimé les coefficients de la dérive par régression sur T2.

Ici, la circonstance est différente. Nous souhaitons comparer moyenne et variance partout où cela est possible, sur un maximum de tubes, la contrainte étant que pour chaque couple (m,σ) utilisé, moyenne et écart-type soient calculés sur les mêmes effectifs. On peut en conséquence aborder directement T2.

2 Par classes de distances à la route

En première approche, on examine les moyennes et variances expérimentales du NO_2 . Aucune correction des moyennes n'est faite.

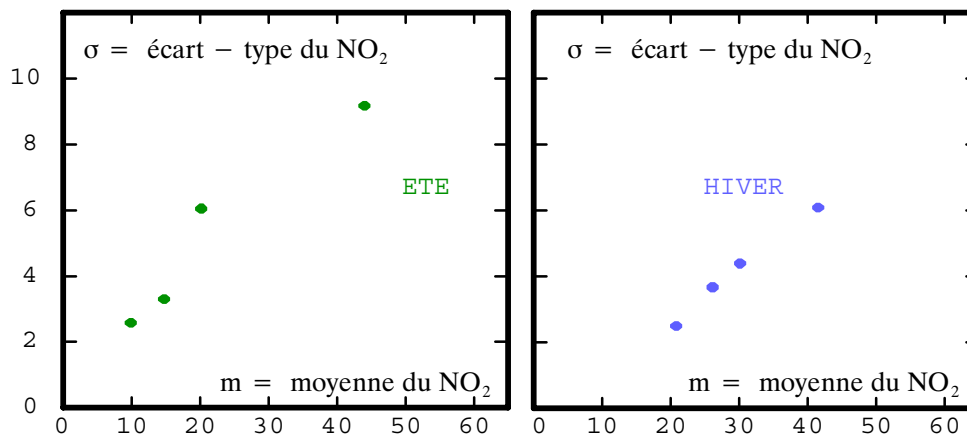


Figure 27 – Ecart-types du NO₂ fonction des moyennes du NO₂ calculées par classe de distance à la route

En hiver, l'alignement des points est marqué, il est moins évident en été. On choisit la simplicité et l'on calcule, par régression, les coefficients d'une relation de type $\sigma = d_1 m + d_2$

	d ₁	d ₂
ÉTÉ	0.191	1.034
HIVER	0.1705	-0.89

Tableau 11 – Coefficients des régressions linéaires avec constante des écart-types expérimentaux du NO₂ fonction des moyennes expérimentales du NO₂ calculées par classe de distance à la route

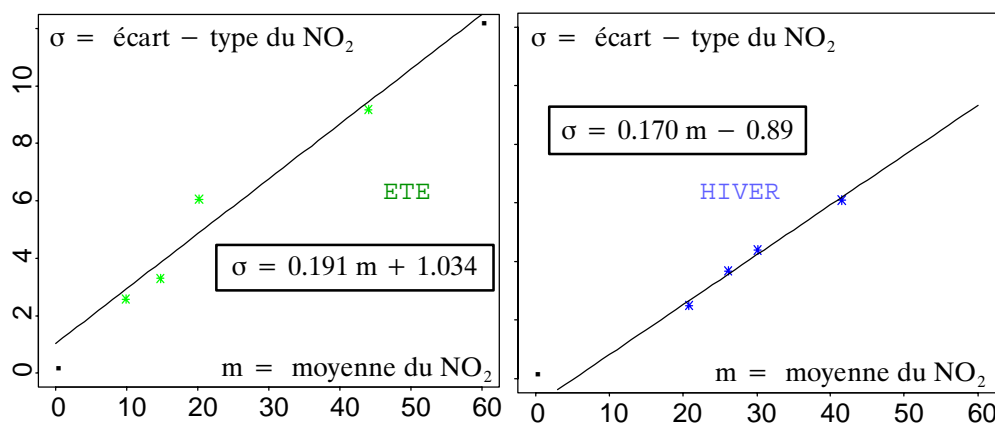


Figure 28 – Régressions linéaires avec constante des écart-types expérimentaux du NO₂ fonction des moyennes expérimentales du NO₂ calculées par classe de distance à la route

Ce modèle conduit en été à un écart-type non nul en 0, qui plus est négatif en hiver. Ces valeurs étant faibles, on peut envisager une régression sans constante, et l'on obtient les coefficients et ajustements suivants, pour une relation qui se simplifie à $\sigma = d m$

	d
ÉTÉ	0.2256
HIVER	0.1422

Tableau 12 – Coefficients des régressions linéaires sans constante des écart-types expérimentaux du NO₂ fonctions des moyennes expérimentales du NO₂ calculées par classe de distance à la route

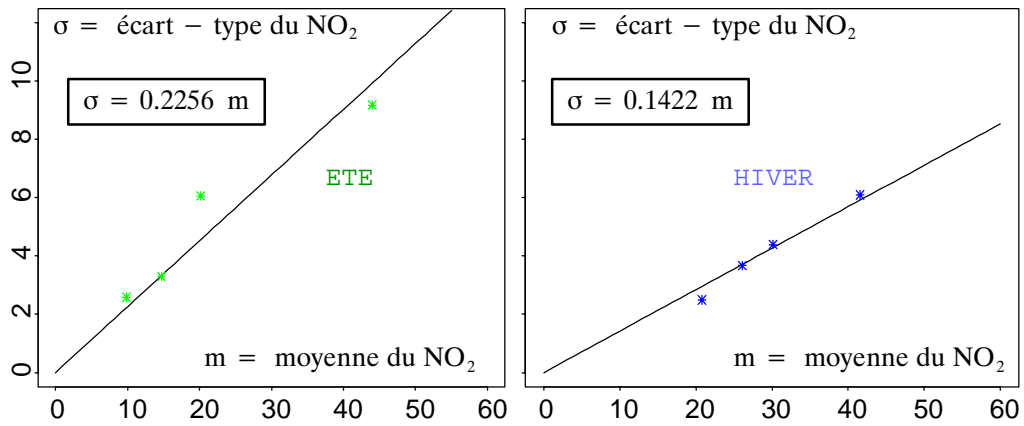


Figure 29 – Régressions linéaires sans constante des écart-types expérimentaux du NO₂ fonction des moyennes expérimentales du NO₂ calculées par classe de distance à la route

Quelle que soit la qualité de ces ajustements, on peut s'interroger sur le bien-fondé de cette première approche. Les écart-types sont centrés sur une moyenne locale qui n'a pas de valeur objective puisque la moyenne n'est pas stationnaire en y. Ces écart-types quantifient une dispersion absolue où l'on est intéressé par une dispersion autour des régressions linéaires. Il faut donc étudier les dispersions autour de la moyenne modélisée en fonction de cette dernière.

3 Par les résidus

3-1 Remarque préliminaire

Examinons les résidus $R(x,y)$ définis par (5) en fonction de $m(x,y)$

$$R(x,y) = \text{NO}_2(x,y) - m(x,y) \quad (5\text{-rappel})$$

Sont représentés en bleu les 5 tubes en bord de route et sur transect, et en noir les 6 tubes isolés.

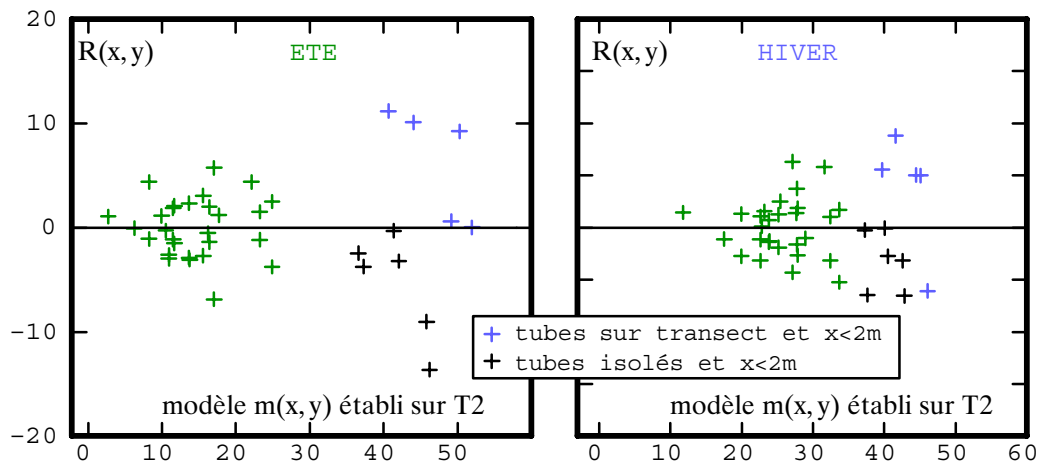


Figure 30 – Résidus expérimentaux fonction du modèle de moyenne établi sur les 39 tubes de T2

En été comme en hiver, la dispersion des résidus augmente avec la dérive. Aux fortes valeurs de $m(x,y)$, le modèle de dérive assure la médiation entre les ensembles de tubes bleus et noirs.

Si maintenant on examine les résidus calculés avec le modèle $m(x,y)$ établi sur T1...

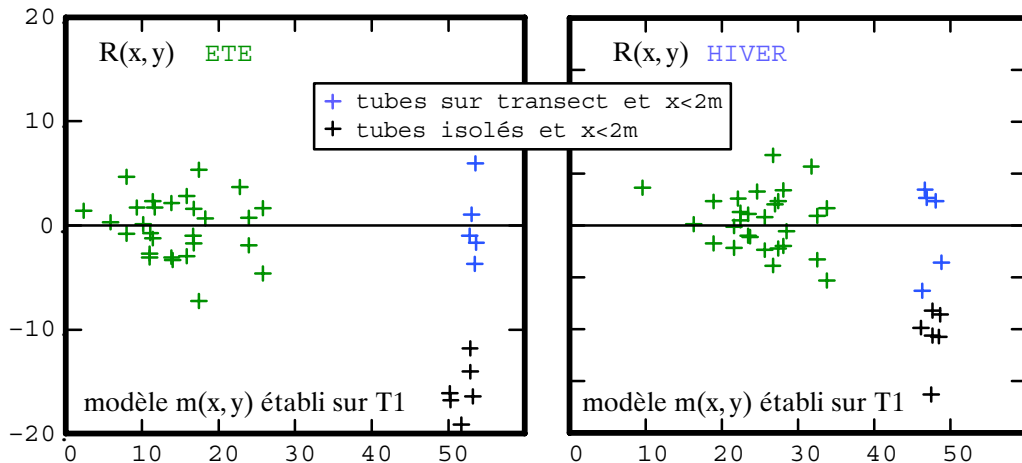


Figure 31 – Résidus expérimentaux fonction du modèle de moyenne établi sur les 26 tubes de T1

... on constate qu'en s'alignant sur les tubes bleus, le modèle de dérive engendre des résidus aux tubes noirs plus importants. Toute la problématique des tubes isolés est illustrée par ces nuages. En choisissant d'estimer $m(x,y)$ sur T2, on incorpore dans la dérive une partie de la variance qui apparaîtrait si l'on se contentait d'utiliser pour T2 la fonction $m(x,y)$ établie sur T1.

3-2 Modélisation

Comme il n'est pas stationnaire, l'écart-type doit être modélisé. Pour l'été, on regroupe les moyennes saisonnières en classes de $m(x,y)$ d'effectif 5 ou 6 pour les 5 premières classes et 11 pour la dernière (les 11 tubes forment un nuage isolé). En hiver, on construit 7 classes d'effectif 5 ou 6. Diminuer les classes (pour augmenter les effectifs par classe) nous pénaliserait quand ultérieurement un ajustement de ces moyennes par classe sera conduit. L'écart-type des résidus ayant pour expression théorique $\sigma(x,y) = \sqrt{E[R^2(x,y)]}$, il est important de ne pas recentrer les résidus autour de leur moyenne réelle (même si elle est négligeable) car nous nous plaçons dans le contexte du modèle (la correction biaiserait l'analyse). L'ajustement de la relation entre écart-type et moyenne se fait en tenant compte de l'effectif de chaque classe. Deux modèles sont ajustés par régression :

$$\begin{aligned} \text{linéaire} \quad \sigma(x,y) &= d m(x,y) \\ \text{puissance} \quad \sigma(x,y) &= d m^p(x,y) \end{aligned} \quad (6)$$

L'éventualité d'autres modèles est discutée plus loin. Les figures suivantes montrent les ajustements obtenus.

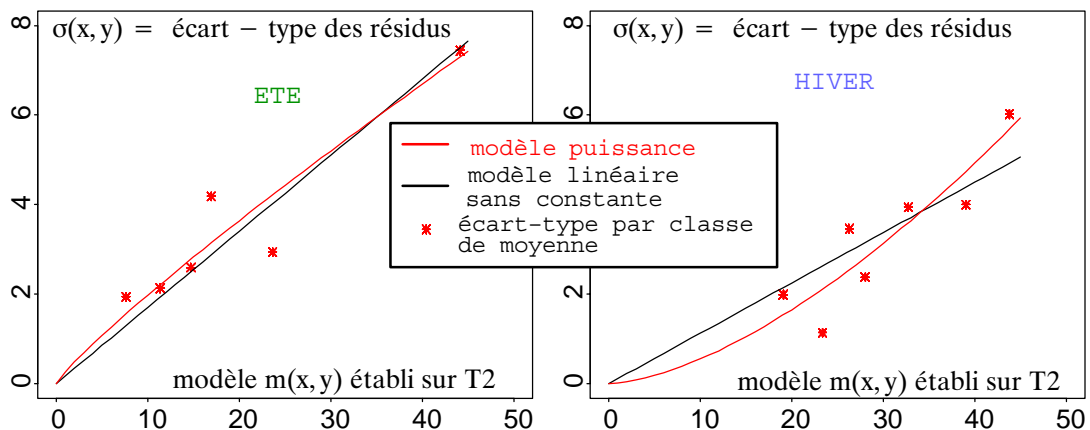


Figure 32 – Régressions des écart-types expérimentaux des résidus fonction des moyennes modélisées

Dans le tableau des coefficients obtenus, la quantité “S” représente l’écart-type à la courbe de régression

	modèle linéaire	modèle puissance
ETE	d=0.170 S=0.66	d=0.261 p=0.879 S=0.62
HIVER	d=0.112 S=0.82	d=0.015 p=1.57 S=0.66

Tableau 13 – Performances des modèles linéaires et puissance

Concernant les modèles linéaires, les pentes, si elles sont différentes de celles obtenues en première approche (tableau 12), leur restent comparables. Parce que les régressions le long de la route donnent des pentes faibles, le lien entre moyennes et variances expérimentales est préservé par la modélisation (ce qui après coup donne un certain intérêt à la première approche).

Le critère de choix est S, l’écart-type à la courbe de régression. Pour l’été, le score est peu différent, pour l’hiver, la différence est sensible, dans les deux cas, l’avantage revient au modèle puissance (6) que nous adoptons dans la suite.

Pour conclure, on examine les résidus normés $RN(x,y)$ définis par

$$RN(x,y) = \frac{NO2(x,y) - m(x,y)}{\sigma(x,y)} \quad (7)$$

Ce calcul est possible en tout tube puisqu’il n’en existe aucun où la moyenne (et donc la variance) est nulle.

Pour juger du travail accompli, et surtout de l’importance de la normation par $\sigma(x,y)$, les échelles de représentation sont les mêmes que celles utilisées pour $R(x,y)$.

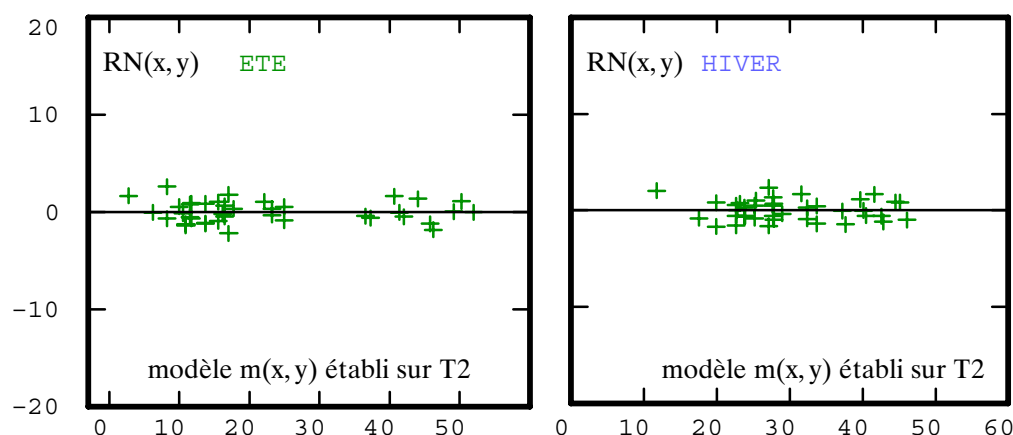


Figure 33 – Résidus expérimentaux normés fonction du modèle de moyenne établi sur les 39 tubes de T2

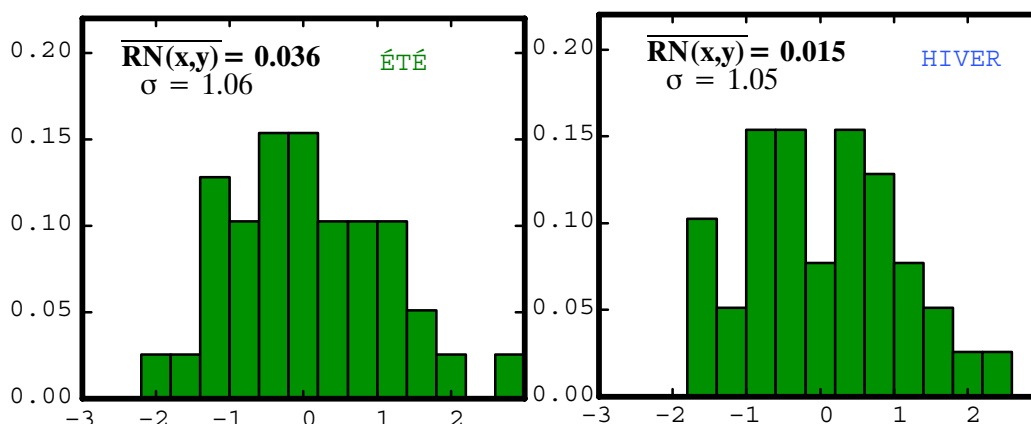


Figure 34 – Histogrammes des résidus normés par l’écart-type modélisé

Les moyennes des résidus normés sont négligeables, leurs écart-types proches de 1.

4 Discussion

4-1 Autres modèles

D'autres pourraient envisager des modèles plus simples. Examinons par exemple la constante et l'escalier (on n'illustre que l'été).

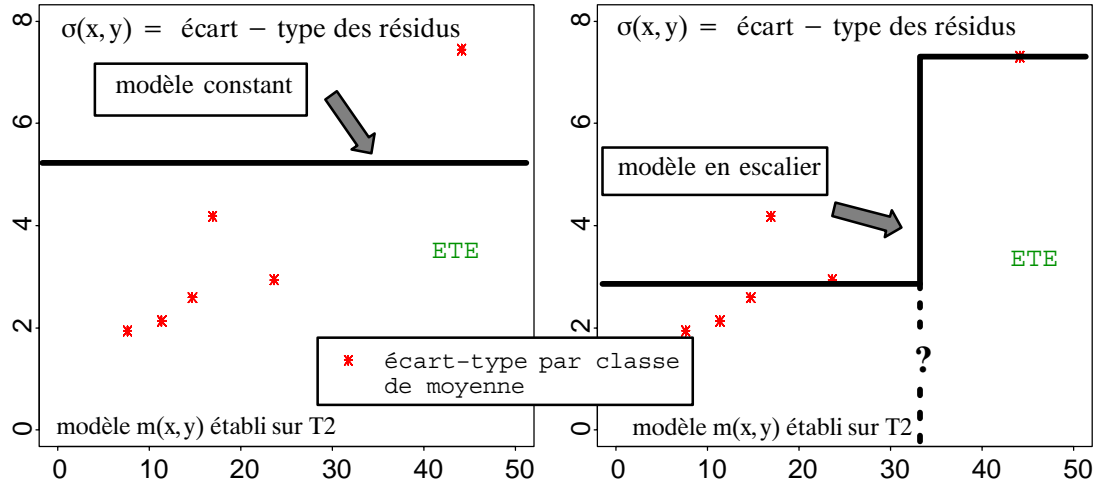


Figure 35 – Autres modèles possibles d'expression de l'écart-type fonction de la moyenne. (a) Modèle constant, l'écart-type ne dépend pas de la moyenne. (b) Modèle en escalier, un écart-type est affecté à des groupes de valeurs

Le modèle constant donne une variance stationnaire. Aucune normation des résidus n'est nécessaire et la variographie (notre prochaine étape) est conduite sur les résidus. Nous l'avons testé. On présente à titre d'exemple les variogrammes transversaux des résidus non normés.

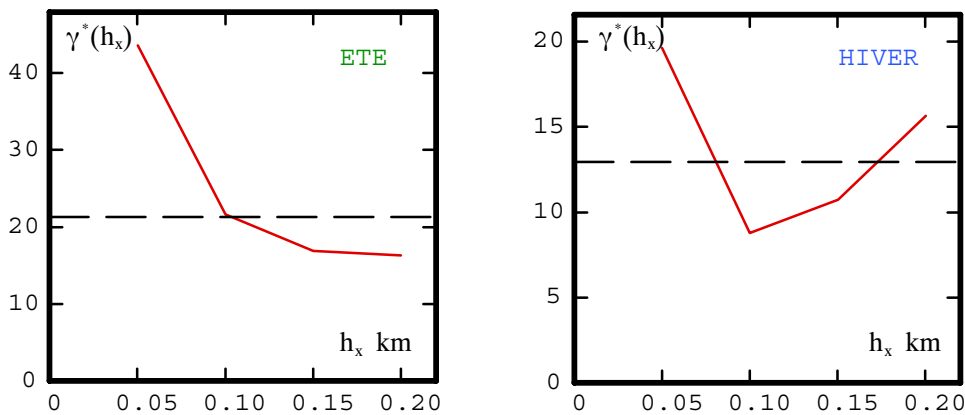


Figure 36 – Variogrammes transversaux des résidus non normés

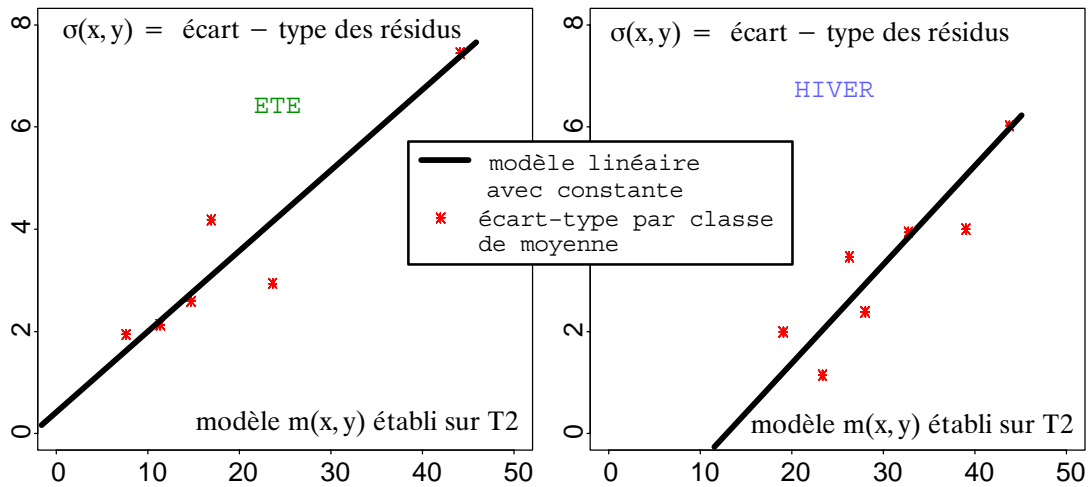
En été comme en hiver, la forte valeur au premier pas (50m) est due au contraste de variabilité entre les données en bordure ($\sigma(x,y)$ associée à $m(x,y)=45$ en été) et les autres mesures. Les autres valeurs du variogramme sont plus sages car dans leur calcul, l'importance des mesures en bordure de route est moindre. Ignorer la non stationnarité de la variance (et donc ne pas normer les résidus) revient à affecter aux tubes en bordure une variance trop faible, et aux autres tubes une variance trop forte. La normation des résidus est nécessaire pour aller plus loin.

Le modèle en escalier revient à donner aux points en bordure de route une variance différente des autres. Mais il faut que $\sigma(x,y)$ couvre l'ensemble du domaine des valeurs prises par $m(x,y)$, et le problème se pose de savoir pour quelle valeur de $m(x,y)$ la variance doit changer : c'est un troisième paramètre où le modèle en puissance n'en nécessite que deux.

On peut imaginer d'autres fonctions encore, mais on voit bien qu'en terme d'économie de paramètres, le modèle $\sigma(x,y)$ proposé est un bon compromis.

4-2 Variance nulle

Le modèle $\sigma(x,y)$ adopté s'annule avec la dérive. Ce choix entraîne peut-être une perte de précision dans l'ajustement de $\sigma(x,y)$ par rapport à $m(x,y)$ car une régression linéaire avec constante donne un meilleur résultat au sens des moindres-carrés. Effectuons ces ajustements.



Ils conduisent en été à une écart-type de l'ordre de 0.5 pour $m(x,y)=0$, et en hiver à un écart-type négatif pour les valeurs de la dérive inférieures à 10.

Pour l'été, 0.5 est négligeable pour un gain de précision tout autant négligeable.

En hiver, la différence est à peine plus importante, du fait du comportement assimilable comme linéaire du modèle puissance pour des valeurs de $m(x,y)$ supérieures à 20. Si l'on adopte ce modèle pour l'hiver, on est obligé de limiter l'utilisation du modèle aux valeurs positives de l'écart-type (valeurs de $m(x,y)$ supérieures à 10), ce qui revient à restreindre la cartographie du NO_2 à des distances_route inférieures à 500m environ (il n'y a pas de limite sur l'ordonnée curviligne). Mais cette contrainte résultait déjà de la discussion sur $m(x,y)$ et de ce fait ne constitue pas un critère déterminant. C'est essentiellement parce que le gain de précision est faible que nous avons opté pour la cohérence théorique et donc le modèle puissance.

On en conviendra, il est toujours difficile d'envisager des variances négatives...

Chapitre IV – Variographie des résidus normés

0 Résumé du chapitre

Les résidus de la dérive sont normés par le modèle $\sigma(x, y)$ établi précédemment et l'on obtient des quantités en théorie stationnaires, d'espérance nulle et de variance unité. On examine si ces résidus sont corrélés spatialement en calculant leurs variogrammes, après avoir exposé dans le détail la manière dont ces calculs sont faits.

Que ce soit le long de la route ou perpendiculairement, aucune structure évidente n'apparaît pour cause probable d'effectifs statistiques insuffisants. Le premier modèle qui vient pour les variogrammes est donc un effet de pépite pur (bruit blanc). Cependant, on propose aussi pour l'été un modèle pépitique + exponentiel à anisotropie géométrique, et pour l'hiver (et sous toutes réserves!) un sphérique à anisotropie géométrique.

1 Remarques préliminaires

1-1 Séparer les parties gauche et droite ?

Les inférences de $m(x, y)$ et $\sigma(x, y)$ sont basées sur la valeur absolue de l'abscisse et n'ont pas différencié ce qui est à droite de ce qui est à gauche de la route. Ce choix résulte du besoin d'augmenter les effectifs statistiques et s'appuie sur le constat qu'aucune dissymétrie évidente n'apparaît quand on examine le comportement transversal moyen. Pour calculer les variogrammes des résidus normés, la question se pose de savoir si l'on continue dans cette voie. Dans la suite, les qualificatifs "**gauche**" (respectivement "**droit**") désignent les variogrammes qui n'utilisent que des couples de points situés à gauche (respectivement à droite) de la route.

1-2 Traverser la route ?

Centrée autour de la moyenne $m(x, y)$ et normée par l'écart-type $\sigma(x, y)$, la variable $RN(x, y)$ est en théorie stationnaire, mais est-ce suffisant pour pouvoir, à une même valeur numérique h donnée, associer des couples $(RN(x+h, y), RN(x, y))$ chevauchant la route à des couples situés du même côté ? Dans la suite, le qualificatif "**traversant**" désigne les variogrammes qui font intervenir des couples de points chevauchant la route.

1-3 Tolérance angulaire

Compte tenu de l'anisotropie de l'échantillonnage, seules deux directions sont possibles pour les variogrammes expérimentaux : le long de la route (par convention, la direction 90 degrés) et perpendiculairement (0 degrés). Pour stabiliser les calculs, on augmente le nombre de couples en utilisant l'artifice des tolérances angulaires. Dans la figure qui suit, sont représentés les 39 tubes dans une échelle dilatée en largeur d'un facteur 10. La discussion porte sur les variogrammes traversants.

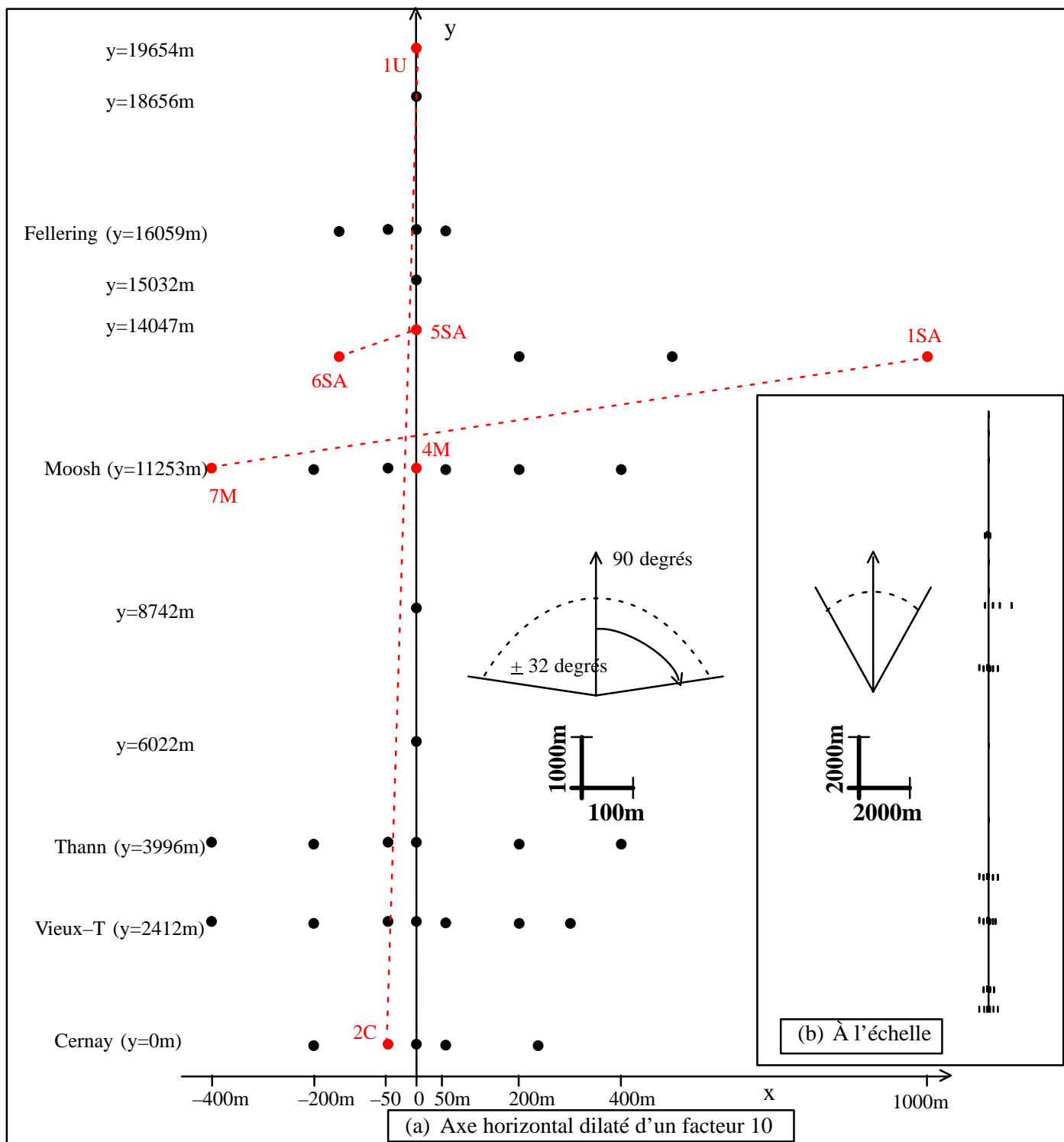


Figure 38 – Les 39 tubes exploitables (ensemble T2) dans le repère de travail. Le couple de tubes (7M,ISA) permet de déterminer la tolérance angulaire maximale le long de la route (32 degrés). En (b) l'échelle est respectée. Il est convenu d'appeler "traversants" les variogrammes qui font intervenir des couples de tubes situés de part et d'autre de la route. On convient également de désigner par "variogramme gauche" (resp. "droit") les variogrammes calculés sur les tubes d'abscisse négative ou nulle (resp. d'abscisse positive ou nulle)

Tolérance autour de la direction 90 degrés. Pour que cette tolérance permette d'incorporer un maximum de couples de points, il faut trouver le couple de points, non alignés sur transect, pour lequel le rapport $\frac{\Delta x}{\Delta y}$ est maximal (Δx et Δy sont les différences entre respectivement les abscisses et les ordonnées). Il s'agit du couple (7M, 1SA). L'angle associé est l'arc-tangente du rapport $\frac{\Delta x}{\Delta y}$, soit 32 degrés. Comme la distance entre ces deux points est inférieure à 10km (demi-longueur de la route couverte par les mesures), on peut accepter la valeur de ± 32 degrés comme tolérance autour de la direction 90 pour le calcul de $\gamma^*(h_y)$, variogramme le long de la route. 1SA étant le seul tube situé à plus de 500 m de la route, cette tolérance ne permet de "gagner" qu'un seul couple. Les autres couples inter-transect donnent des angles inférieurs à 28 degrés.

Tolérance autour de la direction 0 degrés. L'un des 2 couples de points non alignés parallèlement à la route et qui rend le rapport $\frac{\Delta y}{\Delta x}$ maximal est (2C, 1U). Mais la distance entre ces points excède de loin la distance à la route maximale admissible qui est 400m (1SA étant unique, la largeur du champ en x qui doit être considérée est 800m). Les deux points les plus proches et non alignés sont 5SA et 6SA, (leur distance est 633m, encore trop grande), et l'on voit bien qu'une tolérance autour de la direction 0 degrés est impossible.

Conclusion. Pour les variogrammes traversants, on se contente d'une tolérance en y de ± 32 degrés autour de 90 degrés, et aucune en x. Pour les variogrammes droit, la tolérance autour de 90 degrés est ± 24 pour récupérer le couple (4M, 1SA), les autres couples inter-transect donnant des angles inférieurs à 14 degrés. Pour les variogrammes gauche, la tolérance est ± 14 degrés.

1-4 Statut des données en bordure de route

Nous allons entre autre examiner des variogrammes calculés à gauche et à droite de la route, ce qui pose la question du statut des mesures en bordure de route. Elles sont données pour être à 1 ou 2m, mais nous ignorons de quel côté. Nous avons choisi de considérer les données en bordure comme centrales et elles interviendront dans les calculs gauches et droits. La route étant un axe de symétrie (dont la largeur est inconnue) et les effectifs statistiques étant faibles, nous n'avons pas d'autre choix.

2 Variogrammes expérimentaux – Eté

2-1 En h_x : traversant

Les calculs sont conduits jusqu'à la moitié du champs couvert transversalement (800m). Ce variogramme utilise des données à cheval sur la route. À 50 et 100m, $\gamma^*(h_x)$ présente des valeurs plus élevées qu'aux autres distances. L'examen de la nuée variographique montre que la cause est d'importants résidus aux tubes 5M et 1. Quand on masque ces tubes, on obtient une allure plus satisfaisante. Mais quelle est la valeur de ce résultat ? Il n'existe aucune raison objective pour masquer ces deux tubes. Si ils perturbent le variogramme, c'est peut-être simplement parce que nous manquons d'effectifs.

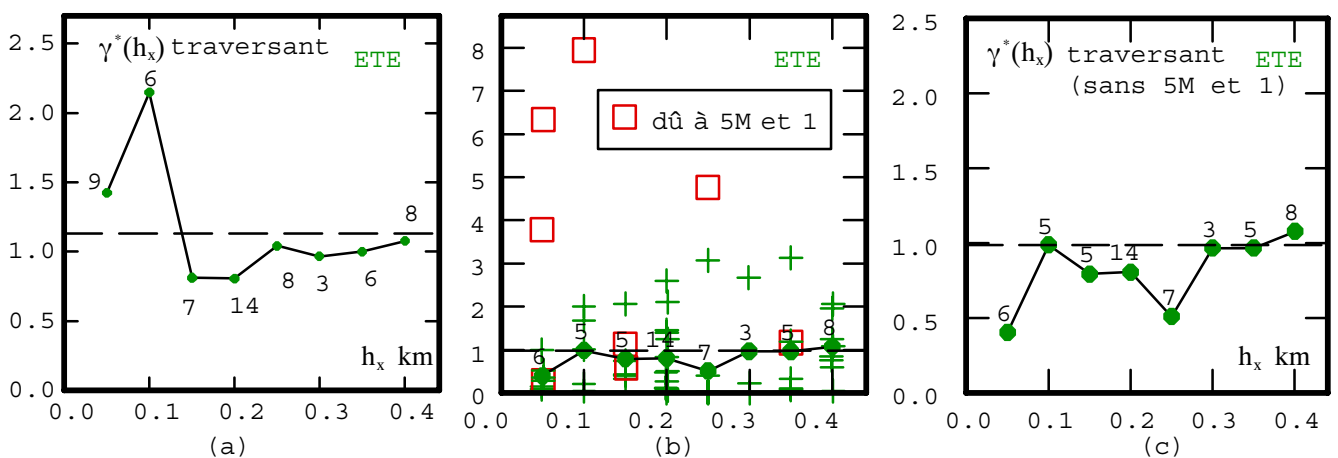


Figure 39 – Variogramme transversal traversant. (a) Calculé sur les 39 tubes. (b) Nuée variographique mettant en évidence l'influence des tubes 5M et 1. (c) Calculé sans 5M et 1. Les chiffres représentent les nombres de couples.

2-2 En h_x : gauche, droit, moyen, traversant

Pour savoir si il faut ou non utiliser les couples de points à cheval sur la route, nous avons calculé les variogrammes gauche, droit, et leur moyenne (pondérée par les effectifs). Les données de bord de route sont utilisées à gauche et à droite. Les calculs sont faits sans les tubes 5M et 1 pour pouvoir comparer au variogramme traversant précédent. Les variogrammes gauche et droit ne sont calculés que jusqu'à 200m.

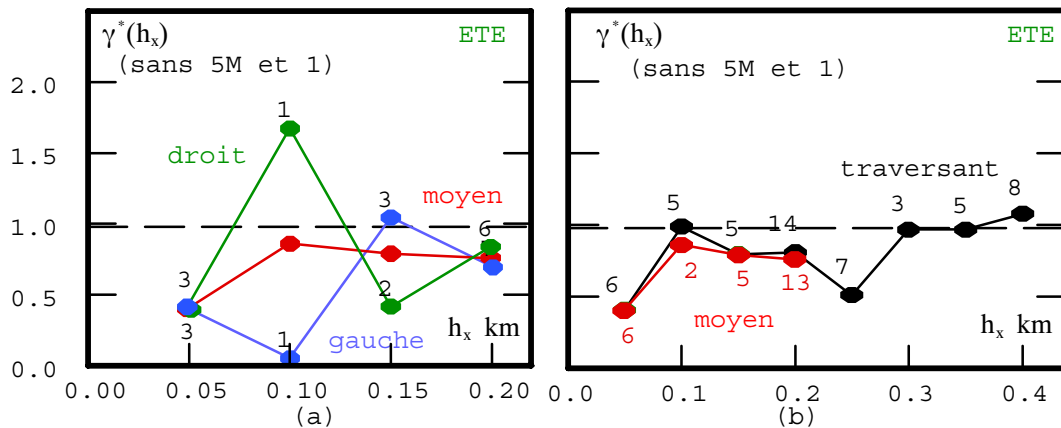


Figure 40 – Variogrammes transversaux. (a) À gauche ou à droite de la route, et leur moyenne pondérée par les effectifs. (b) Moyen et traversant

La comparaison entre γ^* moyen et γ^* traversant ne montre pas de grandes différences, mais on ne peut rien en déduire pour autant car les couples impliqués dans l'un et l'autre cas sont quasiment les mêmes et de toutes manières d'un effectif trop faible.

2-3 En h_y : gauche, droit, moyen, traversant

Le long de la route, tous les tubes sont utilisés. Les variogrammes gauche et droits incorporent les tubes situés en bord de route (et considérés comme étant sur la route). En conséquence, le variogramme moyen fait intervenir deux fois les couples de tube situés sur la route. Les effectifs qui lui sont affectés peuvent alors être supérieurs à ceux du variogramme traversant qui, lui, ne fait intervenir ces couples qu'une fois.

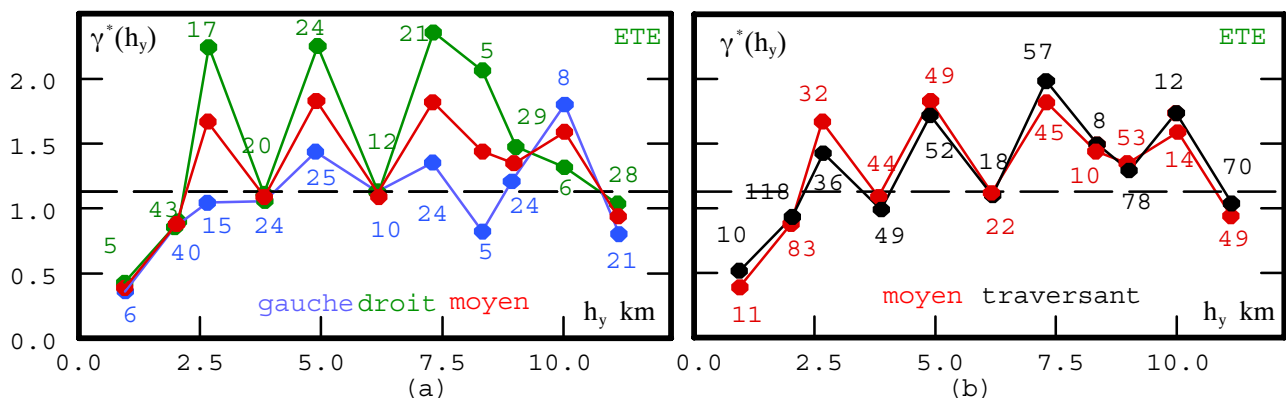


Figure 41 – Variogrammes longitudinaux. (a) À gauche ou à droite de la route, et leur moyenne pondérée par les effectifs. (b) Moyen et traversant

Tous les variogrammes présentent une composante cyclique. Une analyse plus fine montre que cet effet est dû aux seuls tubes sur la route : leur position sur la route alterne tube isolé (moyenne faible) et tube sur transect (moyenne plus élevée), produisant l'effet périodique constaté. L'importance de l'effet périodique nous amène à réfléchir une fois de plus sur l'opportunité d'inclure ou non les 6 tubes isolés dans l'analyse. Si on les garde, on perturbe le variogramme, si on les enlève, les couples utilisés dans les calculs se réduisent fortement, entraînant d'importantes fluctuations statistiques. Ce point est développé dans le chapitre VI – Conclusions. Dans l'ensemble, les variogrammes moyen et traversant sont cohérents. Pour un éventuel ajustement, on choisira le variogramme traversant car il ne fait intervenir qu'une fois les couples de tubes sur la route. Le premier point de ce variogramme ($h_y=1000m$) ne fait intervenir que 10 couples alors que le suivant ($h_y=2200m$) utilise 118 couples. Il n'est probablement pas représentatif.

3 Variogrammes expérimentaux – Hiver

3-1 En h_x : traversant

Ici, c'est à 350m et 400m que des variations importantes sont constatées. La nuée variographique imposerait d'au moins masquer les tubes 6T, 2T, et 1, tout aussi arbitrairement qu'en été, et pour n'obtenir aucune structure identifiable.

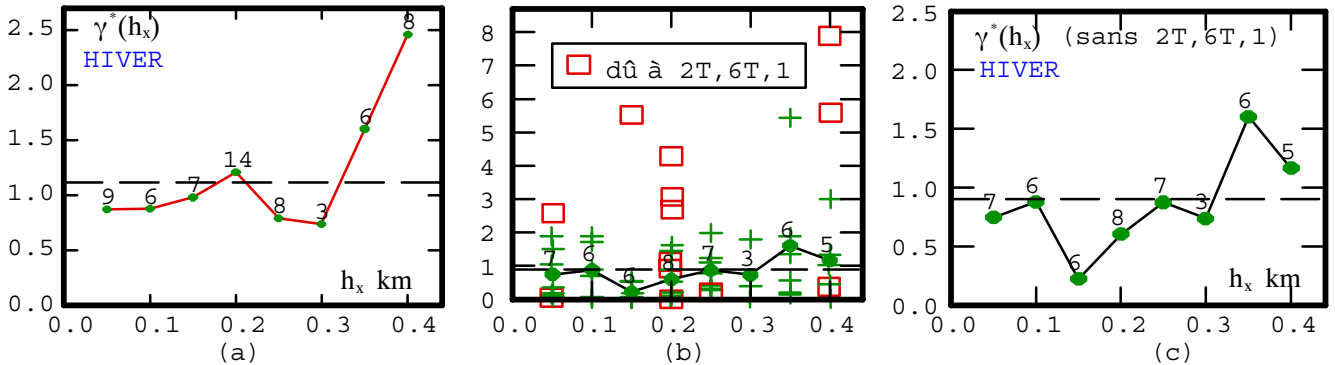


Figure 42 – Variogramme transversal traversant. (a) Calculé avec les 39 tubes. (b) Nuée variographique mettant en évidence l'influence des tubes 2T, 6T et 1. (c) Calculé sans 2T, 6T et 1

3-2 En h_x : gauche, droit, moyen, traversant

Comme en été, on ne peut constater aucune incohérence entre γ^* _moyen et γ^* _traversant pour des calculs faisant intervenir un nombre faible de couples

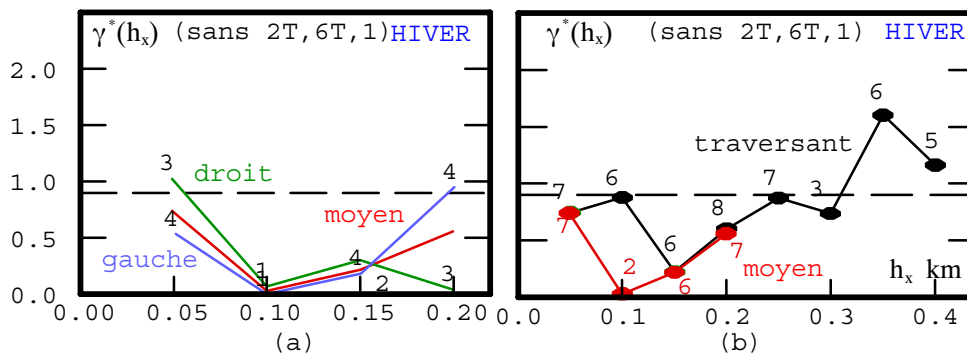


Figure 43 – Variogrammes transversaux. (a) À gauche ou à droite de la route, et leur moyenne pondérée par les effectifs. (b) Moyen et traversant

3-3 En h_y : gauche, droit, moyen, traversant

La différence gauche-droit est moins importante qu'en été et la cohérence moyen-traversant toujours bonne. On retient le variogramme traversant pour les mêmes raisons qu'en été.

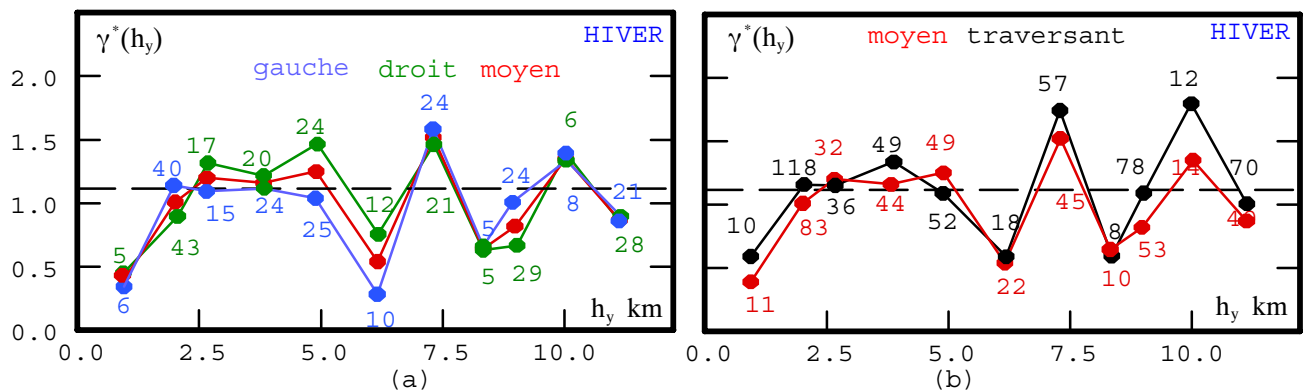


Figure 44 – Variogrammes longitudinaux. (a) À gauche ou à droite de la route, et leur moyenne pondérée par les effectifs. (b) Moyen et traversant

4 Ajustements

Les variogrammes expérimentaux sont interprétés de deux manières : une approche prudente ou ils sont ajustés par un effet de pépite pur, une approche plus discutabile qui consiste à voir une structure et la modéliser. On utilise les variogrammes traversants.

4-1 Été : par un effet de pépite

Transversalement (en h_x), on considère qu'il n'y a aucune raison de masquer les tubes 5M et 1, et les fortes valeurs du variogramme aux deux premiers pas sont attribués aux fluctuations résultantes d'un nombre de couples trop faible. On ne peut donc voir qu'un effet de pépite ou une structure (inaccessible) d'une portée inférieure à 50m.

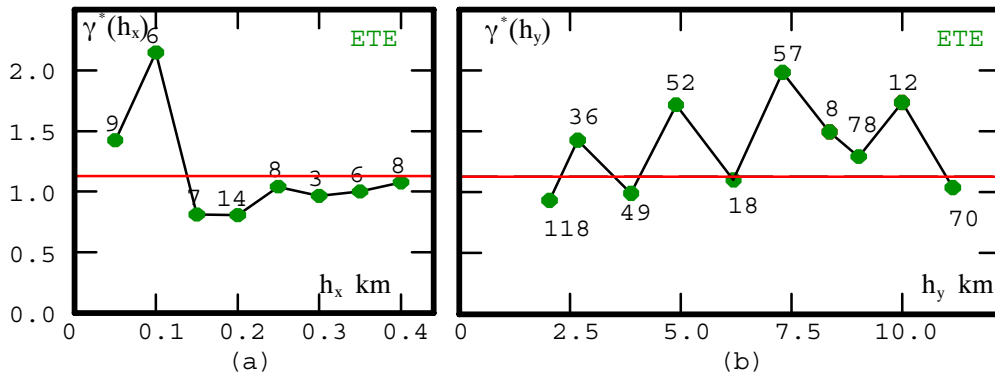


Figure 45 – Ajustement par un effet de pépite égal à la variance des résidus normés. (a) Variogramme transversal. (b) Variogramme longitudinal

Le long de la route (en h_y), on se refuse à considérer le premier point du variogramme ($h_y=1000m$) car il n'est calculé qu'avec 10 couples, alors que la plupart des autres points utilisent de 36 à 118 couples. Ce premier point étant masqué, les autres points nous conduisent à un effet de pépite ou une structure d'une portée (inaccessible) inférieure à 2200m. N'ayant aucune indication sur les portées rendues inaccessibles par l'échantillonnage, le premier modèle retenu est un effet de pépite pur de palier 1.13, variance des résidus normés.

4-2 Été : par une structure non purement pépitique

Le long de la route (figure 46–b), le variogramme considéré inclut tous les tubes de T2 (notamment 5M et 1). On garde le premier point ($h_y=1000m$) car bien qu'il soit calculé avec peu de couples, il indique peut-être une tendance qui mérite d'être exploitée. On aboutit à une structure gigogne, somme d'un effet de pépite de valeur 0.3 et d'un modèle exponentiel de palier 1.4 et de portée pratique 10000 m.

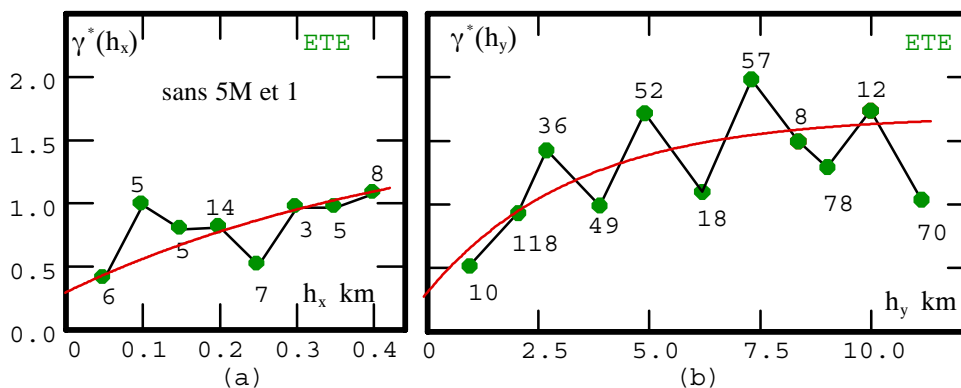


Figure 46 – Ajustement par un modèle pépitique + exponentiel à anisotropie géométrique. (a) Portée transversale : 1450 m. (b) Portée longitudinale 10 km

Transversalement (figure 46–a), on considère le variogramme calculé sans les tubes 5M et 1 qui perturbent les deux premières valeurs du variogramme. En se basant sur le modèle longitudinal, on trouve une portée de 1450 m. Le modèle global retenu est un effet de pépite et un schéma exponentiel à anisotropie géométrique de portées 1450 m en x et 10000 m en "y".

4-3 Hiver : par un effet de pépité

Transversalement, si il y a une structure, elle est de portée inférieure à 50 m.

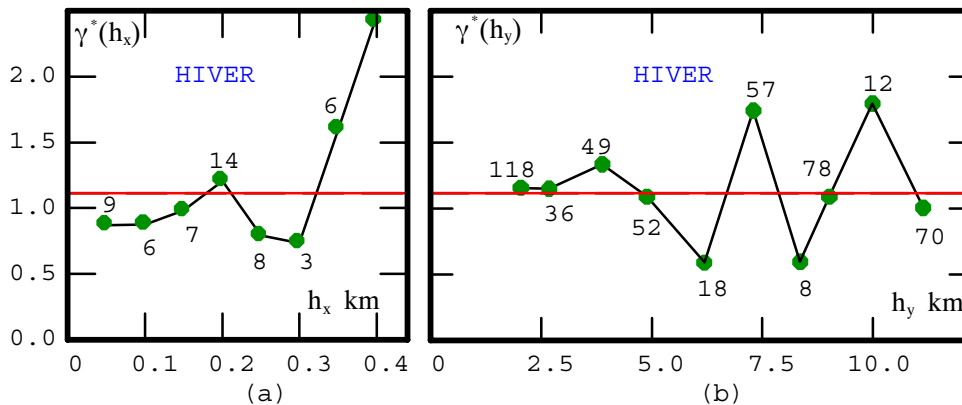


Figure 47 – Ajustement par un effet de pépité égal à la variance des résidus normés.
(a) Variogramme transversal. (b) Variogramme longitudinal

Le long de la route (en h_y), si il y a une structure, elle est de portée inférieure à 2200 m.

N'ayant aucune indication sur les portées rendues inaccessibles par l'échantillonnage, le modèle global retenu est un effet de pépité de palier 1.11, variance des résidus normés.

4-4 Hiver : par une structure non pépitique

Transversalement (en h_x), on s'appuie sur le premier point pour identifier une portée de 80 m associé au choix totalement arbitraire d'un schéma sphérique.

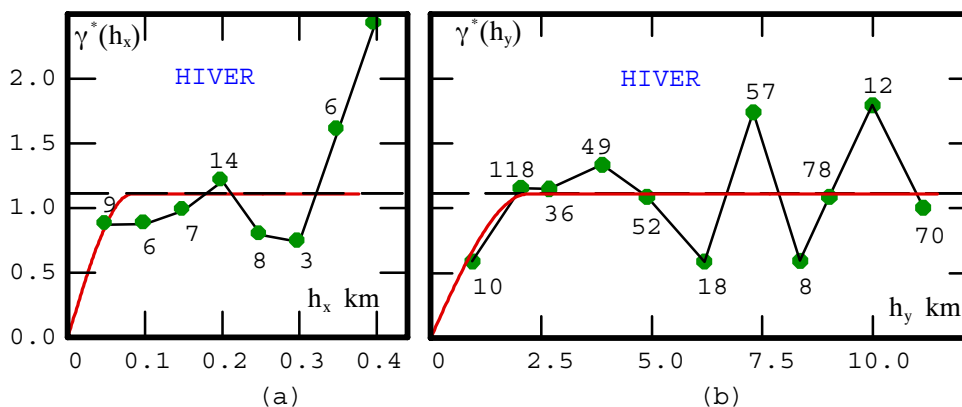


Figure 48 – Ajustement par un modèle sphérique à anisotropie géométrique.
(a) Portée transversale : 80 m. (b) Portée longitudinale 2200 m

Longitudinalement, si l'on garde le premier point et que l'on choisit un schéma sphérique, la portée est 2200m. Le modèle global retenu (sous toutes réserves !) est un schéma sphérique à anisotropie géométrique de portée 80 m en x et 2200 m en "y".

Chapitre VI – Estimations

0 Résumé du chapitre

Nous construisons les équations du krigeage simple des résidus normés, et l'on considère deux cas :

- Cas pépitique : le meilleur estimateur du NO₂ est la dérive m(x,y) obtenue par moindres-carrés
- Cas non pépitique : le système de krigeage simple est conduit en voisinage unique pour être comparé au cas précédent.

L'inférence des modèles de variogramme s'est faite dans des conditions extrêmement difficiles : pour l'été (figure 46), on peut y croire car la tendance des points est marquée (tout du moins longitudinalement). Pour l'hiver (figure 48), le modèle est très arbitraire. Mais quel que soit le crédit apporté à ces ajustements, la comparaison des estimations résultantes avec celles issues des régressions permet d'illustrer l'apport indiscutable du krigeage quand un ajustement robuste des variogrammes est possible. De plus, en comparant l'estimateur krige en été à l'estimateur krige en hiver (estimateurs pour les quels les portées des variogrammes sont très différentes), on montre l'influence des portées du variogramme sur l'allure des cartes.

On propose en annexe d'autres modèles possibles de krigeage. Ils n'ont pas été testés, faute de temps.

1 Rappels

Les résidus normés RN(x,y)

$$RN(x,y) = \frac{NO2(x,y) - m(x,y)}{\sigma(x,y)} \quad (7\text{-rappel})$$

sont supposés stationnaires d'ordre 2, d'espérance nulle et de covariance C(h).

La dérive m(x,y) a pour expression

$$m(x,y) = c_0 + c_1 e^{-\frac{|x|}{18}} + c_2 e^{-\frac{|x|}{550}} + c_3 y e^{-\frac{|x|}{18}} + c_4 y e^{-\frac{|x|}{550}} \quad (4\text{-été-rappel})$$

$$m(x,y) = c_0 + c_1 e^{-\frac{|x|}{14}} + c_2 e^{-\frac{|x|}{940}} + c_3 y e^{-\frac{|x|}{14}} + c_4 y e^{-\frac{|x|}{940}} \quad (4\text{-hiver-rappel})$$

avec les coefficients suivants

	c ₀	c ₁	c ₂	c ₃	c ₄
ETE	0.338	28.23	24.97	0.66 10 ⁻⁴	-7.70 10 ⁻⁴
HIVER	3.52	11.93	31.46	1.78 10 ⁻⁴	-5.64 10 ⁻⁴

Tableau 14 – Paramètres de la dérive été et hiver

L'écart-type σ(x,y) est donné par

$$\sigma(x,y) = d m^p(x,y) \quad (6\text{-rappel})$$

avec les coefficients suivants

	d	p
ETE	0.261	0.879
HIVER	0.015	1.57

Tableau 15 – Paramètres de l'écart-type été et hiver

2 Equations du krigeage simple (KS) des résidus

On choisit d'estimer les résidus normés en des points (x₀,y₀) où ils sont inconnus par une combinaison linéaire de mesures

$$RN^*(x_0,y_0) = \sum_{\alpha} \lambda_{\alpha} RN(x_{\alpha},y_{\alpha}) \quad (8\text{-a})$$

La minimisation de la variance d'erreur donne un système classique de krigeage simple

$$\sum_{\alpha} \lambda^{\alpha} C_{\alpha\beta} = C_{\beta 0} \quad \forall \beta : 1, n \quad (8-b)$$

La variance de krigeage s'exprime par

$$\sigma_{RN}^{2KS}(x_0, y_0) = \text{VAR}[RN^*(x_0, y_0) - RN(x_0, y_0)] = C(0) - \sum_{\alpha} \lambda^{\alpha} C_{\alpha 0} \quad (8-c)$$

$C(0)$ étant égal à σ_{RN}^2 , la variance théorique des résidus normés. En hiver, elle est égale à 1.11, variance expérimentale des résidus normés, puisque les modèles de variogramme ont cette variance pour palier. En été, et pour le modèle non pépitique, cette variance est 1.7 (au lieu de 1.13, variance expérimentale).

On construit l'estimateur du NO_2 en posant

$$NO_2^*(x_0, y_0) = \sigma(x_0, y_0) RN^*(x_0, y_0) + m(x_0, y_0) \quad (8-d)$$

et la variance d'estimation du NO_2 est

$$\sigma_{NO_2}^{2KS}(x_0, y_0) = \sigma^2(x_0, y_0) \sigma_{RN}^{2KS}(x_0, y_0) = \sigma^2(x_0, y_0) C(0) - \sum_{\alpha} \lambda^{\alpha} \sigma^2(x_0, y_0) C_{\alpha 0} \quad (8-e)$$

Cette variance d'estimation ne tient pas compte des incertitudes sur l'estimation de $m(x, y)$ et $\sigma(x, y)$, ni d'ailleurs celles du variogramme, chose courante en géostatistique. Ces incertitudes sont liées à l'échantillonnage. Pour les quantifier, il faudrait recourir à des modèles de distribution que nous n'avons pas le moyen d'inférer. Dans l'expression (8-e), négliger ces incertitudes suppose implicitement que l'échantillonnage est bien réparti dans l'espace et d'un effectif suffisant.

3 Cas pépitique

Quand la covariance $C(h)$ est pépitique, tous les pondérateurs λ^{α} sont nuls et l'estimateur se réduit à la dérive:

$$NO_2^*(x_0, y_0) = m(x_0, y_0)$$

On retrouve un fait connu : le krigeage avec effet de pépité équivaut à la seule estimation de la dérive (obtenue ici par régression en moindres-carrés).

La dérive est calculée en chaque noeud (x_0, y_0) d'une grille choisie par l'utilisateur, mais n'excédant pas environ 500 m en x. Dans ce qui suit, le pas de notre grille est 25m en x et 250 m en "y".

En chaque noeud (x_0, y_0) , la variance d'estimation associée est

$$\sigma_{NO_2}^{2KS}(x_0, y_0) = \sigma^2(x_0, y_0) C(0)$$

3-1 Été

On présente l'estimation en 3D et en projection.

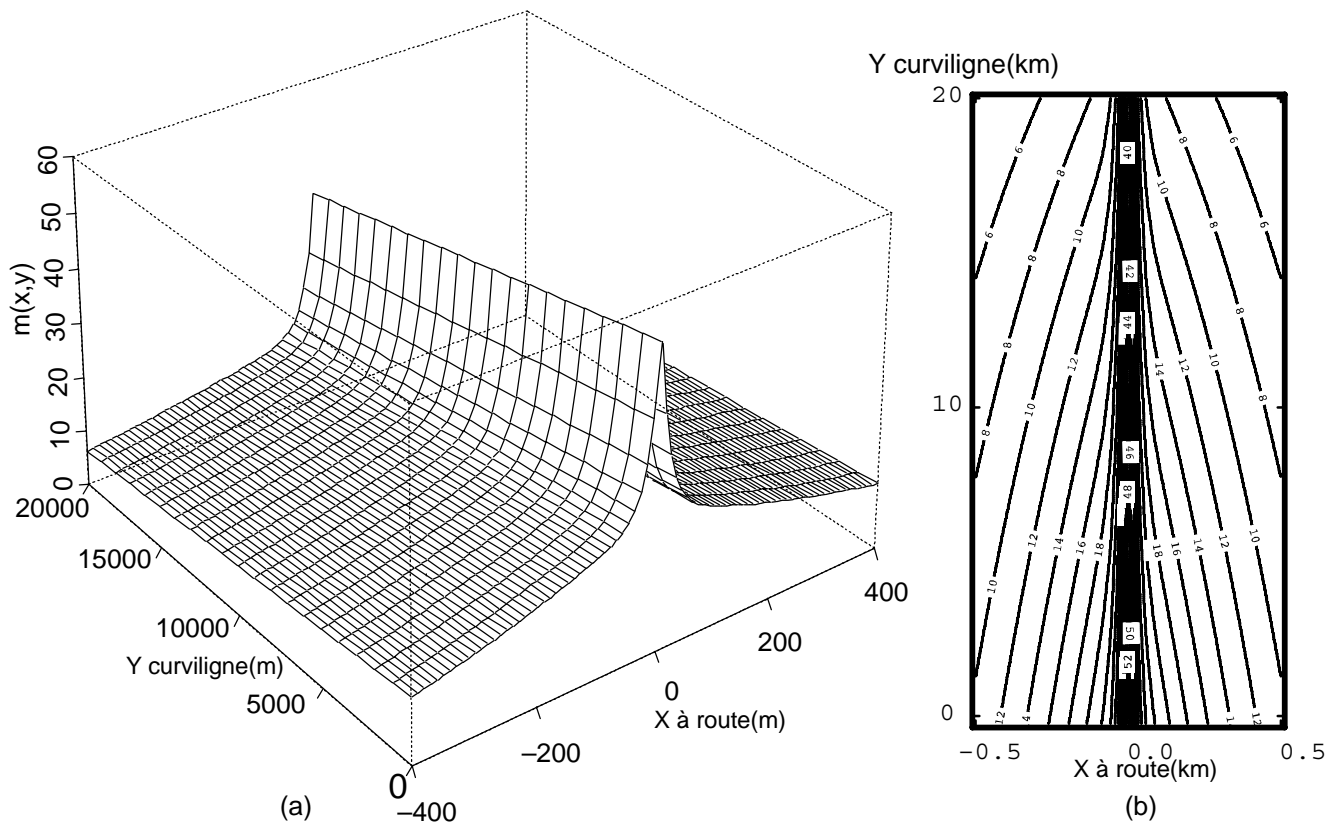


Figure 49 – Krigeage simple des résidus normés sous covariance pépitique : estimateur égal à la dérive.
(a) Représentation par éléments finis. Les lignes sont parallèles à la route. (b) Projection en iso-valeurs.

Sur la projection, la représentation en lignes d'iso-valeurs (pas 2) permet d'apprécier le resserrement de la dérive quand on s'enfonce dans la vallée. Cette projection illustre la remarque 2-4 du chapitre II concernant l'importance relative des coefficients de la dérive. La bande centrale noire ne permet pas de dissocier les iso-lignes proches de la route tant le gradient est important.

3-2 Hiver

On présente l'estimation en 3D et en projection.

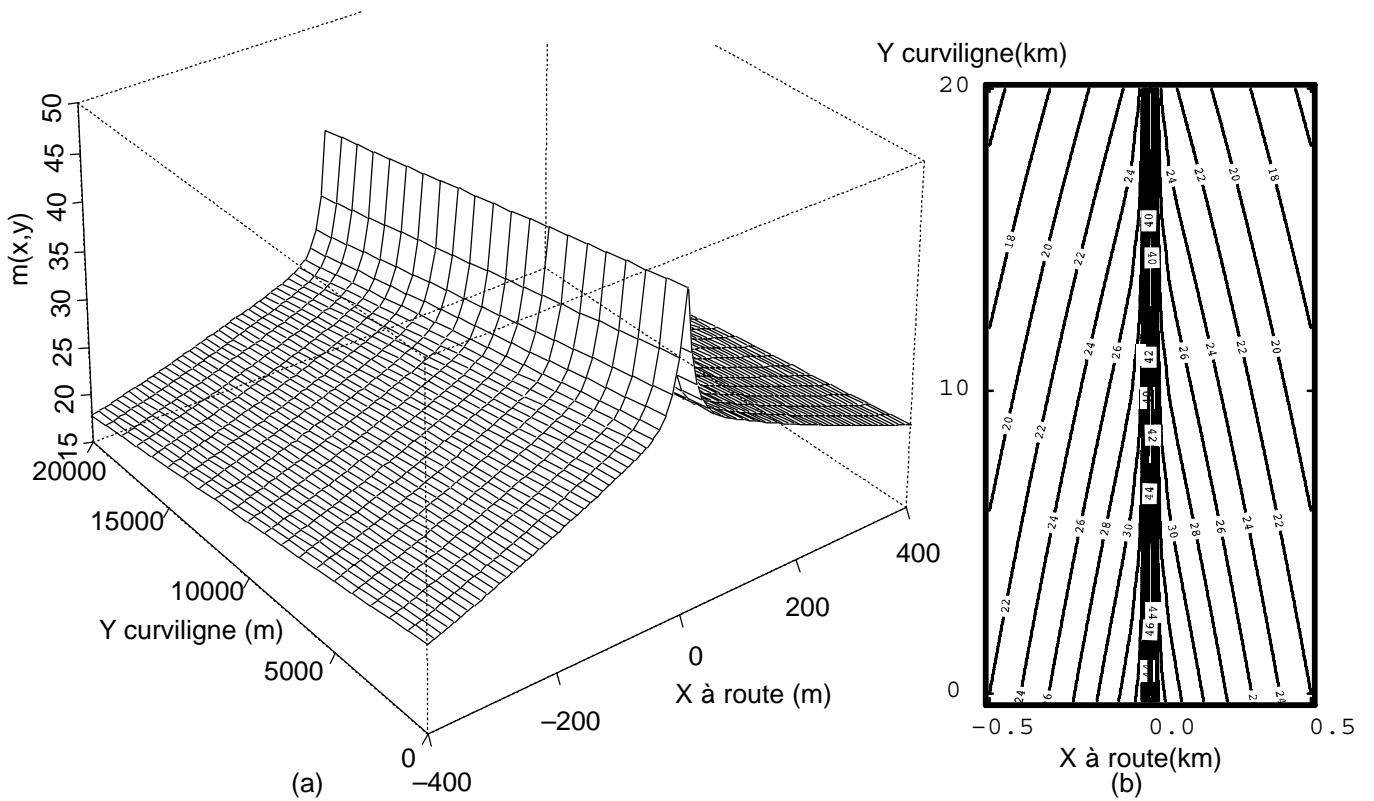


Figure 50 – Krigeage simple des résidus normés sous covariance pépitique : estimateur égal à la dérive
(a) Représentation par éléments finis (b) Projection en iso-valeurs.

4 Cas non purement pépitique

Ce krigeage, comparé au cas précédent, permet d'évaluer l'intérêt d'une modélisation du variogramme. Pour la comparaison, l'estimation est conduite avec un voisinage unique puisque tel est le cas pour $m(x,y)$, mais ce système peut aussi être mis en oeuvre avec un voisinage glissant (de même qu'une régression. C'est d'ailleurs en procédant par voisinage glissant que la dérive a été construite).

4-1 Été

On présente tout d'abord la carte des résidus estimés $RN^*(x,y)$ donnés par (8-a), avec λ^α solution du système (8-b), et la carte de l'écart-type de krigeage $\sigma_{RN}^{KS}(x,y)$ déduit de (8-e).

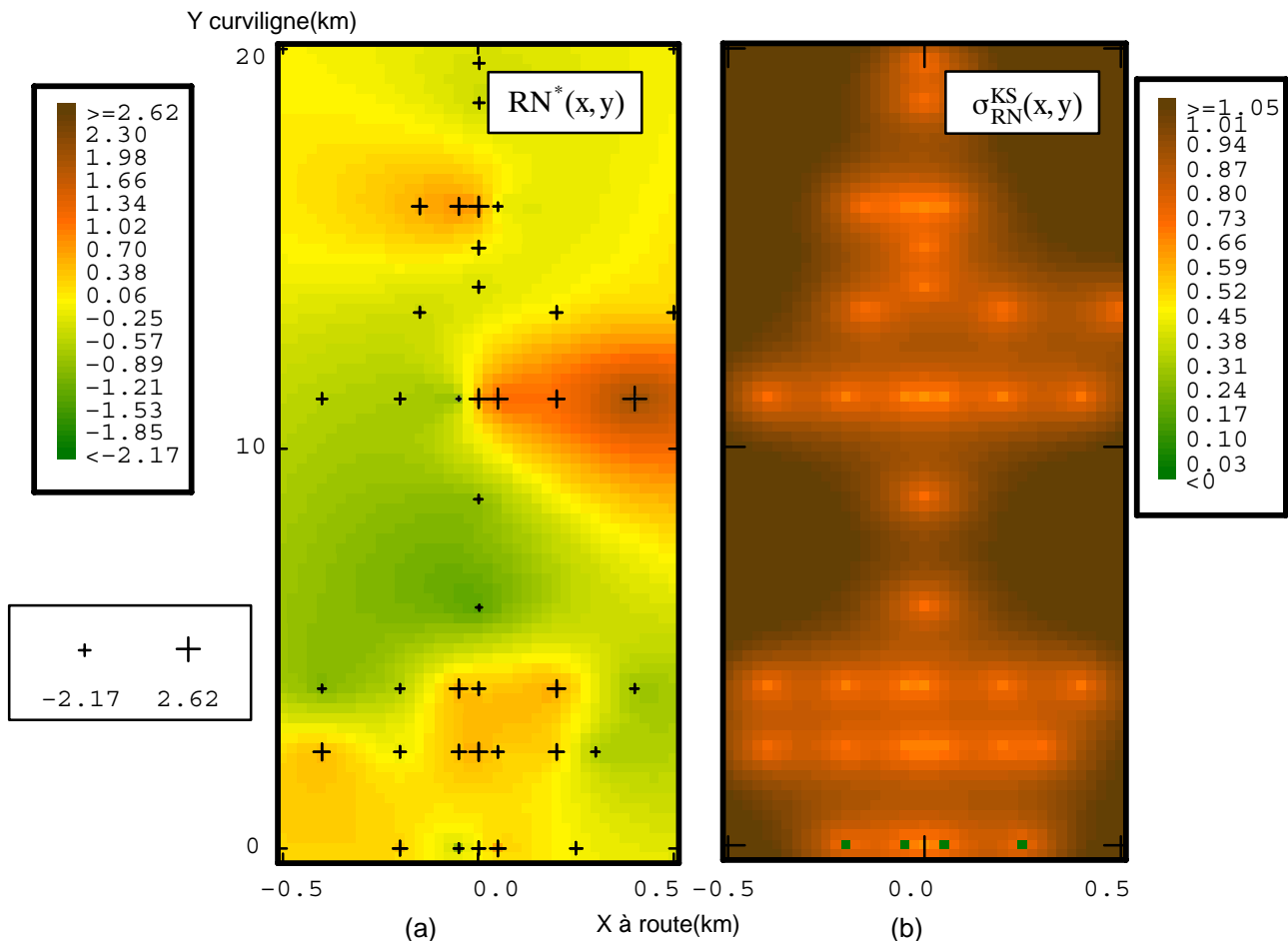


Figure 51 – Krigeage simple des résidus normés sous covariance exponentielle+pépitique. (a) Résidus estimés en couleurs et résidus aux tubes (croix proportionnelles aux valeurs). (b) Écart-types d'estimation

Sur la carte des résidus estimés, les résidus expérimentaux calculés aux tubes sont représentés par des croix dont la taille est proportionnelle à la valeur. Pour évaluer l'intérêt du krigeage comparé à la simple régression, et comprendre comment il fonctionne, il faut comparer cette carte à l'estimateur des résidus associé à la régression : une valeur nulle partout (nappe uniformément jaune). Par krigeage, la surface plane et égale à 0 se déforme pour s'adapter aux valeurs expérimentales (positives et négatives), faisant apparaître des linéaments spatiaux (lignes "d'iso-couleurs") relativement étendus.

La carte des écart-types d'estimation montre comment, dans le modèle, l'estimateur se dégrade quand on s'éloigne de l'information.

À l'aide de ces résidus, on construit l'estimateur NO_2^* du dioxyde d'azote en appliquant (8-d), l'écart-type de krigeage étant donné par (8-e). On présente en (b) la carte obtenue.

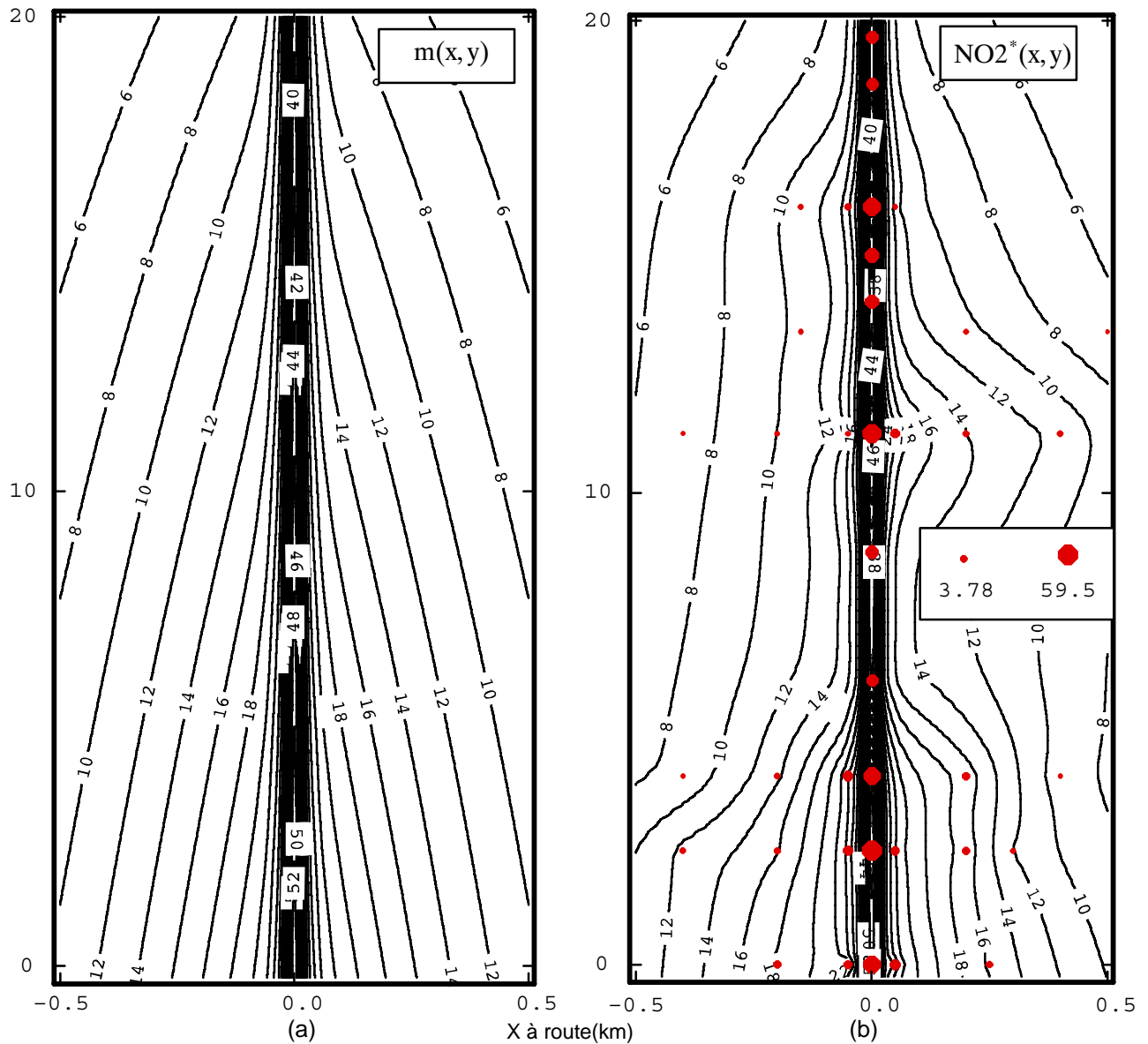


Figure 52 – (a) NO_2 estimé à partir du krigeage simple des résidus sous covariance pépitique (dérive)
 (b) NO_2 estimé à partir du krigeage simple des résidus sous covariance exponentielle+pépitique.
 En rouge les valeurs aux tubes (taille proportionnelle à la valeur)

La comparaison avec la carte de $m(x,y)$ montre les déformations subies par les lignes d'iso-valeurs pour s'adapter aux valeurs expérimentales représentées par des points rouges (taille proportionnelle à la valeur).

4-2 Hiver

On présente la carte des résidus estimés $RN^*(x, y)$ et celle de l'écart-type de krigeage $\sigma_{RN}^{KS}(x, y)$.

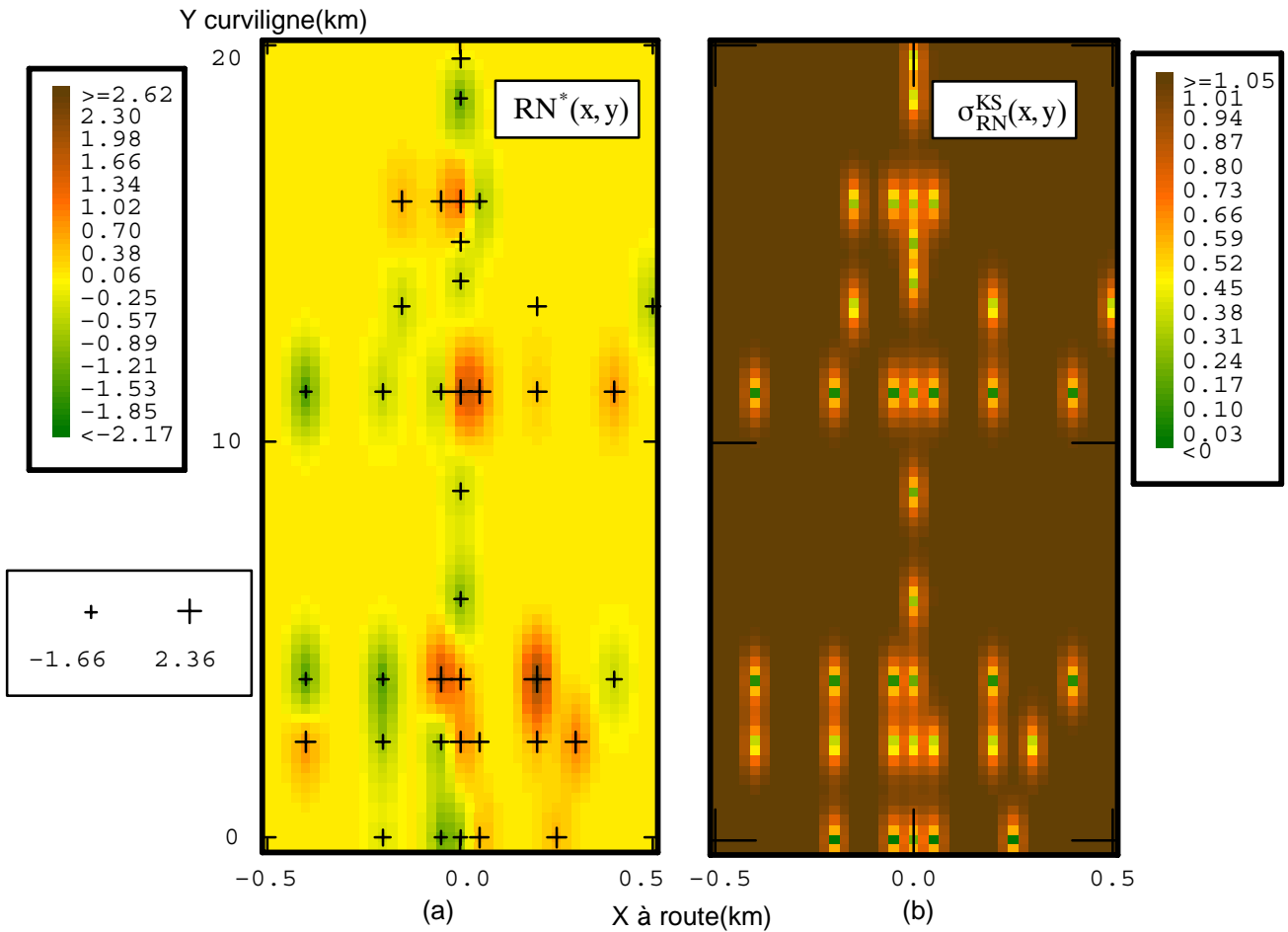


Figure 53 – Krigeage simple des résidus normés sous covariance sphérique (a) Résidus estimés en couleurs et résidus aux tubes (croix proportionnelles aux valeurs) (b) Écart-types d'estimation

Comparativement à l'été, la carte des résidus estimés présente des structures d'iso-couleurs moins étendues. La raison est liée aux portées du modèle de variogramme : 80 m en x (au lieu de 1450 m en été), 2200 m en "y" (au lieu de 10000 m en été). Ce constat montre, s'il en était besoin, l'importance des portées du variogramme sur le résultat.

Terminons par la carte de l'estimateur du dioxyde d'azote comparée à la carte de la dérive.

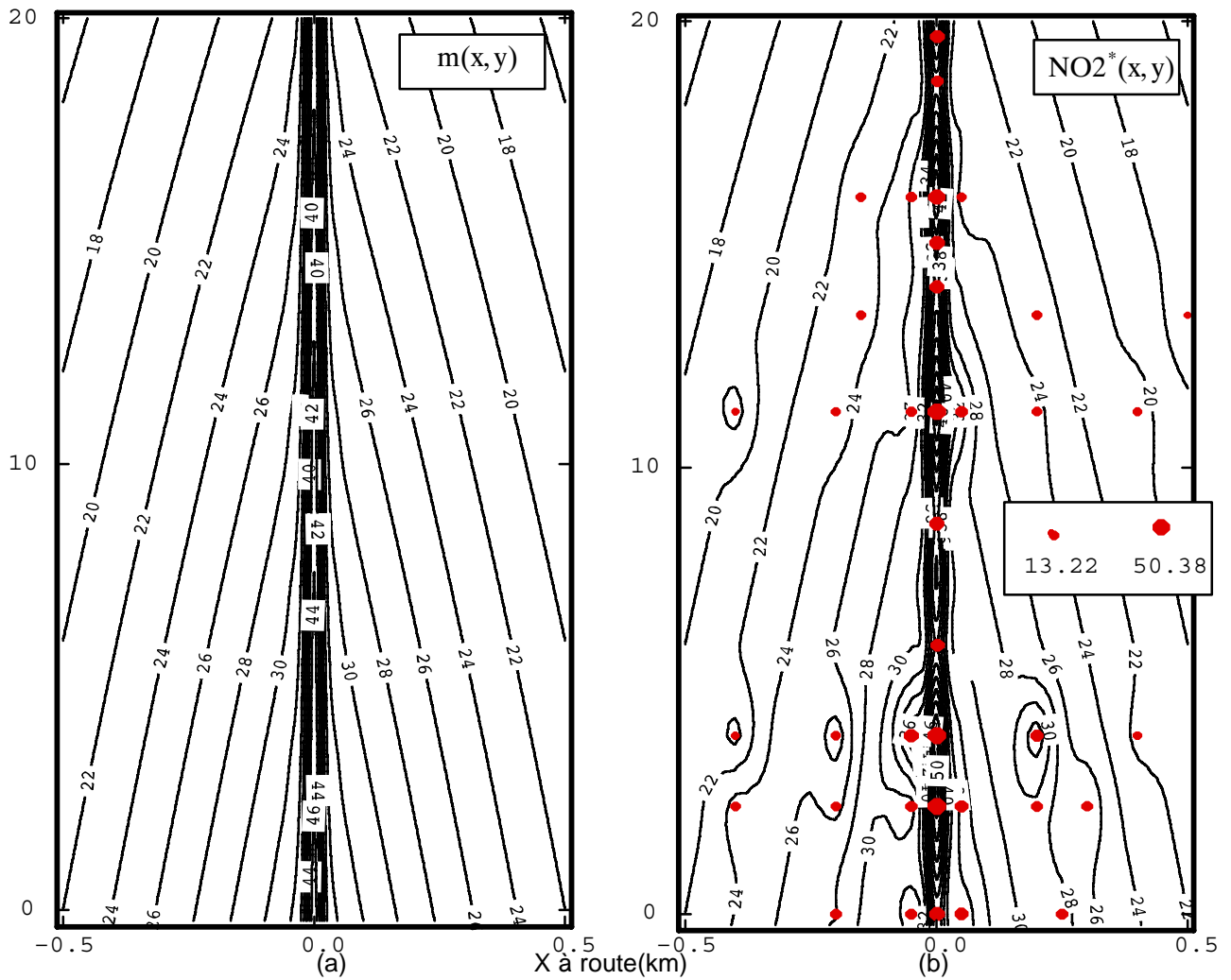


Figure 54 – (a) NO_2 estimé à partir du krigeage simple des résidus sous covariance pépitique (dérive).
 (b) NO_2 estimé à partir du krigeage simple des résidus sous covariance sphérique.
 En rouge les valeurs aux tubes (taille proportionnelle à la valeur)

Pour les noeuds de la grille qui ne sont pas aux abords immédiats de la route, le krigeage (carte de $NO_2^*(x,y)$) déforme moins la dérive (carte de $m(x,y)$) qu'il ne le faisait en été. Explications en trois points :

- Les résidus normés estimés sont moins dispersés en hiver.

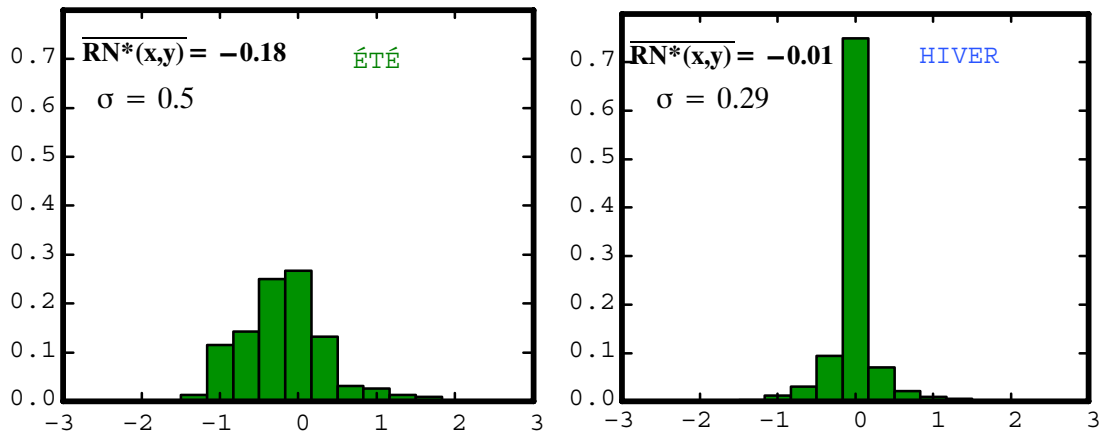
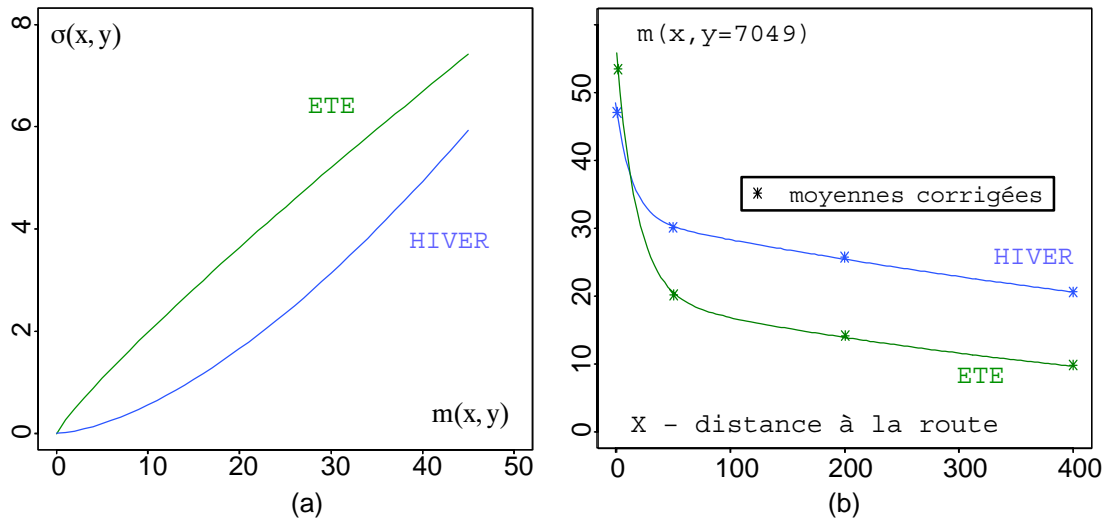


Figure 55 – Histogrammes des résidus normés estimés par krigeage simple non purement pépitique

- Le modèle $\sigma(x,y)$ (formule (6)) associe, à une valeur fixée de $m(x,y)$, un écart-type moins important en hiver qu'en été.



- Les valeurs de $m(x,y)$ sont en moyenne plus importantes en hiver quand la distance à la route est supérieure à 20 m (figure 56-b).

En conséquence de ces trois points, dans la formule de reconstruction de l'estimateur NO_2^*

$$\text{NO}_2^*(x_0, y_0) = \sigma(x_0, y_0) \text{RN}^*(x_0, y_0) + m(x_0, y_0) \quad (8\text{-d-rappel})$$

l'importance du produit $\sigma(x_0, y_0)\text{RN}^*(x_0, y_0)$, relativement à $m(x_0, y_0)$, est moins grande en hiver quand la distance à la route est supérieure à 20m. D'où les déformations moindres.

Dernière remarque. Dans le krigeage, l'influence de chaque point de mesure est une ellipse centrée sur ce point et possédant pour petit et grand axe les portées en x et en "y" du modèle de variogramme. Tout noeud de la grille situé hors de l'ellipse ne subira pas l'influence de ce point. Les portées du modèle de variogramme d'hiver (80m en x, 2200m en "y") étant moins importantes que celles du variogramme d'été (1440m en x, 10000m en "y"), l'ensemble des noeuds de la grille qui ne sont touchés par aucun point de mesure est plus important en hiver. En ces points isolés de toute mesure, le krigeage se comporte comme s'il était purement pépétique : le résidu estimé est nul et la valeur estimée du dioxyde est la dérive. Le nombre de noeuds où le résidu estimé est nul étant plus important en hiver (ce qui se voit sur l'histogramme des résidus estimés – plus de 70% de valeurs nulles), la carte d'hiver se rapprochera d'autant de la carte de la dérive.

Chapitre VII – Conclusions

1 Synthèse

1-1 Objectif

Au début de l'étude, nous ignorions que la campagne devait quantifier l'impact environnemental d'une route nationale. Il était question de cartographier la pollution de fond au dioxyde d'azote dans la vallée de la Thur et il s'agissait de voir si les variables auxiliaires permettaient d'améliorer les résultats. Les analyses n'ont pas permis de dégager de corrélation, et nous nous sommes orientés vers la résolution d'un problème de non-stationnarité. C'est à ce moment que nous avons appris l'objectif de la campagne de mesures, ce qui nous a confortés dans nos choix, et corrélativement prouvé que la campagne de mesure a été bien faite puisqu'elle a su en grande partie isoler les capteurs des sources polluantes non routières.

1-2 Les étapes

La non-stationnarité est double : elle concerne la moyenne du dioxyde d'azote, et sa variance, d'où les quatre grandes étapes de cette étude :

- Modélisation de la dérive $m(x,y)$
 - Modélisation de l'écart-type $\sigma(x,y)$
 - Variographie
 - Estimation
- **Modélisation de $m(x,y)$** – Le modèle s'appuie sur des fonctions de base choisies en examinant des sections le long de la route et transversalement. Pour pouvoir construire un modèle 3D cohérent avec ce qui est observé sur des sections planes de la surface inconnue $m(x,y)$, l'information utilisée devait être la même dans chaque direction. En première approche, nous nous sommes donc limités à 5 transects. La forme analytique de $m(x,y)$ déterminée, les coefficients multiplicatifs et additifs qu'elle contient ont été calculés par moindres-carrés sur l'ensemble des tubes exploitables.
- **Modélisation de $\sigma(x,y)$** – La dérive nous a permis de construire des résidus à moyenne stationnaire nulle. On constate alors que la variance des résidus n'est pas stationnaire et l'on établit une relation simple entre moyenne et variance. L'objet de l'étude se reporte alors sur les résidus centrés réduits, stationnaires en moyenne et en variance.
- **Variographie** – Les variogrammes des résidus centrés réduits conduisent à deux approches pour la modélisation : une approche prudente avec un modèle purement pépitique, et une approche moins évidente avec des modèles exponentiels ou sphériques dont les portées pratiques sont en hiver moins importantes qu'en été.
- **Estimation** – Quand le modèle de variogramme est purement pépitique, l'estimateur du NO_2 est la dérive elle-même et l'écart-type $\sigma(x,y)$ exprime l'écart-type d'estimation. L'estimateur est identique à la régression par moindres-carrés. Quand un modèle non purement pépitique est utilisé, c'est un krigeage simple des résidus normés qui est conduit. Les modèles non pépitiques procédant d'un certain arbitraire (surtout en hiver), l'intérêt de cette deuxième approche est essentiellement méthodologique : la comparaison avec la régression permet de décomposer les mécanismes du krigeage et montre son importance quand sa mise en oeuvre est possible.

1-3 Portée de l'étude

Que ce soit avec ou sans krigeage, l'enchaînement méthodologique proposé constitue une alternative pour résoudre le problème de la non-stationnarité spatiale. Cette étude constitue un premier pas qui mérite d'être poursuivi, peut-être dans une étude pilote où la campagne de mesures serait faite conjointement avec le géostatisticien, afin de notamment optimiser le réseau des tubes, mais aussi pour guider la modélisation par un dialogue permanent avec les chargés de campagne. Ce pourrait être l'impact environnemental d'une route, ou mieux un aéroport ou un pôle industriel, autant de nouveaux exemples qui permettraient de dégager d'autres fonctions de dérive que celles utilisées ici, et peut-être même des fonctions de covariance adaptées au phénomène. La non-stationnarité spatiale est fréquente en pollution de l'air, il me semble que cette voie de recherche doit être activement poursuivie.

2 Discussion

Abordons pour conclure certains points qui sont apparus au fil de l'étude et que nous avons laissé de côté pour ne pas alourdir l'exposé. Il s'agit de :

- L'habitat et les points isolés
- L'altitude
- Les rond-points
- Les portions de route courbes
- Le retour à l'espace géographique
- La stationnarité temporelle

2-1 Habitat et points isolés

- Des tubes du transect Bitschwiller ont été écartés dès le début car ils présentent des valeurs anormalement basses par rapport aux autres transects. L'explication donnée ultérieurement concerne l'habitat, "dense, de type rue". Si plusieurs transects de ce type avaient été dans le jeu de données, il aurait bien fallu les utiliser, ce qui aurait sans doute modifié de manière notable la dérive et peut-être même ses fonctions de base. Tel n'est pas le cas, et c'est une chance car cela permet d'identifier la portée du modèle de dérive : les calculs de cette étude ne sont valables que pour un habitat qui n'est pas du type de celui associé à Bitschwiller. La grille d'estimation autour de la route donne une image idéalisée de la répartition du NO₂ associée à quelque chose comme un habitat "moyen" (qui reste à définir). Mais la réalité du terrain ne correspond pas à cette image idéalisée : en de nombreux endroits autour de la route se trouvent des rues et des immeubles qui freinent la diffusion du NO₂. En ces endroits, l'estimateur proposé est sans doute autant éloigné de la réalité que l'est le transect Bitschwiller des autres transects. Il faudrait appliquer à l'estimateur une fonction, dépendante de l'habitat, qui modifie localement l'estimation. Ce qui nous ramène à la grille des variables auxiliaires : 1 km est un pas trop lâche pour permettre de corriger une estimation faite tous les 25 m en x (par exemple). Un moyen d'intégrer l'habitat pourrait être simplement de remettre dans le jeu de données les transects "anormaux", et ce au moment du krigeage seulement. Ils influenceraient alors localement l'estimation.

- Les 6 tubes isolés en bord de route ont une moyenne nettement plus faible que celle des 5 autres tubes, également en bord de route, mais sur transect. Ces tubes isolés ont conditionné d'une manière importante les coefficients de la dérive $m(x,y)$, le modèle de variance $\sigma(x,y)$, et ils ont induit une périodicité dans le variogramme des résidus normés. Fallait-il les utiliser dans l'estimation ? Oui car ils informent des ordonnées curvilignes où par ailleurs nous n'avons pas d'autres mesures ; non car ils induisent un excès de variance directement lié au choix de la dérive. Si nous avons modélisé cette périodicité au niveau de la dérive $m(x,y)$ (en introduisant par exemple des fonctions trigonométriques actives seulement en bordure de route), la variance $\sigma(x,y)$ des résidus, et les variogrammes des résidus normés, auraient peut-être été plus fiables. Fallait-il procéder ainsi ? Tout dépend de l'explication. S'il y a effectivement une périodicité, ce qui est peu probable, il faut la modéliser. Si la cause est par exemple liée au type d'habitat (comme pour Bitschwiller), alors il ne fallait pas utiliser ces tubes pour élaborer l'estimateur et ils ne devaient être utilisés qu'au moment de l'estimation.

2-2 Altitude

Parmi les variables auxiliaires autres que la distance à la route, seule l'altitude a pu être corrélée au NO₂. En premier lieu, on peut se demander si cette corrélation ne s'est pas dégagée simplement parce que le pas de la grille de l'altitude est 250m alors qu'il est de 1 km pour les autres variables auxiliaires. Ensuite, nous avons dans cette étude bénéficié d'une circonstance exceptionnelle : l'altitude augmente quand on s'enfonce dans la vallée, ce qui fait qu'en régionalisant le NO₂ dans le repère (distance_route, y_curviligne), on a intégré une partie de cette corrélation avec l'altitude. Ce cas de figure très particulier ne doit pas nous faire oublier que la chimie de diffusion du NO₂ dépend de l'altitude, de nombreuses études le montrent. Dans un grand nombre de cas, l'altitude devra être utilisée comme variable explicative.

2-3 Rond-points

Certains tubes présentent des concentrations anormalement fortes car ils se situent à proximité de rond-points (tube 4B par exemple). Pour ces tubes, la distance à la route (mesurée probablement à partir du centre de la route) n'est pas représentative car c'est le bord de la source de l'émission qui compte (le bord du rond-point). Mais même si la référence est le bord de la source, il est probable que la diffusion du NO_2 à partir de ces sources présente un comportement différent des autres points car les rond-points sont des zones où les voitures sont en moyenne plus nombreuses, roulent plus lentement, freinent et accélèrent. De plus, l'existence d'un rond-point signifie souvent la présence de routes transverses qui peuvent influencer les tubes à proximité. Enfin, quand une étude d'impact d'une route est faite, ces zones de fortes émissions sont importantes et doivent être prises en compte. Alors comment procéder, sachant qu'un seul rond-point dans le jeu de donnée peut à lui seul brouiller les analyses ?

Si l'étude ne comprend qu'un nombre statistiquement faible de tels points, ils doivent être écartés lors de l'inférence du modèle. Si l'inférence permet d'aboutir à une modélisation du variogramme des résidus normés (et donc qu'un krigeage non purement pépitique peut être conduit), il faut réintégrer ces points dans l'estimation et utiliser un voisinage glissant dont le diamètre permettra de limiter leur influence. Si le krigeage n'est pas possible, on peut envisager de la même manière des régressions glissantes qui incorporent ces points.

Si maintenant l'étude comprend un nombre non négligeable de rond-points, il faudra les utiliser lors de l'inférence du modèle, à l'image de ce qui a été fait pour les tubes isolés en bordure de route, avec les mêmes conséquences sur les variances et variogrammes.

Mais quelles que soient les circonstances, l'estimateur ne collera pas à la réalité aux rond-points où aucune mesure n'est disponible. Comme pour l'habitat, la solution passe peut-être par l'application, sur l'estimateur, d'une fonction dépendant d'une codification (ici, une typologie de la route).

2-4 Portions de route courbes

Cette étude a bénéficié d'une autre circonstance favorable : aux transects, la route peut être considérée comme rectiligne. Il en résulte que mis à part les problèmes de rond-points et d'habitat, toutes les concentrations sont comparables.

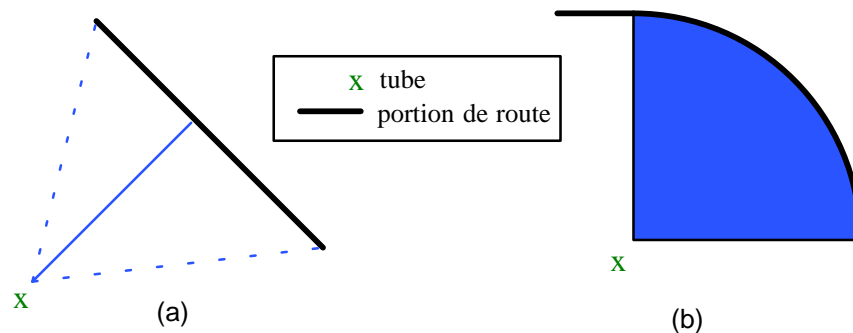


Figure 57 – (a) Route localement rectiligne. (b) Route courbe et zone d'influence par vent nul (en bleu)

Si la route présente une forte courbure, un tube situé à une distance x de la route enregistrera une concentration plus forte qu'un tube situé à la même distance x d'une route rectiligne. Les deux mesures ne sont pas directement comparables. Ce problème n'est pas identique à celui des rond-points qui présentent une émission sur la route différente et une distance_{route} non représentative.

Comment procéder alors ? Si les endroits courbes sont peu nombreux, on peut les omettre pour l'inférence, puis les intégrer au moment du krigeage, comme pour les rond-points. Si les endroits courbes sont plus nombreux, les utiliser avec les autres mesures et ignorer le problème risque de rendre les analyses difficiles. Il faut peut-être mieux remonter à la physique du phénomène et considérer chaque mesure comme une fonction d'une intégrale curviligne pondérée par une mesure de type inverse de distance (ou exponentielle négative), à l'image des potentiels gravimétrique ou magnétique. Il faudra alors définir la forme de cette fonction.

2-5 Retour à l'espace géographique

Compte tenu de la portée essentiellement méthodologique de cette étude, nous n'avons pas exprimé l'estimation dans l'espace géographique initial car cela ne présentait pas de réel intérêt. Comment procéder ? Tout d'abord, définir une grille dans l'espace géographique initial. En chaque noeud, calculer les distance_route et y_curviligne, puis calculer aux noeuds les valeurs de la dérive et de l'écart-type. L'estimation est ensuite conduite dans le repère de travail en présentant la grille déformée comme un simple ensemble de points. Il suffit alors d'exprimer l'estimateur dans ses coordonnées géographiques initiales pour obtenir le résultat.

2-6 Stationnarité temporelle

Dans cette étude, il a été convenu d'étudier le dioxyde d'azote saisonnier, sans autre considération sur la nature de cette moyenne saisonnière. Examinons l'évolution dans le temps du NO₂ journalier.

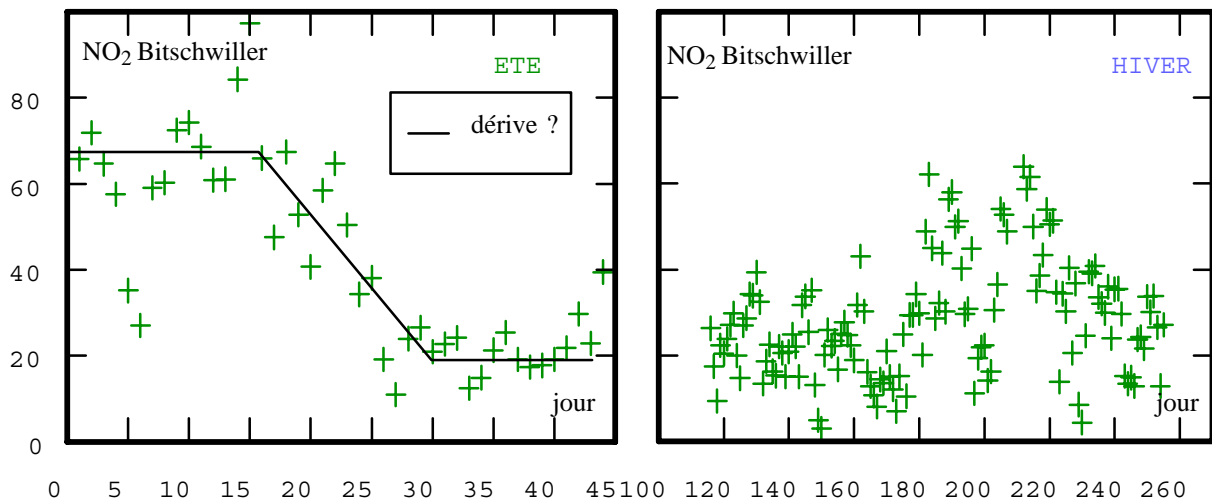


Figure 58 – NO₂ journalier mesuré à la station Bitschwiller (45 jours en été, 140 jours en hiver)

Ces graphiques représentent les mesures prises aux stations pendant la campagne. Il ne s'agit plus de tubes passifs, mais ces mesures, réputées plus fiables, donnent une indication du comportement temporel. (On rappelle que les mesures des tubes passifs en été couvrent les jours 1 à 85, et en hiver les jours 171 à 255).

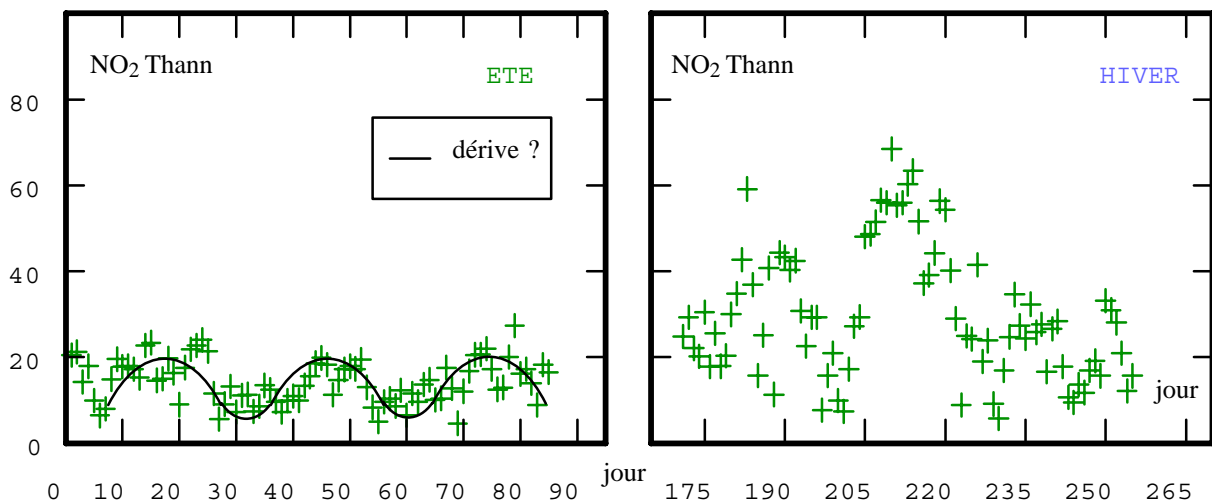


Figure 59 – NO₂ journalier mesuré à la station Thann (85 jours en été, 85 jours en hiver)

Pour Bitschwiller en été, on constate une non stationnarité à cette échelle d'observation. Si on la voit comme une “dérive” (au sens de comportement moyen autour duquel les mesures fluctuent, comme pour $m(x,y)$, mais ici dans le temps), cette dérive temporelle serait majoritairement linéaire, précédée et suivie de paliers. Pour Thann en été, on peut y voir un phénomène cyclique. En hiver et pour ces deux stations, le comportement est moins évident (attention, les plages de temps couvertes par les stations sont différentes d'une figure

à l'autre). Dans tous les cas, on peut s'interroger sur la représentativité d'une moyenne saisonnière sur des données qui ne sont pas stationnaires dans le temps, sauf dans le cas périodique, et encore : il suffit de diminuer ou d'augmenter d'une demi-période le domaine sur lequel s'effectue la moyenne saisonnière pour obtenir un résultat très différent. On peut se dire que les moyennes affectées aux tubes ont été faites partout dans l'espace sur les mêmes 85 jours en été et les mêmes 85 jours en hiver, ce qui les rend comparables entre elles et justifie leur cartographie spatiale. Mais dans la mesure où, dans cette approche de type dérive, la dérive temporelle n'est pas du même type d'un point à l'autre de l'espace, est-ce vraiment le cas ? Quand la dérive temporelle est linéaire, on peut considérer la moyenne saisonnière comme la quantité associée au jour moyen situé au centre de l'intervalle temporel utilisé. Quand on a affaire à une dérive périodique, on peut s'imposer de moyenniser sur un nombre fini de périodes et interpréter la moyenne saisonnière comme la constante autour de laquelle le NO_2 oscille. Mais quand, en vue d'une cartographie spatiale, on utilise simultanément des mesures à tendance temporelle périodique et d'autres à tendance temporelle linéaire, quelle interprétation peut-on avoir du résultat ?

Ce problème est particulièrement délicat et ne saurait être résolu en quelques lignes. J'ai choisi de conclure par ces interrogations car je me demande s'il ne faudrait pas étendre à l'espace-temps l'approche spatiale de la non-stationnarité illustrée dans cette étude.

Annexe– Modèles possibles de krigeage

Selon le degré de confiance accordé à la dérive $m(x,y)$, nous proposons deux modèles de krigeage autres que le krigeage simple des résidus :

- 1) Krigeage Universel du NO_2 (KU)
- 2) Krigeage Universel plus Général du NO_2 (KUG)

On adopte l'écriture suivante

$$f(x_0, y_0) = f_0$$

Les résidus normés $\text{RN}(x,y)$ donnés par (7) sont supposés stationnaires d'ordre 2, d'espérance nulle et de covariance $C(h)$.

$m(x,y)$ et $\sigma(x,y)$ sont donnés par (4) et (6).

1 KU : Krigeage Universel du NO_2

La variable d'intérêt est exprimée en fonction des résidus normés

$$\text{NO}_2(x,y) = \sigma(x,y) \text{RN}(x,y) + a_0 + a_1 m(x,y) \quad (9)$$

où a_0 et a_1 sont inconnus. On choisit d'estimer la variable NO_2 en des points (x_0, y_0) où elle n'est pas connue par

$$\text{NO}_2^* = \sum_{\alpha} \lambda^{\alpha} \text{NO}_2_{\alpha} \quad (10)$$

moyenne pondérée de données du voisinage, l'erreur commise devant satisfaire les habituelles contraintes de non-biais et de variance minimum.

L'erreur d'estimation a pour espérance mathématique

$$E[\text{NO}_2^* - \text{NO}_2_0] = a_0 \left(\sum_{\alpha} \lambda^{\alpha} - 1 \right) + a_1 \left(\sum_{\alpha} \lambda^{\alpha} m_{\alpha} - m_0 \right) \quad (11)$$

On impose en conséquence

$$\sum_{\alpha} \lambda^{\alpha} = 1 \quad (12-a)$$

$$\sum_{\alpha} \lambda^{\alpha} m_{\alpha} = m_0 \quad (12-b)$$

et sous ces contraintes, la variance de l'erreur a pour expression

$$E\left[\left(\sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \text{RN}_{\alpha} - \sigma_0 \text{RN}_0 \right)^2 \right] = \sum_{\alpha} \sum_{\beta} \lambda^{\alpha} \lambda^{\beta} \sigma_{\alpha} \sigma_{\beta} C_{\alpha\beta} - 2 \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \sigma_0 C_{\alpha 0} + \sigma_0^2 C(0)$$

Après introduction des paramètres de Lagrange μ_0 et μ_1 , la minimisation de cette variance sous les contraintes (12-a) et (12-b) conduit au système de krigeage

$$\sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \sigma_{\beta} C_{\alpha\beta} + \mu_0 + \mu_1 m_{\beta} = \sigma_{\beta} \sigma_0 C_{\beta 0} \quad \forall \beta : 1, n \quad (12-c)$$

et la variance de l'erreur vaut

$$\sigma_{\text{NO}_2}^{\text{KU}}(x_0, y_0) = \sigma_0^2 C(0) - \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \sigma_0 C_{\alpha 0} - \mu_0 - \mu_1 m_0 \quad (12-d)$$

Ce système de krigeage se présente comme un krigeage universel classique dans lequel la covariance est affectée d'une correction régionale. Il n'est pas équivalent à un krigeage simple des résidus car c'est une fonction affine de la dérive $m(x,y)$ qui est estimée à l'échelle du voisinage de krigeage. On aura tout intérêt à l'utiliser en voisinage glissant. La positivité de la covariance non stationnaire est assurée par le fait que $C(h)$ est de type positif.

2 KUG : Krigeage Universel plus Général du NO₂

On peut aussi considérer que $m(x,y)$ est une combinaison linéaire de l'ensemble $\{X^{1,l:1,5}\} = \{1, e^{-\frac{|x|}{14}}, e^{-\frac{|x|}{940}}, ye^{-\frac{|x|}{14}}, ye^{-\frac{|x|}{940}}\}$ pour l'hiver par exemple et choisir un estimateur linéaire qui réestime implicitement les coefficients multiplicatifs c_l des formules (4). La variance théorique d'estimation va augmenter, mais l'estimation pourrait s'avérer en pratique meilleure car la dérive possèdera un degré de liberté plus important pour s'adapter aux fluctuations locales (voisinage glissant).

Les contraintes sont

$$\sum_{\alpha} \lambda^{\alpha} X_{\alpha}^1 = X_0^1 \quad \forall l : 1, 4 \quad (13-a)$$

les équations deviennent

$$\sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \sigma_{\beta} C_{\alpha\beta} + \mu_l X_{\beta}^1 = \sigma_{\beta} \sigma_0 C_{\beta 0} \quad \forall \beta : 1, n \quad (13-b)$$

et la variance de l'erreur vaut

$$\sigma_{NO_2}^{2KUG}(x_0, y_0) = \sigma_0^2 C(0) - \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \sigma_0 C_{\alpha 0} - \sum_{l=1}^5 \mu_l X_0^1 \quad (13-c)$$

la covariance utilisée étant la même que précédemment. La stricte positivité conditionnelle de la covariance non stationnaire $\sigma_{\alpha} \sigma_{\beta} C_{\alpha\beta}$ ayant déjà été examinée précédemment, il faut ici considérer l'indépendance linéaire des fonctions X^1 sur les données pour s'assurer de l'existence et l'unicité d'une solution au système (13).

Soit à vérifier que si, $\forall \alpha : 1, n$ on a $c_l X_{\alpha}^1 = 0$, alors $c_l = 0 \quad \forall l : 1, 4$.

Sur les données du voisinage, aucune des fonctions ne doit s'exprimer linéairement en fonction des autres, et ce pour éviter une dégénérescence du système linéaire (lignes ou colonnes identiques, à un facteur multiplicatif près). Cette circonstance peut arriver si tous les points du voisinage appartiennent à un seul transect, une configuration à éviter et un problème qui disparaît quand au moins un point d'un autre transect est utilisé. Concernant les problèmes de stabilité numérique (dépendance linéaire *numérique*), les deux exponentielles utilisées ont des paramètres suffisamment différents pour ne jamais donner des lignes ou des colonnes dans la matrice qui soient numériquement proches.

Index des figures et tableaux

Chapitre I – Problématique

1	NO ₂ moyen, variable d'intérêt	
	Figure 1 Carte du relief de l'Alsace et zone d'étude	5
	Figure 2 55 tubes et 3 stations	6
2	Données auxiliaires	
	Figure 3 Projection des variables dans les plans factoriels	7
4	Sélection et regroupement des tubes	
	Figure 4 Décalage entre route SIG et position des tubes	8
5	Repère de travail	
	Figure 5 Les 39 tubes régionalisés dans le repère de travail	9
	Figure 6 Nuage de corrélation entre la distance sur la route et l'altitude	9
	Figure 7 (a) projection de toutes les variables retenues dans le plan factoriel	10
	(b) Dioxyde_été dans le plan factoriel des coordonnées de travail	
	(c) Dioxyde_hiver dans le plan factoriel des coordonnées de travail	

Chapitre II – Modélisation de la dérive $m(x,y)$

1	Symétrie axiale	
	Figure 8 Projection des valeurs dans un plan perpendiculaire à la route	11
2	Construction de la dérive	
	2-1 Comportement en y, ordonnée curviligne	
	Figure 9 Régressions linéaires par classe de distance à la route (été)	12
	Tableau 1 Pentes et constantes des régressions linéaires par classe de distance à la route	12
	2-2 Comportement en x, distance à la route	
	Figure 10 Valeurs du dioxyde_été aux transects	12
	Figure 11 Régressions linéaires par classe de distance à la route (été)	13
	Figure 12 Moyennes du dioxyde par classe de distance à la route (été)	13
	(a) Moyenne corrigées pour se ramener à l'ordonnée curviligne 7049m	
	(b) Moyennes des logarithmes du dioxyde	
	Figure 13 Ajustement, par le modèle bi-exponentiel, des moyennes corrigées (été)	14
	Figure 14 Régressions bi-exponentielles à ordonnée curviligne fixée (été)	14,15
	Tableau 2 Coefficients obtenus pour chaque ordonnée curviligne (été)	15
	2-3 Expression de $m(x,y)$	
	Figure 15 Coefficients obtenus pour chaque ordonnée curviligne	16
	Tableau 3 Coefficients obtenus par régression linéaire sur $c_1(y)$ et $c_2(y)$	17
	Figure 16 Régressions linéaires sur $c_1(y)$ et $c_2(y)$	17
	Figure 17 Modèle bi-exponentiel issu des régressions transversales, confronté aux	

pentes $a(x)$ et constante $b(x)$ des régressions linéaires longitudinales	17
2–4 Synthèse : $m(x,y)$ sur T1	
Tableau 4 Coefficients de la dérive selon trois approches	18
(1) Approche par sections transversale et longitudinale	
(2) et (3) Coefficients obtenus en une seule régression sur les 26 tubes de T1	
2–5 Remarques	
Figure 18 Régressions linéaires sur $c_1(y)$ et $c_2(y)$ (été)	19
Figure 19 Importances relatives de $c_1(y)$ et $c_2(y)$ par rapport à leur somme	19
3 $m(x,y)$ sur T2	
Figure 20 Influence des tubes à moins de 2m de la route et isolés	20
Tableau 5 Pentes $a(x)$ et constantes $b(x)$ des régressions linéaires longitudinales sur les tubes situés à moins de 2m de la route	20
Tableau 6 Coefficients obtenus par régression globale sur deux ensembles de tubes	20
(1) Régression globale sur T2, ensemble des 39 tubes exploitables	
(2) Régression globale sur T1, ensemble des 26 tubes sur transect	
4 Hiver	
4–1 Construction	
Figure 21 Ajustement des moyennes corrigées par le modèle bi-exponentiel	21
Figure 22 Importances relatives de $c_1(y)$ et $c_2(y)$ par rapport à leur somme	21
Figure 23 Modèle bi-exponentiel issu des régressions transversales, confronté aux pentes $a(x)$ et constante $b(x)$ des régressions linéaires longitudinales	21
Tableau 7 Coefficients obtenus par régression linéaire sur les coefficients $c_1(y)$ et $c_2(y)$ issus des régressions bi-exponentielles sur transect	22
Tableau 8 Coefficients de la dérive selon trois approches	22
(1) Approche par sections transversale et longitudinale	
(2) et (3) Coefficients obtenus en une seule régression sur les 26 tubes de T1	
4–3 $m(x,y)$ sur T2	
Figure 24 Régressions linéaires par classe de distance à la route (Hiver)	22
Tableau 9 Coefficients de la dérive par régression globale sur deux ensembles de tubes	23
(1) Régression globale sur T2, ensemble des 39 tubes exploitables	
(2) Régression globale sur T1, ensemble des 26 tubes sur transect	
Figure 25 Histogrammes des résidus été et hiver centrés sur la moyenne modélisée	23
6 Discussion	
6–2 Dérive nulle	
Tableau 10 Constantes de la dérive en été et en hiver, sur deux ensembles de tubes	24
(1) Par régression globale sur T2, ensemble des 39 tubes exploitables	
(2) Par régression globale sur T1, ensemble des 26 tubes sur transect	

Chapitre III – Modélisation de l'écart–type

1 Remarque préliminaire

Figure 26 Moyennes (non corrigées) du NO₂ par classes de distances à la route 25

2 Par classes de distances à la route

Figure 27 Ecart–types du NO₂ fonction des moyennes du NO₂ 26

Tableau 11 Coefficients des régressions avec constante des écart–types expérimentaux
fonction des moyennes expérimentales du NO₂ 26

Figure 28 Régressions linéaires avec constante des écart–types expérimentaux du NO₂
fonction des moyennes expérimentales du NO₂ 26

Tableau 12 Coefficients des régressions sans constante des écart–types expérimentaux
fonction des moyennes expérimentales du NO₂ 26

Figure 29 Régressions linéaires sans constante des écart–types expérimentaux du NO₂
fonction des moyennes expérimentales du NO₂ 27

3 Par les résidus

3–1 Remarque préliminaire

Figure 30 Résidus fonction du modèle de moyenne établi sur les 39 tubes de T2 27

Figure 31 Résidus fonction du modèle de moyenne établi sur les 26 tubes de T1 28

3–2 Modélisation

Figure 32 Régressions des écart–types expérimentaux des résidus 28

Tableau 13 Performances des modèles linéaires et puissance 29

Figure 33 Résidus expérimentaux normés fonction du modèle de moyenne 29

Figure 34 Histogrammes des résidus normés par l'écart–type modélisé 29

4 Discussion

4–1 Autres modèles

Figure 35 Autres modèles possibles d'expression de l'écart–type fonction de la moyenne 30
(a) Modèle constant, l'écart–type ne dépend pas de la moyenne
(b) Modèle en escalier, un écart–type est affecté à des groupes de valeurs

Figure 36 Variogrammes transversaux des résidus non normés 30

4–2 Variance nulle

Figure 37 Régressions avec constante des écart–types des résidus fonction des moyennes 31

Chapitre IV – Variographie des résidus normés

1 Remarques préliminaires

1–3 Tolérance angulaire

Figure 38 Les 39 tubes exploitables (ensemble T2) dans le repère de travail 33

2 Variogrammes expérimentaux – Été

2–1 En h_x : traversant

Figure 39 Variogrammes transversaux traversants. (a) Calculé sur les 39 tubes.

	(b) Nuée mettant en évidence l'influence de 2 tubes. (c) Calculé sans.	34
2-2	En h_x : gauche, droit, moyen, traversant	
Figure 40	Variogrammes transversaux. (a) À gauche ou à droite de la route, et leur moyenne pondérée par les effectifs. (b) Moyen et traversant	35
2-3	En h_y : gauche, droit, moyen, traversant	
Figure 41	Variogrammes longitudinaux. (a) À gauche ou à droite de la route, et leur moyenne pondérée par les effectifs. (b) Moyen et traversant	35
3 Variogrammes expérimentaux – Hiver		
3-1	En h_x : traversant	
Figure 42	Variogrammes transversaux traversants. (a) Calculé avec les 39 tubes. (b) Nuée mettant en évidence l'influence de 3 tubes. (c) Calculé sans.	36
3-2	En h_x : gauche, droit, moyen, traversant	
Figure 43	Variogrammes transversaux. (a) À gauche ou à droite de la route, et leur moyenne pondérée par les effectifs. (b) Moyen et traversant	36
Figure 44	Variogrammes longitudinaux. (a) À gauche ou à droite de la route, et leur moyenne pondérée par les effectifs. (b) Moyen et traversant	36
4 Ajustements		
4-1	Été : par un effet de pépite	
Figure 45	Ajustement par un effet de pépite égal à la variance des résidus normés (a) Variogramme transversal. (b) Variogramme longitudinal	37
4-2	Été : par une structure non purement pépitique	
Figure 46	Ajustement par un modèle pépitique + exponentiel à anisotropie géométrique (a) Portée transversale : 1450 m. (b) Portée longitudinale 10 km	37
4-3	Hiver : par un effet de pépite	
Figure 47	Ajustement par un effet de pépite égal à la variance des résidus normés (a) Variogramme transversal. (b) Variogramme longitudinal	38
4-4	Hiver : par une structure non pépitique	
Figure 48	Ajustement par un modèle sphérique à anisotropie géométrique (a) Portée transversale : 80 m. (b) Portée longitudinale 2200 m	38

Chapitre V – Estimations

1 Rappels

Tableau 14	Paramètres de la dérive été et hiver	39
Tableau 15	Paramètres de l'écart-type été et hiver	39

3 Cas pépitique

3-1 Été

Figure 49	Krigeage simple des résidus normés sous covariance pépitique (a) Représentation par éléments finis. (b) Projection en iso-valeurs	41
------------------	---	----

3-2 Hiver

Figure 50	Krigeage simple des résidus normés sous covariance pépitique	
------------------	--	--

	(a) Représentation par éléments finis. (b) Projection en iso-valeurs	42
4 Cas non purement pépitique		
4-1 Été		
Figure 51	Krigeage simple des résidus normés sous covariance exponentielle + pépitique (a) Résidus estimés et résidus aux tubes. (b) Écart-types d'estimation	43
Figure 52	(a) NO ₂ estimé à partir du krigeage simple des résidus sous covariance pépitique. (b) NO ₂ estimé à partir du krigeage sous covariance pépitique + exponentielle	44
4-2 Hiver		
Figure 53	Krigeage simple des résidus normés sous covariance sphérique (a) Résidus estimés et résidus aux tubes. (b) Écart-types d'estimation	45
Figure 54	(a) NO ₂ estimé à partir du krigeage des résidus sous covariance pépitique (b) NO ₂ estimé à partir du krigeage simple des résidus sous covariance sphérique	46
Figure 55	Histogrammes des résidus normés estimés par krigeage non purement pépitique	46
Figure 56	Été / Hiver. (a) Modèles d'écart-types. (b) Modèles de dérive	47

Chapitre VI – Conclusions

2 Discussion

2-4 Portions de route courbes

Figure 57	(a) Route localement rectiligne. (b) Route courbe et zone d'influence	50
------------------	---	----

2-6 Stationnarité temporelle

Figure 58	NO ₂ journalier mesuré à Bitschwiller (45 jours en été, 140 jours en hiver)	51
Figure 59	NO ₂ journalier mesuré à Thann (85 jours en été, 85 jours en hiver)	51

Index des formules

Chapitre II – Modélisation de la dérive $m(x,y)$

2 Construction de la dérive

2-1 Comportement en y , ordonnée curviligne

$$(1) \quad m(x_i, y) = a(x_i) y + b(x_i) \quad 12$$

2-2 Comportement en x , distance à la route

$$(2) \quad m(x, y_j) = c_1(y_j) e^{-\frac{|x|}{18}} + c_2(y_j) e^{-\frac{|x|}{550}} \quad 14$$

2-4 Synthèse : $m(x,y)$ sur T1

$$(3\text{-été}) \quad m(x,y) = c_1 e^{-\frac{|x|}{18}} + c_2 e^{-\frac{|x|}{550}} + c_3 y e^{-\frac{|x|}{18}} + c_4 y e^{-\frac{|x|}{550}} \quad 18$$

$$(4\text{-été}) \quad m(x,y) = c_0 + c_1 e^{-\frac{|x|}{18}} + c_2 e^{-\frac{|x|}{550}} + c_3 y e^{-\frac{|x|}{18}} + c_4 y e^{-\frac{|x|}{550}} \quad 18$$

4 Hiver

4-1 Construction

$$(3\text{-hiver}) \quad m(x,y) = c_1 e^{-\frac{|x|}{14}} + c_2 e^{-\frac{|x|}{940}} + c_3 y e^{-\frac{|x|}{14}} + c_4 y e^{-\frac{|x|}{940}} \quad 22$$

4-2 synthèse : $m(x,y)$ sur T1

$$(4\text{-hiver}) \quad m(x,y) = c_0 + c_1 e^{-\frac{|x|}{14}} + c_2 e^{-\frac{|x|}{940}} + c_3 y e^{-\frac{|x|}{14}} + c_4 y e^{-\frac{|x|}{940}} \quad 22$$

5 Histogrammes des résidus

$$(5) \quad R(x,y) = \text{NO2}(x,y) - m(x,y) \quad 23$$

Chapitre III – Modélisation de l'écart-type

3 Par les résidus

$$(6) \quad \sigma(x,y) = d m^p(x,y) \quad 28$$

$$(7) \quad \text{RN}(x,y) = \frac{\text{NO2}(x,y) - m(x,y)}{\sigma(x,y)} \quad 29$$

Chapitre V – Estimations

2 Equations du krigeage simple (KS) des résidus

$$(8\text{-a}) \quad \text{RN}^*(x_0, y_0) = \sum_{\alpha} \lambda_{\alpha} \text{RN}(x_{\alpha}, y_{\alpha}) \quad 39$$

$$(8\text{-b}) \quad \sum_{\alpha} \lambda_{\alpha} C_{\alpha\beta} = C_{\beta 0} \quad \forall \beta : 1, n \quad 40$$

$$(8\text{-c}) \quad \sigma_{\text{RN}}^{\text{KS}}(x_0, y_0) = \text{VAR}[\text{RN}^*(x_0, y_0) - \text{RN}(x_0, y_0)] = C(0) - \sum_{\alpha} \lambda_{\alpha} C_{\alpha 0} \quad 40$$

$$(8-d) \text{NO2}^*(x_0, y_0) = \sigma(x_0, y_0) \text{RN}^*(x_0, y_0) + m(x_0, y_0) \quad 40$$

$$(8-e) \sigma_{\text{NO2}}^{2\text{KS}}(x_0, y_0) = \sigma^2(x_0, y_0) \sigma_{\text{RN}}^{\text{KS}}(x_0, y_0) = \sigma^2(x_0, y_0) C(0) - \sum_{\alpha} \lambda^{\alpha} \sigma^2(x_0, y_0) C_{\alpha 0} \quad 40$$

Annexe – Modèles possibles de krigeage

1 KU : Krigeage Universel du NO₂

$$(9) \text{NO2}(x, y) = \sigma(x, y) \text{RN}(x, y) + a_0 + a_1 m(x, y) \quad 53$$

$$(10) \text{NO2}_0^* = \sum_{\alpha} \lambda^{\alpha} \text{NO2}_{\alpha} \quad 53$$

$$(11) E[\text{NO2}_0^* - \text{NO2}_0] = a_0 \left(\sum_{\alpha} \lambda^{\alpha} - 1 \right) + a_1 \left(\sum_{\alpha} \lambda^{\alpha} m_{\alpha} - m_0 \right) + \quad 53$$

$$(12-a) \sum_{\alpha} \lambda^{\alpha} = 1 \quad 53$$

$$(12-b) \sum_{\alpha} \lambda^{\alpha} m_{\alpha} = m_0 \quad 53$$

$$(12-c) \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \sigma_{\beta} C_{\alpha\beta} + \mu_0 + \mu_1 m_{\beta} = \sigma_{\beta} \sigma_0 C_{\beta 0} \quad \forall \beta : 1, n \quad 53$$

$$(12-d) \sigma_{\text{NO2}}^{2\text{KU}}(x_0, y_0) = \sigma_0^2 C(0) - \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \sigma_0 C_{\alpha 0} - \mu_0 - \mu_1 m_0 \quad 53$$

2 KUG : Krigeage Universel plus Général du NO₂

$$(13-a) \sum_{\alpha} \lambda^{\alpha} X_{\alpha}^l = X_0^l \quad \forall l : 1, 4 \quad 54$$

$$(13-b) \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \sigma_{\beta} C_{\alpha\beta} + \mu_1 X_{\beta}^l = \sigma_{\beta} \sigma_0 C_{\beta 0} \quad \forall \beta : 1, n \quad 54$$

$$(13-c) \sigma_{\text{NO2}}^{2\text{KUG}}(x_0, y_0) = \sigma_0^2 C(0) - \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha} \sigma_0 C_{\alpha 0} - \sum_{l=1}^5 \mu_l X_0^l \quad 54$$