



Laboratoire Central de Surveillance de la Qualité de l'Air



Traitements numériques

Application de méthodes géostatistiques pour la détermination de zones de représentativité en concentration et la cartographie des dépassements de seuils

Décembre 2011

Programme 2010

Maxime Beauchamp, Laure Malherbe, Laurent Létinois





PREAMBULE

Le Laboratoire Central de Surveillance de la Qualité de l'Air

Le Laboratoire Central de Surveillance de la Qualité de l'Air est constitué de laboratoires de l'École des Mines de Douai, de l'INERIS et du LNE. Il mène depuis 1991 des études et des recherches finalisées à la demande du Ministère chargé de l'environnement, et en concertation avec les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA). Ces travaux en matière de pollution atmosphérique ont été financés par la Direction Générale de l'Énergie et du Climat (bureau de la qualité de l'air) du Ministère de l'Écologie, du Développement durable et de l'Énergie. Ils sont réalisés avec le souci constant d'améliorer le dispositif de surveillance de la qualité de l'air en France en apportant un appui scientifique et technique au MEDDE et aux AASQA.

L'objectif principal du LCSQA est de participer à l'amélioration de la qualité des mesures effectuées dans l'air ambiant, depuis le prélèvement des échantillons jusqu'au traitement des données issues des mesures. Cette action est menée dans le cadre des réglementations nationales et européennes mais aussi dans un cadre plus prospectif destiné à fournir aux AASQA de nouveaux outils permettant d'anticiper les évolutions futures.



Application de méthodes géostatistiques pour la détermination de zones de représentativité en concentration et la cartographie des dépassements de seuils

Laboratoire Central de Surveillance
de la Qualité de l'Air

Traitements numériques

Programme financé par la
Direction Générale de l'Énergie et du Climat (DGEC)

2011

Maxime Beauchamp – Laure Malherbe – Laurent Létinois

Ce document comporte 53 pages (hors couverture et annexes)


	Rédaction	Vérification	Approbation
NOM	M. Beauchamp L. Malherbe	B. Bessagnet	L. Rouïl
Qualité	Ingénieurs de l'unité MOCA Direction des Risques Chroniques	Responsable de l'unité MOCA Direction des Risques Chroniques	Responsable du pôle DECI Direction des Risques Chroniques
Visa			

TABLE DES MATIÈRES

RESUME	9
REMERCIEMENTS	11
1. INTRODUCTION	12
1.1 Contexte et objectif.....	12
1.2 Analyse des exigences législatives et réglementaires.....	12
1.3 Approche proposée	14
1.4 Contenu du rapport	17
2. CARTOGRAPHIE DES CONCENTRATIONS DE FOND ET DE PROXIMITE A L'ECHELLE URBAINE	18
2.1 Remarques préliminaires	18
2.2 Organisation de la méthodologie.....	18
2.3 Application.....	20
2.3.1 Cartographie des concentrations de fond	20
2.3.2 Prise en compte de la pollution de proximité	21
2.3.3 Cartographie incluant pollution de fond et pollution de proximité.....	24
3. REPRESENTATIVITE EN CONCENTRATION	26
3.1 Remarques préliminaires	26
3.2 Définition proposée	26
3.3 Organisation de la méthodologie.....	27
3.4 Critères de rattachement aux stations et stabilité temporelle	28
3.5 Application.....	30
3.5.1 Représentativité et pollution de fond.....	30
3.5.2 Exploitation complémentaire des résultats	32
3.5.3 Représentativité et pollution intégrant fond et proximité	33
4. DEPASSEMENTS DE SEUIL	35
4.1 Remarques préliminaires	35
4.2 Approche probabiliste.....	35
4.3 Organisation de la méthodologie.....	36
4.4 Application.....	36
4.4.1 Dépassement et pollution de fond	36
4.4.2 Risques de première et seconde espèces.....	37
4.4.3 Dépassement et pollution intégrant fond et proximité	38
4.4.4 Discussion sur les risques de première et seconde espèces	38

5. RECALAGE ANNUEL DES RESULTATS	40
5.1 Cartographie des concentrations de fond.....	40
5.2 Représentativité en concentration et dépassements de seuil	42
6. SYNTHESE ET CONCLUSION	45
7. REFERENCES	50
8. LISTE DES ANNEXES	53

RESUME

Conformément aux Directives européennes sur la qualité de l'air et à leur transposition en droit français, les AASQA doivent évaluer la représentativité spatiale de leurs sites de mesure. De plus, si un dépassement de seuil réglementaire est constaté en un ou plusieurs sites, elles doivent estimer la surface et la population exposée à ce dépassement dans toute la zone de surveillance.

Ces exigences nécessitent de disposer d'une information sur la distribution spatiale des concentrations, selon un pas de temps adapté à l'échelle temporelle et au seuil considérés. Différents moyens d'obtenir une telle information, qui reposent sur la modélisation et/ou le traitement de données de campagnes, sont envisageables.

La présente étude est consacrée au NO₂ en moyenne annuelle. Elle propose une méthodologie probabiliste fondée sur la réalisation de campagnes d'échantillonnage par tubes à diffusion passive. L'usage complémentaire de résultats de modélisation fera l'objet d'une étape ultérieure. Les calculs s'organisent en trois parties. La première fait appel aux techniques d'estimation de la géostatistique. Elle consiste à cartographier les concentrations moyennes annuelles de NO₂ sur le domaine considéré. Elle constitue un préalable indispensable aux deux parties suivantes, qui peuvent être mises en œuvre indépendamment : l'estimation des zones de représentativité en concentration des stations et la délimitation des zones de dépassement de la valeur limite annuelle (40 µg/m³).

L'approche a été appliquée dans son ensemble à deux cas d'étude : Montpellier et Troyes, où Air Languedoc-Roussillon et ATMO Champagne-Ardenne ont réalisé des campagnes d'échantillonnage (respectivement en 2007 et 2009). Les concentrations moyennes annuelles de NO₂ ont été cartographiées sur ces deux agglomérations en se limitant d'abord à la pollution de fond puis en intégrant l'influence des émissions routières. Les critères et paramètres définissant la représentativité ont été évalués et ajustés à l'aide de tests de sensibilité. Pour Montpellier comme pour Troyes, l'estimation de la représentativité spatiale des stations de fond montre une bonne couverture du domaine de surveillance. La disposition des zones diffère toutefois selon la ville. En ce qui concerne Montpellier, les deux zones de représentativité sont disjointes et se complètent : l'une correspond au centre de l'agglomération, l'autre forme une couronne autour de ce centre. En revanche, la représentativité de la station trafic de Montpellier n'a pu être établie, faute d'une précision d'estimation suffisante aux abords de ce site. Dans le cas de Troyes, les zones de représentativité associées aux deux stations de fond sont imbriquées. Dans les deux agglomérations, de multiples points situés le long des axes ont été classés en zone de dépassement. Il serait intéressant de contrôler ces estimations par des données complémentaires de mesure.

Les méthodes présentées reprennent des développements antérieurs, issus de travaux cités en introduction. Elles ont été mises au point avec le logiciel R. Si elles offrent une réponse efficace aux questions de la représentativité et des dépassements, les exemples étudiés en indiquent aussi les limites, à savoir une sensibilité au plan d'échantillonnage et à certains paramètres de calcul (écart de concentration définissant la représentativité, risques statistiques). Les zones de représentativité et de dépassement seront délimitées d'autant plus finement que la stratégie de mesure permet une estimation plus fiable des concentrations de fond et une modélisation plus précise du surcroît de concentration aux abords des routes. Qu'il s'agisse de l'échantillonnage spatial ou temporel, de nombreuses préconisations ont déjà été émises (LCSQA, GT plans d'échantillonnage) ; il est recommandé de s'y référer.

Une synthèse des méthodes utilisées et des résultats obtenus, assortie d'indications et de précautions d'usage, est fournie en conclusion. Des travaux complémentaires et approfondissements – réalisés ou en cours d'étude - y sont également mentionnés : analyse de la représentativité et des zones de dépassement pour les PM_{10} et le benzène, utilisation de la modélisation urbaine, quantification de la population exposée.

REMERCIEMENTS

Les auteurs remercient les associations Air Languedoc-Roussillon et ATMO Champagne-Ardenne pour les données fournies et leur collaboration. Ils remercient également le Centre de Géosciences (équipe géostatistique) de Mines ParisTech pour son appui dans les développements méthodologiques.

1. INTRODUCTION

1.1 CONTEXTE ET OBJECTIF

Les concepts de **représentativité spatiale d'une station de mesure** et de **zone de dépassement de seuil** apparaissent à plusieurs reprises dans la législation européenne et française. Tous deux se rattachent principalement au problème de l'exposition : à quels niveaux de pollution sont exposées les populations en dehors des sites de mesure ? Les données des stations incluses dans une zone de surveillance, par exemple une agglomération, décrivent-elles entièrement les concentrations sur ce domaine ? A ces interrogations s'en ajoutent deux autres, dès lors que les seuils établis pour la protection de la santé humaine sont franchis en un ou plusieurs sites : sur quelle surface s'étend le dépassement ? Combien d'habitants sont touchés ?

Avant toute question d'ordre méthodologique, la première difficulté soulevée par ces concepts réside dans leur définition même : s'ils sont plusieurs fois mentionnés dans la législation, celle-ci n'en fournit aucune définition précise.

Différents travaux sur ces sujets, dont il sera fait état plus loin, ont déjà été engagés par les AASQA et le LCSQA. Tirant parti de ces expériences, la présente étude a pour objet de clarifier les éléments qui, **dans le cadre imposé par la surveillance réglementaire et le rapportage**, peuvent définir une zone de représentativité spatiale et une zone de dépassement, et de décrire une méthodologie permettant de délimiter de telles zones.

1.2 ANALYSE DES EXIGENCES LEGISLATIVES ET REGLEMENTAIRES

Préalablement à tout travail de définition et de méthodologie, il convient de revenir aux textes législatifs et réglementaires et d'examiner dans quelles situations ceux-ci imposent de caractériser la représentativité spatiale d'une station de mesure et de délimiter une zone de dépassement.

Les extraits de la législation européenne et française qui abordent ces sujets sont référencés en annexe (2). De cet état des lieux, ressortent les points suivants:

Représentativité spatiale :

- La représentativité spatiale est établie **pour chaque station de mesure**.
- La zone de représentativité correspond au minimum à un **ensemble de points situés au voisinage de la station**. Pour les stations de fond et les stations industrielles, elle forme une surface (km²) dont l'étendue, en ordre de grandeur, dépend de la typologie du site. Pour les stations de trafic, elle est

assimilée, selon les cas à une portion d'axe (km) (directive 2008/50/CE) ou à une surface (directive 2004/107/CE). A cet ensemble, peuvent s'ajouter des **points plus éloignés similaires au site de mesure.**

- La directive de 2008 ne précise pas sur quelle base temporelle cette représentativité doit être évaluée. Elle ne dit pas non plus de quelle façon est appréciée la similarité de deux sites.

Une étude réalisée pour la Commission Européenne par l'Agence environnementale autrichienne (Spangl et al., 2007) propose de définir la zone de représentativité d'une station par polluant et par statistique annuelle de concentration (moyenne, centile), selon plusieurs critères:

- homogénéité spatiale des concentrations : la différence de concentration par rapport à la station est inférieure à une certaine valeur, fonction du polluant et de la statistique considérée.
- homogénéité des conditions de dispersion (déterminées par la situation topographique et climatique) et des émissions de polluants ;
- étendue maximale (l'étude mentionne une distance maximale de l'ordre de 100 km autour de la station).

Ce document, cité à titre indicatif, n'a pas valeur de référence.

Zone de dépassement :

- Selon la **directive 2008/50/CE** , les zones de dépassement sont établies **pour les zones de surveillance et les agglomérations dans lesquelles des dépassements de seuils réglementaires ont été mesurés par les stations fixes.** Les seuils concernés sont :
 - les valeurs limites et les valeurs cibles, dans le cas de la déclaration annuelle sur la qualité de l'air ainsi que des plans et programmes ;
 - les objectifs de long terme et les seuils d'évaluation inférieurs et supérieurs, dans le cas de la déclaration annuelle sur la qualité de l'air ;
 - les seuils d'information et d'alerte, pour l'information du public.
- Selon l'**arrêté français du 21 octobre 2010**, elles sont établies **pour les stations de mesures fixes où un dépassement des valeurs limites ou des valeurs cibles a été constaté.**
- Les zones de dépassement se définissent
 - de manière synthétique, comme une simple valeur: une superficie en km² (cas général) ou une distance en km lorsque les dépassements sont liés à la pollution d'origine routière ;
 - sous forme de carte.

- A une zone de dépassement, est associée, selon le polluant et le seuil de concentration considérés, une information sur la population exposée (en nombre d'habitants) ou sur la superficie de végétation ou d'écosystème exposée au dépassement (en km²).

1.3 APPROCHE PROPOSEE

Nous proposons de distinguer de la même façon les notions de représentativité spatiale et de dépassement et de les aborder de la manière décrite ci-après. Notons que des simplifications théoriques seront volontairement opérées afin de faciliter l'application des méthodes dans un contexte opérationnel.

Représentativité spatiale :

A la première question: *les stations présentes dans un domaine de surveillance caractérisent-elles d'une manière exhaustive, ou que l'on peut considérer comme telle, les concentrations ?*, nous répondrons par la délimitation de **zones de représentativité en concentration**. Ces zones seront établies **par station, par polluant, et pour une statistique de concentration donnée** (moyenne annuelle, saisonnière, journalière, etc.).

Comme dans Spangl et al. (2007), une zone de représentativité en concentration sera constituée des points où la concentration du polluant est estimée proche de la valeur mesurée à la station.

Elle sera délimitée selon une approche probabiliste, à partir d'une estimation spatiale des concentrations (krigeage) et de l'écart-type de l'erreur d'estimation associée. Cette méthodologie repose sur des travaux réalisés par Air Normand, en collaboration avec l'équipe de géostatistique de Mines ParisTech (Cori, 2005 ; Bobbia et al., 2008) et sur une étude LCSQA qui en reprend les développements (Cárdenas et Malherbe, 2007).

Deux étapes seront considérées: **estimation individuelle** de la zone de représentativité d'une station ; affinement de cette zone de représentativité **en tenant compte des stations environnantes**.

Zone de dépassement :

Si un dépassement de seuil est constaté en un site de mesure, quelle est l'étendue géographique de ce dépassement ? La zone de représentativité en concentration pour le polluant incriminé n'apporte pas de réponse exacte et complète à cette question. Il se peut qu'elle surestime ou sous-estime la zone de dépassement, comme l'illustrent les schémas ci-après (Figure 1 a) à d)).

Pour répondre aux exigences de la directive 2008/50/CE, une **approche globale** est proposée. Si dans une zone de surveillance ou une agglomération, une ou plusieurs stations de mesure enregistrent un dépassement, on cherchera à délimiter la zone touchée **sur l'ensemble du domaine**, sans en rattacher les différentes composantes (susceptibles de se recouper) à chaque station en particulier. Les zones de dépassement seront établies **par polluant** et pour **chaque valeur seuil considérée**.

Comme pour la représentativité spatiale, une approche probabiliste est retenue, fondée sur l'exploitation de résultats de krigeage. **Une cartographie délimitant la zone de dépassement** en sera le résultat, conformément à l'arrêté du 21 octobre 2010 (article 10.IV).

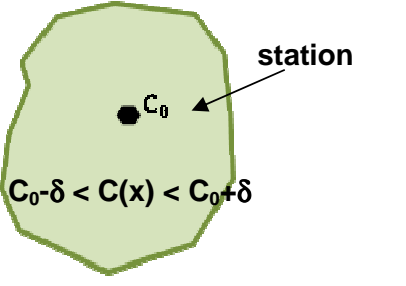
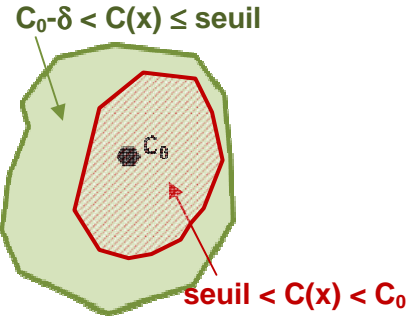
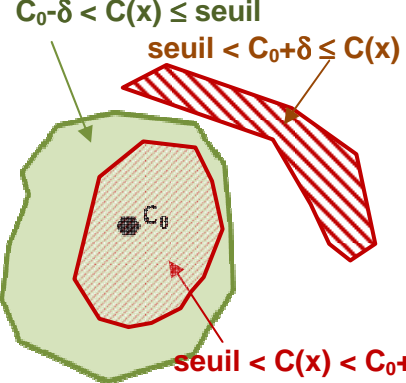
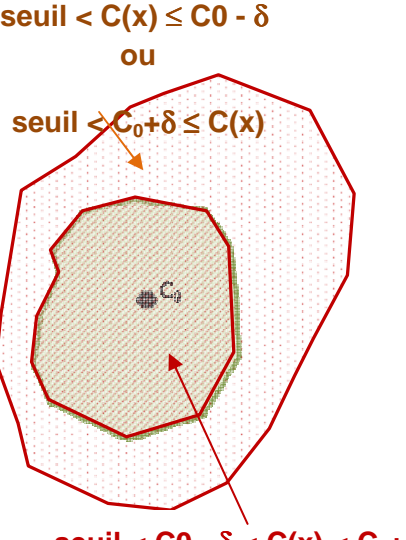
 <p>station</p> <p>C_0</p> <p>$C_0 - \delta < C(x) < C_0 + \delta$</p>	<p>a) Zone de représentativité en concentration : ensemble des points x dont la concentration s'écarte d'au maximum une certaine valeur δ de la concentration mesurée C_0 :</p> $C_0 - \delta < C(x) < C_0 + \delta$
 <p>$C_0 - \delta < C(x) \leq \text{seuil}$</p> <p>$\text{seuil} < C(x) < C_0 + \delta$</p>	<p>b) La concentration C_0 mesurée par la station dépasse légèrement le seuil ($C_0 > \text{seuil} > C_0 - \delta$). La zone de dépassement se restreint à une surface incluse dans la zone de représentativité en concentration. Sur le reste de cette zone de représentativité et à l'extérieur de celle-ci, les concentrations sont inférieures au seuil.</p>
 <p>$C_0 - \delta < C(x) \leq \text{seuil}$</p> <p>$\text{seuil} < C_0 + \delta \leq C(x)$</p> <p>$\text{seuil} < C(x) < C_0 + \delta$</p>	<p>c) La concentration C_0 mesurée par la station dépasse légèrement le seuil ($C_0 > \text{seuil} > C_0 - \delta$). C'est la même situation qu'en b) à ceci près qu'à l'extérieur de la zone de représentativité, s'étend un secteur pollué dont les niveaux de concentration excèdent nettement C_0. La zone de dépassement est composée d'une surface incluse dans la zone de représentativité et du secteur plus pollué.</p>
 <p>$\text{seuil} < C(x) \leq C_0 - \delta$ ou $\text{seuil} < C_0 + \delta \leq C(x)$</p> <p>$\text{seuil} < C_0 - \delta < C(x) < C_0 + \delta$</p>	<p>d) La concentration C_0 mesurée par la station est très supérieure au seuil ($C_0 \gg \text{seuil}$). La zone de dépassement couvre toute la zone de représentativité en concentration et s'étend également au-delà, incluant :</p> <ul style="list-style-type: none"> - des secteurs particulièrement pollués dans lesquels les concentrations sont supérieures à $C_0 + \delta$; - des secteurs moins pollués dans lesquels les concentrations, quoique inférieures à $C_0 - \delta$, dépassent le seuil.

Figure 1: a) Représentation schématisée d'une zone de représentativité en concentration. b) à d) Illustration pour quelques exemples des différences possibles entre zone de représentativité en concentration et zone de dépassement de seuil.

1.4 CONTENU DU RAPPORT

L'approche proposée est développée à **l'échelle d'une agglomération**, avec l'exemple **du calcul des concentrations de NO₂ en moyenne annuelle**.

Le rapport s'organise en trois parties principales incluant éléments méthodologiques et illustrations sur des cas concrets :

- La première porte sur la cartographie des concentrations, étape préalable à la définition des zones de représentativité aussi bien qu'à l'étude des zones de dépassement. Cette étape consiste à estimer les valeurs de fond par krigeage, à partir de données de campagnes (échantillonnage passif) et de variables auxiliaires. Si l'on s'intéresse à la pollution de proximité routière, elle inclut l'ajout d'un surplus de concentration le long des axes. A chaque valeur estimée elle fait correspondre un écart-type d'erreur.
- La seconde partie montre comment les résultats de cette estimation et leur incertitude associée permettent d'évaluer, moyennant certaines hypothèses, la représentativité spatiale des stations de mesure.
- Enfin, la troisième partie décrit la façon dont ces mêmes résultats peuvent servir à estimer des probabilités de dépassement de seuil et délimiter des zones de dépassement.

Les agglomérations de Montpellier et de Troyes ont été retenues comme domaines d'application. Des campagnes de mesure du NO₂ s'y sont respectivement déroulées en 2007 (Air Languedoc-Roussillon) et 2009 (ATMO Champagne-Ardenne). Les représentations relatives à Montpellier figurent dans le corps du rapport, les résultats obtenus sur Troyes sont fournis en annexe (4).

Les développements méthodologiques s'accompagnent de programmes informatiques conçus avec le logiciel R. A partir des résultats de l'étape préliminaire de cartographie (concentrations estimées et incertitude associée) et de paramètres d'entrée, ces programmes ont pour fonction de délimiter des zones de représentativité spatiale et de dépassement de seuil (cf. annexe 7). Ils seront mis à la disposition des AASQA sur le site Internet du LCSQA, en complément du présent rapport.

2. CARTOGRAPHIE DES CONCENTRATIONS DE FOND ET DE PROXIMITE A L'ECHELLE URBAINE

2.1 REMARQUES PRELIMINAIRES

Les méthodes présentées dans la suite du rapport pour la définition de la représentativité spatiale des stations fixes (chapitre 3) et pour l'estimation des surfaces en dépassement (chapitre 0) supposent la connaissance des concentrations de NO₂ - ici en moyenne annuelle - en tout point du domaine d'étude.

Il convient donc au préalable d'estimer spatialement ces concentrations. Dans cette étude, des données d'échantillonnage issues de campagnes par tubes à diffusion passive sont exploitées par des techniques géostatistiques de krigeage.

Notons que la plupart des dépassements de la valeur limite annuelle relative au NO₂ sont observés sur des sites de proximité : aussi l'estimation des concentrations de NO₂ doit elle tenir compte à la fois de la pollution de fond et du surplus de concentration lié au trafic routier.

2.2 ORGANISATION DE LA METHODOLOGIE

On recherche une procédure générale qui permette de cartographier les concentrations moyennes annuelles de NO₂ en incluant à la fois la pollution de fond et la pollution de proximité. La méthodologie s'appuie sur des publications antérieures du LCSQA (Malherbe *et al.*, 2008, 2006, 2005). Elle se développe en trois étapes (Tableau 1).

La première consiste à exploiter les données issues d'une ou plusieurs campagnes de mesure par tubes à diffusion passive afin d'estimer les concentrations de fond sur le domaine d'étude.

La seconde a pour objet d'évaluer les concentrations de NO₂ à proximité des routes. Celles-ci dépendent de plusieurs facteurs, en particulier les émissions de NO_x, les transformations chimiques subies par les gaz d'échappement ainsi que les conditions de dispersion liées au bâti. Kiesewetter *et al.* (2012) proposent ainsi d'établir un modèle paramétrique simple de chimie (Düring *et al.*, 2011), qui rend compte de la conversion NO_x/NO₂ au niveau des axes routiers. Dans la présente étude, une méthode sensiblement différente est retenue, qui consiste à estimer le surplus de concentration lié au trafic automobile à l'aide de variables explicatives. Un modèle de régression linéaire est construit en fonction de variables telles que les émissions de NO_x dans un rayon fixé ou encore la distance par rapport aux axes routiers.

La troisième étape réalise la combinaison des deux types d'informations : le modèle de surplus est appliqué aux axes routiers du domaine et le résultat de cette modélisation, ajouté à l'estimation des concentrations de fond.

Tableau 1 : Etapes de la méthodologie de cartographie des concentrations de NO₂

Etape 1	Echelle : urbaine	Polluant : NO ₂
Objectif : Cartographie de la pollution de fond		
Moyen : Application du krigeage aux données d'échantillonnage. La technique de krigeage proposée est celle de la dérive externe qui permet de prendre en compte des variables auxiliaires dans l'estimation.		
Données requises :		
<ul style="list-style-type: none"> - Données de concentration issues d'une campagne par tubes à diffusion passive : prise en compte des observations uniquement sur les sites de fond. - Connaissance des variables auxiliaires prises en compte par le krigeage sur la grille d'estimation et aux points d'observations. 		

Etape 2	Echelle : locale	Polluant : NO ₂
Objectif : Modélisation du surplus de concentration lié au trafic routier		
Moyen : Mise en place d'un modèle linéaire prédictif du surplus de concentration lié au trafic en fonction de variables jugées significatives lors de la construction du modèle. On notera le surplus Δ_{TF} .		
Données requises :		
<ul style="list-style-type: none"> - Données de concentration issues d'une campagne par tubes à diffusion passive : prise en compte des observations uniquement sur les sites de fond. - Connaissance des variables explicatives du surplus de concentration sur les sites de proximité 		

Etape 3	Echelle : urbaine + locale	Polluant : NO ₂
Objectif : Cartographie de la pollution intégrant le fond et le trafic		
Moyen : Calcul de Δ_{TF} le long des axes routiers et ajout des valeurs calculées à la carte de fond.		
Données requises :		
<ul style="list-style-type: none"> - Cartographie de la pollution de fond (valeurs estimées et incertitude) : elle est obtenue par l'étape 1 - Formule de prédiction du Δ_{TF} : elle est obtenue par l'étape 2 - Localisation des points du domaine où le modèle de Δ_{TF} doit être appliqué et connaissance en ces points des variables utilisées dans le modèle de Δ_{TF} 		

2.3 APPLICATION

2.3.1 CARTOGRAPHIE DES CONCENTRATIONS DE FOND

On cherche à estimer la pollution de fond sur l'ensemble du domaine à partir des données de tubes (sites de fond). La méthode employée est celle du krigeage avec dérive externe, qui permet d'introduire des informations secondaires dans l'estimation (voir par exemple de Fouquet et al., 2007).

A Montpellier, l'échantillonnage comprend **82 sites de fond** (cf. Figure 3), équipés en tubes durant **8 périodes de 2 semaines** (d'août à fin décembre 2009). Les concentrations moyennes annuelles de NO₂ en ces points ont été calculées par Air Languedoc-Roussillon. Elles intègrent une correction statistique (établie à l'aide des données de stations fixes) destinée à compenser les effets de l'échantillonnage temporel¹.

Différentes variables auxiliaires - émissions de NO_x, densité de population, taux d'urbanisation, ainsi que leurs transformées logarithmiques, dans un rayon de 500m, 1km et 2km - ont été préalablement calculées aux points expérimentaux. Une seule est finalement retenue dans le krigeage : le logarithme translaté des émissions de NO_x dans un rayon de 2 km. Ce choix découle d'une étude de corrélation et de régression entre la concentration moyenne annuelle de NO₂ et ces variables. Il est conforme à des travaux antérieurs sur le NO₂ dans lesquels le rayon de 2km se révèle généralement le plus adapté à la cartographie de fond (Malherbe et Wroblewski, 2006).

La figure ci-après (Figure 2) présente les concentrations de fond de NO₂ estimées en moyenne annuelle (2007).

Tous les calculs ont été effectués à la fois avec les logiciels Isatis (Géovariations) et R (cran.r-project.org). Les résultats obtenus avec ces deux outils concordent ; la comparaison est fournie en annexe (6).

¹ D'après les stations fixes, la moyenne des mesures sur la durée de la campagne sous-estime d'environ 12% la moyenne sur l'année complète. En conséquence, un facteur multiplicatif de 1/0,88 a été appliqué uniformément à l'ensemble des moyennes brutes issues de l'échantillonnage. Nous avons, à titre indicatif, comparé ces résultats à une reconstitution site par site de la moyenne : l'outil de reconstitution par régression, disponible sur le site du LCSQA (<http://www.lcsqa.org/outil/missions-diverses/strategie-surveillance/reconstitution-donnees>), a été utilisé à cette fin. Les résultats diffèrent peu des valeurs calculées par Air Languedoc-Roussillon ; ces dernières ont été conservées dans la suite de l'étude.

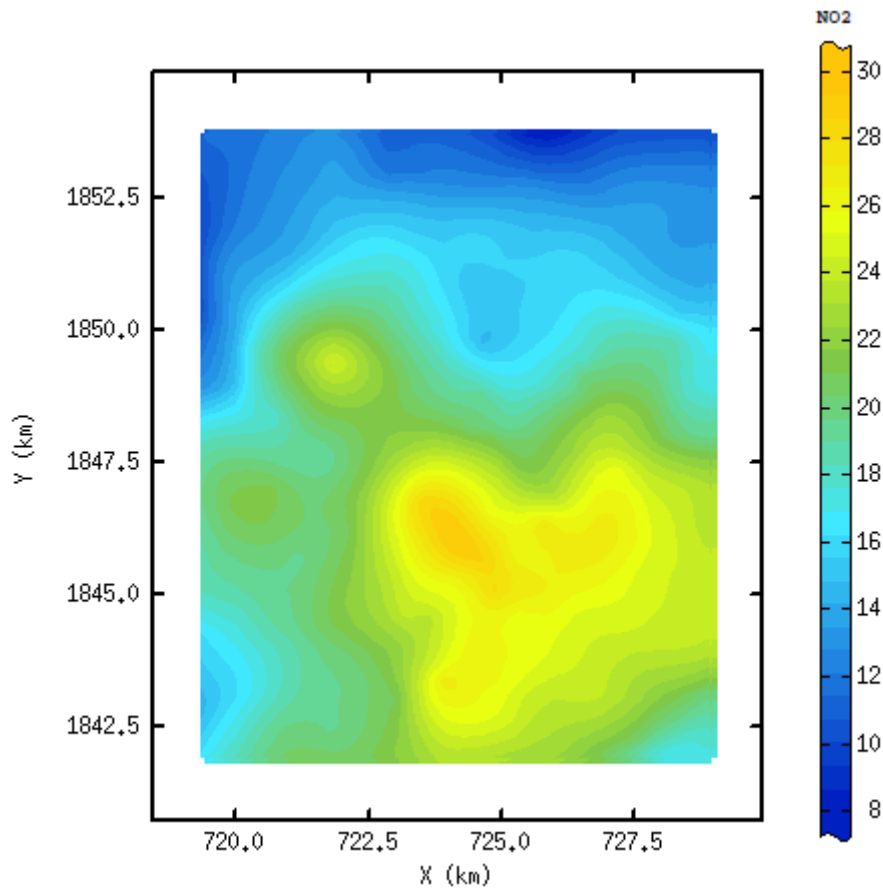


Figure 2 : Concentrations de fond de NO₂ estimées par krigeage avec dérive externe sur Montpellier. Moyenne annuelle de 2007.

2.3.2 PRISE EN COMPTE DE LA POLLUTION DE PROXIMITE

2.3.2.1 ESTIMATION DES CONCENTRATIONS DE FOND SUR LES SITES DE PROXIMITE

Afin de pouvoir calculer le surplus de pollution lié au trafic routier sur les sites de proximité, il est nécessaire, dans un premier temps, d'estimer la pollution de fond en ces points.

Dans le cas de Montpellier, la campagne d'échantillonnage incluait une dizaine de sites de proximité. On y estime, toujours par krigeage avec dérive externe, les concentrations de fond.

On peut ainsi représenter simultanément les concentrations de proximité et de fond respectivement mesurées et estimées en ces points: il s'agit des points "doubles" sur la cartographie ci-dessous (Figure 3). Ceux-ci se répartissent préférentiellement le long de la seconde ligne de tramway, la campagne par tubes ayant été mise en place pour quantifier l'impact de cette nouvelle ligne sur la qualité de l'air.

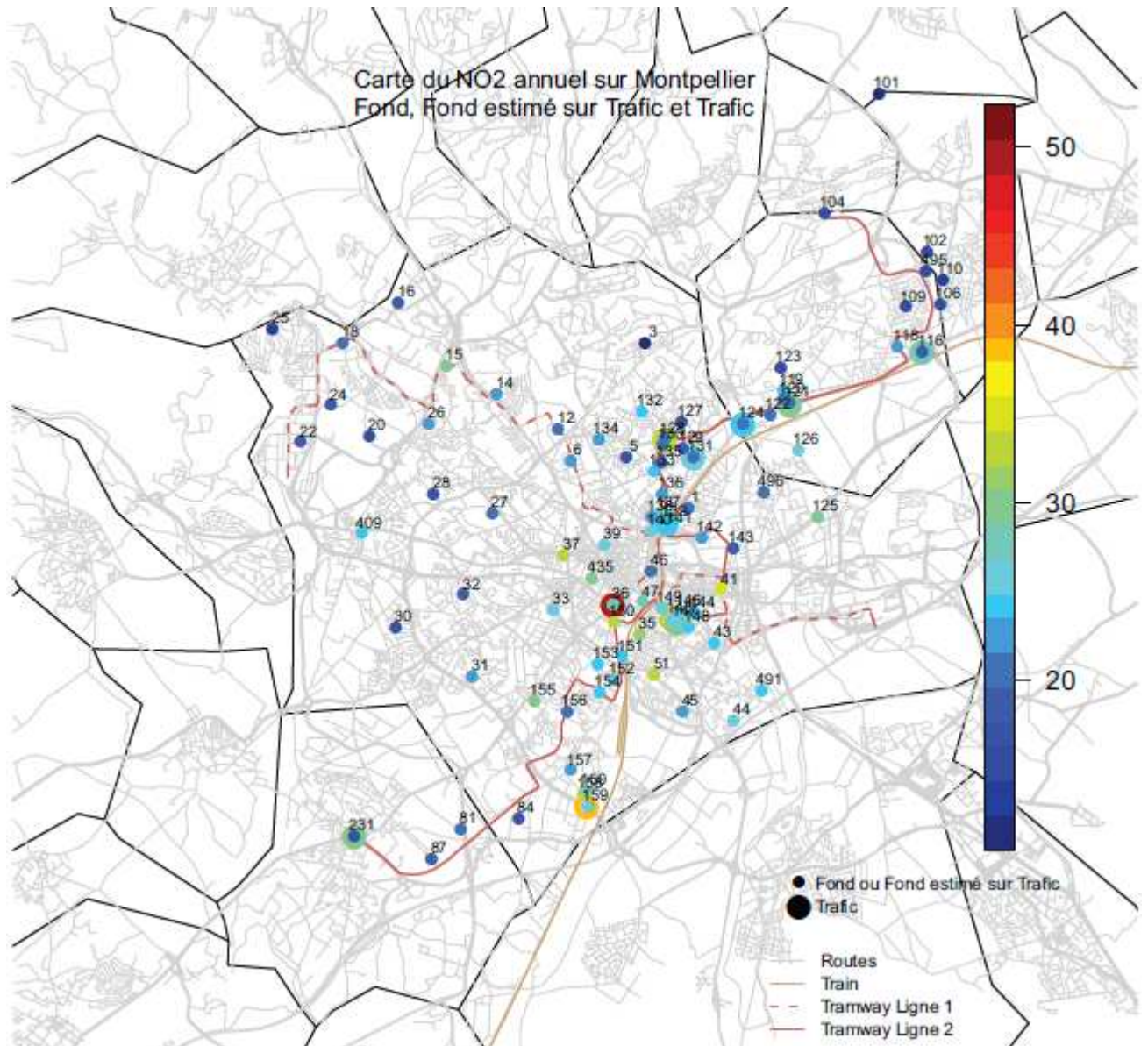


Figure 3 : Cartographie des concentrations de fond sur les sites de fond, de l'estimation du fond sur les sites de proximité et des valeurs réelles de concentration mesurées en ces sites de proximité. Moyenne annuelle de 2007.

Cette représentation met en évidence l'impact sensible du trafic routier sur les concentrations des sites de proximité. Pour chacun de ces sites, on calcule la différence entre la concentration réelle et la concentration de fond estimée : cette différence est appelée Delta Trafic-Fond ; elle sera notée Δ_{TF} .

2.3.2.2 MODELISATION STATISTIQUE DE Δ_{TF} .

L'objectif est de définir un modèle de surplus de concentration attribuable au trafic automobile et de l'appliquer le long des principaux axes routiers de la zone étudiée. A cette fin, on modélise statistiquement Δ_{TF} par régression linéaire sur les variables les plus susceptibles d'expliquer ce surplus.

Cette étape montre le caractère déterminant du choix des sites de proximité utilisés dans la construction du modèle. L'idéal serait de disposer de mesures de proximité en des points où des variables explicatives telles que les émissions de NO_x liées au trafic ou la distance par rapport à l'axe routier le plus proche peuvent avoir une influence significative sur la valeur de Δ_{TF} . Un échantillonnage réparti le long des axes mais également perpendiculairement à ceux-ci, selon une série de « transects », pourrait présenter de telles propriétés. Un schéma d'échantillonnage de ce type a été utilisé sur l'agglomération de Troyes. Les résultats correspondants sont fournis en annexe (4).

Sur Montpellier, l'implantation des sites de proximité s'est révélée insuffisante pour la modélisation de Δ_{TF} parce qu'elle n'était pas conçue à cet effet (l'échantillonnage était destiné à quantifier l'impact de la nouvelle ligne du tramway). Il est apparu que l'exploitation des seules données de proximité dans le calcul d'un tel modèle rendait ce dernier incohérent d'un point de vue physique (diminution du surplus pour un accroissement des émissions de NO_x). Aussi a-t-on pris le parti, afin d'obtenir une corrélation exploitable entre Δ_{TF} et une ou plusieurs variables, et construire un modèle cohérent, d'ajouter à l'échantillon d'apprentissage des sites de fond rattachés aux mêmes axes routiers que les sites de proximité. Il a été alors possible, par régression linéaire, de modéliser le surplus de concentration Δ_{TF} comme une fonction décroissante de la distance par rapport à l'axe :

$$\Delta_{TF}(x) = 16.59 - 3.30 * \log(\text{dist}(x, \text{axe}) + 1) \quad (\text{eq. 2.1})$$

où x est un point sous influence directe du trafic. $\log(\text{dist}(x, \text{axe}))$ est le logarithme népérien de la distance la plus courte (en m) entre le point x et la route. En l'absence de lien significatif avec les émissions de NO_x celles-ci n'ont pu être intégrées comme variable explicative.

Les graphiques ci-dessous présentent :

- à gauche : la corrélation entre le Δ_{TF} mesuré et le Δ_{TF} estimé par le modèle (1)
On notera que pour les sites de fond, le Δ_{TF} mesuré est nul. Ce résultat est logique puisque le krigeage est un interpolateur exact (la valeur de fond estimée en ces points est donc égale à la valeur mesurée).
- à droite : la corrélation entre les mesures de proximité et leurs estimations.

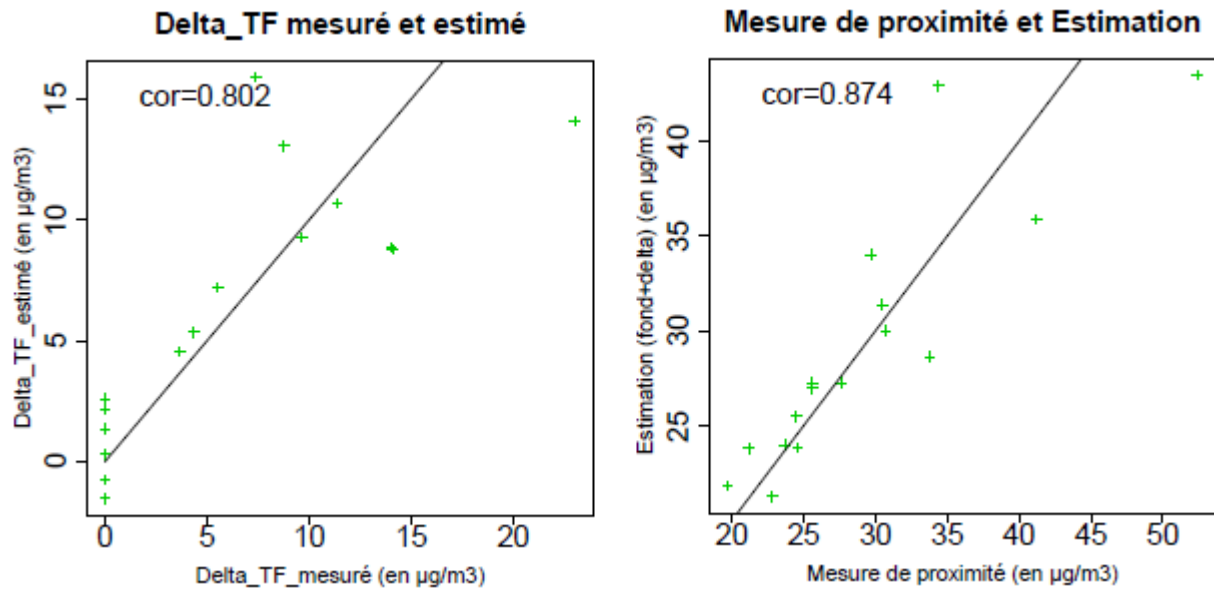


Figure 4 : Graphes et corrélations Mesure-Modèle pour le Δ_{TF} et les concentrations de proximité. Les concentrations de proximité estimées représentent la somme du fond estimé et du Δ_{TF} estimé

Si ce modèle semble satisfaisant (Figure 4), il reste très simplificateur, notamment parce qu'il ne tient pas compte des émissions de NO_x ; aussi ne devrait-il fournir qu'une approximation assez grossière des concentrations de proximité sur l'ensemble de l'agglomération. En outre, il ne permet pas, à cause de l'échantillon d'apprentissage, de rendre compte de certaines valeurs extrêmes. On prendra l'exemple du site 36 (implanté sur la station de trafic 08004 de Montpellier), où la concentration mesurée excède $50 \mu\text{g}/\text{m}^3$, valeur qui n'est observée sur aucun des autres sites de proximité du centre-ville. En ce point, le surplus Δ_{TF} et la concentration totale se retrouvent sous-estimés d'environ $8 \mu\text{g}/\text{m}^3$.

Afin d'illustrer la méthodologie dans son ensemble, et malgré sa simplicité, le modèle 2.1 est mis en œuvre sur tous les grands axes de la ville.

2.3.3 CARTOGRAPHIE INCLUANT POLLUTION DE FOND ET POLLUTION DE PROXIMITÉ

Aux réserves près exprimées ci-dessus, les résultats des sections précédentes permettent de construire une cartographie qui intègre la pollution de fond et le surplus de concentration lié au trafic.

Le résultat obtenu est conforme à nos attentes (Figure 5) : le centre ville de Montpellier apparaît comme étant la zone présentant les concentrations de NO_2 les plus élevées. De plus, le modèle de Δ_{TF} permet de prendre en compte les axes routiers dans des zones périurbaines ou rurales, rehaussant sensiblement les concentrations des points proches de ces axes.

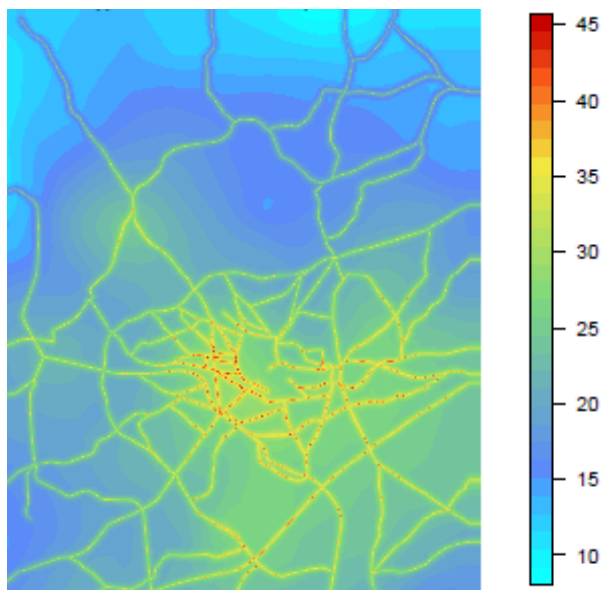


Figure 5 : Cartographie des concentrations moyennes annuelles de NO₂ incluant fond et proximité

Des discontinuités dans les concentrations sont perceptibles le long de certains axes ; elles s'expliquent par les limites du modèle de surplus précédemment évoquées, et en particulier par l'absence des émissions de NO_x dans ce modèle. Comme cela a été expliqué, les valeurs extrêmes ne sont pas présentes, car impossibles à reproduire par le modèle statistique. Ainsi, même en centre-ville, la concentration n'excède jamais 45 µg/m³. Notons qu'à l'avenir ce type d'approche pourra bénéficier de l'information apportée par la modélisation à l'échelle de la ville et de la rue, dont les AASQA font de plus en plus usage.

Les limites identifiées ci-dessus ne sont pas sensibles pour Troyes (annexe 4) où l'échantillonnage a été spécifiquement conçu pour l'application d'une telle méthodologie de cartographie (implantation de sites de fond et de proximité, avec des points de mesure le long des axes de circulation et perpendiculairement à ceux-ci). Il en résulte un modèle de surplus plus précis.

3. REPRESENTATIVITE EN CONCENTRATION

3.1 REMARQUES PRELIMINAIRES

La représentativité spatiale des mesures de stations fixes est un concept récurrent dans la réglementation européenne ; cependant, aucun texte ne le définit précisément (cf. remarques §1.1).

Ce chapitre s'attache à la représentativité spatiale d'une station fixe du point de vue des concentrations. A la définition proposée (§3.2), répond l'élaboration d'une méthodologie ayant pour objet de délimiter, pour chaque station, la zone de représentativité associée. L'approche développée (§3.3) a été mise au point avec le concours du Centre de Géosciences de Mines ParisTech à partir des travaux cités en introduction. Différentes hypothèses de calcul ont été testées (de Fouquet, 2011). Ne sont ici décrites que les hypothèses retenues pour la présente étude. Des approfondissements sont fournis au paragraphe 3.4 et un exemple d'application au paragraphe 3.5.

3.2 DEFINITION PROPOSEE

Dans l'approche développée, un point sera considéré comme appartenant à la zone de représentativité d'une station si l'écart de concentration entre le point et la station est inférieur à une valeur donnée, qu'on qualifiera d'écart maximal toléré par rapport à la station.

Etant donné un site de mesure, **sa représentativité doit être établie pour chaque variable de concentration considérée.**

Dans cette étude, on s'intéresse à la concentration moyenne annuelle² de NO₂.

Soient,

$Z(x)$: la concentration moyenne annuelle de NO₂ au point x

$Z(x_0)$: la concentration moyenne annuelle de la station située en x_0

δ : l'écart maximal de concentration toléré entre le point x et la station fixe située en x_0

Le point x appartient à la zone de représentativité de la station située en x_0 si :

$$|Z(x) - Z(x_0)| < \delta \quad (\text{eq. 3.1})$$

Cette définition, qui reste simple, suppose une connaissance de la concentration en tout point x du domaine d'étude, ce qui est impossible. On utilise en conséquence les estimations obtenues par krigeage selon la méthodologie définie en 3.2.

Cela permet de se placer dans un cadre probabiliste et de définir la zone de représentativité à un risque statistique η près, suivant une certaine hypothèse sur la distribution des erreurs d'estimation, et en tenant compte de l'écart-type de krigeage.

² ou éventuellement pluriannuelle, si des données sont disponibles sur plusieurs années.

Cette approche revient à corriger statistiquement la valeur de δ . En supposant que l'erreur suit une distribution de loi gaussienne, de moyenne nulle et d'écart-type égal à l'écart-type de krigeage, la condition (1) devient ainsi :

$$|Z^*(x) - Z(x_0)| < \delta - \sigma_K(x) * q_{1-\frac{\eta}{2}} \quad (\text{eq. 3.2})$$

$Z^*(x)$: concentration estimée par krigeage au point x

$\sigma_K(x)$: écart-type de krigeage au point x

η : risque statistique toléré, i.e. probabilité maximale autorisée qu'un point x soit inclus par erreur dans la zone de représentativité de la station

$q_{1-\eta/2}$: quantile d'ordre $(1-\eta/2)$ de la loi normale

Pour un risque η inférieur à 50%, ce quantile est positif, ce qui a pour effet de diminuer δ : le critère de rattachement d'un point à une station est donc plus strict.

Pour plus de précisions sur la méthode probabiliste associée, on pourra se reporter à l'annexe 3 qui présente les divers éléments théoriques.

Enfin, cette définition, fondée sur les écarts de concentration, pose le problème des recouvrements (un point peut appartenir à plusieurs zones de représentativité) et de la stabilité temporelle de ces zones. Ces questions seront discutées dans la suite de cette section.

3.3 ORGANISATION DE LA METHODOLOGIE

Le calcul des zones de représentativité est immédiat, à condition que toutes les données d'entrée soient disponibles. Si des zones de représentativité se recoupent, le traitement des intersections requiert une étape supplémentaire.

Etape 1	Echelle : locale	Polluant : NO ₂
Objectif : Délimiter des zones de représentativité en concentration des stations de mesure		
Moyen : Méthodologie probabiliste fondée sur une analyse des écarts de concentration par rapport aux stations (programme développé avec le langage R)		
Données requises, valeurs de paramètres et critères à préciser :		
<ul style="list-style-type: none"> - Estimation des concentrations par krigeage sur tout le domaine et écart-type de krigeage associé à chaque concentration estimée. - Ecart maximal toléré δ, risque statistique η, hypothèse sur la distribution des erreurs d'estimation. 		

Etape 2	Echelle : locale	Polluant : NO ₂
Objectif : Supprimer les recouvrements entre plusieurs zones de représentativité de manière à d'obtenir une partition du domaine.		
Moyen : Plusieurs méthodes possibles. Moyen ici retenu : un point d'une zone de recouvrement est rattaché préférentiellement à la station dont il est le plus proche en concentration.		
Données requises, valeurs de paramètres et critères à préciser :		
Estimation des concentrations par krigeage sur tout le domaine.		

3.4 CRITERES DE RATTACHEMENT AUX STATIONS ET STABILITE TEMPORELLE

Comme il a été mentionné, l'un des problèmes posés par cette méthode est la question des zones de recouvrement. Une information individuelle sur la représentativité spatiale d'une station constitue une donnée essentielle pour pouvoir apprécier, à partir d'une mesure ponctuelle, la qualité de l'air environnant. Dans une seconde étape, il peut être intéressant de considérer un réseau de mesure dans son ensemble, à l'échelle d'une ville par exemple, et de déterminer de manière relative l'importance de chaque station dans l'évaluation de la qualité de l'air de ce domaine. Une méthode est d'établir une partition du domaine dans laquelle un point ne se trouve rattaché qu'à une seule station.

On propose d'étudier différents critères de rattachement préférentiel d'un point à une station fixe dans le cas où ce point appartiendrait à l'intersection de plusieurs zones de représentativité.

Les quatre critères proposés sont les suivants : distance minimale par rapport à la station (1), écart minimal de concentration par rapport la station (2) et similarité de milieu avec la station. La similarité de milieu pourra être vue sous deux angles différents : en examinant les distances entre des vecteurs dits de milieu (3) ou bien à l'aide d'une ACP sur les variables dites de milieu (4). Là encore, pour des définitions plus formelles, on se reportera à l'annexe 3.

L'objectif est de ne conserver qu'un seul critère de rattachement aux stations. Cette partie a ainsi pour objet d'évaluer l'effet de chaque critère et de comparer les résultats correspondants en vue de déterminer, s'il est possible, le critère le plus adapté. Afin de pouvoir multiplier les cas de figure (différentes zones géographiques et différentes années), ces tests ont été réalisés avec les données de concentration de NO₂ issues des cartes analysées nationales.

Sur la région de Montpellier, on propose d'étudier les résultats de la méthode en considérant dans un premier temps les résultats des cartes analysées comme vraies, c'est-à-dire qu'on ne définit pas de risque statistique pour corriger l'écart maximal toléré entre un point et une station (δ). Ainsi, les zones de recouvrement seront plus vastes, puisque la condition de rattachement à une zone sera moins stricte, et l'on pourra analyser plus facilement les résultats obtenus avec les différents critères.

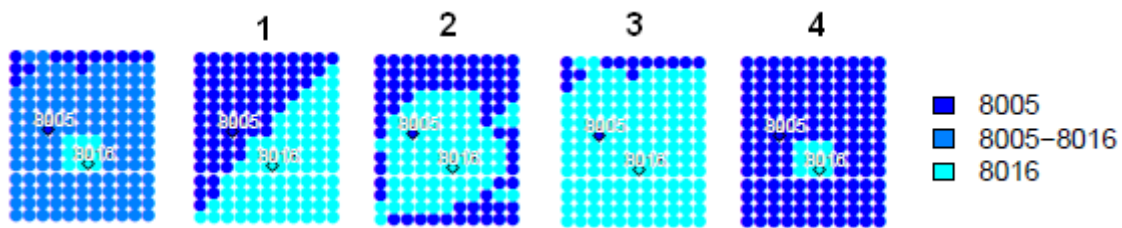


Figure 6 : Zones de représentativité en concentration des stations 8005 et 8016 pour l'année 2007
 graphe de gauche : visualisation des recouvrements
 graphes 1 à 4 : suppression des recouvrements avec les différents critères

Ces résultats, obtenus avec un écart maximal toléré par rapport aux stations de $10\mu\text{g}/\text{m}^3$ mettent en lumière la grande difficulté à définir le critère de suppression des recouvrements : la distance favorise la localisation géographique, l'écart minimal de concentration est le plus cohérent avec la définition qui est proposée de la zone de représentativité (3.2.2 et annexe 5) ; de manière surprenante, la similarité de milieu offre des résultats très peu stables suivant la définition qui en est donnée. Dans ce cas précis, la zone de recouvrement est entièrement rattachée à la station 8005 par l'ACP, et à la station 8016 par l'analyse des distances entre les vecteurs de milieu.

Il est évident que les zones de représentativité seront très différentes selon le critère choisi. De ce fait, l'une des options facilement exploitables pour fixer le critère est d'étudier la stabilité temporelle de la méthode pour chacun des critères possibles. Cette analyse est réalisée pour les années 2005 à 2009.

		2005	2006	2007	2008	2009
Concentrations aux stations (en $\mu\text{g}/\text{m}^3$)	8005	16	17	17	16	17
	8016	26	25	27	24	22
Avec recouvrement						
distance						
écart min						
Similarité de milieu par Vecteur						
Similarité de milieu par ACP						

Figure 7 : Zones de représentativité en concentration des stations 8005 et 8016 pour les années 2005 à 2009 sur Montpellier

Si les années 2006 à 2008 n'apportent pas d'indication pour le choix du critère, en revanche les années 2005 et 2009, qui présentent des zones de recouvrement sensiblement différentes, permettent de tirer quelques conclusions.

Dans le cas de l'année 2009 où l'écart de concentration entre les deux stations est plus faible, la zone de recouvrement des deux stations est répartie sur l'ensemble du domaine d'étude. Le critère d'écart minimal par rapport à la station produit des zones de représentativité assez similaires à celles des trois années précédentes tandis que les critères de similarité de milieu rattachent tous les points à une seule et unique station.

Pour l'année 2005, le recouvrement est moindre mais la définition d'une zone de représentativité propre à la station 8016 est impossible sans suppression préalable de ce recouvrement. L'écart minimal de concentration par rapport la station permet de rattacher des points à chacune des deux stations alors qu'une fois encore, les critères de similarité de milieu ne favorisent qu'une seule d'entre elles.

Enfin, le critère de distance est abandonné parce qu'il accorde trop d'importance à la localisation géographique des points. Ainsi, un déplacement même minime de la station modifierait grandement les zones de représentativité, et cela d'autant plus si le nombre de stations dans le domaine concerné est plus important. Ce critère aurait été probablement plus approprié si les zones de recouvrement étaient d'amplitude restreinte, ce qui sera rarement le cas pour un écart maximal toléré δ de $10\mu\text{g}/\text{m}^3$.

Des tests similaires ont été réalisés pour les agglomérations de Troyes, Rouen et Lille, dont les résultats sont fournis en annexe (5). L'influence des critères observée sur chacune de ces villes aboutit aux mêmes conclusions que pour Montpellier.

En conclusion, compte tenu de la cohérence spatiale et de la stabilité temporelle des zones de représentativité, le critère d'écart minimal de concentration par rapport à la station est désormais utilisé pour supprimer les recouvrements.

3.5 APPLICATION

3.5.1 REPRESENTATIVITE ET POLLUTION DE FOND

La méthodologie mise au point avec les données des cartes analysées nationales de NO_2 est appliquée sur la région de Montpellier. Les concentrations moyennes annuelles de NO_2 (année 2007) y ont été précédemment estimées, grâce à l'exploitation de données d'échantillonnage (voir chapitre 2). On présente dans un premier temps les résultats obtenus avec la cartographie de fond. Les conséquences du surplus de concentration lié au trafic sur la méthodologie sont ensuite discutées.

Dans cette application, la carte de représentativité en concentration tient compte de l'incertitude sur les estimations, i.e. de l'écart-type de krigeage. De plus, comme cela a été introduit en 3.2, le risque statistique du modèle probabiliste est associé à une

hypothèse sur la distribution des erreurs de krigeage. Conventionnellement, cette loi des erreurs est supposée gaussienne, mais d'autres modèles peuvent être testés.

Au risque statistique de 10% et pour un écart maximal toléré par rapport aux stations de $10 \mu\text{g}/\text{m}^3$, les zones de représentativité en concentration se présentent ainsi :

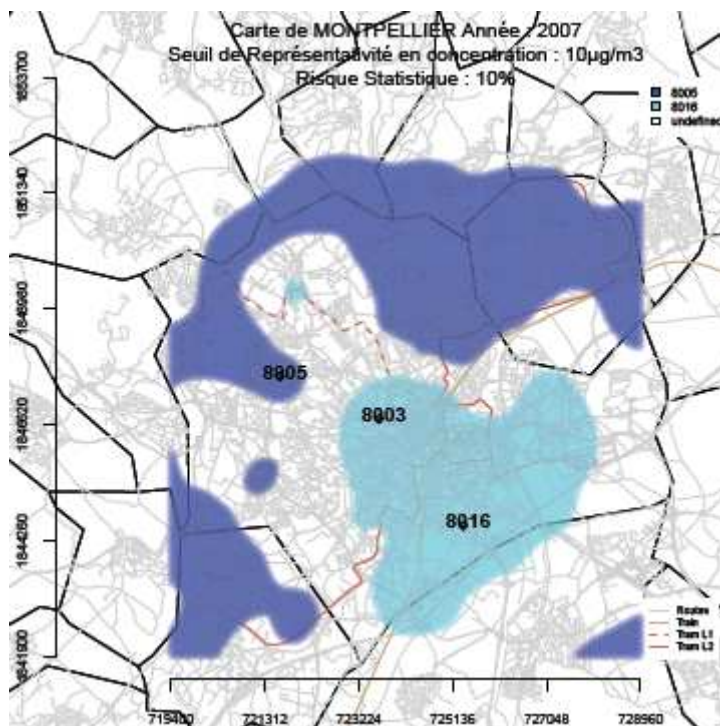


Figure 8 : Zones de représentativité en concentration des stations 8005 et 8016 (stations de fond) sur l'agglomération de Montpellier. Superficie de la zone de représentativité de la station 8005 : $\sim 35 \text{ km}^2$. Superficie de la zone de représentativité de la station 8016 : $\sim 19 \text{ km}^2$

Cette carte est comparable à celle qui a été obtenue pour la même année dans la section précédente (avec les données des cartes analysées nationales). Les zones de représentativité des stations 8005 et 8016 n'ont pas subi de changement notable, si ce n'est que l'estimation, affinée par la campagne de tubes, permet à la station 8005 d'être incluse dans sa propre zone de représentativité, ce qui est plus cohérent. L'incertitude d'estimation étant ici prise en compte, il devient impossible, au risque statistique η_R , de rattacher certains points à une station.

Les zones de représentativité des deux stations sont disjointes. La station 8016 apparaît représentative d'un fond urbain propre au centre-ville tandis que la station 8005 caractérise surtout la couronne périurbaine.

Aucune zone de représentativité n'a pu être définie pour la station de fond 8003 (Chaptal). La concentration moyenne annuelle de NO_2 observée sur ce site est en effet bien supérieure aux concentrations de fond mesurées ou estimées par krigeage sur le reste de l'agglomération. Les données de cette station ne semblent donc rendre compte que très localement des concentrations de NO_2 . Des données de tubes supplémentaires au voisinage de ce site seraient utiles pour valider une telle hypothèse.

L'application de la méthode montre en outre la difficulté de fixer des valeurs pour les différents paramètres : l'écart maximal toléré δ et le risque statistique η . Ces valeurs ont en effet une influence déterminante sur l'amplitude des zones de représentativité, comme l'illustrent les courbes ci-après :

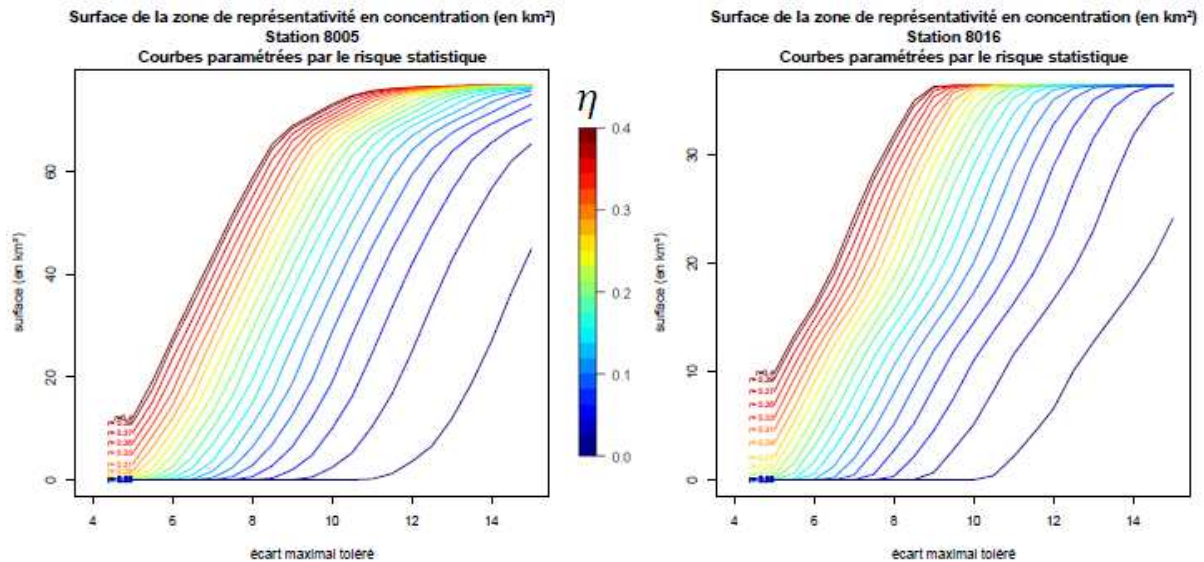


Figure 9 : Montpellier, année 2007. Tests de sensibilité aux valeurs des paramètres. Surface des zones de représentativité en concentration en fonction de l'écart maximal toléré δ (écart de concentration par rapport à la station). Courbes paramétrées par le risque statistique η (valeurs représentées par l'échelle de couleurs)

Un assouplissement de l'une ou l'autre des conditions qui définissent la zone de représentativité d'une station - i.e. de tolérer un plus grand écart en concentration par rapport à la station (δ) ou d'accepter d'un plus grand risque d'erreur (η) - accroît la surface de cette zone. Les deux valeurs ici retenues - $\delta=10 \mu\text{g}/\text{m}^3$ et $\eta=10\%$ - correspondent approximativement au point central de chaque ensemble de courbes (Figure 9). Une valeur de 10% est un risque communément admis en statistique.

3.5.2 EXPLOITATION COMPLEMENTAIRE DES RESULTATS

La Figure 8 met en évidence tout un espace, compris entre le centre-ville et la périphérie de l'agglomération de Montpellier, dont les points ne se trouvent rattachés à aucune des trois stations de fond du domaine. L'analyse ci-après présente une autre exploitation possible des données de tubes et du calcul des zones de représentativité : le choix d'un site supplémentaire pour la mesure fixe du NO_2 .

L'idée est d'identifier, parmi les différents points d'échantillonnage, celui qui représente au mieux les concentrations moyennes annuelles de NO_2 dans l'espace susmentionné : dans cet exemple, le point situé à proximité du boulevard Paul Valéry (numéroté 8900) se révèle comme l'un des plus appropriés (Figure 11).

On verra dans la section consacrée aux techniques de recalage annuel des concentrations (0) que les zones de représentativité peuvent évoluer d'une année

sur l'autre. Ainsi, l'information qui était manquante en 2007 ne l'est plus en 2009 du fait de la baisse des concentrations en centre-ville. Dans le cas où l'on chercherait à implanter une nouvelle station fixe, il faudrait donc s'intéresser au préalable aux variations annuelles des concentrations moyennes qui sont observées sur les stations déjà en place.

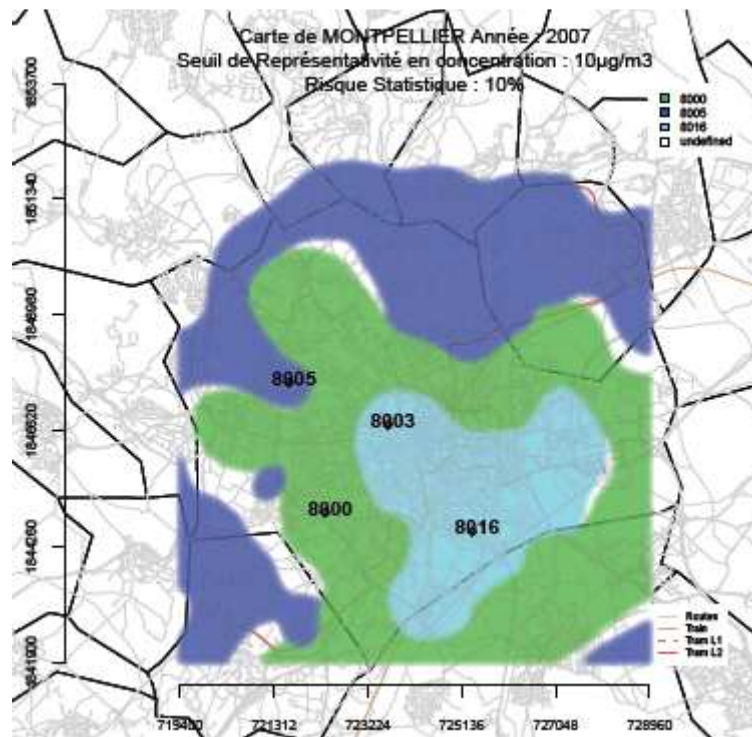


Figure 10 : Zones de représentativité en concentration des stations 8005, 8016 (stations de fond) et du tube 31 (Bld Paul Valéry) assimilé à la station fictive 8000 sur l'agglomération de Montpellier

3.5.3 REPRESENTATIVITE ET POLLUTION INTEGRANT FOND ET PROXIMITE

L'objectif est d'affiner les zones de représentativité précédemment définies en incluant le surplus de concentration routier dans les calculs.

Outre l'erreur d'estimation du fond, il convient de prendre en compte l'erreur d'estimation de ce surplus.

Aux points de proximité, l'hypothèse de distribution de l'erreur d'estimation reste celle d'une distribution gaussienne, d'écart-type correspondant à l'écart-type de krigeage. Cependant, au lieu d'être centrée en 0, cette distribution est centrée sur l'écart-type des résidus de la régression en validation croisée du Δ_{TF} , ce qui revient à ajouter un terme d'erreur constant à l'erreur d'estimation de la pollution de fond.

En fonction de la fiabilité du modèle, et donc de la valeur prise par l'écart-type des résidus en validation croisée du Δ_{TF} , les corrections statistiques seront plus ou moins importantes, et les zones de représentativité plus ou moins grandes.

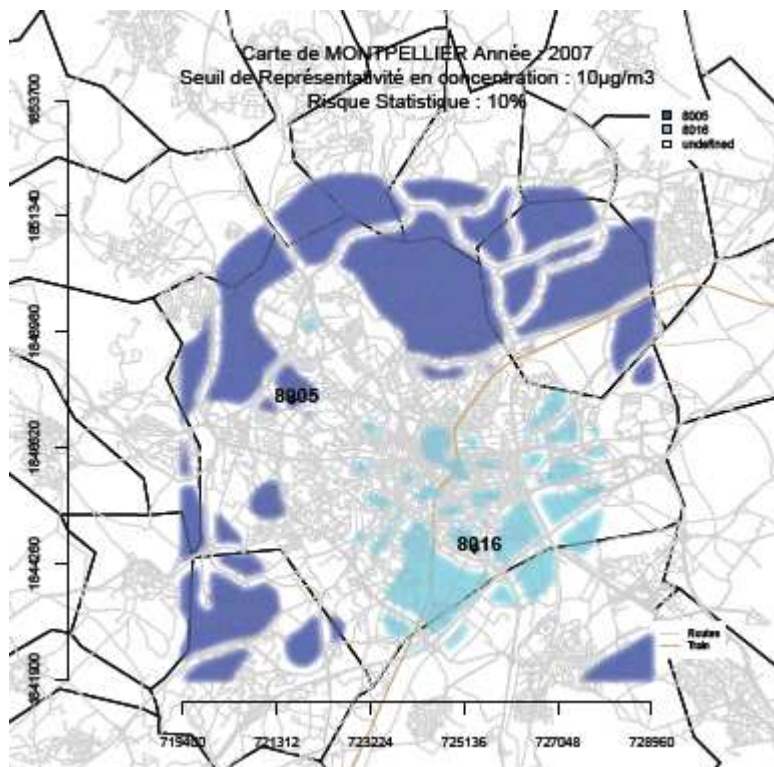


Figure 11 : Zones de représentativité en concentration des stations 8005 et 8016 (stations de fond) sur l'agglomération de Montpellier après ajout de la pollution proximité

Les zones de représentativité des deux stations de fond ressemblent à celles qui ont été précédemment délimitées (2.3.1) mais à cause du rehaussement des concentrations le long des axes routiers, elles excluent ces derniers. On aurait pu s'attendre, en incluant la proximité, à voir apparaître des zones de représentativité pour la station de trafic 8004 ou encore pour la station de fond 8003 dont les concentrations sont assez élevées. Toutefois, l'erreur associée au modèle de surplus est trop importante, et donc, après correction statistique (cf. 3.2), l'écart maximal toléré est trop faible (il peut même devenir négatif) pour que des points puissent être rattachés à ces stations.

Ces résultats posent la question de la représentativité spatiale des stations de trafic. Le modèle statistique relatif à la pollution de proximité se révèle trop imprécis pour que l'on puisse identifier les points du domaine représentés par ce type de station. L'ajout de points d'échantillonnage à la fois le long des axes routiers et perpendiculairement à ceux-ci et la prise en compte de variables d'influence supplémentaires, comme le bâti, pourraient contribuer à l'amélioration de ce modèle statistique.

4. DEPASSEMENTS DE SEUIL

4.1 REMARQUES PRELIMINAIRES

Le cadre théorique et les hypothèses retenues sont identiques à ceux du chapitre 3 mais l'optique est différente. Il ne s'agit pas de délimiter les zones dans lesquelles les concentrations sont proches des valeurs mesurées sur les stations fixes mais d'identifier les régions dans lesquelles les concentrations sont supérieures à un seuil réglementaire, ici la valeur limite annuelle relative au NO₂.

4.2 APPROCHE PROBABILISTE

Dans le cadre conventionnel déjà utilisé pour la représentativité (cf. §3.2), on admet que l'erreur d'estimation $\varepsilon(x)$ sur la concentration $Z(x)$ suit une distribution gaussienne dont l'écart-type est égal à l'écart-type de krigeage en x .

En conséquence, il est possible d'évaluer en tout point x la probabilité que la concentration $Z(x)$ dépasse le seuil :

$$\mathbb{P}[Z(x) > s] = \mathbb{P}\left[T > \frac{s-Z^*(x)}{\sigma_K(x)}\right] = 1-F\left(\frac{s-Z^*(x)}{\sigma_K(x)}\right) \quad (\text{eq. 4.1})$$

où F est la fonction de répartition de la loi normale.

Dans une seconde partie, on cherche à délimiter, conformément à la demande réglementaire (cf. §1.2 et annexe 2), les zones vraisemblablement exposées à un dépassement. Pour ce faire, deux types de risque statistique sont pris en compte:

- le risque *alpha* de première espèce, qui représente la probabilité de conclure par erreur à un non dépassement, alors que la concentration excède le seuil.
- le risque *beta* de seconde espèce, qui représente la probabilité de conclure par erreur à un dépassement, alors que la concentration est inférieure au seuil.

Par précaution, on choisira de préférence un faible risque *alpha* et un risque *beta* plus élevé. En effet, dans une perspective de réduction de l'exposition, manquer un dépassement paraît plus préjudiciable qu'en signaler un à tort.

Prendre un risque *alpha* conventionnel de 5% et un risque *beta* conventionnel de 34% (soit environ une chance sur trois de se tromper) (Cori, 2005) conduit à identifier les deux types de zones suivants :

- **Zone de non-dépassement** : ensemble des points x tels que

$$\mathbb{P}[Z(x) > s] < \alpha \Leftrightarrow Z^*(x) < s - \sigma_K(x) * q_{1-\alpha} \Leftrightarrow Z^*(x) < s - 1.65 * \sigma_K(x) \quad (\text{eq. 4.2})$$

- **Zone de dépassement** : ensemble des points x tels que

$$\mathbb{P}[Z(x) \leq s] < \beta \Leftrightarrow Z^*(x) > s - \sigma_K(x) * q_{\beta} \Leftrightarrow Z^*(x) > s + 0.41 * \sigma_K(x) \quad (\text{eq. 4.3})$$

Les points ne vérifiant aucun des deux critères précédents définissent une **zone dite d'incertitude**.

L'ensemble des calculs et des éléments théoriques de la méthodologie est présenté en annexe (3).

4.3 ORGANISATION DE LA METHODOLOGIE

Etape 1	Echelle : locale	Polluant : NO ₂
Objectif : Déterminer les zones de dépassement de la valeur limite annuelle		
Moyen : Méthodologie probabiliste fondée sur une analyse des écarts de concentration par rapport à la valeur limite (programme développé avec le langage R).		
Données requises :		
<ul style="list-style-type: none">- Estimation des concentrations par krigeage sur tout le domaine et écart-type de krigeage associé à chacune des estimations.- Seuil de concentration, risques statistiques de première et de seconde espèce.		

4.4 APPLICATION

4.4.1 DEPASSEMENT ET POLLUTION DE FOND

Le seuil généralement considéré pour le NO₂ est de 40 µg/m³ en moyenne annuelle, valeur limite la plus contraignante. Cependant, en dehors de l'Île-de-France, il est rare que les concentrations de fond soient supérieures à cette valeur. Afin de pouvoir illustrer la méthode, nous fixerons donc le seuil à 32 µg/m³ (seuil d'évaluation supérieur, dont le dépassement rend obligatoire la mesure fixe).

La carte ci-après représente la probabilité de dépassement du seuil de 32 µg/m³ sur l'agglomération de Montpellier. Cette probabilité est faible, notamment à cause d'écarts-types de krigeage assez élevés (entre 4 et 5 µg/m³). Même dans les zones où les concentrations estimées sont proches de la valeur limite (voir la cartographie de la pollution de fond en Figure 2), elle n'excède pas 25%.

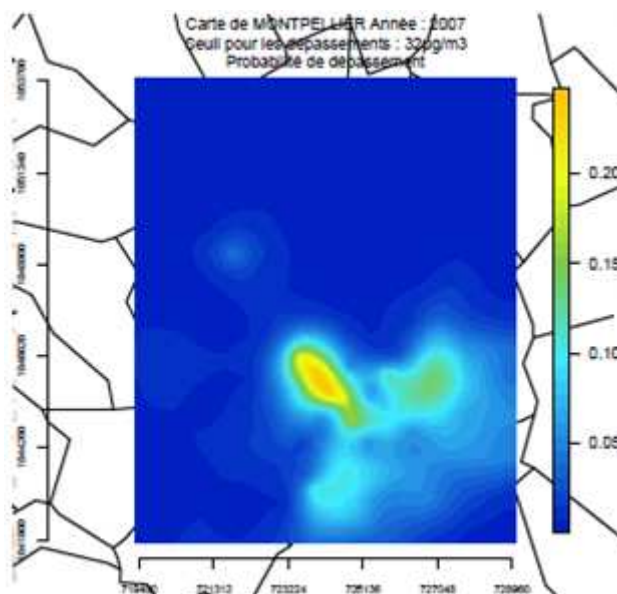


Figure 12 : Carte des probabilités de dépasser le seuil de 32 µg/m³ pour l'année 2007

4.4.2 RISQUES DE PREMIERE ET SECONDE ESPECES

Les équations 4.2 et 4.3 sont appliquées sur l'agglomération de Montpellier, pour la valeur seuil s de 32 µg/m³. Pour un risque de première espèce de 5%, sont ainsi définies des zones de non dépassement (en vert sur les cartes), où il est quasi certain que la concentration est inférieure au seuil. Pour un risque de seconde espèce de 34%, aucune zone de dépassement n'est identifiée (Figure 13, carte de gauche). Mettre en évidence une zone de dépassement nécessiterait d'admettre un risque élevé (supérieur à 50%) d'y inclure des points par erreur (Figure 13, carte de droite).

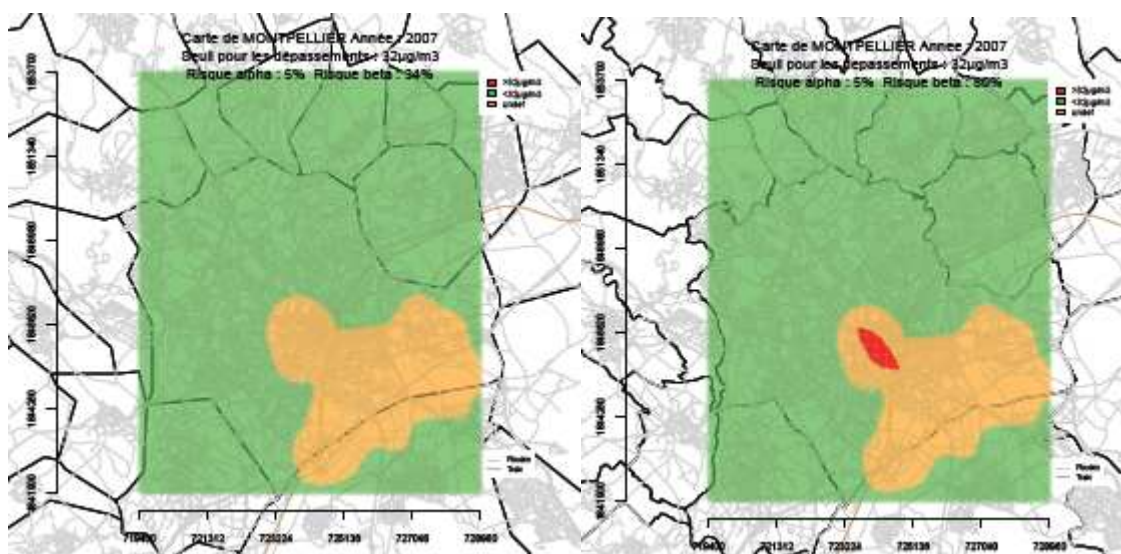


Figure 13 : Carte de dépassement du seuil de 32µg/m³ intégrant les notions de risques de 1^{ère} et 2^e espèce. En rouge : dépassement au risque 34% (fig. de gauche) et 80%(fig. de droite), en vert : non-dépassement au risque 5%, en orange, statut indéterminé

4.4.3 DEPASSEMENT ET POLLUTION INTEGRANT FOND ET PROXIMITE

On s'intéresse à la question du dépassement de seuil lorsque la pollution de proximité est prise en compte. Les concentrations atteignant des valeurs plus élevées, le seuil est fixé à $40 \mu\text{g}/\text{m}^3$, valeur limite réglementaire.

La méthode employée précédemment est appliquée de manière identique. Comme cela a été présenté (3.5.3), l'erreur d'estimation pour les points de proximité prend en compte l'erreur d'estimation du surplus liée au modèle statistique.

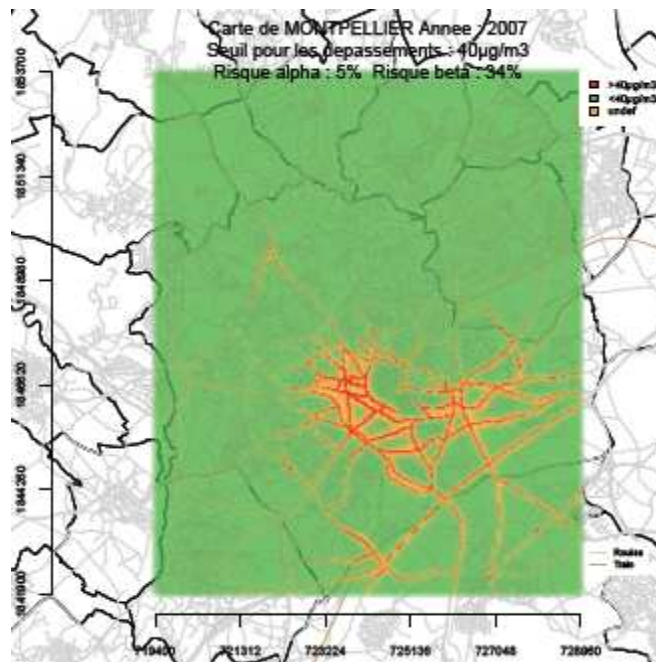


Figure 14 : Carte de dépassement du seuil de $40 \mu\text{g}/\text{m}^3$ intégrant les notions de risques de 1^{ère} et 2^e espèce. En rouge : dépassement au risque 34%, en vert : non-dépassement au risque 5%, en orange, statut indéterminé

Dans les zones éloignées des principaux axes routiers, la probabilité de dépassement du seuil de $40 \mu\text{g}/\text{m}^3$ est faible ; la condition de « non dépassement » (eq. 4.2) est vérifiée (en vert sur la carte). En revanche, dans le centre-ville, et même autour de celui-ci, de nombreux points situés à proximité des axes sont classés comme étant en dépassement (inégalité 4.3 vérifiée, points en rouge).

Enfin, la méthodologie ne permet pas de se prononcer au sujet de certains axes situés en périphérie de la ville (en orange). Les valeurs estimées sur ces axes sont suffisamment élevées pour mettre en défaut le critère de non dépassement mais l'estimation est trop incertaine pour qu'il soit possible de conclure fermement.

4.4.4 DISCUSSION SUR LES RISQUES DE PREMIERE ET SECONDE ESPECES

Dans les calculs précédents, une valeur conventionnelle de 5% et une valeur plus élevée de 34% ont été attribuées respectivement aux risques de première et seconde espèces (cf. §4.2).

Si l'utilisateur a grande confiance dans l'estimation des concentrations de fond et de proximité, il pourra tolérer un risque de première espèce plus élevé (probabilité d'inclure un point de dépassement dans une zone de non dépassement) et, à l'inverse, se montrer plus strict dans le choix du risque de seconde espèce (probabilité d'inclure un point de non dépassement dans une zone de dépassement). Ces modifications auront respectivement pour conséquence d'étendre les zones de non dépassement et de restreindre celles de dépassement. La Figure 15 illustre l'effet d'une augmentation du risque de seconde espèce : par rapport à la Figure 15Figure 13, les portions d'axes classées en dépassement se réduisent.

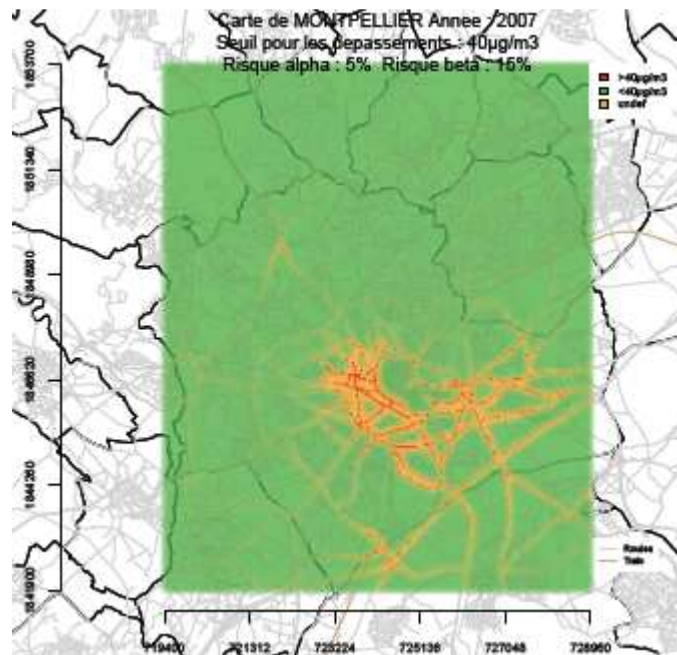


Figure 15 : Carte de dépassement du seuil de 40µg/m³ intégrant les notions de risques de 1^{ère} et 2^e espèce. En rouge : dépassement au risque de seconde espèce de 15%, en vert : non-dépassement au risque de première espèce de 5%, en orange, statut indéterminé

5. RECALAGE ANNUEL DES RESULTATS

5.1 CARTOGRAPHIE DES CONCENTRATIONS DE FOND

Les ressources disponibles ne permettent généralement pas de reconduire annuellement sur une même zone une campagne d'échantillonnage par tubes à diffusion passive.

Pour une mise à jour annuelle des zones de représentativité et de dépassement, il convient de recalculer les moyennes annuelles aux points d'échantillonnage. Sur Montpellier, il s'agit de réajuster pour différentes années les concentrations moyennes annuelles de NO₂ mesurées en 2007 (année de référence).

Une approche géostatistique consiste à effectuer un cokrigage entre les mesures fixes de l'année considérée (variable principale) et les valeurs aux tubes de l'année de référence (variable secondaire) (de Fouquet et al., 2011). Cependant, le nombre de stations fixes est en général trop réduit pour appliquer cette technique dans des conditions satisfaisantes (trois stations de fond en 2007 à Montpellier, dont une de représentativité incertaine -cf. §3.5.1- ; deux stations de fond en 2008 à Troyes). Des tests effectués sur Montpellier l'ont confirmé.

Une autre méthode a été explorée, fondée sur l'application d'un facteur correctif variable dans l'espace. Elle s'organise selon le schéma suivant :

- 1- Interpolation des concentrations de NO₂ sur les points d'échantillonnage à partir des cartes analysées nationales³, pour l'année à corriger et pour l'année de référence.
 - 2- Correction des données en chaque point d'échantillonnage. Le facteur correctif à appliquer en chaque point est égal au rapport entre la valeur interpolée pour l'année à corriger et la valeur interpolée pour l'année de référence :
- $$NO2(\text{tube})_{\text{année à corriger}} = NO2(\text{tube})_{\text{année de référence}} * \frac{\text{Interpolation}(\text{tube})_{\text{année à corriger}}}{\text{Interpolation}(\text{tube})_{\text{année de référence}}}$$
- 3- Krigage avec dérive externe à partir des données de tubes corrigées suivant la même procédure que celle décrite en 3.1.2.

Cette méthode a été appliquée aux années 2005, 2006, 2008 et 2009 en prenant comme référence la campagne de tubes de 2007 :

³ Cartes élaborées par krigage avec dérive externe, selon un maillage kilométrique, en utilisant les observations sur les stations fixes de fond, les données de simulation fournies par CHIMERE, la densité d'émissions de NOx issue de l'INS et la densité de population.

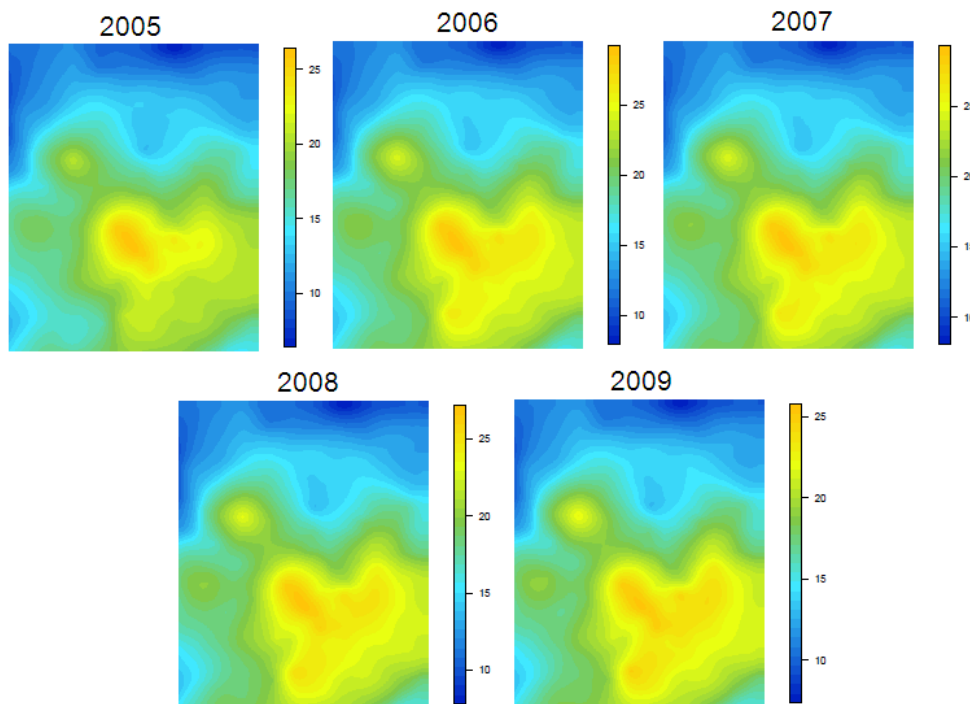


Figure 16 : Recalage annuel par un facteur correctif sur les tubes à partir des données Chimère

Les cartographies sont relativement stables dans le temps, avec toutefois de légers changements dans l'amplitude des concentrations.

Pour s'assurer du bien-fondé de la méthode, on a comparé les moyennes mesurées par tubes sur l'année de référence (sites de fond) et les moyennes estimées en ces mêmes points à partir des cartes analysées nationales de NO₂. Sur Montpellier, les deux séries de valeurs sont cohérentes, avec une corrélation de 0.645 (Figure 17). Les estimations peuvent être vues comme un lissage des mesures réelles, avec une tendance à surestimer les concentrations dans la plupart des cas, et à sous-estimer les valeurs les plus fortes.

Il est probable que ce lissage, lié à la résolution des cartes analysées (maille kilométrique), se répète d'année en année. Ainsi le facteur correctif, qui représente un rapport de concentrations entre deux années différentes, ne devrait-il pas en être sensiblement affecté. La corrélation entre les concentrations mesurées en 2007 et les concentrations corrigées pour une année différente reste très élevée (Figure 17).

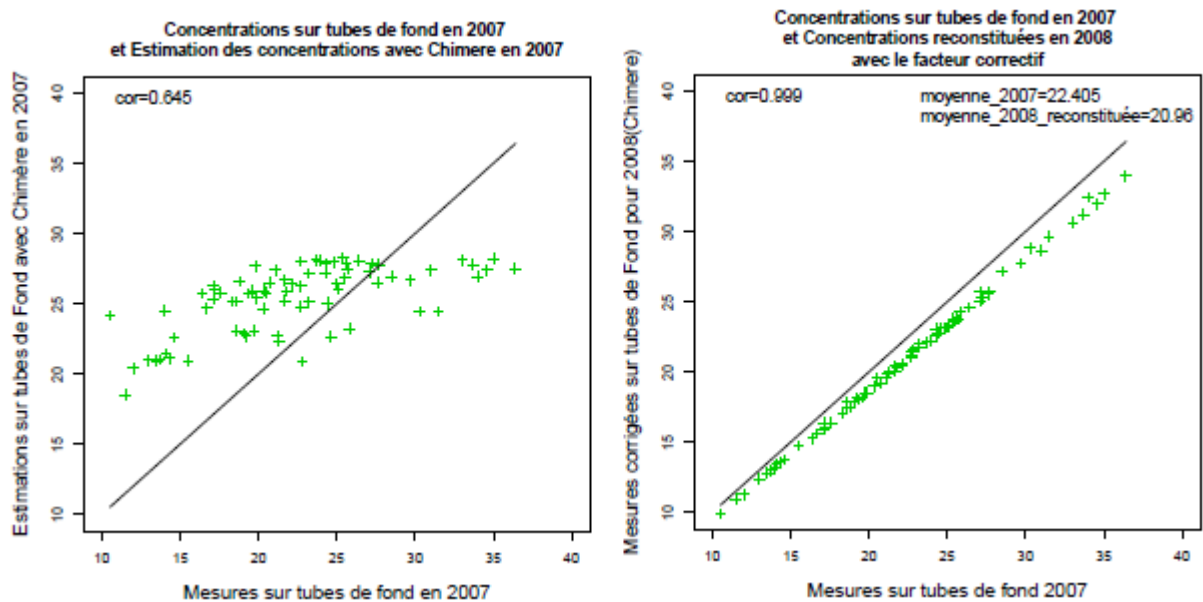


Figure 17 : A gauche, nuage de points entre les concentrations estimées à partir des cartes analysées nationales et les mesures de fond de 2007 et estimation en 2007 avec Chimère à droite,. A droite, nuage de points entre les mesures corrigées pour l'année 2008 et les mesures du fond en 2007, année de référence

L'idéal serait de pouvoir comparer les mesures corrigées (pour une année différente de 2007) à des mesures réelles. Sur Montpellier, nous ne disposons pas de campagne supplémentaire susceptible de servir de validation.

5.2 REPRESENTATIVITE EN CONCENTRATION ET DEPASSEMENTS DE SEUIL

Le recalage annuel de la pollution de fond permet, même si l'on ne dispose pas de données de campagne propres à l'année étudiée, de cartographier les zones de représentativité en concentration et les dépassements de seuil en situation de fond.

La méthodologie est appliquée sur Montpellier pour les années 2005 à 2009:

	Concentrations des stations (en $\mu\text{g}/\text{m}^3$)		Représentativité en concentration	Probabilités de dépassements
	8005	8016		
2005	16	26		

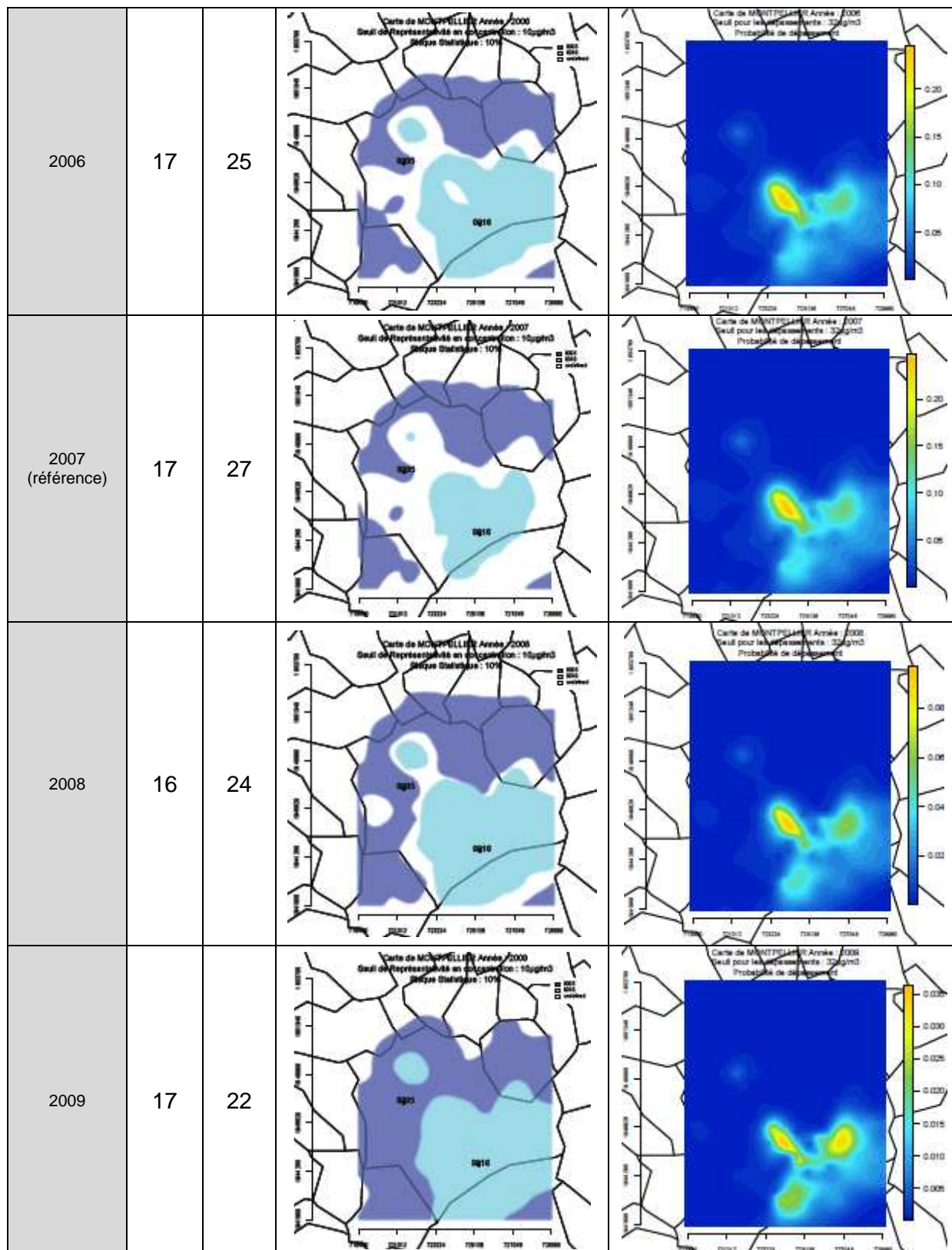


Figure 18 : Zones de représentativité en concentration des stations 8005 à 8016
 Probabilités de dépassements du seuil 30µg/m³ pour les années 2005 à 2009 sur Montpellier

La zone de représentativité de la station 8005 apparaît relativement stable ; les concentrations de NO₂ mesurées sur cette station varient peu, en moyenne, d'une

année à l'autre. Il en va différemment pour la station 8016 dont les mesures évoluent à la baisse : ainsi la moyenne annuelle de 2009 est-elle inférieure de $5 \mu\text{g}/\text{m}^3$ à celle de 2007, ce qui conduit à une zone de représentativité plus étendue.

Les cartographies des probabilités de dépassement du seuil de $32 \mu\text{g}/\text{m}^3$ présentent elles aussi des différences interannuelles. En particulier ces probabilités sont plus élevées en 2006 et 2007, engendrant des surfaces de dépassement plus importantes.

D'autre part, les résultats relatifs à l'année 2005 présentent des différences notables par rapport aux autres années, ce que l'on pouvait déjà constater sur la cartographie de la pollution de fond (Figure 16). Ce résultat s'explique par l'absence de donnée pour la station 8003 cette année-là, qui affecte sensiblement la carte analysée nationale de NO_2 utilisée dans le calcul du facteur correctif.

Il semble donc plus prudent de n'appliquer cette méthode de réajustement par facteur correctif qu'en considérant un parc de stations constant. La station 8003 qui, d'après les évaluations précédentes (§3.5.1), est représentative de phénomènes localisés, pourrait être retirée du calcul des cartes analysées.

6. SYNTHÈSE ET CONCLUSION

Les Directives européennes relatives à la qualité de l'air et leur transposition en droit français imposent de caractériser la représentativité spatiale des stations de mesure et de cartographier les zones exposées à des dépassements de valeurs limites ou de valeurs cibles. Ces exigences soulèvent des questions de définition et de méthode qui nécessitent d'être abordées par polluant, en fonction des moyens de surveillance disponibles.

Cette étude traite le problème de la représentativité spatiale et des zones de dépassement pour le **NO₂ en moyenne annuelle**. Elle développe une approche fondée sur l'exploitation de campagnes de mesure par tubes à diffusion passive. La méthodologie proposée se découpe en trois parties. La première constitue un préalable indispensable aux deux autres, qui peuvent être mises en œuvre séparément.

1) Cartographie des concentrations

La première partie consiste à cartographier les concentrations moyennes annuelles de NO₂ sur le domaine d'étude (dans les exemples traités, les agglomérations de Montpellier et de Troyes). Les valeurs de fond sont estimées par krigeage avec dérive externe. Outre les données de mesure, cette technique géostatistique permet de prendre en compte différentes variables potentiellement liées aux concentrations, comme les émissions de NO_x ou la densité de population. La cartographie peut être ensuite affinée en ajoutant le long des axes un surplus de concentration, modélisé statistiquement comme une fonction linéaire de variables explicatives (émissions, distance par rapport à l'axe...). Les exemples étudiés mettent en exergue le rôle essentiel de la stratégie d'échantillonnage spatial et temporel. Celle-ci détermine en grande partie la précision des estimations et, par suite, la fiabilité des résultats qui en découlent, à savoir l'évaluation des zones de représentativité et la délimitation des zones de dépassement. On peut à ce propos, émettre les recommandations suivantes.

- ***Echantillonnage spatial :***

Les sites de fond doivent se répartir dans l'espace géographique mais également dans celui des variables auxiliaires, c'est-à-dire qu'ils doivent se distribuer à la fois dans les zones de faibles, moyennes et fortes valeurs de ces variables. L'objectif est que le modèle géostatistique établi pour le krigeage rende compte des relations entre les concentrations et les variables auxiliaires sur l'ensemble du domaine et que le krigeage avec dérive externe s'applique dans de bonnes conditions à toute la zone d'étude. Pour des recommandations plus précises, on se reportera notamment aux études suivantes : Wroblewski *et al.* (2006), de Fouquet et Fauchaux (2008) ; Romary et de Fouquet (2011, 2009).

En ce qui concerne la pollution de proximité, l'expérience montre qu'on ne peut expliquer de façon satisfaisante le surplus de concentration lié au trafic avec seulement quelques points : les points d'échantillonnage doivent être disposés à la fois le long des axes et perpendiculairement à ceux-ci, selon des « transects ». Le rapport de Wroblewski et al. (2007) fournit des conseils pour dimensionner ce type d'échantillonnage.

- **Echantillonnage temporel :**

Les périodes de mesure doivent se répartir dans l'année de manière à capter la variabilité temporelle des concentrations. Pour des recommandations sur l'échantillonnage temporel, on se reportera aux travaux du GT Plans d'échantillonnage et reconstitution de données (guide et outils) accessibles à l'adresse suivante : <http://www.lcsqa.org/outil/missions-diverses/strategie-surveillance/planification-temporelle-echantillonnage>.

2) Représentativité en concentration des stations et zones de dépassement

2a) Cadre théorique

La délimitation des zones de représentativité spatiale des stations tout comme celle des zones de dépassement s'inscrivent dans un même cadre probabiliste qui tient compte des incertitudes associées à la cartographie. Elles font intervenir la notion de risque statistique, ou plus explicitement, la probabilité d'inclure par erreur un point dans une zone. Afin de faciliter l'application des méthodes dans un contexte opérationnel, un cadre volontairement simplifié a été adopté : l'approche proposée se limite aux techniques de la géostatistique linéaire et l'erreur d'estimation est supposée se distribuer selon une loi normale, d'écart-type égal à l'écart-type de krigeage. Des tests de rééchantillonnage, effectués sur les données moyennes annuelles de NO₂, ont permis de contrôler la normalité de cette distribution.

2b) Représentativité spatiale des stations de mesure

La représentativité *en concentration* d'une station est définie pour un certain polluant (ici le NO₂) et une certaine variable de concentration (ici la moyenne annuelle). Elle décrit la partie du domaine d'étude dans laquelle cette variable est correctement approchée par la valeur mesurée sur la station.

En pratique, la zone de représentativité d'une station est déterminée en comparant les concentrations préalablement estimées sur le domaine à la concentration mesurée en ce site. On considère qu'un point appartient à une telle zone si la différence de concentration entre le point et la station est inférieure à un écart maximal prédéfini (noté δ), ou plus exactement, à cause de l'incertitude d'estimation, si la probabilité que cette différence excède δ est faible (inférieure à η). L'incertitude de mesure propre à la station fixe est ignorée. Les valeurs des paramètres δ et η doivent être ajustées par l'utilisateur. Les valeurs retenues à l'issue d'une analyse de

sensibilité sont respectivement de 10 $\mu\text{g}/\text{m}^3$ et 10% pour le NO_2 en moyenne annuelle.

A l'intérieur d'un domaine d'étude, **les zones de représentativité en concentration peuvent être définies séparément pour chaque station de mesure.** L'analyse conjointe des résultats permet de s'assurer que ces stations offrent une bonne couverture du domaine, détecter d'éventuelles redondances entre sites ou au contraire, identifier des manques dans la surveillance.

Ces mêmes zones de représentativité peuvent être ensuite redéfinies en considérant le réseau de mesure dans son ensemble. Il est alors nécessaire, comme cela a été fait pour Troyes, de supprimer les recouvrements entre zones, autrement dit, de relier préférentiellement chaque point de ces recouvrements à une unique station. Les résultats obtenus servent à affiner l'analyse précédente, en faisant notamment ressortir le rôle prépondérant de certaines stations. Ainsi est-ce le cas de la station 14033 de Troyes, à laquelle une grande partie de la ville se retrouve rattachée (cf. annexe 4). Plusieurs critères de suppression ont été testés. Le critère fondé sur les écarts minimaux de concentration se révèle le plus satisfaisant.

Dans les exemples de Montpellier et Troyes, l'application de la méthodologie aux données de fond produit des résultats comparables et cohérents : une station apparaît représentative du centre-ville et une autre, du pourtour urbain. Pour les stations de proximité, l'évaluation de la représentativité se révèle plus délicate, en raison des spécificités locales de ces stations et de la difficulté d'estimer avec précision les concentrations aux abords des axes. Ce problème sera approfondi sur d'autres jeux de données dans les travaux 2011-2012.

Dans tous les cas **se pose la question de la stabilité temporelle des zones de représentativité définies.** Il est recommandé, si les données le permettent, d'examiner les résultats sur des années différentes. La réalisation de campagnes de mesure à intervalle de temps régulier (par exemple, toutes les x années) et la mise en œuvre consécutive de la méthodologie en sont un moyen. Entre deux campagnes, et si la distribution spatiale des émissions et des concentrations est jugée relativement stable, un recalage temporel des données d'échantillonnage disponibles sur l'année la plus proche est envisageable. En première approche, celles-ci peuvent être réajustées par un facteur multiplicatif issu des données de stations fixes ou, comme il est ici proposé, de cartes analysées. La possibilité d'un réajustement plus fin (selon la méthode géostatistique du cokrigage) est limitée par le petit nombre de stations.

2c) Zones de dépassement de seuil

Les zones de dépassement de seuil sont définies dans un certain domaine, par exemple une agglomération, pour un polluant et un seuil donnés (ici la valeur limite

annuelle relative au NO₂). Elles le sont obligatoirement dès lors qu'une ou plusieurs stations fixes ont enregistré un dépassement.

Les zones de dépassement ne coïncident pas avec les zones de représentativité en concentration mais l'approche employée pour les définir est similaire. Elles sont déterminées en comparant les concentrations préalablement estimées sur le domaine à la valeur seuil. Du fait de l'incertitude d'estimation, il est proposé de distinguer trois types de zones : des zones de non dépassement, de dépassement, et d'incertitude. En un point donné, on conclut à l'absence de dépassement si la probabilité que la concentration soit supérieure au seuil est faible, plus basse qu'un risque *alpha* dit de *première espèce*. De même, on conclut à un dépassement si la probabilité que la concentration soit inférieure au seuil est faible, plus basse qu'un risque *beta* dit de *seconde espèce*. Les valeurs d'*alpha* et *beta* sont fixées par l'utilisateur, selon la confiance qu'il accorde aux estimations et le risque qu'il est prêt à prendre. La valeur du risque *alpha* utilisée dans cette étude est de 5%, l'omission d'un dépassement étant jugée plus dommageable qu'une fausse alerte ; celle de *beta*, moins pénalisante, est de 34% (soit d'environ 1/3). Les points qui ne répondent à aucune des deux conditions précédentes forment une zone d'incertitude où des investigations complémentaires pourraient être utiles.

La méthode a été appliquée aux estimations intégrant pollution de fond et pollution de proximité. Sur Montpellier comme sur Troyes, elle permet d'identifier des zones de dépassement le long des principaux axes du centre-ville ainsi qu'en certains points de la périphérie.

3) Etudes complémentaires et travaux futurs

- **Nouveaux cas d'étude et comparaison avec la modélisation**

De nouveaux cas d'étude sont en cours d'exploitation (campagnes de mesure du NO₂ sur MADININAIR, LIG'AIR, ATMO Auvergne). Des résultats de modélisation urbaine sont disponibles pour certaines agglomérations, offrant l'opportunité de comparer les deux approches (géostatistique/modélisation urbaine) et d'en évaluer les complémentarités.

- **Autres polluants**

Les méthodes mises au point pour le NO₂ peuvent s'appliquer de la même manière au benzène si le nombre de données disponibles le permet. Des campagnes de mesure du benzène ont été traitées lors de travaux sur l'échantillonnage spatial (Romary et al., 2011). Elles sont actuellement réexploitées dans ce nouvel objectif.

En ce qui concerne les PM_{10} , l'approche développée ne peut être répliquée du fait qu'une cartographie fine des concentrations, notamment au voisinage des routes, est limitée par les moyens de mesure (qui sont plus coûteux), et que les échelles de temps diffèrent (pour les PM_{10} on s'intéresse aussi à la valeur limite en moyenne journalière). Une étude a été spécifiquement engagée sur ce polluant. La première phase a consisté à réaliser des campagnes de mesure en collaboration avec des AASQA. Les résultats et conclusions de ce travail sont consignés dans un document de synthèse indépendant (Beauchamp et al., 2011).

- **Echelle nationale**

Le présent rapport est consacré à l'échelle urbaine. L'étude des zones de dépassement à l'échelle nationale, pour le NO_2 et les PM_{10} , fera l'objet d'une note spécifique (Malherbe et Beauchamp, 2011).

- **Population exposée**

La question de la distribution spatiale de la population et de l'estimation du nombre d'habitants exposés à des dépassements de seuil n'a pas été spécifiquement abordée. Des pistes sont évoquées (de Fouquet et al., 2011). Ce sujet sera approfondi en 2012.

Ces différents travaux serviront à alimenter l'étude LCSQA de 2012 consacrée à la cartographie (bilan des pratiques et élaboration de recommandations).

7. REFERENCES

Bobbia M., Cori A., de Fouquet Ch., 2008. Représentativité spatiale d'une station de mesure de la pollution atmosphérique. Pollution Atmosphérique, n°197, 63-75.

Cárdenas G. et Malherbe L., 2007. Représentativité des stations de mesure du réseau national de surveillance de la qualité de l'air : Application des méthodes géostatistiques à l'évaluation de la représentativité spatiale des stations de mesure de NO₂ et O₃. Rapport LCSQA, ref.DRC-08-85146-01036A, www.lcsqa.org.

Cori A., 2005a. Utilisation de données environnementales pour la cartographie du NO₂ sur la région rouennaise. Rapport de stage, première partie. Ecole Nationale Supérieure des Mines de Paris.

Cori A., 2005b. Représentativité spatiale des stations de mesure de la concentration moyenne annuelle en NO₂. Rapport de stage, deuxième partie. Ecole Nationale Supérieure des Mines de Paris.

Cori A., 2005c. Définition de zones homogènes vis-à-vis du dépassement de seuil pour la concentration en ozone. Rapport de stage, troisième partie. Ecole Nationale Supérieure des Mines de Paris.

Directive 2008/50/CE du Parlement Européen et du Conseil du 21 mai 2008 concernant la qualité de l'air ambiant et un air pur pour l'Europe.

Directive 2004/107/CE du Parlement Européen et du Conseil du 15 décembre 2004 concernant l'arsenic, le cadmium, le mercure, le nickel et les hydrocarbures aromatiques polycycliques dans l'air ambiant.

Décret n° 2010-1250 du 21 octobre 2010 relatif à la qualité de l'air: ce décret transpose la directive 2008/50/CE du Parlement européen et du Conseil du 21 mai 2008 concernant la qualité de l'air ambiant et un air pur pour l'Europe. Il modifie la partie consacrée à l'air et à l'atmosphère du code de l'environnement (consultable sur le site de Légifrance : <http://www.legifrance.gouv.fr>).

Arrêté du 21 octobre 2010 relatif aux modalités de surveillance de la qualité de l'air et à l'information du public.

Décision de la Commission du 20 février 2004 (2004-224-CE) fixant les modalités de transmission d'informations sur les plans ou les programmes exigés par la directive 96/62/CE du Conseil en ce qui concerne les valeurs limites de certains polluants dans l'air ambiant.

Décision de la Commission du 29 avril 2004 (2004-461-CE) établissant un questionnaire à utiliser pour la déclaration annuelle concernant l'évaluation de la qualité de l'air ambiant au titre des directives du Conseil 96/62/CE et 1999/30/CE ainsi que des directives du Parlement européen et du Conseil 2000/69/CE et 2002/3/CE.

De Fouquet Ch., Beauchamp M., Malherbe L., 2011. Remarques sur la représentativité des stations de mesure de qualité de l'air. Rapport n° R2011-334CFOU, Centre de Géosciences, Géostatistique, Ecole des Mines de Paris. Rapport réalisé pour le LCSQA, www.lcsqa.org.

De Fouquet Ch., Fauchaux C., 2008. Recommandations pour les schémas d'échantillonnage des campagnes de mesure de la qualité de l'air. Annexe 1 du rapport : Malherbe L. et Cárdenas G., 2008. Echantillonnage et reconstitution de données dans l'espace et le temps. Rapport LCSQA, ref. DRC-08-94308-15147A, www.lcsqa.org.

De Fouquet Ch., Gallois D., Perron G., 2007. Geostatistical characterization of the nitrogen dioxide concentration in an urban area: Part I: Spatial variability and cartography of the annual concentration. Atmospheric Environment, Vol. 41(32), 6701-6714.

Düring I., Bächlin W., Ketzel M., Baum A., Friedrich U. and Wurzler S., 2011. A new simplified NO/NO₂ conversion model under consideration of direct NO₂ emissions. Meteorologische Zeitschrift, Vol. 20, No. 1, pp. 067-073.

GT *Plans d'échantillonnage et reconstitution de données*, 2008. Elaboration de plans d'échantillonnage temporel et reconstitution de données. Guide pratique. <http://www.lcsqa.org/echantillonnage-reconstitution-donnees/documentation>.

Kiesewetter G., Borken-Kleefeld J., Heyes C. Schoepp W., Amann M., Thunis P., Cuvelier K., Bessagnet B., Terrenoire E., 2012. Urban to local scales: NO₂ traffic station modeling. EC4MACS workshop, Paris, 8 mars 2012, <http://gains.iiasa.ac.at/index.php/meetings/pastmeetings/385-ec4macs-workshop-in-paris>.

Malherbe L., Cárdenas G., Colin P., Yahyaoui A., 2008. Using different spatial scale measurements in a geostatistically based approach for mapping atmospheric nitrogen dioxide concentrations. Application to the French Centre region. Environmetrics, Vol. 19(7), 751-764.

Malherbe L. et Wroblewski A., 2006. Synthèse des techniques de cartographie du dioxyde d'azote. Valorisation des travaux 2003-2005. Rapport LCSQA, ref. DRC-07-74777-03941A, www.lcsqa.org.

Malherbe L. et Cárdenas G., 2005. Application des méthodes géostatistiques pour l'exploitation conjointe des mesures de fond et de proximité. Rapport LCSQA, ref. INERIS – DRC/MECO – 05 - 65015 – LMa / N°155, www.lcsqa.org.

Romary T., Fouquet C. (de), Malherbe L., 2011. Sampling design for air quality measurement surveys: an optimization approach. Atmospheric Environment, Vol. 45, n°21, 3613-3620.

Romary T., de Fouquet Ch., 2009. Recommandations sur le schéma d'échantillonnage pour les campagnes de mesures de la qualité de l'air : une approche par optimisation. Annexe 3 du rapport : Malherbe L. et Debry E., 2009.

Assistance à l'exploitation de données de campagnes et à la réalisation de cartographies. Rapport LCSQA, ref. DRC-10-103349-00553A, www.lcsqa.org.

Wroblewski A., Riffault V., Perdrix E., Malherbe L., 2007. Adaptation des plans d'échantillonnage aux objectifs des campagnes : Echantillonnage spatial – Guide de recommandations. Rapport LCSQA, ref. DRC-07-84894-17725A, www.lcsqa.org.

Wroblewski A., Riffault V., Perdrix E., Malherbe L., 2006. Adaptation des plans d'échantillonnage aux objectifs des campagnes : Echantillonnage spatial – Guide de recommandations. Rapport LCSQA, ref. DRC-07-74775-03939A, www.lcsqa.org.

8. LISTE DES ANNEXES

Repère	Désignation	Nombre de pages
Annexe 1	Fiche LCSQA	6
Annexe 2	Contexte législatif et réglementaire (Europe et France)	10
Annexe 3	Eléments théoriques sur la méthodologie proposée	10
Annexe 4	Application de la méthodologie sur l'agglomération de Troyes	14
Annexe 5	Sensibilité aux critères de suppression des recouvrements	3
Annexe 6	Comparaison des résultats sous R et Isatis	2
Annexe 7	Notice d'utilisation des programmes	4

ANNEXE 1 : FICHE LCSQA

THEME 6 : Modélisation et traitements numériques

ETUDE N° 6/5 : METHODOLOGIE DE DETERMINATION DES ZONES GEOGRAPHIQUES CONCERNEES PAR LES DEPASSEMENTS DE SEUILS

Responsable de l'étude : INERIS

Contexte et Objectifs

Cette étude a pour objet de fournir aux AASQA des éléments méthodologiques qui leur permettent, à l'aide de données complémentaires dont elles disposent par ailleurs, de qualifier les zones géographiques réellement concernées par les dépassements de seuils enregistrés de manière ponctuelle par leurs stations permanentes.

Il s'agit de répondre à une demande réglementaire (Directive sur la qualité de l'air) relative au rapportage sur les dépassements des valeurs limites de concentrations de polluants. Cette disposition concerne en priorité les concentrations de PM10, de dioxyde d'azote et de benzène. Chaque état membre doit qualifier, et expliquer les dépassements ponctuels constatés. En particulier l'aire et la localisation des zones géographiques « réelles » concernées ainsi que la population exposée doivent être estimées, qu'il s'agisse d'un seuil de court ou de long terme.

Les AASQA sont donc régulièrement sollicitées par le Ministère en charge de l'Ecologie via les DRIRE pour fournir ces éléments quantitatifs à faire remonter à la Commission Européenne. Mais à ce jour, aucune méthode harmonisée n'a cours pour traiter cette question.

L'objectif de cette fiche est de fournir un support méthodologique à cette activité qui constitue une priorité des pouvoirs publics compte tenu des courts délais imposés par la Commission pour le rapportage.

Ce travail a été planifié sur au minimum deux ans (2009-2010) avec, en première partie, la mise au point d'une approche simplifiée, ciblée sur les PM10, afin de répondre à l'urgence de la demande, et en seconde partie, la définition d'une méthodologie plus élaborée incluant également le NO2 et le benzène.

Travaux antérieurs

En 2009, après consultation de l'ensemble des AASQA concernées, une méthodologie simplifiée a été mise au point pour l'évaluation des zones géographiques touchées par les dépassements des valeurs limites de PM10.

Elle s'organise par typologie de site : pour le fond, une méthode de cartographie journalière selon une résolution kilométrique a été mise en place à partir des données d'observation de la BDQA et des simulations du modèle CHIMERE ; pour la proximité routière et industrielle, un arbre de décision, qui tient compte des informations locales potentiellement disponibles, a été développé.

La précision de cette méthodologie a été limitée par le manque de données d'observation disponibles. En effet, les données de campagnes mobiles sont peu nombreuses, l'usage d'appareils portatifs n'étant pas encore opérationnel pour les PM10.

Travaux proposés pour 2010

Les outils et données disponibles nous amènent à distinguer deux échelles d'étude, traitées avec des approches différentes (cf travaux menés sur les PM10 en 2009) :

- l'échelle nationale/régionale représentative de la pollution de fond, y compris la pollution urbaine de fond)
- la petite échelle, incluant la pollution de proximité.

1) Echelle nationale / régionale

PM10

En cohérence avec la méthodologie établie par le LCSQA en 2009, les concentrations moyennes annuelles de fond (régional et urbain) et les dépassements de valeurs limites seront cartographiés pour les années 2008 et 2009 sur l'ensemble de la France. Ces cartographies et les données numériques associées seront mises à disposition des AASQA sur le site du LCSQA.

NO₂

La méthode mise au point pour les PM10 sera réajustée pour le NO₂. On recherchera en particulier des variables auxiliaires adaptées à ce polluant. Des cartes de concentrations moyennes annuelles de fond et de dépassements de valeurs limites seront établies de manière similaire pour les années 2008 et 2009.

2) Proximité

PM10

La méthodologie proposée en 2009, simple et peu coûteuse, permet de délimiter des zones de dépassement, de façon plus ou moins précise selon les informations disponibles, mais elle ne permet pas d'attribuer une zone de représentativité à chaque station en particulier. L'enjeu était de proposer une méthode qui puisse être appliquée dans différentes situations quelles que soient les données d'évaluation disponibles. Mais l'une des conclusions de cette étude met en évidence le besoin de disposer de données de mesure complémentaires, issues de campagnes par exemple, pour évaluer correctement la zone de représentativité d'une station de mesure.

Il est proposé en 2010 d'élaborer une méthode fondée sur la disponibilité de données complémentaires, issues de la réalisation d'une campagne de mesure ciblée sur la caractérisation de la représentativité spatiale d'une station. Pour la mettre au point, une campagne de mesure test d'environ un mois sera réalisée par le LCSQA au voisinage d'un site de trafic restant à définir. Plusieurs techniques de mesure seront utilisées simultanément et en plusieurs points autour du site : des préleveurs de type MicroVol et des indicateurs optiques, relativement souples à mettre en œuvre, ainsi que des analyseurs automatiques.

Cette phase expérimentale nécessitera une réflexion préalable sur la stratégie d'échantillonnage temporel et spatial. La campagne devra être notamment définie dans l'intention d'établir la représentativité spatiale de la station :

- en moyenne sur une année
- sur des épisodes conduisant à des dépassements du seuil journalier de $50 \mu\text{g}/\text{m}^3$.

La conception de cette stratégie et la méthode d'exploitation des données s'appuieront sur l'expérience du LCSQA et sur une revue bibliographique. Des recommandations pour la réalisation de campagnes futures seront émises.

NO2

Pour le NO₂, les informations spatiales localement disponibles sont plus denses que pour les PM10, notamment grâce aux campagnes de mesure par tubes à échantillonnage passif. En moyenne annuelle, l'évaluation de la représentativité spatiale d'une station peut donc se fonder sur une exploitation géostatistique des données.

Le travail prévu pour 2010 exploitera les travaux du LCSQA sur la représentativité spatiale d'une station (2007). L'approche étudiée en 2007 permet en effet de délimiter, avec un niveau de confiance donné, la zone de représentativité d'une station. Elle permet aussi d'estimer des zones où la concentration dépasse potentiellement une valeur limite. La méthode de calcul utilisée suppose que l'on ait pu estimer la concentration du polluant sur tout le domaine ainsi que la précision associée. Tout l'enjeu réside donc dans l'obtention de

ces estimations, qui requiert de nombreuses données de mesure, idéalement issues de campagnes.

Dans les cas traités en 2006 (Rouen et Lille), seule la pollution de fond a été considérée. Or les dépassements de la valeur limite annuelle pour le NO₂ concernent majoritairement des stations de trafic. En 2010, la méthode sera donc préférentiellement appliquée à des cartes d'estimation du NO₂ qui combinent pollution de fond et pollution de proximité. Les travaux réalisés sur ce sujet en 2009 seront exploités. Des cartes d'estimation produites par les AASQA pourront être également utilisées.

La méthode de calcul des zones de représentativité et de dépassement sera automatisée sous la forme d'une macro SIG.

Benzène

On suivra la même démarche que pour le NO₂. Compte tenu des niveaux de benzène mesurés en France, des cas d'étude autres que des situations de dépassement seront considérés.

L'application au benzène de la méthode pourra se heurter à un manque de données. Les jeux de données de benzène issus de campagnes par tubes sont en effet moins nombreux et généralement moins denses que les jeux de données disponibles pour le NO₂. S'il est besoin, des propositions pour la réalisation d'une campagne de mesure du benzène adaptée à l'objectif de cette étude seront émises.

3) Valorisation des travaux d'étude effectués

L'ensemble des méthodes mises au point sera décrit dans un document technique qui en précisera les données d'entrée, le principe, les conditions d'utilisation et les limites.

Les méthodes seront testées par le LCSQA, en collaboration avec des AASQA volontaires, sur un ensemble de stations qui enregistrent des dépassements préoccupants des valeurs limites, afin de vérifier leur précision et leur robustesse.

La réalisation et l'exploitation de la campagne de mesure PM en 2010 et les travaux d'exploitation des campagnes de mesure NO₂ et benzène existantes permettront de définir les spécifications d'une campagne type dont le but est d'évaluer la représentativité spatiale d'une future station, préalablement à son implantation (nombre et répartition des points d'échantillonnage, type de mesure, ...), et les informations qu'il convient d'extraire de ces mesures. Ainsi suite à cette analyse il est proposé d'établir un cahier des charges pour la qualification de l'implantation de nouvelles stations de mesure afin de définir a priori sa zone de représentativité (aire géographique, population impactée, etc ;...) et de disposer des données techniques qui faciliteront le rapportage des données mesurées à la Commission. Ce cahier des charges sera rédigé en collaboration avec l'ADEME, qui instruit les demandes d'investissement des AASQA, et en concertation avec un panel d'AASQA volontaires.

Plan de travail pour l'année 2010

Fond	Proximité	
	Analyse bibliographique	Fin janvier
	Conception d'une campagne de mesure des PM10 : <ul style="list-style-type: none"> - choix du site de mesure - choix des moyens de mesure - définition de la stratégie temporelle et spatiale d'échantillonnage 	Fin février
	Réalisation des mesures de PM10	Un mois susceptible d'inclure des épisodes de pollution
Mise à jour des cartographies nationales de PM10 Cartographies nationales de NO ₂		Fin juin
	Rédaction d'une note méthodologique pour le NO ₂ et le benzène	Octobre
	Application des méthodes sur un ensemble de stations prioritaires	Octobre
	Exploitation des données de la campagne PM10	Fin décembre
	Rédaction d'une note méthodologique pour les PM10	Fin décembre
	Dossier type pour une campagne	Fin décembre

Renseignements synthétiques

Titre de l'étude	Méthodologie de détermination des zones géographiques concernées par les dépassements de seuils		
Personne responsable de l'étude	Laurent Létinois, Laure Malherbe, Laura Chiappini (INERIS)		
Travaux	Pluri-annuels		
Durée des travaux pluriannuels	2 ans (2009-2010) ou 3 ans (2009-2011)		
Collaboration AASQA	OUI avec toutes		
Heures d'ingénieur	EMD :	INERIS : 1500	LNE :
Heures de technicien	EMD :	INERIS : 300	LNE :
Document de sortie attendu	Note méthodologique		
Lien avec le tableau de suivi CPT	Thème : Modélisation - traitements numériques 6.6		
Lien avec un groupe de travail			
Matériel acquis pour l'étude			

ANNEXE 2 : CONTEXTE LEGISLATIF ET REGLEMENTAIRE (EUROPE ET FRANCE)

Tableau 2 : Dispositions européennes et françaises concernant la représentativité spatiale et les zones de dépassement

	Extraits des textes de loi et de réglementation	Référence	Remarques	Correspondance avec les formulaires de déclaration
Représentativité spatiale	<p>ANNEXE III</p> <p>Évaluation de la qualité de l'air ambiant et emplacement des points de prélèvement</p> <p>B. Macro-implantation des points de prélèvements</p> <p>1. Protection de la santé humaine</p> <p>a) Les points de prélèvement visant à assurer la protection de la santé humaine sont implantés de manière à fournir des renseignements sur:</p> <ul style="list-style-type: none"> • les endroits des zones et des agglomérations où s'observent les plus fortes concentrations auxquelles la population est susceptible d'être directement ou indirectement exposée pendant une période significative par rapport à la période considérée pour le calcul de la moyenne de la ou des valeurs limites, • les niveaux dans d'autres endroits à l'intérieur de zones ou d'agglomérations qui sont représentatifs de l'exposition de la population en général. <p>b) D'une manière générale, les points de prélèvement sont implantés de façon à éviter de mesurer les concentrations dans des microenvironnements se trouvant à proximité</p>	<p>Droit européen</p> <p>Directive 2008/50/CE</p>	<p>Pas d'exigence particulière quant à la géométrie de la zone de représentativité : au minimum, elle se compose de points situés à proximité immédiate de la station, sur une certaine distance ou une certaine surface ; au mieux, elle inclut également, des sites similaires non nécessairement proches.</p> <p>La Directive ne précise pas selon quel(s) critère(s) on peut affirmer qu'une station est représentative d'un point donné de l'espace, ni selon quelle échelle de temps cette représentativité doit être déterminée.</p> <p>Elle ne précise pas non plus selon quels critères deux sites peuvent être jugés similaires.</p>	

Extraits des textes de loi et de réglementation	Référence	Remarques	Correspondance avec les formulaires de déclaration
<p>immédiate. Autrement dit, un point de prélèvement doit être implanté de manière à ce que l'air prélevé soit représentatif de la qualité de l'air sur une portion de rue d'au moins 100 m de long pour les sites liés à la circulation et d'au moins 250 × 250 m pour les sites industriels, dans la mesure du possible.</p> <p>c) Les emplacements consacrés à la pollution de fond urbaine (...). Les points de prélèvement sont, en règle générale, représentatifs de plusieurs kilomètres carrés.</p> <p>(...)</p> <p>f) Les points de prélèvement sont, dans la mesure du possible, également représentatifs de sites similaires ne se trouvant pas à proximité immédiate.</p> <p>2. Protection de la végétation et des écosystèmes naturels</p> <p>Les points de prélèvement visant à assurer la protection de la végétation et des écosystèmes naturels sont implantés à plus de 20 km des agglomérations ou à plus de 5 km d'une autre zone bâtie, d'une installation industrielle, d'une autoroute ou d'une route principale sur laquelle le trafic est supérieur à 50 000 véhicules par jour. Autrement dit, un point de prélèvement doit être implanté de manière à ce que l'air prélevé soit représentatif de la qualité de l'air dans une</p>			

	Extraits des textes de loi et de réglementation	Référence	Remarques	Correspondance avec les formulaires de déclaration
	zone environnante d'au moins 1 000 km ² . Les États membres peuvent prévoir qu'un point de prélèvement sera implanté à une distance plus rapprochée ou qu'il sera représentatif de la qualité de l'air dans une zone moins étendue, compte tenu des conditions géographiques ou des possibilités de protection des zones particulièrement vulnérables.			
Représentativité spatiale	<p>ANNEXE VIII</p> <p>Critères de classification et d'implantation des points de prélèvement pour l'évaluation des concentrations d'ozone</p> <p>Type de station - Représentativité (1)</p> <p>Urbaine - Quelques km²</p> <p>Périurbaine - Quelques dizaines de km²</p> <p>Rurale - Quelques centaines de km²</p> <p>Rurale de fond - De 1000 à 10000 km²</p> <p>(1) Les points de prélèvement devraient, dans la mesure du possible, être également représentatifs de sites similaires ne se trouvant pas à proximité immédiate.</p>	<p>Droit européen</p> <p>Directive 2008/50/CE</p>	<p>Cette annexe précise les points b) et c) de l'annexe III.B pour les stations de mesure de l'ozone.</p>	

	Extraits des textes de loi et de réglementation	Référence	Remarques	Correspondance avec les formulaires de déclaration
Représentativité spatiale	<p>ANNEXE III</p> <p>Emplacement et nombre minimal des points de prélèvement pour la mesure des concentrations dans l'air ambiant et des taux de dépôt</p> <p>Les points de prélèvement devraient en général être situés de façon à éviter de mesurer des concentrations liées à des micro-environnements très petits se trouvant à proximité immédiate. À titre d'orientation, un point de prélèvement devrait être représentatif de la qualité de l'air dans une zone environnante d'au moins 200 m² pour les sites axés sur le trafic, d'au moins 250 m x 250 m pour les sites industriels lorsque cela est faisable, et de plusieurs kilomètres carrés pour les sites urbains de fond.</p> <p>(...)</p> <p>Les points de prélèvement devraient, dans la mesure du possible, être également représentatifs de sites similaires qui ne se trouvent pas à proximité immédiate. Le cas échéant, il convient de les implanter au même endroit que les points de prélèvement pour PM10.</p>	<p>Droit européen</p> <p>Directive 2004/107/CE</p>		

	Extraits des textes de loi et de réglementation	Référence	Remarques	Correspondance avec les formulaires de déclaration
Représentativité spatiale	<p>Article 6</p> <p>Pour chaque station de mesure dont ils ont la charge, les organismes agréés de surveillance de la qualité de l'air établissent une documentation exhaustive qui permet de vérifier que les critères d'implantation sont respectés. (...)</p> <p>Cette documentation comprend notamment les éléments requis aux paragraphes B, C et D de l'annexe III de la directive 2008/50/CE susvisée.</p>	<p>Droit français</p> <p>Arrêté du 21 octobre 2010</p>	<p>Conformément au point B détaillé plus haut, les éléments requis pour chacune des stations de mesure comprennent des informations sur la représentativité spatiale.</p>	
Dépassement de seuil	<p>ANNEXE I</p> <p>Objectifs de qualité des données</p> <p>B. Résultats de l'évaluation de la qualité de l'air</p> <p>Les informations ci-après sont réunies pour les zones ou agglomérations pour lesquelles d'autres sources de renseignements complètent les données fournies par la mesure ou sont les seuls moyens d'évaluation de la qualité de l'air:</p> <ul style="list-style-type: none"> — (...) — description des résultats, y compris 	<p>Droit européen</p> <p>Directive 2008/50/CE</p>	<p>Contexte : Evaluation de la qualité de l'air à l'aide de moyens complémentaires de la mesure fixe</p> <p>Seuils considérés : valeurs limites, valeurs cibles, objectifs à long terme, seuils d'évaluation inférieur et supérieur</p>	<p>Rapportage <i>Zonage</i></p> <p>Formulaire 19 : Résultats tabulaires de l'évaluation complémentaire et méthodes employées (cf. Tableau 3)</p>

	Extraits des textes de loi et de réglementation	Référence	Remarques	Correspondance avec les formulaires de déclaration
	<p>les incertitudes et, en particulier, indication de l'étendue de tout site («extent of any <u>area</u>» dans le texte anglais) ou, le cas échéant, de la longueur de route à l'intérieur de la zone ou de l'agglomération où les concentrations dépassent une valeur limite, une valeur cible ou un objectif à long terme majoré, le cas échéant, de la marge de dépassement, et l'étendue de tout site (« area ») à l'intérieur duquel les concentrations dépassent le seuil d'évaluation supérieur ou le seuil d'évaluation inférieur,</p> <p>— la population potentiellement exposée à des niveaux dépassant une valeur limite pour la protection de la santé humaine.</p>			
Dépassement de seuil	<p>Article 10</p> <p>IV. Pour les stations de mesure fixes où un dépassement des valeurs limites ou des valeurs cibles a été constaté, les organismes agréés de la surveillance de la qualité de l'air réalisent une carte délimitant l'aire de la zone en dépassement et précisant la valeur mesurée, sur fond de carte IGN à l'échelle appropriée. La population et la superficie exposée au dépassement sont également</p>	<p>Droit français</p> <p>Arrêté du 21 octobre 2010</p>	<p>Contexte : Evaluation de la qualité de l'air</p> <p>Seuils considérés : valeurs limites, valeurs cibles</p>	

	Extraits des textes de loi et de réglementation	Référence	Remarques	Correspondance avec les formulaires de déclaration
	indiquées en précisant les méthodologies utilisées au regard des recommandations du Laboratoire de surveillance de la qualité de l'air.			
Dépassement de seuil	<p>ANNEXE XV</p> <p>Informations devant figurer dans les plans relatifs à la qualité de l'air locaux, régionaux ou nationaux destinés à améliorer la qualité de l'air ambiant</p> <p>A. Informations à communiquer au titre de l'article 23 (plans)</p> <p>1. Lieu du dépassement (...)</p> <p>2. Informations générales</p> <p>a) type de zone (ville, zone industrielle ou rurale);</p> <p>b) estimation de la superficie polluée (en km²) et de la population exposée à la pollution;</p> <p>(...)</p>	<p>Droit européen</p> <p>Directive 2008/50/CE</p>	<p>Contexte : Plans relatifs à la qualité de l'air</p> <p>Situations dans lesquelles sont établis de tels plans : « Lorsque, dans une zone ou agglomération donnée, les niveaux de polluants dans l'air ambiant dépassent toute valeur limite ou toute valeur cible, majorée dans chaque cas de toute marge de dépassement » (Art. 23)</p> <p>Seuils considérés : valeurs limites et valeurs cibles</p>	<p>Rapportage <i>Plans et Programmes</i></p> <p>Formulaire n°2 : description du dépassement de la valeur limite, points l, m et n (cf. Tableau 3)</p>
Dépassement de seuil	<p>Doivent être couvertes par un plan de protection de l'atmosphère :</p> <p>1° Les agglomérations de plus de 250 000 habitants ;</p> <p>2° Les zones dans lesquelles le niveau dans l'air ambiant de l'un au moins des polluants, évalué conformément aux dispositions des articles R.</p>	<p>Droit français</p> <p>Section 2 du chapitre II du titre II du livre II du code de l'environnement (partie réglementaire,</p>	<p>Contexte : Plans relatifs à la qualité de l'air (plans de protection de l'atmosphère)</p>	

	Extraits des textes de loi et de réglementation	Référence	Remarques	Correspondance avec les formulaires de déclaration
	221-1 à R. 221-3, dépasse ou risque de dépasser une valeur limite ou une valeur cible mentionnée à l'article R.221-1. (...)	article R. 222-13) modifiée par le Décret du 21 octobre 2010 (Art 5)		
Dépassement de seuil	<p>ANNEXE XVI</p> <p>INFORMATION DU PUBLIC</p> <p>4. Les États membres veillent à ce que le public soit informé en temps utile des dépassements constatés ou prévus en ce qui concerne les seuils d'alerte et les seuils d'information. Les renseignements fournis comportent au moins les informations suivantes:</p> <p>a) des informations sur le ou les dépassements observés:</p> <ul style="list-style-type: none"> - lieu ou zone (« area ») du dépassement, - type de seuil dépassé (seuil (...)) <p>b) des prévisions pour l'après-midi ou le ou les jours suivants:</p> <ul style="list-style-type: none"> - zone géographique où sont prévus des dépassements du seuil d'information et/ou d'alerte, (...) 	<p>Droit européen</p> <p>Directive 2008/50/CE</p>	<p>Contexte : information du public</p> <p>Seuils considérés : seuils d'information et d'alerte</p>	

Tableau 3 : Détail des formulaires pour la transmission d'informations relatives aux dépassements

<p>Rapportage Plans et Programmes</p> <p>Formulaire n°2 : description du dépassement de la valeur limite, points l, m et n</p>	<p>Informations à fournir par zone, dans les situations de dépassement concernées :</p> <ul style="list-style-type: none"> • l. Estimation de la superficie (km²) où le niveau était supérieur à la VL dans l'année de référence • m. Estimation de la distance (km) sur laquelle le niveau était supérieur à la VL dans l'année de référence • n. Estimation de la population totale exposée à un niveau dépassant la VL dans l'année de référence 	
<p>Rapportage Zonage</p> <p>Formulaire 19 : Résultats tabulaires de l'évaluation complémentaire et méthodes employées</p>	<p>Selon les seuils concernés (valeurs limites, valeurs cibles, objectifs à long terme):</p> <p>SO₂, O₃</p> <p>Secteur concerné (en km²) et population exposée</p> <p>Secteur concerné (en km²) et superficie d'écosystème exposée</p> <p>NO₂/NO_x</p> <p>Secteur concerné (en km²), longueur des tronçons routiers concernés (km) et population exposée</p> <p>Secteur concerné (en km²) et superficie de végétation exposée</p> <p>PM10, Pb, C6H6, CO</p> <p>Secteur concerné (en km²), longueur des tronçons routiers concernés (km) et population exposée</p>	<p>Le formulaire 19 peut être complété par des cartes indiquant la répartition des concentrations. Il est recommandé à l'État membre, dans la mesure du possible, de compiler les cartes représentant la répartition des concentrations dans chaque zone et agglomération. Il est également recommandé de fournir les isolignes des concentrations pour les paramètres dans lesquels les seuils de qualité sont exprimés (voir tableau 3) au moyen d'isolignes situées à des intervalles de 10 % du seuil.</p>

ANNEXE 3 : ELEMENTS THEORIQUES SUR LA METHODOLOGIE

A.3.1. REPRESENTATIVITE EN CONCENTRATION

A.3.1.1 PREMIERE DEFINITION

On cherche à mettre en place une méthode permettant de définir la représentativité spatiale d'une station fixe pour la moyenne annuelle de NO₂. On considère qu'un point du domaine d'étude se situe dans la zone de représentativité en concentration de la station si l'écart entre leurs concentrations respectives est faible.

On définira alors un écart-maximal toléré par rapport à la station, auquel on comparera la différence de concentration pour statuer sur l'appartenance du point à la zone de représentativité.

On note :

$Z(x)$: la concentration moyenne annuelle de NO₂ au point x

$Z(x_0)$: la concentration moyenne annuelle de la station située en x_0

δ : écart maximal toléré entre le point x et la station fixe située en x_0

Alors le point x appartient à la zone de représentativité en concentration de la station située en x_0 si :

$$|Z(x) - Z(x_0)| < \delta \quad (\text{Def.1})$$

La zone de représentativité en concentration de $Z(x_0)$, notée A_0 peut donc se définir formellement par :

$$A_0 = \{x : |Z(x) - Z(x_0)| < \delta\} \quad (\text{Def.2})$$

A.3.1.2 MODELISATION PROBABILISTE AVEC HYPOTHESE D'ACCROISSEMENT GAUSSIEN

La concentration de NO₂ n'étant connue qu'aux points de mesure, i.e. aux stations fixes, le recours à la modélisation est indispensable.

On considère alors la concentration moyenne annuelle de NO₂ comme la réalisation d'une fonction aléatoire Z .

Cette approche permet d'étendre la définition d'une zone de représentativité, afin qu'elle soit vérifiée en espérance :

$$A_0 = \{x : E(|Z(x) - Z(x_0)|) < \delta\} \quad (\text{Def.4})$$

Cette définition n'est toutefois pas assez restrictive pour le calcul des zones puisqu'appliquer cette inégalité à l'espérance de l'écart signifie que cet écart peut être dépassé avec une probabilité importante.

On cherche donc à adapter cette définition en introduisant la notion de risque statistique - noté η_{RC} - que l'écart à la station d'un point appartenant à A_0 au sens de la définition 2, dépasse δ .

$$\mathbb{P}[|Z(x) - Z(x_0)| \geq \delta] < \eta_{RC} \quad (Eq.5)$$

Il est possible de majorer cette équation par l'inégalité de Bienaymé-Tchebychev, qui sera cependant très large (cette majoration vaut pour toutes les lois de probabilité), et va donc nous fournir des zones de représentativité de tailles excessivement restreintes.

Cependant, dans notre contexte d'étude, il est possible de supposer conventionnellement la loi de l'accroissement $\Delta_{x-x_0} = Z(x) - Z(x_0)$ comme gaussienne.

Soit Δ_{x-x_0} de loi normale, centrée et de variance σ^2 , alors, si l'on note q_α le quantile d'ordre α d'une gaussienne réduite, on a :

$$\mathbb{P}\left[\frac{\Delta_{x-x_0}}{\sigma} < q_\alpha\right] = \alpha \text{ et } \mathbb{P}\left[\frac{\Delta_{x-x_0}}{\sigma} \geq q_\alpha\right] = 1 - \alpha \quad (Eq.6)$$

Et d'après la symétrie de la gaussienne centrée, on peut dire :

$$\mathbb{P}\left[\frac{|\Delta_{x-x_0}|}{\sigma} < q_\alpha\right] = 2\alpha - 1 \text{ et } \mathbb{P}\left[\frac{|\Delta_{x-x_0}|}{\sigma} \geq q_\alpha\right] = 2(1 - \alpha) \quad (Eq.7)$$

Ainsi, pour un risque statistique η_{RC} fixé, on prendra $2(1 - \alpha) = \eta_{RC}$, i.e. $\alpha = 1 - \frac{\eta_{RC}}{2}$. Dans la pratique, on prendra généralement un risque statistique de 5% ou 10%.

A.3.1.3 MODELISATION CONDITIONNELLEMENT AUX MESURES

Dans la pratique, la concentration au point x n'est pas connue, on utilise donc la valeur estimée par krigeage sur les données disponibles (obtenues à partir d'une campagne par tubes à diffusion passive par exemple), que l'on note $Z^*(x)$, et on connaît l'écart-type de krigeage $\sigma_K(x)$ associée à cette estimation.

Notons alors ε l'erreur d'estimation, $Z = Z^* + \varepsilon$ (Def.8)

La définition 1 devient $|Z^*(x) + \varepsilon(x) - Z(x_0)| < \delta$, (on négligera alors l'erreur de mesure à la station. Ainsi, une condition suffisante pour que la définition 1 soit vérifiée est :

$$|\varepsilon(x)| < \delta - |Z^*(x) - Z(x_0)| \quad (Eq.9)$$

Comme décrit dans la section précédente, on introduit le risque statistique que cette inégalité ne soit pas vérifiée :

$$\mathbb{P}[|\varepsilon(x)| \geq \delta - |Z^*(x) - Z(x_0)|] < \eta_{RC} \quad (\text{Eq.10})$$

Conventionnellement, on choisit $\varepsilon(x)$ de loi normale centrée et dont l'écart-type est égal à l'écart-type de krigeage en x , et alors l'équation 11 se traduit par :

$$|Z^*(x) - Z^*(x_0)| < \delta - \sigma_K(x) * q_{1-\frac{\eta_{RC}}{2}} \quad (\text{Eq.11})$$

Le choix du risque statistique est très important dans le calcul des zones de représentativité en concentration : en effet diminuer le risque statistique revient à corriger de façon plus importante l'écart maximal toléré par rapport à la station. Le critère de rattachement à une station sera donc plus strict et les surfaces des zones de représentativité s'en verront diminuées.

Au risque 5% de dépasser l'écart maximal toléré, l'équation 12 devient :

$$|Z^*(x) - Z^*(x_0)| < \delta - 1.96 * \sigma_K(x)$$

Au risque 10% de dépasser l'écart maximal toléré, l'équation 12 devient :

$$|Z^*(x) - Z^*(x_0)| < \delta - 1.65 * \sigma_K(x)$$

A.3.1.4 PROBLEME DES INTERSECTIONS

Lors du calcul des zones de représentativité en concentration, il est possible que certaines zones se recouvrent, notamment si elles présentent des concentrations de NO_2 assez proches.

Afin d'obtenir une partition du domaine d'étude ne faisant pas intervenir d'intersections dans les zones de représentativité, il peut apparaître judicieux de rattacher chaque point à une seule et unique station. Pour un point appartenant à plusieurs zones de représentativité, différents critères de rattachement pourront être testés afin de voir lequel de ces critères apparaît le plus cohérent spatialement, mais aussi lequel offre la meilleure stabilité temporelle.

On propose donc une nouvelle définition de la zone de représentativité en concentration en fonction de ces différents critères :

Critère n°1 : distance à la station la plus faible.

Critère n°2 : écart de concentration par rapport à la station le plus faible. La zone de représentativité A_0 de la station fixe en x_0 , en notant $j=1, \dots, n$ les stations fixes, se définit maintenant par :

$$A_0 = \{x : |Z(x) - Z(x_0)| \leq \delta\} \cap \{x : x_0 = \operatorname{argmin} |Z(x) - Z(x_0)|\}$$

Si on introduit le risque statistique de dépassement de seuil, alors :

$$A_0 = \left\{x / |Z^*(x) - Z^*(x_0)| < \delta - \sigma_K(x) * q_{1-\frac{\eta_{RC}}{2}}\right\} \cap \{x / x_0 = \operatorname{argmin} |Z(x) - Z(x_0)|\}$$

Critère n°3 : similarité de milieu avec les stations. Pour ce critère, deux approches sont envisagées.

On pourra comparer les distances euclidiennes entre les vecteurs de milieu des stations et celui du point appartenant à la zone de recouvrement. Un vecteur de milieu sera constitué des valeurs des variables choisies pour représenter le milieu d'un point.

Si $x \in A_i \cap A_j$, zone de recouvrement des stations fixes en x_i et x_j , alors A_k , la zone de rattachement du point x , se calculera comme suit :

$$k = \operatorname{argmin} (\|f(x) - f(x_k)\|) \text{ avec } k \in (i, j)$$

où $f(x) = (f_1(x), f_2(x), \dots, f_n(x))^t$ avec $f_j(x)$, valeur de la $j^{\text{ième}}$ variable de milieu en x

On pourra également réaliser une Analyse en Composantes principales (ACP) sur les variables de milieu, et comparer les écarts entre les stations et le point x sur l'axe Δ_1 d'inertie minimale résultant de l'ACP (si on projette les variables de milieu sur Δ_1 , on aura l'image la moins déformée possible du nuage des individus sur un axe).

$$\begin{aligned} &\Delta_1 \text{ axe d'inertie minimale, } h_{\Delta_1}(x) \text{ projection orthogonale de } x \text{ sur } \Delta_1 \\ &\text{alors, } k = \operatorname{argmin} (|h_{\Delta_1}(x) - h_{\Delta_1}(x_k)|) \text{ avec } k \in (i, j) \end{aligned}$$

A.3.2 ESTIMATION LOCALE ET DEPASSEMENTS DE SEUIL

A.3.2.1 PREMIERE APPROCHE

En conservant le même cadre théorique que précédemment, on peut, indépendamment des zones de représentativité de concentration définies dans les paragraphes précédents, évaluer en tout point du domaine le dépassement de seuil. Soit s le seuil limite, il est possible, tout comme pour le calcul des zones de représentativité, de prendre en compte l'erreur d'estimation du krigeage pour corriger ce seuil statistiquement.

On a déjà vu que $Z(x) = Z^*(x) + \varepsilon(x)$ avec, dans le cadre conventionnel, $\varepsilon(x)$, gaussienne centrée dont l'écart-type est égal à l'écart-type de krigeage en x , ce qui se traduit par :

$$\begin{aligned} \varepsilon(x) &= Z(x) - Z^*(x) = \sigma_K(x).T \\ &\text{où } T \sim \mathcal{N}(0,1) \end{aligned} \tag{Eq.12}$$

Formellement, déterminer si la concentration en un point du domaine dépasse le seuil s'écrit :

$$\begin{aligned} Z(x) > s &\Leftrightarrow Z^*(x) + \sigma_K(x).T > s \\ \Leftrightarrow T &> \frac{s - Z^*(x)}{\sigma_K(x)} \end{aligned} \tag{Eq.13}$$

Introduisons désormais le risque statistique η_{DS} , que la concentration d'un point qui ne vérifie pas l'équation 13, i.e. qui n'est pas considéré dans la zone de dépassement, dépasse en réalité le seuil :

$$\begin{aligned} & \mathbb{P}[Z(x) > s] < \eta_{DS} \\ \Leftrightarrow & \mathbb{P}\left[T > \frac{s - Z^*(x)}{\sigma_K(x)}\right] < \eta_{DS} \\ \Leftrightarrow & \mathbb{P}\left[T \leq \frac{s - Z^*(x)}{\sigma_K(x)}\right] > 1 - \eta_{DS} \end{aligned} \quad (Eq. 14)$$

La loi de T étant fixée, on connaît ses quantiles, et l'équation 14 est vérifiée sous la condition suivante :

$$\begin{aligned} & \mathbb{P}\left[T \leq \frac{s - Z^*(x)}{\sigma_K(x)}\right] > 1 - \eta_{DS} \\ \Leftrightarrow & q_{1-\eta_{DS}} < \frac{s - Z^*(x)}{\sigma_K(x)} \\ \Leftrightarrow & Z^*(x) < s - \sigma_K(x) * q_{1-\eta_{DS}} \end{aligned} \quad (Eq. 15)$$

Il est possible de visualiser le passage de l'équation 14 au critère 15 en mettant en relation la densité de T et sa fonction de répartition.

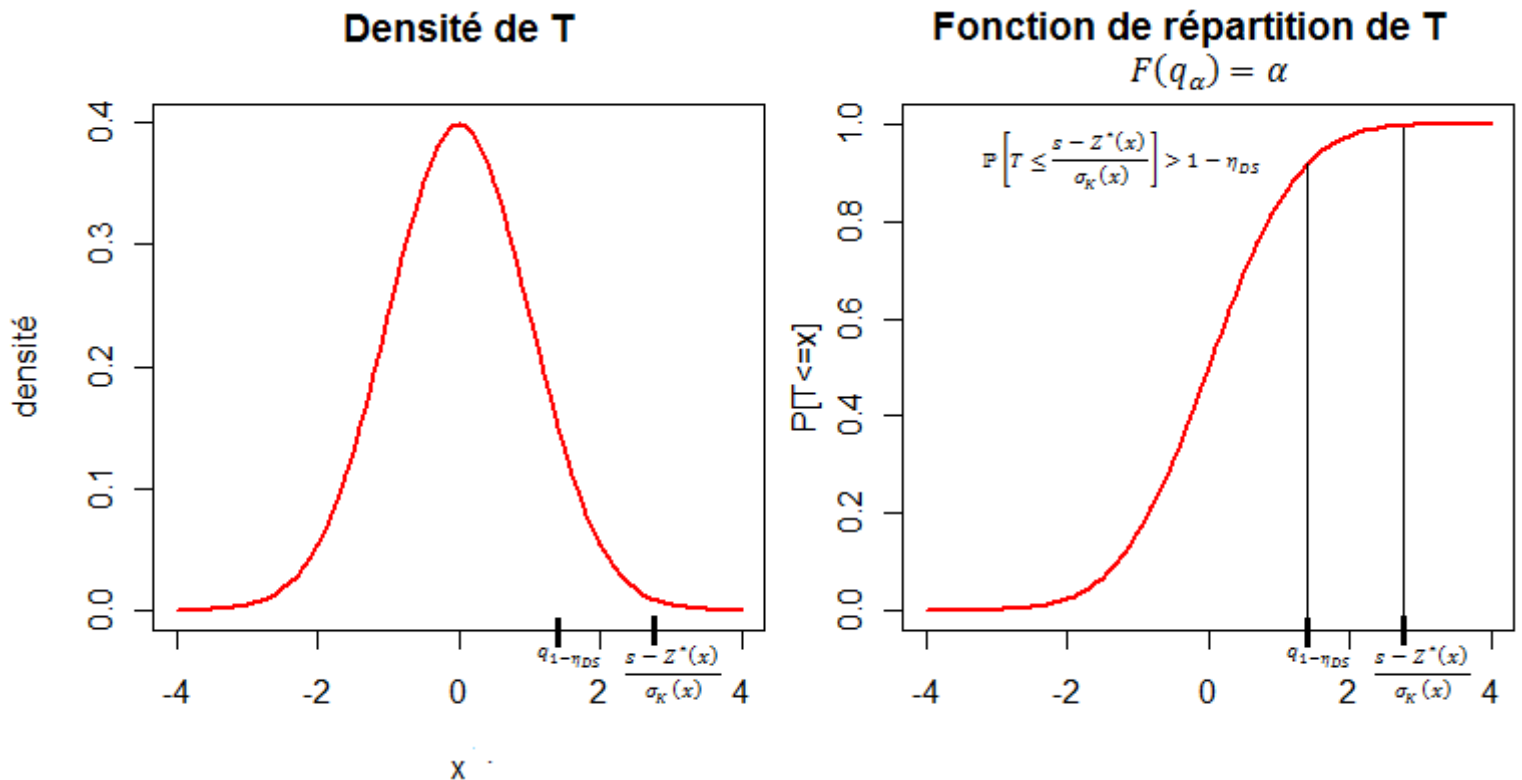


Figure 19 : Illustrations du passage de l'équation 14 au critère 15 de non-sélection d'un point en dépassement

Cette méthode permet de réduire le risque de non-détection des points dont la concentration excède le seuil limite, en fixant le critère de non-dépassement au seuil corrigé $s - \sigma_K(x)q_{1-\eta_{DS}}$, inférieur à s et fonction de la localisation de la maille – plus

un point est éloigné des mesures, plus la qualité de l'estimation sera faible et plus le seuil sera corrigé, donc faible.

Au risque 5% qu'un point qui n'est pas considéré en dépassement dépasse en réalité le seuil, la condition 15 se traduit :

$$Z^*(x) < s - 1.65 * \sigma_K(x) \quad (\text{Eq.16})$$

On précise que si l'on considère une hypothèse de distribution normale pour les erreurs, alors choisir un risque statistique de 50% équivaut à ne pas corriger le seuil ($q_{0.5} = 0$).

A.3.2.2 RISQUES DE PREMIERE ET SECONDE ESPECE

La première approche utilisée pour quantifier les dépassements de seuil a pour défaut de construire des zones de dépassement trop réduites si le risque statistique de ne pas dépasser dans la zone de dépassement est trop faible.

On cherche donc dans un deuxième temps à définir des zones où les concentrations ne dépasseront pas la valeur limite avec une probabilité très forte et des zones où les concentrations dépasseront la valeur limite avec une probabilité plus faible. L'idée est la suivante : on préfère considérer un point en dépassement alors qu'il ne l'est pas réellement, plutôt que de ne pas considérer un point en dépassement alors qu'il l'est.

Formellement, on obtient les définitions suivantes :

Zone de non-dépassement : $\mathbb{P}[Z(x) > s] < \alpha$ où α est le risque de première espèce. α doit être faible et est fixé à 5% puisqu'on cherche à ce que la probabilité $1 - \alpha$ que la concentration ne dépasse pas le seuil soit très élevée. La zone est donc construite sur le critère suivant :

$$\begin{aligned} \mathbb{P}\left[T > \frac{s - Z^*(x)}{\sigma_K(x)}\right] &< \alpha \\ \Leftrightarrow \mathbb{P}\left[T \leq \frac{s - Z^*(x)}{\sigma_K(x)}\right] &> 1 - \alpha \\ \Leftrightarrow q_{1-\alpha} &< \frac{s - Z^*(x)}{\sigma_K(x)} \\ \Leftrightarrow Z^*(x) &< s - \sigma_K(x) * q_{1-\alpha} \\ \Leftrightarrow Z^*(x) &< s - 1.65 * \sigma_K(x) \end{aligned} \quad (\text{Eq.17})$$

Par conséquent, **augmenter le risque de première espèce aura pour effet d'augmenter la surface de la zone de non-dépassement de valeur limite.**

Zone de dépassement : $\mathbb{P}[Z(x) \leq s] < \beta$ où β est le risque de seconde espèce. β sera plus souple que le risque de première espèce et sera fixé conventionnellement à 34%. La zone est donc construite sur le critère suivant :

$$\begin{aligned} \mathbb{P}\left[T \leq \frac{s - Z^*(x)}{\sigma_K(x)}\right] &< \beta \\ \Leftrightarrow q_\beta &> \frac{s - Z^*(x)}{\sigma_K(x)} \\ \Leftrightarrow Z^*(x) &> s - \sigma_K(x) * q_\beta \\ \Leftrightarrow Z^*(x) &> s + 0.41 * \sigma_K(x) \quad (\text{Eq. 18}) \end{aligned}$$

Par conséquent, **diminuer le risque de seconde espèce aura pour effet de diminuer la surface de la zone de dépassement de valeur limite.**

A.3.3 PRISE EN COMPTE DE LA POLLUTION DE PROXIMITE

A.3.3.1 AJOUTS THEORIQUES

Soit $\varepsilon_F(x) = Z_F(x) - Z_F^*(x)$ l'erreur commise sur la pollution de fond

Soit $\varepsilon_T(x) = \Delta_{TF}(x) - \Delta_{TF}^*(x)$ l'erreur commise sur le surplus de pollution routier

Alors $\varepsilon(x) = Z(x) - Z^*(x)$

$$= [Z_F(x) + \Delta_{TF}(x)] - [Z_F^*(x) + \Delta_{TF}^*(x)]$$

$$= Z_F(x) - Z_F^*(x) + \Delta_{TF}(x) - \Delta_{TF}^*(x)$$

$$= \varepsilon_F(x) + \varepsilon_T(x)$$

On suppose, comme précédemment, que $\varepsilon_F(x) = \sigma_K(x).T$ où $T \sim \mathcal{N}(0,1)$

Il ne reste qu'à considérer $\varepsilon_T(x)$.

Première possibilité : Le modèle de surplus de pollution étant peu fiable, on considère $\varepsilon_T(x)$ constante. Cette constante sera égale à l'écart-type des résidus en validation croisée du Δ_{TF} , noté σ_S . Alors, on aura :

$$\varepsilon(x) = \sigma_K(x).T + \sigma_S = \sigma_K(x).T' \text{ où } T' \sim \mathcal{N}(\sigma_S, 1)$$

Cette première approche a pour défaut de considérer l'erreur liée au surplus de pollution toujours positive ($\sigma_S > 0$), i.e. on considère que l'on sous-estime toujours ce surplus.

Nouveau critère pour la représentativité en concentration	Nouveau critère pour les dépassements
$\mathbb{P}[\varepsilon_F(x) + \varepsilon_T(x) \geq \delta - Z^*(x) - Z(x_0)] < \eta_{RC}$ $\Leftrightarrow \mathbb{P}[\varepsilon_F(x) + \varepsilon_T(x) < \delta - Z^*(x) - Z(x_0)] > 1 - \eta_{RC}$ $\stackrel{c.s.}{\Rightarrow} \mathbb{P}[\varepsilon_F(x) + \varepsilon_T(x) < \delta - Z^*(x) - Z(x_0)] > 1 - \eta_{RC}$ $\Leftrightarrow \mathbb{P}[\varepsilon_F(x) < \delta - Z^*(x) - Z(x_0) - \sigma_S] > 1 - \eta_{RC}$ $\Leftrightarrow Z^*(x) - Z^*(x_0) < \delta - \sigma_S - \sigma_K(x) * q_{1-\frac{\eta_{RC}}{2}}$	$\varepsilon(x) = \varepsilon_F(x) + \varepsilon_T(x) = \sigma_K(x).T + \sigma_S$ <p>Zone de non-dépassement :</p> $\mathbb{P}[Z(x) > s] < \alpha$ $\stackrel{c.s.}{\Rightarrow} \mathbb{P}\left[T \leq \frac{s - Z^*(x) - \sigma_S}{\sigma_K(x)}\right] > 1 - \alpha$ $\Leftrightarrow Z^*(x) < s - \sigma_S - \sigma_K(x) * q_{1-\alpha}$ <p>Zone de dépassement :</p> $\mathbb{P}[Z(x) \leq s] < \beta \Leftrightarrow Z^*(x) > s - \sigma_S - \sigma_K(x) * q_\beta$

Deuxième possibilité : On fait une hypothèse forte sur les erreurs d'estimations du surplus de pollution lié au trafic en considérant que cette erreur suit une loi gaussienne centrée en 0 et d'écart-type égal à l'écart-type des résidus en validation croisée du Δ_{TF} , noté σ_S . Cette hypothèse n'est pas incohérente si le modèle d'estimation du Δ_{TF} est suffisamment fiable. Alors, on aura :

$$\varepsilon(x) = \sigma_K(x).T + \sigma_S.T = \sqrt{\sigma_K^2(x) + \sigma_S^2}.T \text{ où } T \sim \mathcal{N}(0,1)$$

Nouveau critère pour la représentativité en concentration	Nouveau critère pour les dépassements
$\mathbb{P}[\varepsilon_F(x) + \varepsilon_T(x) \geq \delta - Z^*(x) - Z(x_0)] < \eta_{RC}$ $\Leftrightarrow \mathbb{P}[\varepsilon_F(x) + \varepsilon_T(x) < \delta - Z^*(x) - Z(x_0)] > 1 - \eta_{RC}$ $\Leftrightarrow \mathbb{P}[\varepsilon(x) < \delta - Z^*(x) - Z(x_0)] > 1 - \eta_{RC}$ $\Leftrightarrow Z^*(x) - Z^*(x_0) < \delta - \sqrt{\sigma_K^2(x) + \sigma_S^2} * q_{1-\frac{\eta_{RC}}{2}}$	$\varepsilon(x) = \varepsilon_F(x) + \varepsilon_T(x) = \sigma_K(x).T + \sigma_S.T$ <p>Zone de non-dépassement :</p> $\mathbb{P}[Z(x) > s] < \alpha$ $\Leftrightarrow \mathbb{P}\left[T \leq \frac{s - Z^*(x)}{\sqrt{\sigma_K^2(x) + \sigma_S^2}}\right] > 1 - \alpha$ $\Leftrightarrow Z^*(x) < s - \sqrt{\sigma_K^2(x) + \sigma_S^2} * q_{1-\alpha}$ <p>Zone de dépassement :</p> $\mathbb{P}[Z(x) \leq s] < \beta \Leftrightarrow Z^*(x) > s - \sqrt{\sigma_K^2(x) + \sigma_S^2} * q_\beta$

A.3.3.2 APPLICATION NUMERIQUE

A.3.3.2.1 MONTPELLIER

La cartographie de la représentativité en concentration des stations fixes sur Montpellier est similaire à celle obtenue avec le terme d'erreur constant. Les cartes de dépassement de seuil présentent en revanche des différences significatives.

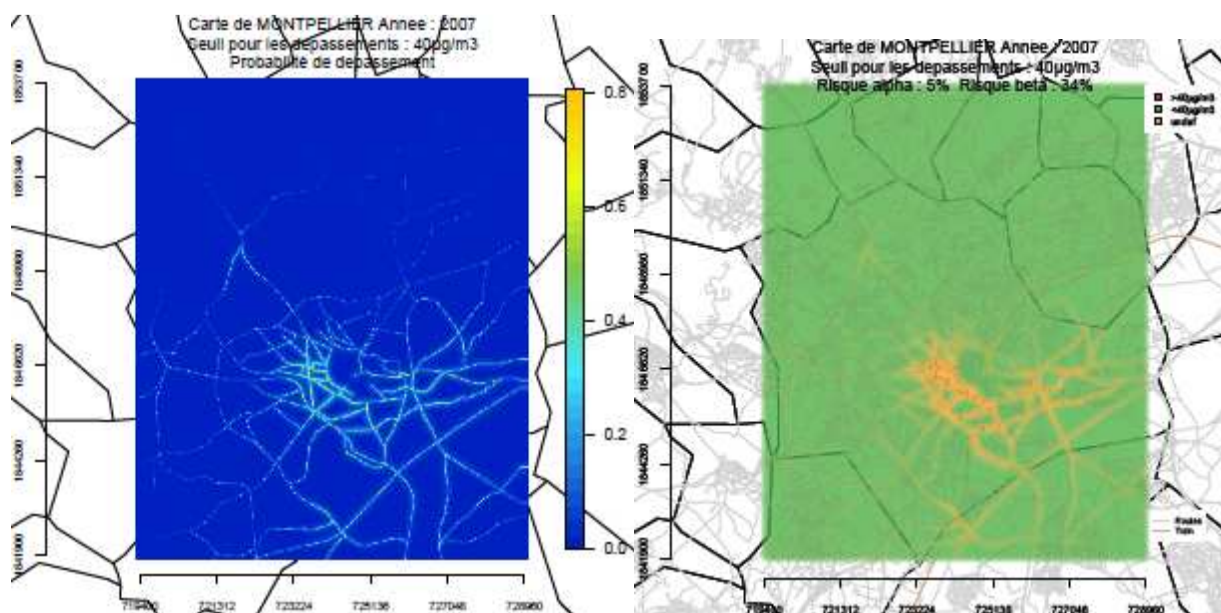


Figure 20 : Probabilités de dépasser les 40µg/m3 sur Montpellier en 2007 et Carte de dépassements pour le seuil 40 µg/m3 et pour les risques de 1^{ère} (5%) et 2^e espèce (34%)

A.3.3.2.2 TROYES

La cartographie de la représentativité en concentration des stations fixes sur Troyes évolue légèrement, mais les remarques présentées en annexe 3 sont toujours valables:

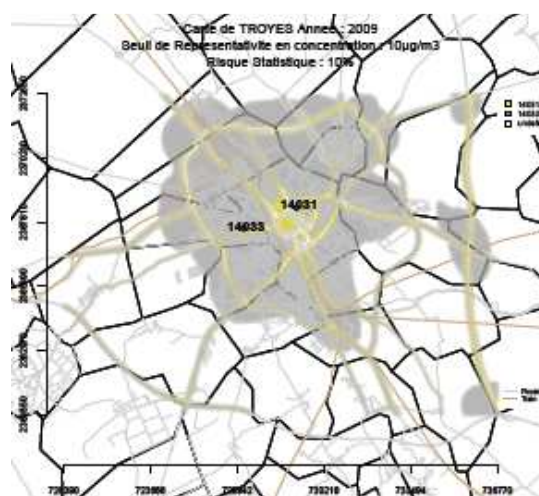


Figure 21 : Zones de représentativité en concentration des stations 14031 et 14033 (stations de fond) sur l'agglomération de Troyes après ajout de la proximité

Les cartes de dépassement de seuil sont différentes elles aussi :

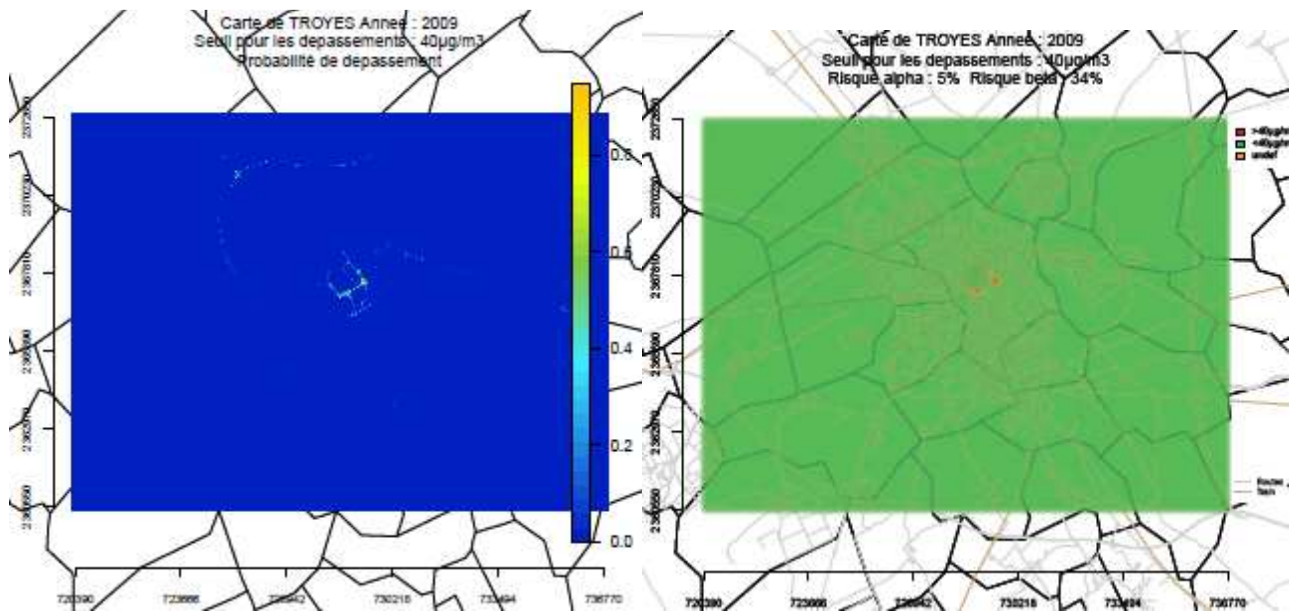


Figure 22 : Probabilités de dépasser les 40µg/m3 sur Troyes en 2009 et Carte de dépassements pour le seuil 40 µg/m3 et pour les risques de 1^{ère} (5%) et 2^e espèce (34%)

En ce qui concerne les dépassements, les mêmes observations peuvent être faites pour les deux agglomérations : l'hypothèse gaussienne de distribution de l'erreur du modèle de surplus de pollution routier a pour effet, à risques statistiques identiques, de réduire la surface des zones de dépassement.

ANNEXE 4 : APPLICATION DE LA METHODOLOGIE SUR L'AGGLOMERATION DE TROYES

A.4.1 Estimation de la pollution de fond

A.4.1.1 Estimation de la pollution de fond

L'application de la méthodologie sur Troyes donne l'occasion de tester l'influence de l'ajustement du variogramme sur les estimations de la pollution de fond et sur les écart-types de krigeage.

La méthode d'ajustement du variogramme utilisée sous R choisit le modèle qui présente, par rapport au variogramme empirique, l'erreur quadratique minimale (les erreurs calculées pour chaque modèle sont pondérées en fonction du nombre de points disponibles pour chaque distance étudiée).

Sur la région de Troyes, on voit que des modèles sphériques ou exponentiels s'ajustent convenablement. On propose d'étudier les résultats obtenus avec un exemple de chacun de ces modèles.

On précise qu'à la différence de Montpellier, la variable utilisée en dérive dans le krigeage est la quantité d'émissions de NO_x et non plus par le logarithme translaté de ces émissions, les données brutes étant mieux corrélées avec les concentrations de NO_2 .

A.4.1.1.a Ajustement du variogramme par un modèle exponentiel

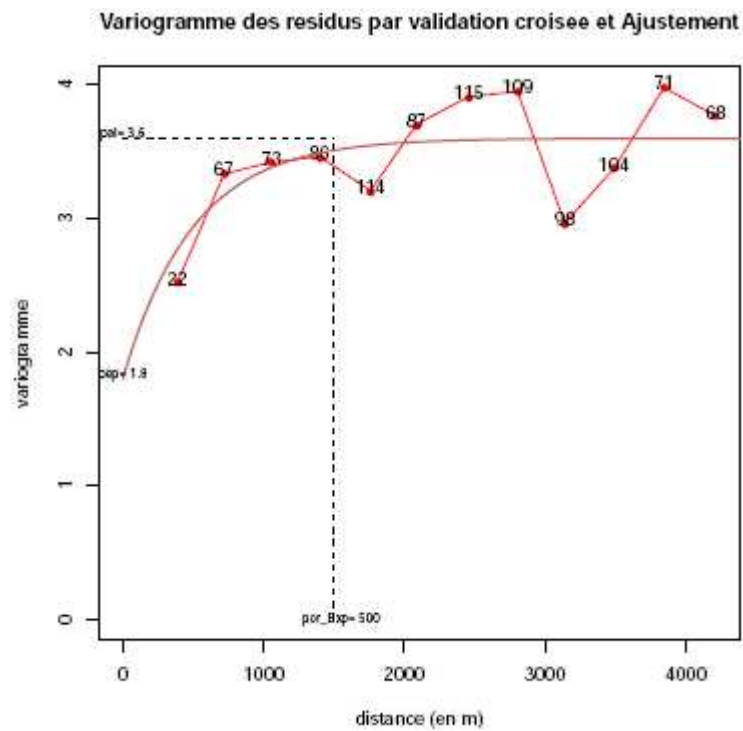


Figure 23 : Variogramme exponentiel ajusté sur le variogramme empirique à partir des données de tubes sur Troyes en 2009 (paramètres du variogramme : Exponentiel ; pépité=1.8 ; palier=1.8 ; portée=500)

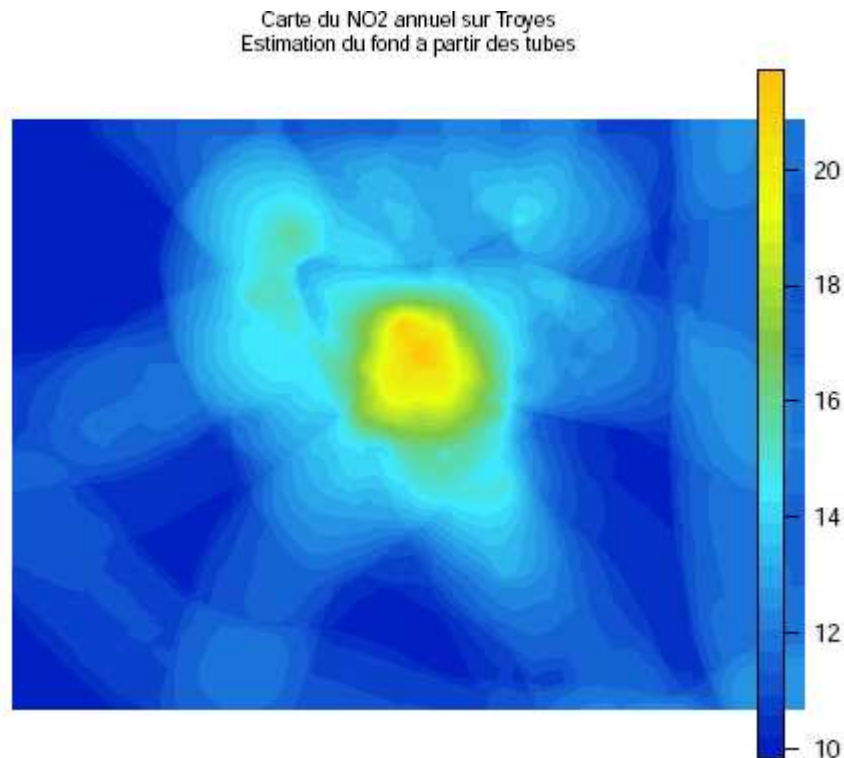


Figure 24 : Estimation de la pollution de fond du NO₂ sur Troyes en 2009 (paramètres du variogramme : Exponentiel ; pépité=1.8 ; palier=1.8 ; portée=500)

Ecart-type de krigage du NO₂ annuel sur Troyes
Estimation du fond à partir des tubes

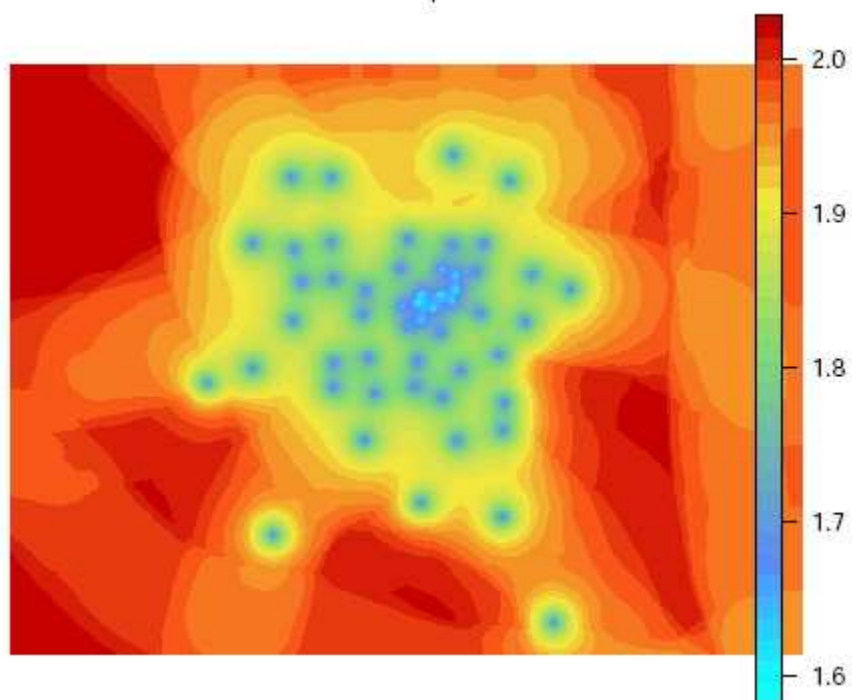


Figure 25 : Ecart-type de krigage pour l'estimation de fond du NO₂ sur Troyes en 2009 (paramètres du variogramme : Exponentiel ; pépité=1.8 ; palier=1.8 ; portée=500)

A.4.1.1.b Ajustement du variogramme par un modèle sphérique

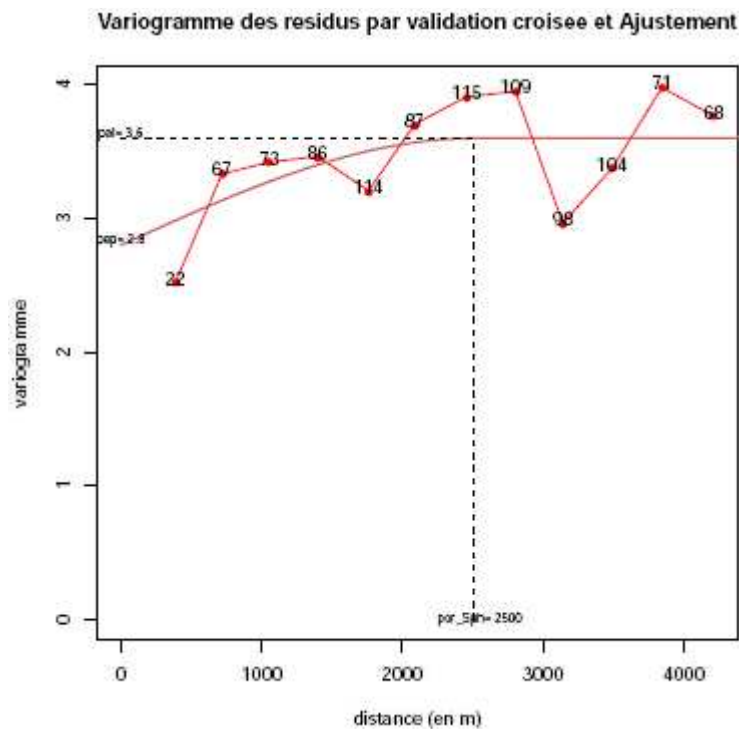


Figure 26 : Variogramme sphérique ajusté sur le variogramme empirique à partir des données de tubes sur Troyes en 2009 (paramètres du variogramme : Sphérique ; pépite=2.8 ; palier=0.8 ; portée=2500)

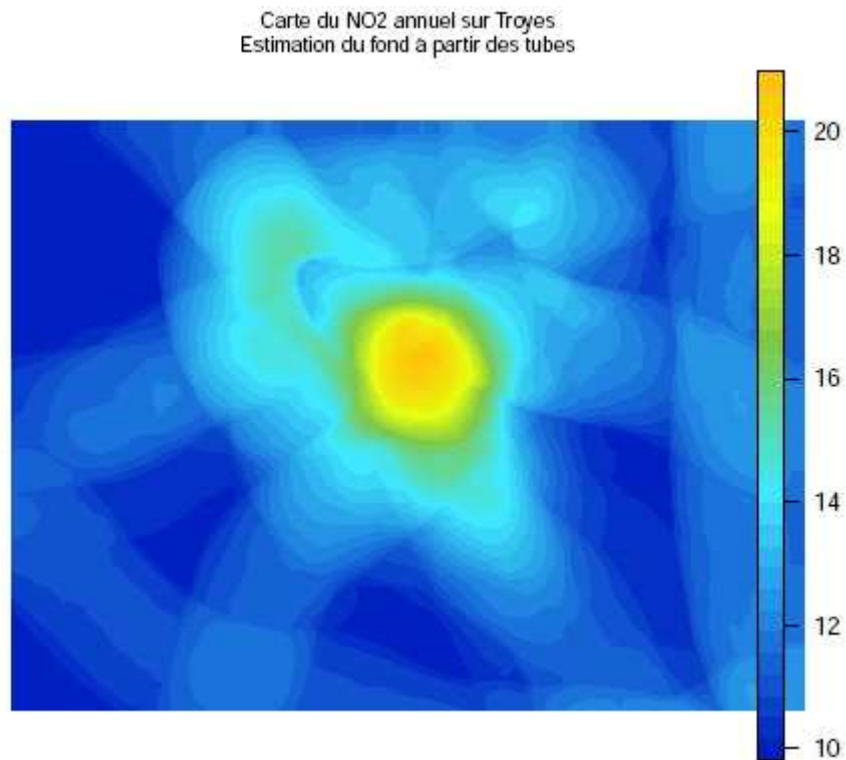


Figure 27 : Estimation de la pollution de fond du NO₂ sur Troyes en 2009 (paramètres du variogramme : Sphérique ; pépite=2.8 ; palier=0.8 ; portée=2500)

A ce stade, on constate que les différences entre les estimations de fond obtenues pour différents ajustements du variogramme sont assez faibles. En revanche, si l'on compare les cartographies des écarts-type de krigeage, on peut noter des différences plus importantes, non pas tant dans la gamme de valeurs que dans la répartition spatiale de ces valeurs.

Le modèle premier modèle a tendance à diminuer l'écart-type de krigeage à très faible distance des tubes tandis que le second tend à uniformiser la distribution des écarts-type dans le centre-ville.

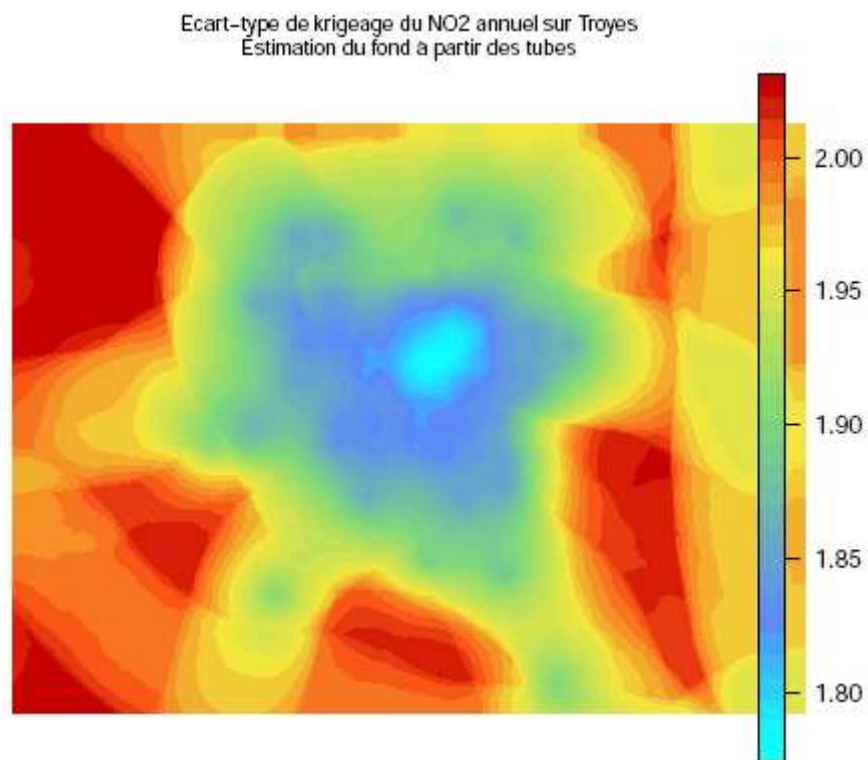


Figure 28 : Ecart-type de krigeage pour l'estimation de fond du NO₂ sur Troyes en 2009 (paramètres du variogramme : Sphérique ; pépite=2.8 ; palier=0.8 ; portée=2500)

Les écarts-type de krigeage, qui interviennent dans l'évaluation de la représentativité spatiale et des dépassements de seuils, sont sensibles à l'ajustement du variogramme. C'est plus particulièrement à l'effet de pépite, relativement plus élevé dans le second modèle, et à la portée que l'on peut imputer ces résultats.

A.4.1.1.c Influence de la portée sur le krigeage

On réalise des tests basés sur le modèle sphérique, à palier et pépite fixés, afin d'observer l'influence de la portée sur le krigeage. Aux résultats d'estimation obtenus pour un ajustement correct, au sens des moindres carrés, on compare les résultats obtenus en augmentant la valeur de la portée de cet ajustement.

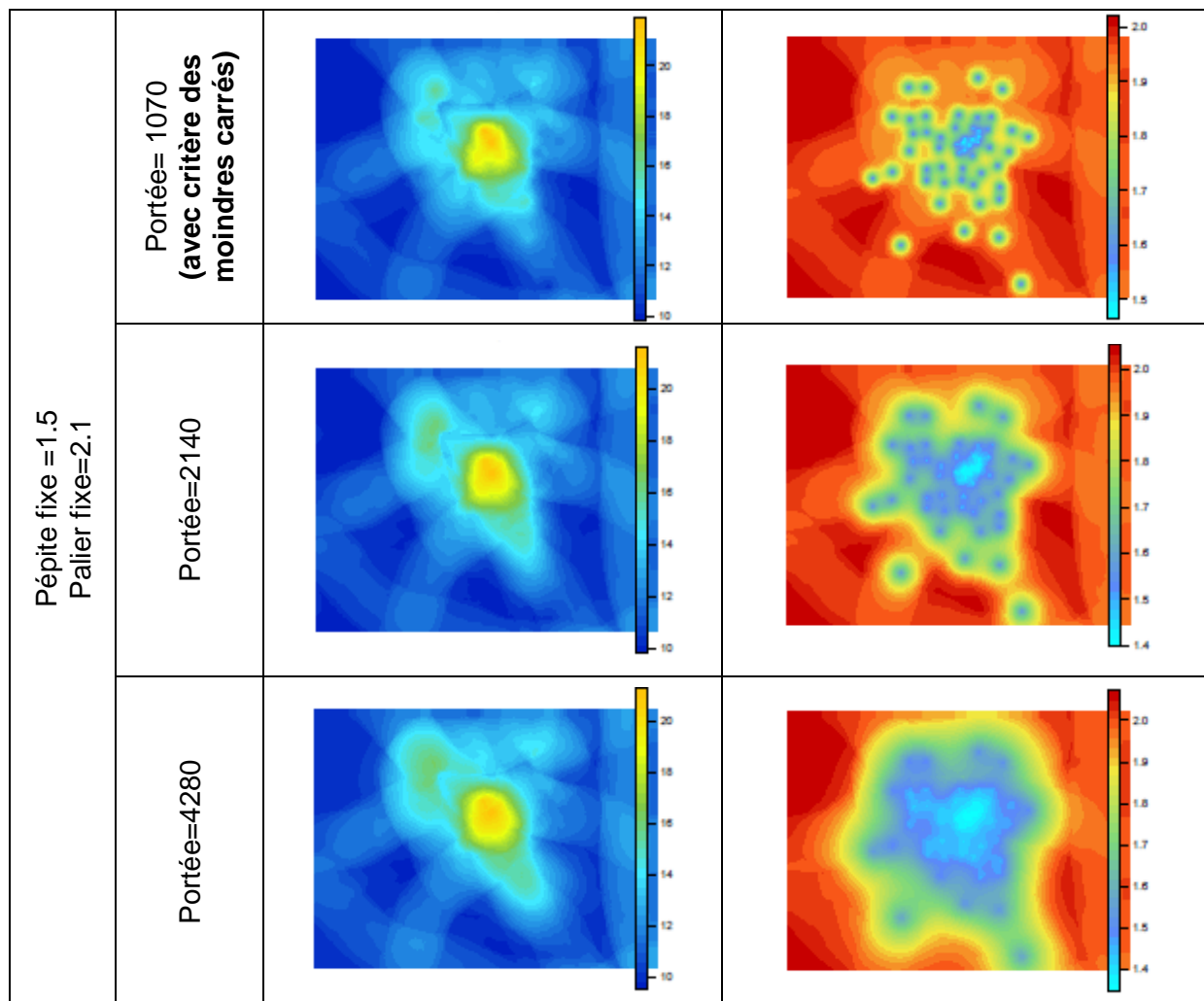


Figure 29 : Estimation de la pollution de fond sur Troyes et Ecart-type de krigeage associé

Augmenter la portée uniformise les concentrations et réduit les écarts-types de krigeage là où la densité de mesures disponibles est forte, ce qui est le cas dans le centre-ville de Troyes.

Toutefois, le variogramme empirique calculé sur les mesures des sites de fond est construit avec suffisamment de données, et possède une structure suffisamment identifiable pour que l'ajustement de cette structure par un modèle qui minimise l'erreur quadratique soit jugé fiable.

Ainsi, sur Troyes, on conservera le modèle ajusté au sens des moindres carrés, dont les paramètres retranscrivent au mieux la variabilité en concentration de la zone.

A.4.1.2 Modélisation statistique de Δ_{TF}

Dans ce qui suit, nous utiliserons comme modèle de variogramme, en accord avec ce qui a été décrit précédemment, un ajustement sphérique avec les paramètres suivants : pépite=1.5, palier=2.1, portée=1070 :

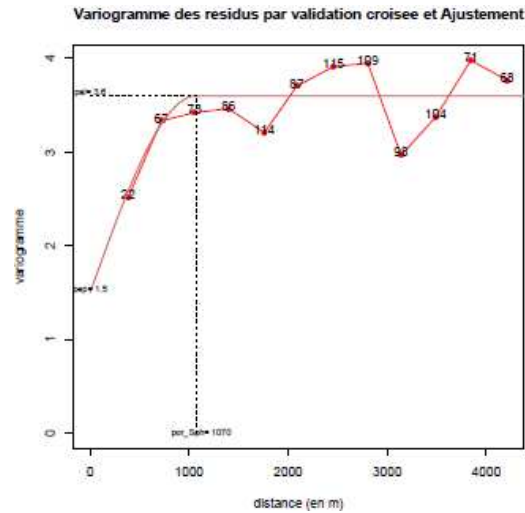


Figure 30 : Variogramme sphérique ajusté sur le variogramme empirique à partir des données de tubes sur Troyes en 2009 (paramètres du variogramme : Sphérique ; pépite=1.5 ; palier=2.1 ; portée=1070)

Comme pour Montpellier, les concentrations de fond sont estimées sur les sites de proximité afin de pouvoir y calculer le surplus de concentration.

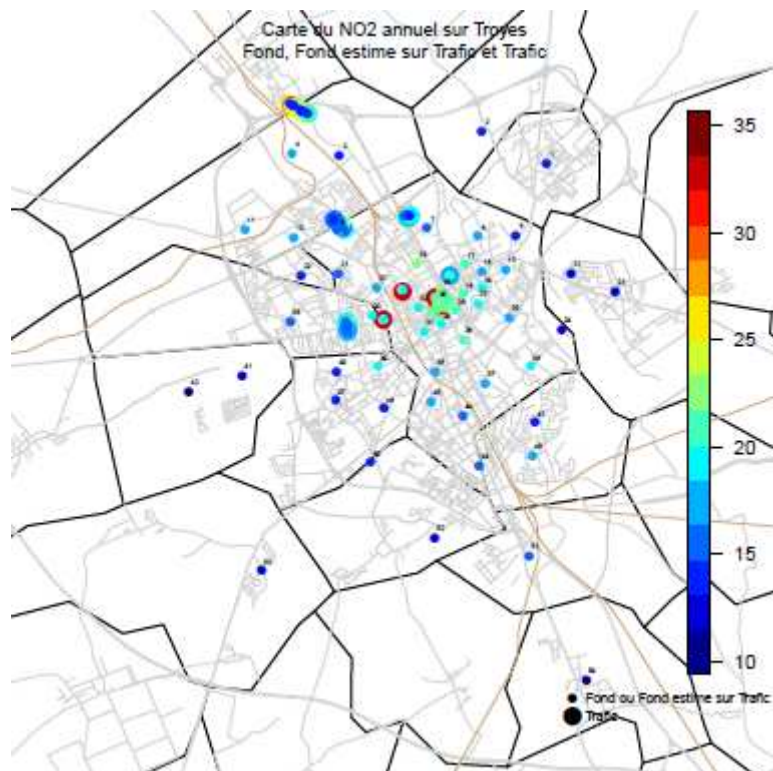


Figure 31 : Répartition des tubes sur Troyes, fond, fond estimé sur sites de Trafic et Trafic

Les variables auxiliaires choisies pour expliquer le surplus de concentration lié au trafic sont le logarithme translaté de la distance à la route et les émissions routières de NO_x dans un rayon de 150m.

$$\Delta_{TF}(x) = 14.66 - 3.02 * \log(\text{dist}(x, \text{axe}) + 1) + 1.90 * \text{Emission_Routières_150m}$$

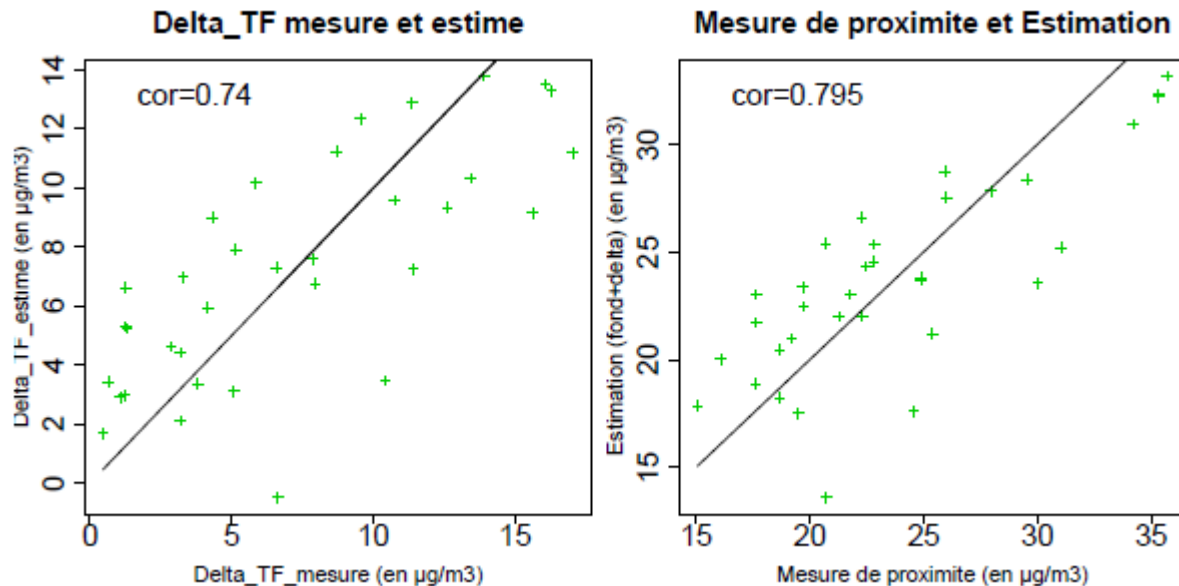
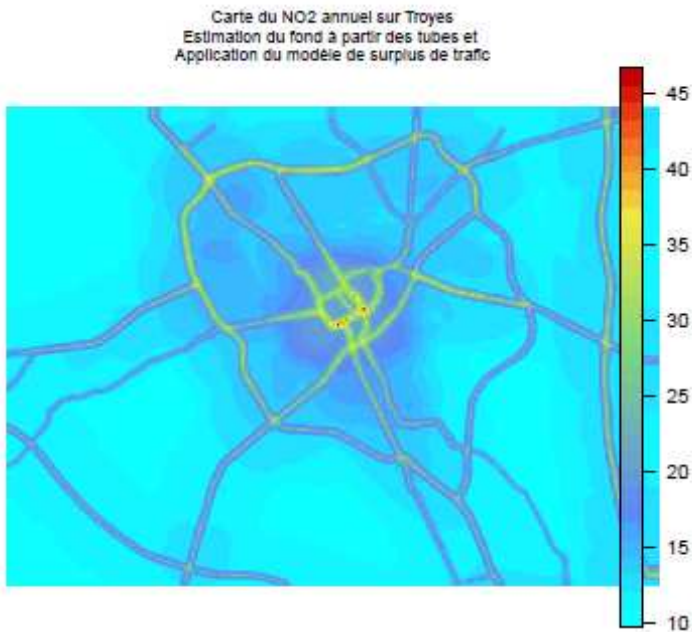


Figure 32 : Graphes et corrélations Mesure-Modèle pour le Δ_{TF} et les concentrations de proximité

Si la corrélation Modèle-Mesure n'est pas parfaite, le modèle semble tout de même apte à fournir une estimation crédible du surplus routier et être utilisable dans une optique de cartographie. On notera que, contrairement à ce qui a déjà été vu sur Montpellier, les valeurs extrêmes sont correctement prises en compte par le modèle : elles ne sont en effet que légèrement sous-estimées. Ainsi, bien que les corrélations soient légèrement plus faibles que sur Montpellier, le modèle peut être jugé plus fiable, ce qui se confirme au regard de l'écart-type des résidus en validation croisée du Δ_{TF} , plus faible sur Troyes (3.8 contre 5).

A.4.1.3 Cartographie incluant pollution de fond et de proximité



Le résultat obtenu est conforme à nos attentes : le centre ville de Troyes ressort de la cartographie comme étant la zone présentant les concentrations de NO₂ les plus élevées. De plus, le modèle de Δ_{TF} permet de prendre en compte les axes routiers dans des zones périurbaines ou rurales et rehaussent fortement les concentrations des points proches de ces axes, sans toutefois leur attribuer des valeurs qui seraient incohérentes avec les concentrations observées sur cette région.

Enfin, certains sites de proximité présentent des concentrations supérieures aux maximums observés sur la campagne de tubes. Il n'est pas possible de savoir si ces prédictions sont vraies, c'est pourquoi il serait intéressant, lors de campagnes ultérieures, d'y effectuer des mesures de NO₂.

Figure 33 : Cartographie incluant fond et proximité.
Concentrations moyennes annuelles de NO₂.

A.4.2 Représentativité en Concentration

A.4.2.1 Représentativité et pollution de fond

Nous conservons les mêmes critères de représentativité en concentration que ceux présentés dans le rapport, soit un écart maximal toléré de 10µg/m³ et un risque statistique de 10%. Les zones de représentativité des deux stations se recoupent (Figure 34). Le critère de différence minimale de concentration est appliqué pour supprimer ces recouvrements (Figure 35).

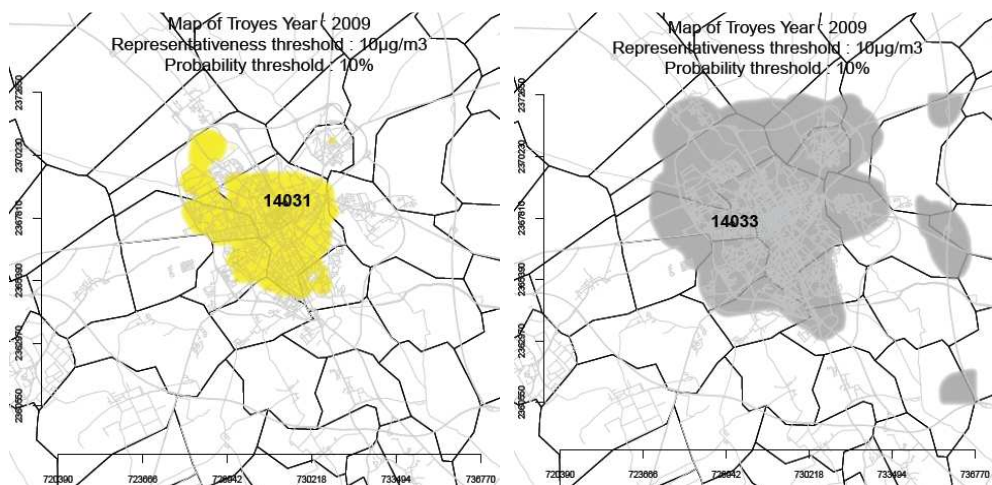


Figure 34 : Zones de représentativité en concentration des stations 14031 et 14033 (stations de fond) sur l'agglomération de Troyes, avant suppression des recouvrements. Superficie de la zone de représentativité de la station 14031 : ~22 km². Superficie de la zone de représentativité de la station 14033 : ~72 km².

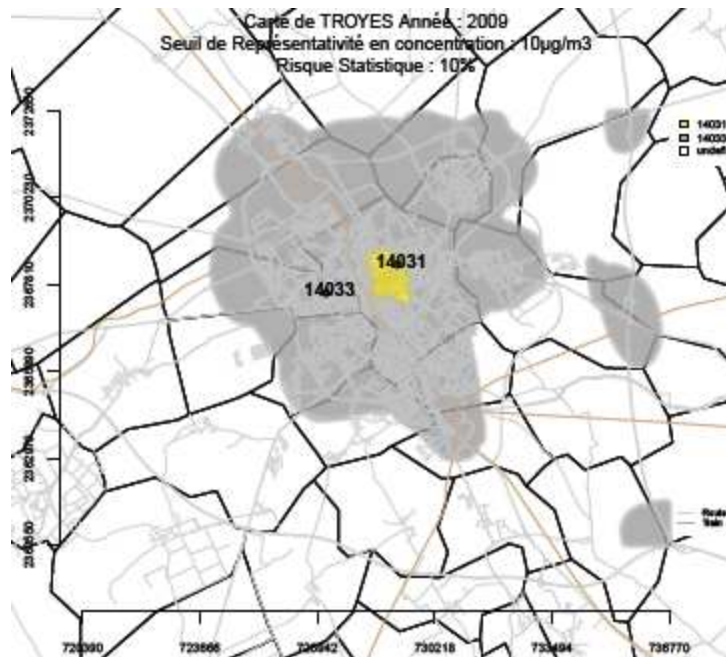


Figure 35 : Zones de représentativité en concentration des stations 14031 et 14033 (stations de fond) sur l'agglomération de Troyes, après suppression des recouvrements.

On peut une nouvelle fois comparer cette carte à celle obtenue avec les données provenant des cartes analysées nationales. On constate que la répartition spatiale des zones est similaire si ce n'est que la faible densité de points d'échantillonnage en périphérie du domaine ainsi que l'introduction du risque statistique empêchent le rattachement de certaines zones périphériques à la station 14033. Bien qu'en 2009, les deux stations aient mesuré des valeurs proches, la précision accrue des estimations permet d'attribuer à chacune une zone de représentativité qui lui est propre (et dans laquelle elle se trouve incluse).

Notons cependant que les concentrations des deux stations étant assez proches (généralement < à 5µg/m3 en moyenne annuelle, sauf en 2005), la zone de représentativité de la station 14031 est toujours de surface très restreinte, et seuls les points qui se situent à une très faible distance de la station (environ 1km) lui sont rattachés. On pourrait alors se poser la question de savoir si l'information fournie par les mesures de ces deux stations n'est pas redondante.

A.4.2.2 Représentativité et pollution intégrant fond et proximité

Le fait de prendre en compte la pollution de proximité dans les calculs de représentativité apporte des informations complémentaires par rapport à ce qui a été dit sur Montpellier. En effet, les écarts-types de krigeage sont bien plus faibles sur Troyes et les corrections appliquées à l'écart maximal toléré δ sont donc moins contraignantes pour le rattachement des points aux stations. Ici, certains points situés à proximité des axes se retrouvent dans la zone de représentativité de la station 14031, voire de la station 14033, qui sont des stations de fond.

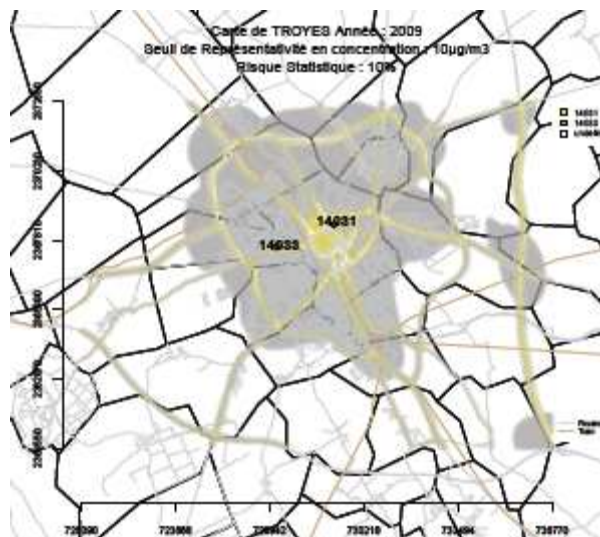


Figure 36 : Zones de représentativité en concentration des stations 14031 et 14033 (stations de fond) sur l'agglomération de Troyes après ajout de la proximité

Ce résultat nous ramène à la question posée dans le corps du rapport, à savoir : dans quelle mesure et selon quels critères est-il admissible de rattacher des points de proximité à une mesure fixe?

A.4.3 Dépassements de seuil

A.4.3.1 Dépassement et pollution de fond

Sur la région de Troyes, et pour le modèle de variogramme finalement choisi, aucun point n'est classé en dépassement de seuil, que ce soit pour une valeur limite de NO_2 de $40\mu\text{g}/\text{m}^3$, ou pour les seuils d'évaluation supérieur et inférieur (respectivement $32\mu\text{g}/\text{m}^3$ et $24\mu\text{g}/\text{m}^3$).

Cela s'explique par le fait que les concentrations estimées les plus élevées ($\sim 21\mu\text{g}/\text{m}^3$) se situent en centre-ville, là où l'écart type de krigeage est le plus faible. En conséquence l'équation 17 (annexe A.3.2.2) est toujours vérifiée.

Afin de poursuivre l'illustration de la méthodologie, on fixe un seuil arbitraire de $18\mu\text{g}/\text{m}^3$, en introduisant, tout comme pour Montpellier, les risques de première et de seconde espèce

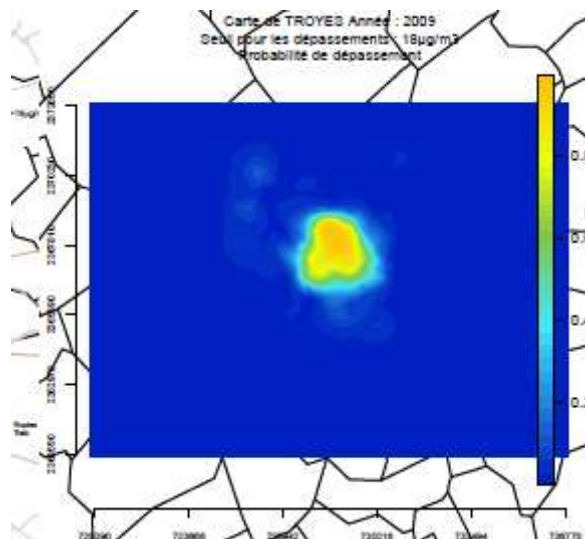


Figure 37 : Carte des probabilités de dépasser le seuil de 18 µg/m³ pour l'année 2009

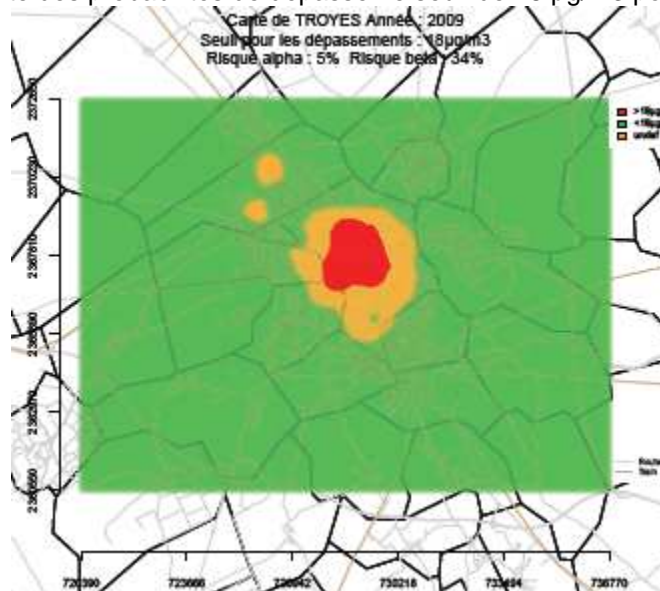


Figure 38 : Carte de dépassement du seuil 18 µg/m³ avec les risques de 1^{ère} (5%) et 2^{nde} espèce (34%)

Sur Troyes, les concentrations les plus élevées sont regroupées dans une zone très restreinte du centre-ville. Ici, on pourrait raisonnablement augmenter le risque de seconde espèce sans que la surface en dépassement n'augmente trop.

A.4.3.2 Dépassement et pollution intégrant fond et proximité

La valeur limite annuelle de 40 µg/m³ est à nouveau prise comme seuil. En intégrant les données de proximité, et en fixant les risques de première et seconde espèces à leurs niveaux conventionnels (resp. 5% et 34%), on peut conclure à l'absence de dépassement sur la majeure partie du domaine. Ne font exception que quelques carrefours du centre-ville et une zone restreinte au nord-ouest, où les concentrations peuvent être considérées comme supérieures au seuil.

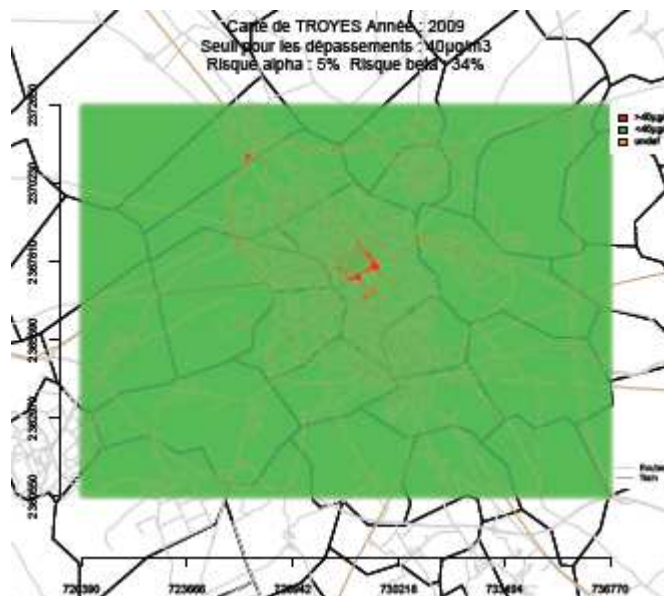


Figure 39 : Figure 40 : Carte de dépassement pour le seuil 40 µg/m³ avec les risques de 1^{ère} (5%) et 2^{nde} espèce (34%)

A.4.4 Recalage Annuel de la pollution de fond

On applique sur Troyes le même procédé que celui mis en place lors de l'étude sur la région de Montpellier, à savoir un recalage annuel des données de fond mesurées sur l'année de référence par application d'un facteur correctif déduit des cartes analysées.

La zone de représentativité de la station 14031 est stable d'une année sur l'autre, alors que celle de la station 14033, dont les concentrations sont plus faibles, est nettement plus variable. En 2005 tous les points en périphérie du domaine peuvent lui être rattachés (on peut observer le même phénomène en annexe 5 pour cette même année, avec les données issues des cartes analysées nationales).

	Concentrations des stations (en µg/m ³)		Pollution de fond	Représentativité en Concentration
	14031	14033		
2005	23	17		

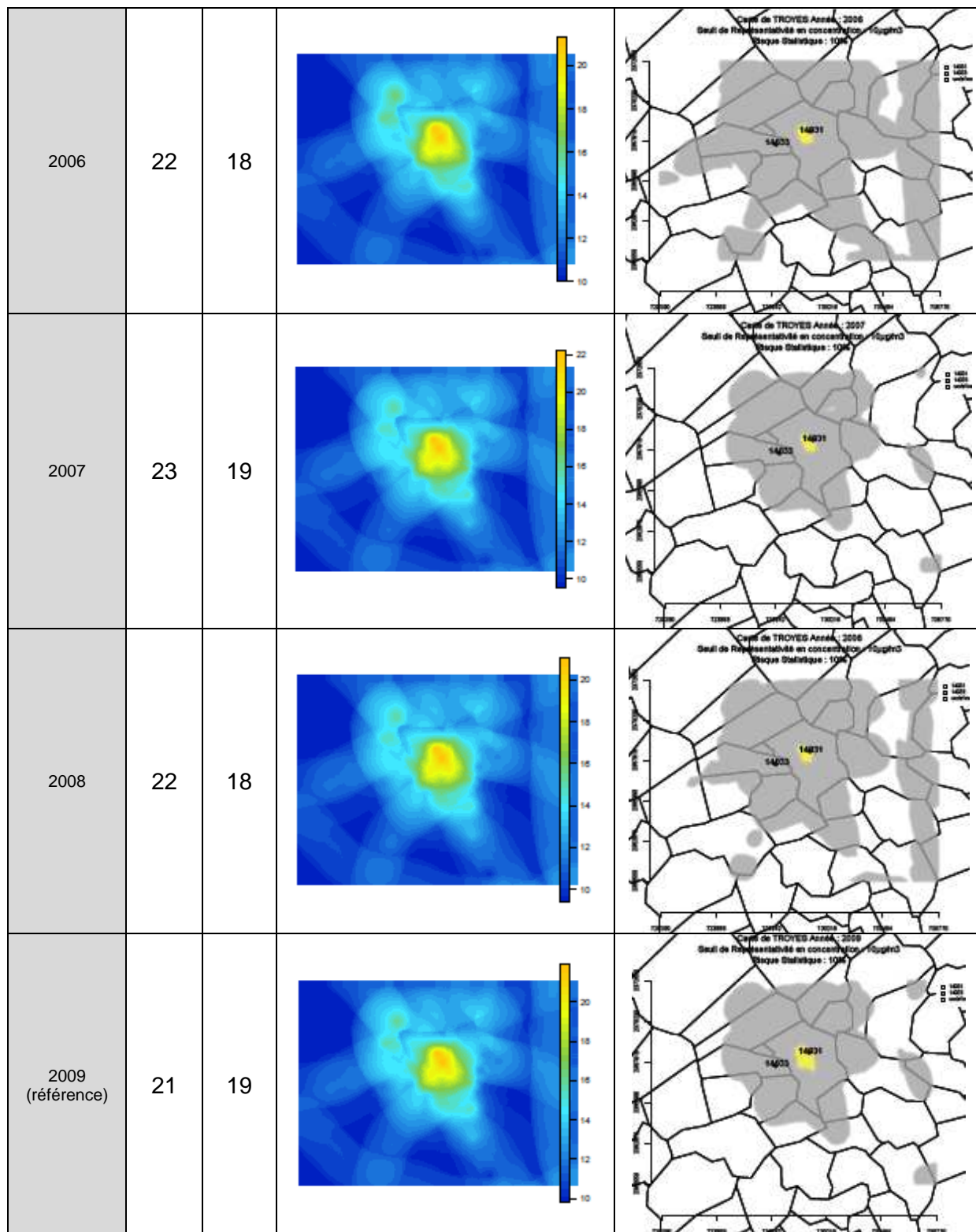


Figure 41 : estimation de la pollution de fond sur Troyes de 2005 à 2009 zones de représentativité en concentration des stations 14031 et 14033

ANNEXE 5 : SENSIBILITE AUX CRITERES DE SUPPRESSION DES RECOUVREMENTS

Dans le rapport, nous avons étudié l'impact des critères de rattachement aux stations en cas de recouvrement, en examinant en particulier, sur la région de Montpellier, la stabilité temporelle des zones de représentativité (§3.4).

Nous présentons les résultats obtenus dans les mêmes conditions sur d'autres régions françaises.

A.5.1 RATTACHEMENT AUX STATIONS ET STABILITE TEMPORELLE : TROYES

Sur la région de Troyes, on constatera que pour certaines années, il est impossible, même en considérant les cartes analysées comme « vraies », et en ne corrigeant donc pas l'écart maximal toléré par rapport aux stations, de rattacher certains points à une station.

		2005	2006	2007	2008	2009
Concentrations des stations (en µg/m3)	14031	23	22	23	22	21
	14033	17	18	19	18	19
Avec recouvrement						
distance						
écart min						
Similarité de milieu par Vecteur						
Similarité de milieu par ACP						

Figure 42 : Zones de représentativité en concentration des stations 14031 et 14033 pour les années 2005 à 2009 sur Troyes

A.5.2 RATTACHEMENT AUX STATIONS ET STABILITE TEMPORELLE : ROUEN

Sur la région rouennaise, les stations 25036, 25043 et 25048 sont concernées par le calcul des zones de représentativité en concentration. Toutefois pour l'année 2005, la concentration de la station 25048 est inconnue : nous ne présenterons donc pas les résultats pour cette année.

De plus, il y a cette fois-ci plus de deux stations et il peut arriver que les zones de recouvrement soient très différentes d'une année sur l'autre.

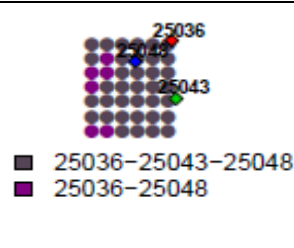
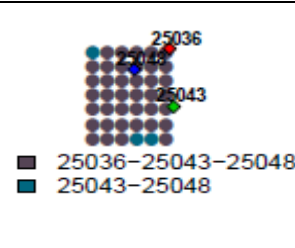
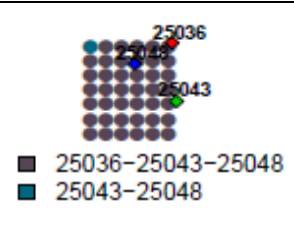
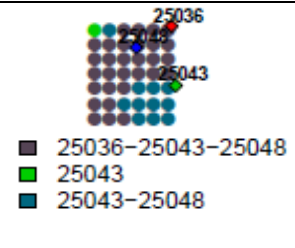
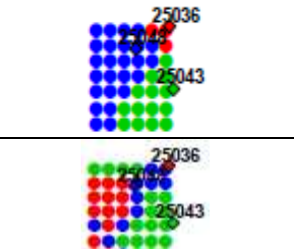
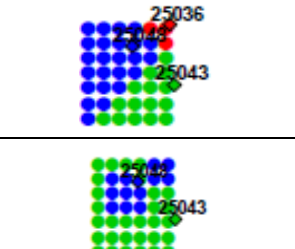
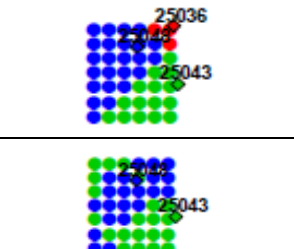
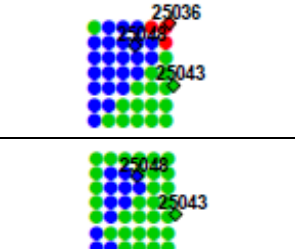
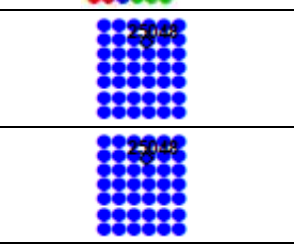
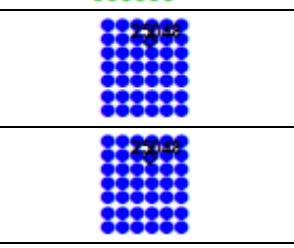
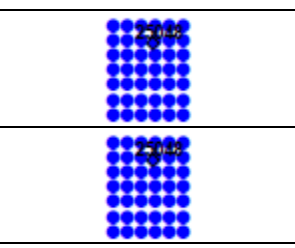
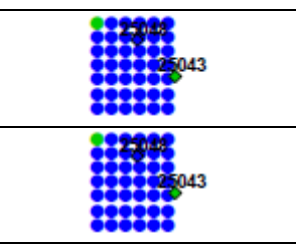
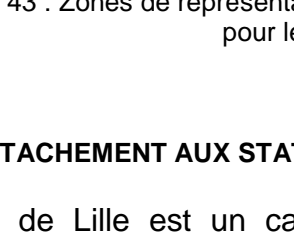
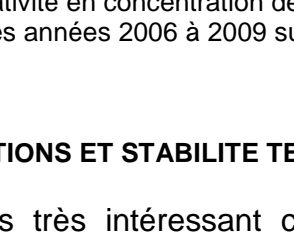
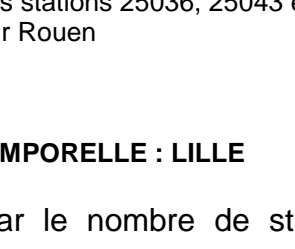
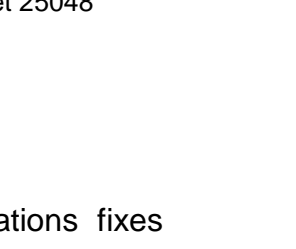
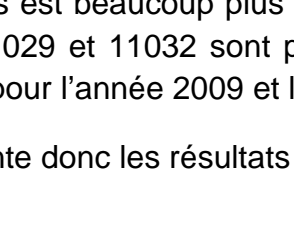
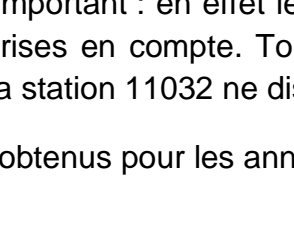
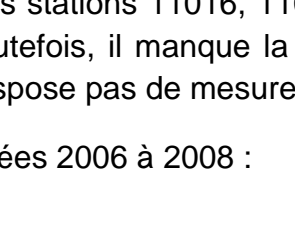
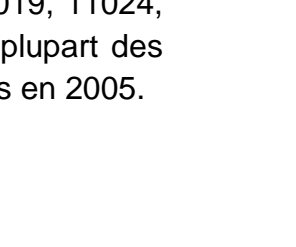
		2006	2007	2008	2009
Concentrations des stations (en µg/m3)	25036	32	32	32	33
	25043	25	24	23	24
	25048	31	31	30	30
Avec recouvrement					
distance					
écart min					
Similarité de milieu par Vecteur					
Similarité de milieu par ACP					

Figure 43 : Zones de représentativité en concentration des stations 25036, 25043 et 25048 pour les années 2006 à 2009 sur Rouen

A.5.3 RATTACHEMENT AUX STATIONS ET STABILITE TEMPORELLE : LILLE

La région de Lille est un cas très intéressant car le nombre de stations fixes impliquées est beaucoup plus important : en effet les stations 11016, 11019, 11024, 11025, 11029 et 11032 sont prises en compte. Toutefois, il manque la plupart des données pour l'année 2009 et la station 11032 ne dispose pas de mesures en 2005.

On présente donc les résultats obtenus pour les années 2006 à 2008 :

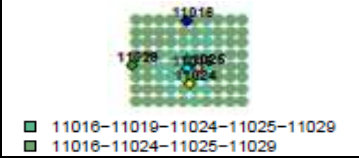
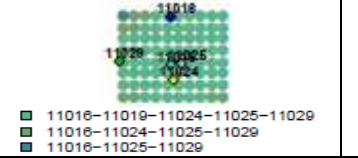
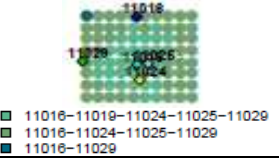
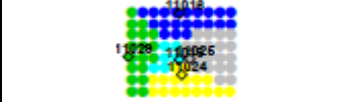
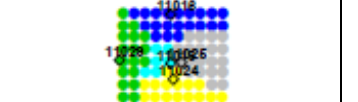

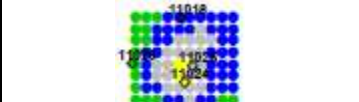
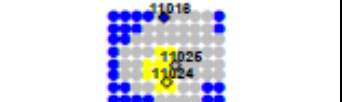
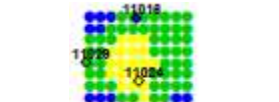
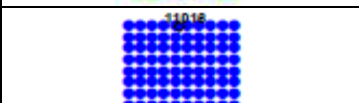
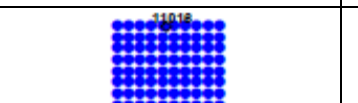
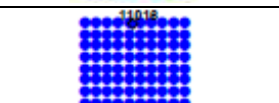
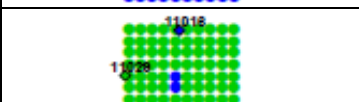
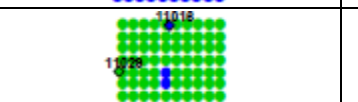
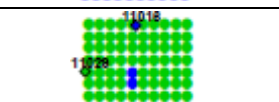
		2006	2007	2008
Concentrations des stations (en µg/m ³)	11016	31	29	26
	11019	42	39	40
	11024	35	35	33
	11025	33	32	33
	11029	28	29	28
Avec recouvrement				
distance				
écart min				
Similarité de milieu par Vecteur				
Similarité de milieu par ACP				

Figure 44 : Zones de représentativité en concentration des stations 11016, 11019, 11024, 11025 et 11029 pour les années 2006 à 2008 sur Lille

ANNEXE 6 : COMPARAISON DES RESULTATS SOUS R ET ISATIS

La section 3.1.2 présente les résultats de l'estimation de la pollution de fond par krigeage en tenant compte des émissions de NO_x dans un rayon de 2km comme dérive.

Pour estimer les concentrations par krigeage, deux outils de calcul ont été mis en œuvre simultanément : un code informatique développé avec le langage R (<http://cran.r-project.org/>) et le logiciel de géostatistique Isatis (Géovariances).

Ces deux outils étant d'usage fréquent dans le domaine de la qualité de l'air, il est intéressant d'en comparer les résultats.

La méthodologie complète a été appliquée pour l'estimation du fond : calcul du variogramme empirique, ajustement du variogramme et krigeage. Les cartographies retournées par les 2 programmes sont très similaires.

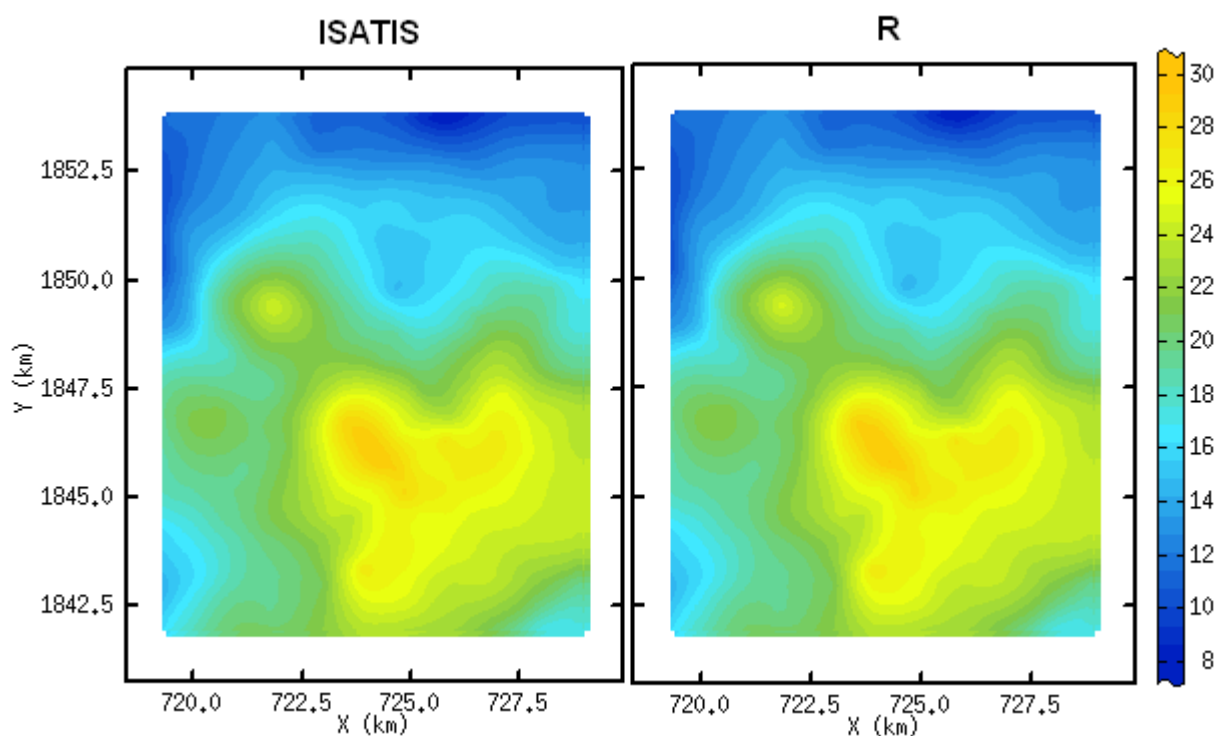


Figure 45 : Prédications de la pollution de fond pour le NO₂
à gauche les prédictions d'Isatis, à droite, les prédictions du logiciel R

Il est d'ailleurs possible de visualiser les écarts d'estimation entre le logiciel Isatis et l'algorithme développé sous R, $\delta_{ISATIS-R} \in [0; 0.25] [\mu\text{g}/\text{m}^3]$:

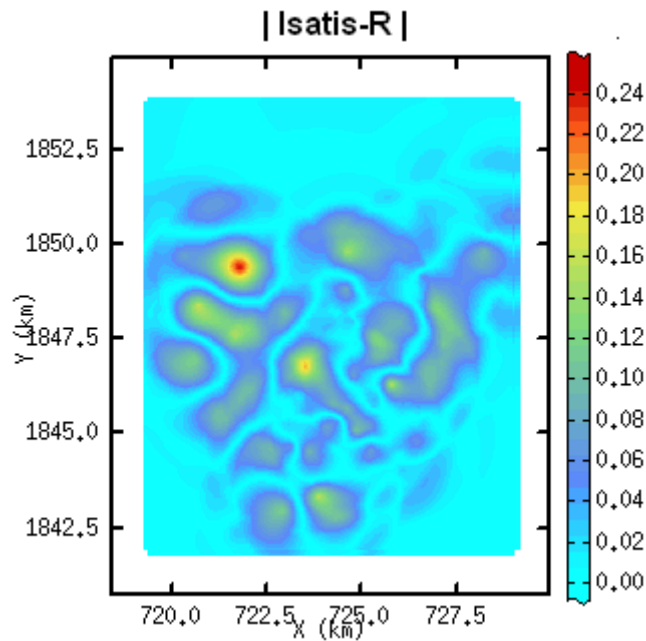


Figure 46 : Ecarts d'estimation [$\mu\text{g}/\text{m}^3$] entre Isatis et R

Les différences observées, très faibles, s'expliquent en grande partie par le variogramme ajusté (ajusté automatiquement par moindres carrés dans R et à la main dans Isatis). En effet, si l'on force R et Isatis à utiliser le même modèle de variogramme dans leur algorithme (ici le variogramme ajusté par moindres carrés), alors les différences d'estimation deviennent négligeables, $\delta_{ISATIS-R} \in [0; 0.023]$:

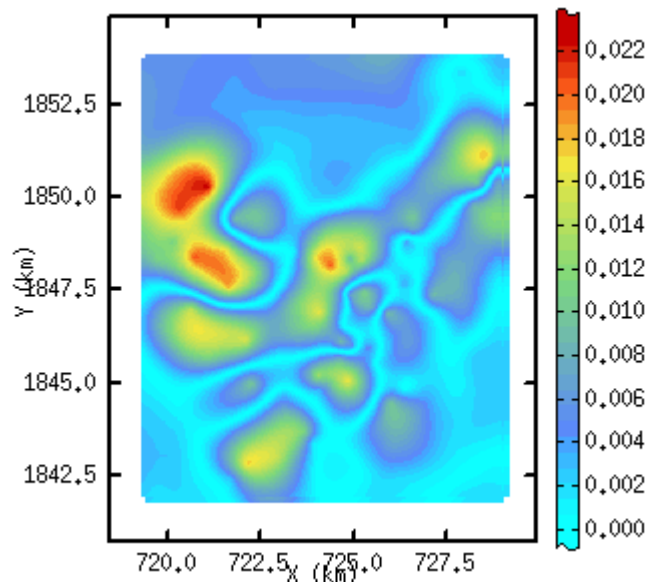


Figure 47 : Ecarts d'estimation entre Isatis et R avec le même ajustement de variogramme

ANNEXE 7 : NOTICE D'UTILISATION DES PROGRAMMES R

A.7.1 ARCHITECTURE DU DOSSIER FOURNI

Dans l'optique d'uniformiser les formats d'entrée et de sortie, tous les fichiers seront au format .csv.

Afin de pouvoir exécuter le programme fourni avec cette étude, il suffit de créer un dossier qui comportera deux répertoires :

- Un répertoire d'entrée (**appelé Input**) où figureront :
 - le fichier contenant les arguments pour la bonne exécution du programme R
 - le fichier qui indique les mesures aux stations pour une année fixée
 - le fichier contenant la grille d'estimation sur le domaine à étudier(On se reportera à la section A.9.2 pour le descriptif de ces fichiers)

- Un répertoire de sortie (**appelé Output**) où figureront :
 - le fichier définissant les zones de représentativité en concentration
 - le fichier définissant les zones de dépassement de seuil(On se reportera à la section A.9.3 pour le descriptif de ces fichiers)



1. Figure 48 : Architecture du dossier fourni contenant le code
2. ainsi que des exemples de fichiers d'entrée et de sortie

A.7.2 FORMAT DES FICHIERS D'ENTREE

1- Fichier contenant les paramètres d'entrée pour le programme R :

Le fichier contenant les arguments en entrée du programme R comporte 11 colonnes décrivant dans l'ordre :

- le répertoire de travail (dossier contenant les répertoires d'entrée et de sortie)
- le nom du fichier des mesures aux stations (dans le répertoire d'entrée)
- le nom du fichier contenant la grille d'estimation (dans le répertoire d'entrée)
- le nom de la zone étudiée (par exemple MONTPELLIER, TROYES, LILLE)
- l'année d'étude
- l'écart maximal toléré par rapport aux stations (en $\mu\text{g}/\text{m}^3$)
- le risque statistique associé à l'écart maximal toléré

- la valeur limite de concentration pour l'évaluation des dépassements
- le risque statistique de première espèce associé à cette valeur limite
- le risque statistique de seconde espèce associé à cette valeur limite
- l'écart-type des résidus (en validation croisée idéalement) pour le modèle de delta trafic (pris en compte dans les calculs sur les sites de proximité avec une hypothèse gaussienne)

work_directory	nom_fichier_station	nom_fichier_grille	region	annee	seuil_RC	risque_RC	seuil_DS	risque1_DS	risque2_DS	stdev_prox
C:/Documents and Settings/...	Input/Donnees_Station	Input/Krigeage_Sur	MONTPELLIE	2006	10	0.1	32	0.05	0.34	5.004

3. Figure 49 : exemple de fichiers de définition des paramètres d'entrée

2- Fichier des mesures aux stations pour l'année concernée :

isitm	X	Y	NO2_AN
8003	723633.197206936	1846736.56151467	42
8004	724333.602128247	1846061.47605285	58
8005	721621.343599693	1847602.33649002	17
8016	725348.791242131	1844562.58887618	25

4. Figure 50 : Fichier des mesures aux stations sur Montpellier en 2006

Il est possible d'ajouter d'autres colonnes à ce fichier, mais ces 4 colonnes, nommées comme sur la figure 49, doivent y figurer.

Pour évaluer les zones de représentativité station par station, sans suppression des recouvrements, autant de fichiers de mesures que de stations présentes dans le domaine seront créés : le calcul sera répété pour chacun de ces fichiers.

3-Grille d'estimation sur le domaine pour l'année concernée :

Le fichier exemple contenant les estimations par krigeage des concentrations de NO₂ pour l'année concernée étant de grande taille, nous n'en présentons ici que les premières lignes :

X	Y	Pred	StDev	Type
719400	1842240	15.46606845	4.948015322	Prox
719400	1842260	15.42059713	4.950217157	Prox
719400	1842280	15.37535005	4.952424628	Prox
719400	1842300	15.33026769	4.954640406	Fond
719400	1842320	15.28555400	4.956854155	Fond
719400	1842340	15.24108091	4.959071867	Fond
719400	1842360	15.19700298	4.961285475	Fond

5. Figure 51 : Premières lignes de la grille d'estimation des concentrations sur Montpellier en 2006

De nouveau, seules 5 colonnes sont requises dans ce fichier : les coordonnées X et Y des mailles, leur estimations par krigeage ainsi que les écarts-types associés (attention à bien respecter le nom des colonnes). La dernière colonne précise la typologie du point : Fond pour fond, Prox pour proximité si on a appliqué le modèle de surplus à ce point.

A.7.3 FORMAT DES FICHIERS DE SORTIE

1- Fichier décrivant les zones de représentativité :

Le nom du fichier de sortie comportera les différents paramètres pris en compte dans le calcul, à savoir le nom de la zone étudiée, ainsi que l'écart maximal toléré par rapport aux stations et le risque statistique associé :

Zone_RC_nom-ville_annee_Seuil_ecart-max-toléré_risque_risque-statpct.csv

EX :Zone_RC_MONTPELLIER_2006_Seuil_10_risque_10pct.csv

X	Y	Zones_RC	Zone_RC	Ecart_Station
719400	1841900	8005	8005	0.648964709692599
719400	1841920	8005	8005	0.740919609838901
719400	1841940	8005	8005	0.828303475755
719400	1841960	8005	8005	0.909294393964199
719400	1841980	8005	8005	0.980046561047502
719400	1842000	8005	8005	1.0317131557141
719400	1842020	8005	8005	1.0639696030312
719400	1842040	8005	8005	1.0999123302592
719400	1842060	8005	8005	1.1383845582973
719400	1842080	8005	8005	1.1789145810994
719400	1842100	8005	8005	1.2211395508589
719400	1842120	8005	8005	1.2642385452529
719400	1842140	8005	8005	1.3078931632831
719400	1842160	8005	8005	1.3522201974275
719400	1842180	8005	8005	1.3970862755462
719400	1842200	8005	8005	1.4425657724122
719400	1842220	8005	8005	1.4883029881993
719400	1842240	8005	8005	1.5339315426698
719400	1842260	8005	8005	1.5794028668808
719400	1842280	8005	8005	1.6246499433669
719400	1842300	8005	8005	1.6697323098461
719400	1842320	8005	8005	1.7144459984081
719400	1842340	8005	8005	1.7589190803874
719400	1842360	8005	8005	1.8029970169099

6. Figure 52 : premières lignes du fichier descriptif des zones de représentativité sur Montpellier en 2006

Une ligne du fichier comportera 5 colonnes décrivant respectivement les coordonnées X et Y d'un point de grille, les stations auxquelles est rattaché le point avant suppression des intersections, la station à laquelle le point est rattaché après suppression des intersections, et l'écart de concentration entre ce point et la station. Si le point n'appartient à aucune zone de représentativité, alors les 3 dernières colonnes seront remplies par des NA.

2- Fichier décrivant les zones de dépassement :

7. Le fichier renseignant les dépassements de seuil a un format assez proche de celui renseignant la représentativité en concentration. Son nom sera de la forme :

Zone_DS_nom-ville_annee_Seuil_seuil-limite_risque1_risque1pct_risque2_risque2pct.csv

EX :Zone_DS_MONTPELLIER_2006_Seuil_40_risque1_5pct_risque2_34pct.csv

X	Y	OVER
719400	1841900	FALSE
719400	1841920	FALSE
719400	1841940	FALSE
719400	1841960	FALSE
719400	1841980	FALSE
719400	1842000	FALSE
719400	1842020	FALSE
719400	1842040	FALSE
719400	1842060	FALSE
719400	1842080	FALSE

8. Figure 53 : premières lignes du fichier descriptif des zones de dépassement sur Montpellier en 2006

Une ligne du fichier comportera 3 colonnes décrivant respectivement les coordonnées X et Y, et le statut du point (TRUE s'il est en dépassement au risque de seconde espèce, FALSE s'il ne l'est pas au risque de première espèce, ??? si son statut n'est pas déterminé).

A.7.4 EXECUTION DU PROGRAMME

Si toutes les étapes décrites précédemment ont été respectées (préparation des fichiers d'entrée), alors la démarche à suivre pour exécuter le code est la suivante :

- 1- Préciser dans le programme R la localisation du fichier de paramètres à la ligne 50. A la réception du code, cette ligne est renseignée par :

```
param=read.csv('tapez l'adresse de votre fichier d'entrée',sep=';')
```

Remplacer là par une ligne du type :

```
param=read.csv('C:/Prog_RC_DS_AASQA/Input/Parameters.csv',sep=';')
```

(Attention : Quel que soit l'OS sous lequel vous travaillez, votre adresse doit être spécifiée avec des /)

- 2- Ouvrir une fenêtre de commande R et charger le programme par une ligne du type :

```
source('C:/Prog_RC_DS_AASQA/RC_DS_AASQA.R')
```

Votre code doit s'exécuter et les résultats en sortie seront dans le fichier Output du dossier que vous avez créé initialement.

A.7.5 REMARQUES

Certaines des cartes présentées dans ce rapport ont été réalisées sous R à partir du même type de fichier de sortie qu'il est possible d'obtenir après l'exécution de ce programme.

Les programmes de construction de ces cartes ne sont pas fournis avec le code car ils requièrent l'installation préalable de packages et le téléchargement de fonds de cartes spécifiques. On laisse donc aux futurs utilisateurs le choix dans la méthode de visualisation des résultats.