

Note technique

ASSISTANCE EN CARTOGRAPHIE ET TRAITEMENT DE DONNEES

Synthèse des actions 2013

Laure Malherbe (INERIS), Maxime Beauchamp (INERIS),
Laurent Létinois (INERIS)

SYNTHESE

Le LCSQA assure chaque année une assistance technique et méthodologique relative au traitement statistique des données et à l'élaboration de cartographies. La présente note synthétise les actions réalisées en 2012 dans ces différents domaines pour le compte des AASQA.

Ces actions sont de plusieurs natures : travail de méthode, aide ponctuelle, formation. Toutes ont pour objet le développement et la mise en œuvre d'une approche cohérente, qui va de la collecte de données à l'analyse statistique ou géostatistique de ces dernières.

Parmi les tâches réalisées en 2013, on relèvera notamment:

- l'organisation d'une session supplémentaire de formation en statistique, en collaboration avec l'université de Nantes ;
- une aide à l'exploitation géostatistique de campagnes d'échantillonnage et à l'élaboration de cartographies avec le logiciel R ;
- un appui sur l'évaluation de la représentativité des stations à partir de cartes de krigeage.

1. INTRODUCTION

La planification de campagnes de mesure, l'exploitation des données de surveillance, qu'il s'agisse de mesures fixes, de mesures temporaires ou de résultats de modélisation, et l'élaboration de cartographies supposent la mise en œuvre de techniques statistiques ou géostatistiques. Chaque année, le LCSQA propose son appui aux AASQA dans l'application des méthodes existantes et l'interprétation des résultats obtenus. Cette assistance peut prendre plusieurs formes : échanges téléphoniques, réunions de travail, réalisation de travaux techniques, organisation de formations. La présente note synthétise les différentes actions conduites en 2013.

2. TRAITEMENT STATISTIQUE DES DONNEES

2.1 Définition de plans d'échantillonnage et reconstitution de données

En 2013 l'assistance dans ce domaine s'est limitée à des conseils ponctuels par téléphone.

2.2 Formation en statistique

Une session de formation en statistique avec le logiciel R a été organisée les 2 et 3 juillet 2013. Animée par le LCSQA, en collaboration avec Frédéric Lavancier, de l'université de Nantes, elle a accueilli cinq participants d'Airparif, de Lig'Air et du LCSQA. Il s'agissait de la sixième édition de la formation conçue en 2010 par le LCSQA avec le concours de F. Lavancier et de la société Stat-Consult¹.

Cette formation s'adresse aux acteurs de la surveillance qui souhaitent développer leurs compétences en analyse et en traitement de données. Alternant points théoriques et cas pratiques avec le logiciel R, elle aborde les sujets suivants :

- *Analyses graphiques* : univariées, bivariées, multivariées (ACP, classification)
- *Tests statistiques et modélisation* : significativité, tests statistiques, comparaison de moyennes, généralisation : analyse de variance (ANOVA), Régression linéaire
- *Aspects temporels* : séries temporelles et dépendance temporelle, analyse des cycles et de la tendance, précautions à prendre en présence de dépendance temporelle

Du temps est réservé aux échanges et à l'étude des jeux de données apportés par les participants.

¹ Malherbe L., 2010. Assistance à l'exploitation de données de campagnes et à la réalisation de cartographies Organisation d'une formation en statistique. Rapport LCSQA, réf. DRC-10-111609-05224A, <http://www.lcsqa.org/rapport/2010/ineris/assistance-exploitation-donnees-campagnes-realisation-cartographies-12>

3. CARTOGRAPHIE

Le LCSQA a répondu aux sollicitations de trois AASQA sur des questions relatives à la cartographie des concentrations et à l'évaluation de la représentativité des stations.

3.1 Cartographie des concentrations avec R

Le LCSQA a assisté Atmo Champagne-Ardenne dans la réalisation d'une cartographie des concentrations de fond en NO_2 sur la ville de Vitry-le-François. Par cette étude, Atmo Champagne-Ardenne, qui utilise habituellement le logiciel Isatis, a souhaité prendre en main les fonctionnalités géostatistiques et cartographiques de R. Les données d'une campagne d'échantillonnage conduite en 2012, complétées par des données d'émissions de NO_x , ont été exploitées par krigeage. A l'issue de travaux préparatoires réalisés par Atmo Champagne-Ardenne et l'INERIS, une journée de travail a été organisée dans l'AASQA en avril 2013 afin de discuter des méthodes, de leur mise en œuvre avec R, des résultats obtenus (exemple en Figure 1 et des voies d'amélioration).

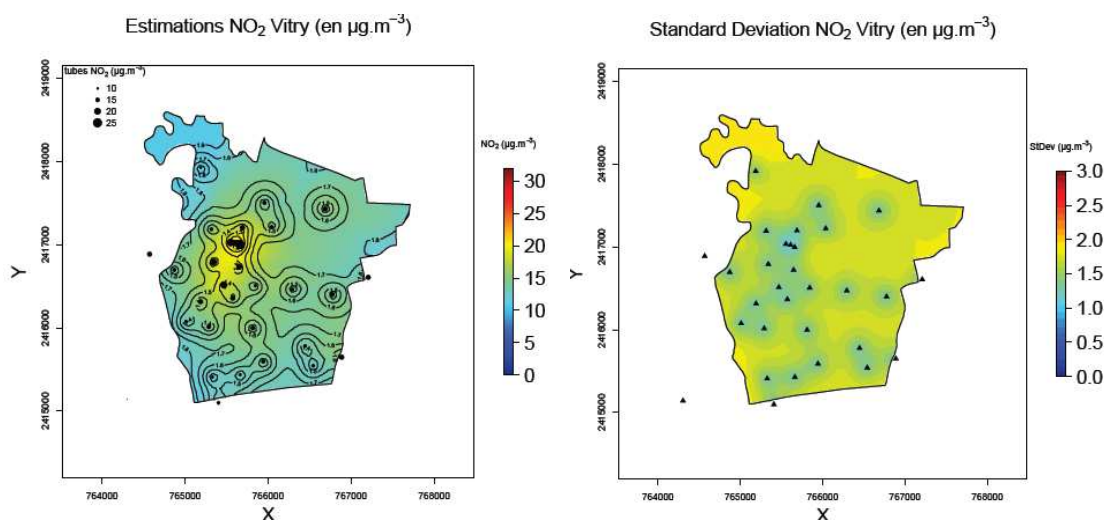


Figure 1 : Exemple de carte d'estimation (krigeage avec émissions de NO_x dans un rayon de 500m en dérive externe) et de carte d'écart-type de krigeage associée

Le travail avec Qualitair Corse commencé en 2012 s'est poursuivi. A la suite des dépassements de la valeur limite annuelle observés en 2010 et 2011 pour le NO_2 sur une station trafic de Bastia, la DREAL a chargé Qualitair Corse d'évaluer les zones potentiellement touchées par un tel dépassement. Dans ce contexte, il a été demandé à l'INERIS de proposer à Qualitair Corse un appui méthodologique pour la réalisation d'une cartographie du NO_2 incluant pollution de fond et pollution de proximité.

En 2012, l'étude a porté sur la cartographie estivale des concentrations de fond². En 2013, les données de la campagne hivernale effectuée par Qualitair Corse ont été incluses afin de

² Malherbe et al., 2012. Assistance en exploitation de données et cartographie. Rapport LCSQA, ref. DRC-13-126730-08778A, <http://www.lcsqa.org/rapport/2012/ineris/assistance-exploitation-donnees-cartographie>

produire une cartographie annuelle. Les mesures réalisées près du trafic routier ont été également prises en compte dans un modèle statistique et intégrées dans la cartographie. Qualitair Corse a passé deux journées à l'INERIS en juillet 2013 afin de pouvoir s'appropriier l'ensemble des développements. Un bilan complet du travail réalisé en 2012 et 2013 est joint en annexe.

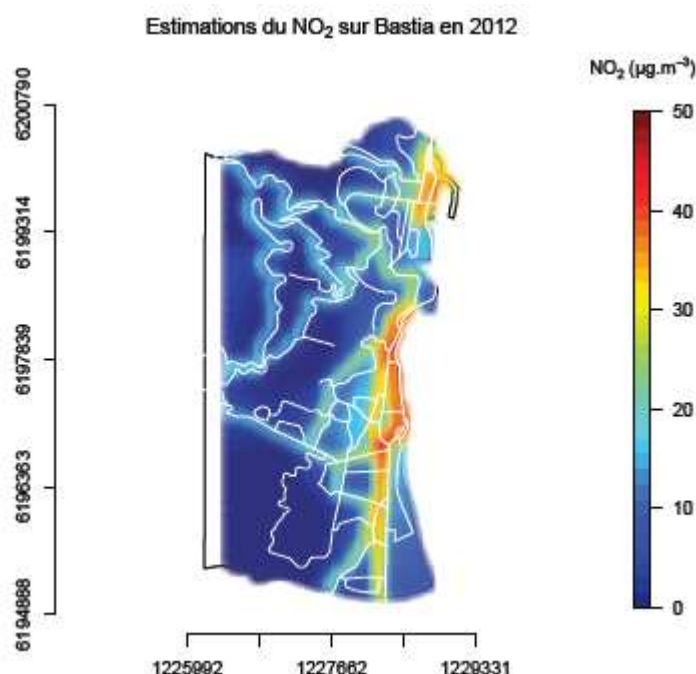


Figure 2 : Cartographie des concentrations de NO₂ en moyenne annuelle sur Bastia (année 2012) intégrant pollution de fond et pollution de proximité automobile

3.2 Etude de la représentativité de sites de mesure

En 2010 et 2011, le LCSQA a travaillé sur une méthodologie destinée à estimer la zone de représentativité spatiale d'une station de mesure (production d'un rapport et d'une note³ et de programmes R associés). Cette méthodologie a pour prérequis une estimation fine des concentrations sur le domaine d'étude et la possibilité d'assortir cette estimation d'une incertitude. Ainsi s'applique-t-elle en particulier aux agglomérations dans lesquelles des campagnes d'échantillonnage relativement denses ont été réalisées : des techniques de krigeage peuvent être alors mises en œuvre dans des conditions favorables.

C'est le cas de la communauté urbaine de Strasbourg dans laquelle, à l'occasion du projet PAISARC, des mesures par tubes à diffusion passive ont été réalisées (hiver 2010-2011 et été 2011). Les données ont été exploitées par l'ASPA qui a comparé différents types de krigeage

³ Beauchamp M., Malherbe L., Létinois L., 2011. Application de méthodes géostatistiques pour la détermination de zones de représentativité en concentration et la cartographie des dépassements de seuils. Rapport LCSQA, réf. DRC-10-111615-05230A, <http://www.lcsqa.org/rapport/2010/ineris/application-methodes-geostatistiques-determination-zones-representativite-concen>.

Beauchamp M., 2012. Cartographie du NO₂ à l'échelle locale, Représentativité des stations, Dépassements de seuils. Note LCSQA, réf. DRC-13-136099-01949A, <http://www.lcsqa.org/rapport/2011/ineris/cartographie-no2-echelle-locale-representativite-stations-depassements-seuils>.

(krigeage avec dérive externe, cokrigeage colocalisé) et étudié l'influence de différentes variables auxiliaires (émissions de NO_x, résultats d'ADMS-Urban). L'ASPA a demandé au LCSQA de l'assister dans l'application de la méthodologie évoquée plus haut et la définition des zones de représentativité spatiale des stations de Strasbourg. A la suite d'échanges préliminaires, l'INERIS et l'ASPA se sont réunis durant une journée au cours de laquelle les points suivants ont été abordés : utilisation d'un modèle urbain comme covariable ou dérive externe dans le krigeage, utilisation des programmes R pour le calcul de la représentativité, analyse des résultats en fonction de l'expérience de terrain de l'ASPA.

4. CONCLUSION

Comme les années précédentes, les actions conduites en 2013 concernent aussi bien le traitement statistique des données que l'élaboration de cartographies. Elles reflètent l'intérêt croissant des AASQA pour le logiciel R qui propose un large éventail de fonctions et se révèle un outil d'exploitation efficace. Ce logiciel tout comme les méthodes qu'il sert à mettre en œuvre nécessitent cependant un certain apprentissage. Compte tenu des avis positifs formulés par les stagiaires, les actions de formation se poursuivront en 2014, en complément de l'offre d'assistance.

ANNEXE I

Bilan de l'assistance apportée à Qualitair Corse sur l'exploitation de campagnes par tubes passifs en 2012-2013

Participants :

Maxime Beauchamp, Laurent Létinois (INERIS)

Nicolas Bernardi, Jean-Luc Savelli (Qualitair Corse)

OBJET : Cartographie du NO₂ sur la ville de Bastia issues de campagnes de mesures par tubes passifs.

Suite à des dépassements de valeurs limites mesurés en 2010 et 2011 en NO₂ sur la station de mesure 41020 de Bastia, l'association Qualitair Corse a présenté à la DREAL un rapport concernant ces dépassements. La DREAL Corse a demandé à Qualitair Corse d'évaluer les zones susceptibles de dépasser les valeurs réglementaires en NO₂. Il a été demandé à l'INERIS, dans le cadre de la fiche technique LCSQA Assistance, de proposer à Qualitair Corse un appui méthodologique pour la réalisation d'une cartographie du NO₂ incluant pollution de fond et de proximité.

1. TRAVAUX PRELIMINAIRES

La méthodologie retenue pour cette étude est l'estimation de la pollution de fond en NO₂ par un krigeage en dérive externe à laquelle est ajoutée la pollution de proximité estimée par l'intermédiaire d'un modèle statistique.

Afin de préparer le déplacement de l'équipe du LCSQA au sein de Qualitair Corse et de centrer l'appui du LCSQA sur la transmission des connaissances, des travaux préparatoires ont été menés en amont du déplacement.

Laurent Létinois était en charge d'extraire et mettre en forme via des logiciels SIG, des variables dites auxiliaires, sur les sites de mesures et sur la grille finale d'estimation. Ces variables prennent une place importante dans les méthodes géostatistiques utilisées pour la cartographie. Différentes variables ont été étudiées :

- Les données d'émissions extraites de l'INS. Certaines corrections ont été apportées à ces données afin de répondre à la fine résolution de l'étude, notamment sur la zone portuaire.
- Les données d'occupation du sol (Corine Land Cover).
- Les données de population sur l'agglomération de Bastia.
- Les données d'altimétrie.

Maxime Beauchamp a développé un programme R permettant la construction de la cartographie. Les différentes tâches réalisées par ce programme sont décrites ci-dessous :

- lecture des données d'observations
- analyse exploratoire des données d'observations
- création d'une grille d'estimation
- calcul du variogramme empirique et modèle associé

- krigeage en dérive externe de la pollution de fond du NO₂
- modèle statistique de proximité (intégré dans le code mais reste à trouver les variables explicatives)
- sorties cartographiques

2. DEROULEMENT DE LA VISITE INERIS - QUALITAIR CORSE (29 AU 31 OCTOBRE 2012)

La visite s'est déroulée sur 2 journées et demie.

29 octobre après-midi :

Une visite des différents sites de mesure s'est déroulée l'après-midi. Celle-ci a permis de mieux appréhender la typologie des sites : Il a été constaté un relief très contrasté entre les différents sites. Ce constat a mené le LCSQA à étudié si une variable d'altimétrie serait utile pour améliorer la modélisation de la répartition spatiale du NO₂ et à vérifier également que celle-ci n'avait pas d'effet sur les concentrations mesurées sur les transects.

30 octobre :

Après un rappel des objectifs et une discussion entre tous les participants, la première journée a principalement été consacrée aux travaux sous SIG. En effet, avant la construction d'une cartographie, il est nécessaire d'évaluer les possibles corrélations entre des variables auxiliaires et les concentrations mesurées lors de la campagne de mesure. Laurent Létinois a donc présenté à N. Bernardi les données utilisées et les méthodes associées pour extraire ces variables auxiliaires. Ces variables sont généralement extraites sur différents rayons autour des points d'intérêt. Certaines limites ont pu être relevées sur les logiciels utilisés au sein de l'association, notamment sur la version d'ArcGIS utilisée. Une version plus récente de ce logiciel ou l'utilisation de MapInfo permettrait une extraction plus simple des données

En parallèle, Maxime Beauchamp a finalisé le script R qui exploite les données extraites sous SIG en complément des données de concentration mesurées par les tubes.

31 octobre :

La deuxième journée a été consacrée à la construction de la cartographie. Le script R étant finalisé, la matinée a été consacrée à la formation de N. Bernardi sur le logiciel R, en déroulant le script ligne par ligne et en lui explicitant les fonctionnalités de ce dernier. Il est bien évident qu'une demi-journée de travail ne peut suffire à intégrer un nouveau langage de programmation, c'est pourquoi un investissement personnel important doit être réalisé au sein de l'AASQA. Le script, déjà rédigé, ainsi que la documentation très fournie sur le langage doit permettre de faire fonctionner de manière autonome le programme, voir de le faire évoluer.

A l'issue de la matinée, la carte de pollution de fond du NO₂ sur la période estivale a été construite.

L'après-midi a constitué une session de travail plus ouverte : les lignes de commande permettant de mettre en place le modèle statistique de proximité ont été présentées. Des premiers tests de mise en place de ce modèle ont été faits, sans qu'ils soient concluants. Cela a permis de préciser à l'AASQA le travail qu'il reste à faire pour mener à bien le travail d'intégration de la pollution de proximité.

En parallèle, des développements annexes ont permis d'ajouter au script des procédures de validation croisée (validation de la cartographie de fond) et une analyse des transects pour évaluer la décroissance des concentrations en fonction de la distance à l'axe.

3. PREMIERS RESULTATS

Ces deux journées de travail ont déjà apporté de nombreux résultats : les principales variables auxiliaires ont été extraites sous SIG (Figure 3) et le programme de construction de la cartographie a été finalisé. La pollution de fond du NO₂ a ainsi pu être cartographiée (Figure 4).

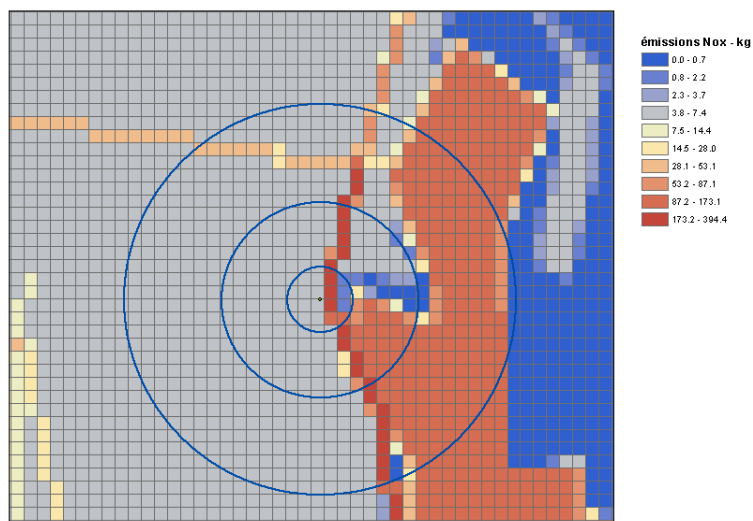


Figure 3 : Extraction des émissions de NO_x sur le site de mesure 41020 sur différents rayons.

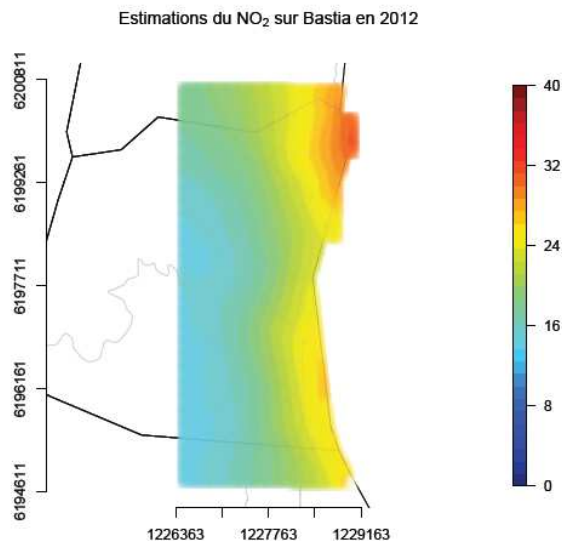


Figure 4 : Cartographie de la pollution de fond du NO₂ été sur Bastia

4. DEROULEMENT DE LA VISITE QUALITAIR CORSE - INERIS (10-11 JUILLET 2013)

Depuis la dernière visite, la campagne de mesure hivernale a été menée (23 novembre – 21 décembre 2012). Pour la cartographie, nous nous intéressons donc désormais à la moyenne annuelle de NO₂. La structure variographique est différente de celle obtenue avec la moyenne été puisqu'on passe d'un modèle linéaire à un modèle sphérique.

La cartographie construite pour la pollution de fond du NO₂ en moyenne annuelle semble plus fiable que celle construite pour la moyenne estivale (voir Figure 5 : nuage de points en validation croisée).

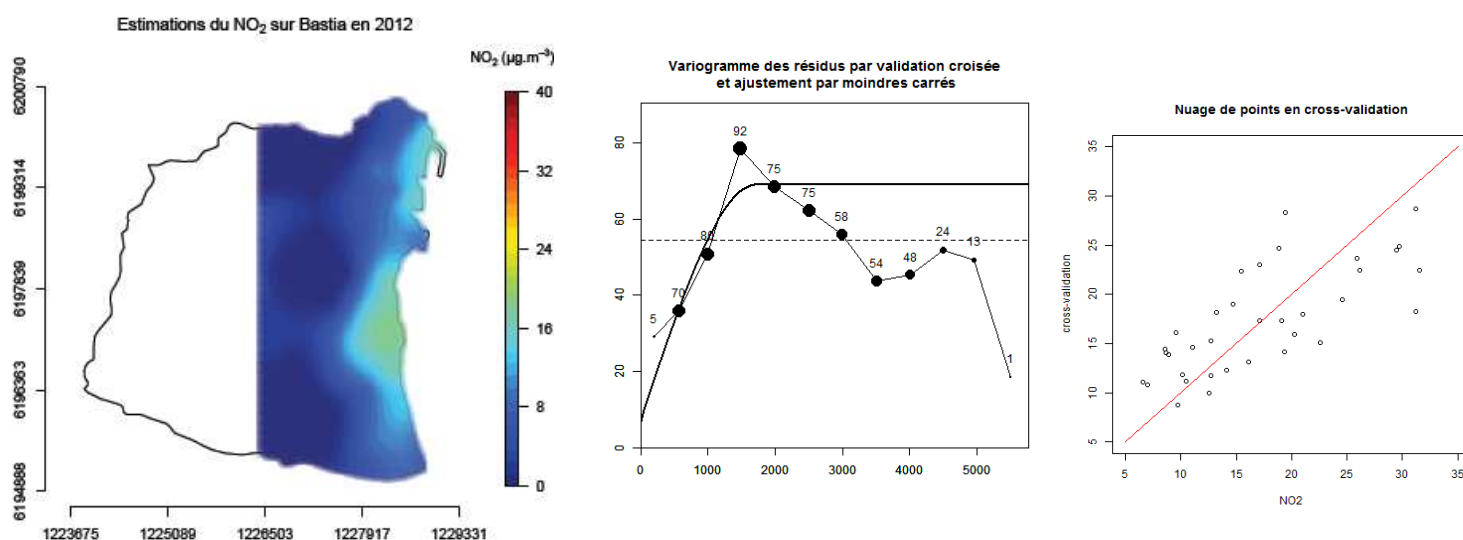


Figure 5 : Cartographie de la pollution de fond du NO₂ en moyenne annuelle sur Bastia

L'objectif principal étant d'identifier les zones en dépassement de VL, on a cherché à intégrer la proximité automobile dans cette cartographie. On utilise la méthode présentée dans le rapport LCSQA 2010⁴, qui consiste à calculer un incrément de pollution sur les sites de proximité en estimant au préalable la pollution de fond par krigeage sur ces sites. On explique ensuite cet incrément par un modèle statistique (régression multilinéaire) qui sera appliqué sur toutes les mailles de la grille d'estimation située à moins de 200m d'un axe routier.

Cette méthode s'est révélée difficile à mettre en place, principalement pour deux raisons liées à l'inventaire des émissions de NO_x routier :

⁴ [Application de méthodes géostatistiques pour la détermination de zones de représentativité en concentration et la cartographie des dépassements de seuils](#). Maxime Beauchamp, Laure Malherbe, Laurent Létinois. 2010

- L'inventaire comporte manifestement des erreurs : certains axes importants présentent des émissions plus faibles que des axes mineurs.
- L'inventaire ne fait pas apparaître certains axes routiers sur lesquels des niveaux de proximité assez importants ont été mesurés par les tubes.

En conséquence, il était impossible de mettre en place le modèle statistique à partir des 53 sites situées à proximité ou en transect d'un axe routier. On a alors développé un algorithme permettant de sélectionner le sous-ensemble de points qui donne à la fois les meilleurs résultats en termes de corrélation modèle-mesure, et la formule de régression affine la plus vraisemblable (coefficient de régression positif pour les émissions de NOx routier dans un rayon de 150m et coefficient de régression négatif pour la distance à l'axe routier le plus proche)

Cette méthode est algorithmique et ne relève d'aucune sélection manuelle des points, ce qui pourrait aussi être fait si l'on souhaite exclure les points concernés par les erreurs de l'inventaire d'émissions. Elle a le mérite de fournir une formule réaliste pour construire la cartographie.

Ci-dessous, la cartographie obtenue avec un sous-ensemble de 20 points de proximité. La formule de régression obtenue est la suivante :

$$\delta(x) = 0.002201 * \text{NOx_routier_150} - 0.140704 * \text{distance axe} + 16.533888$$

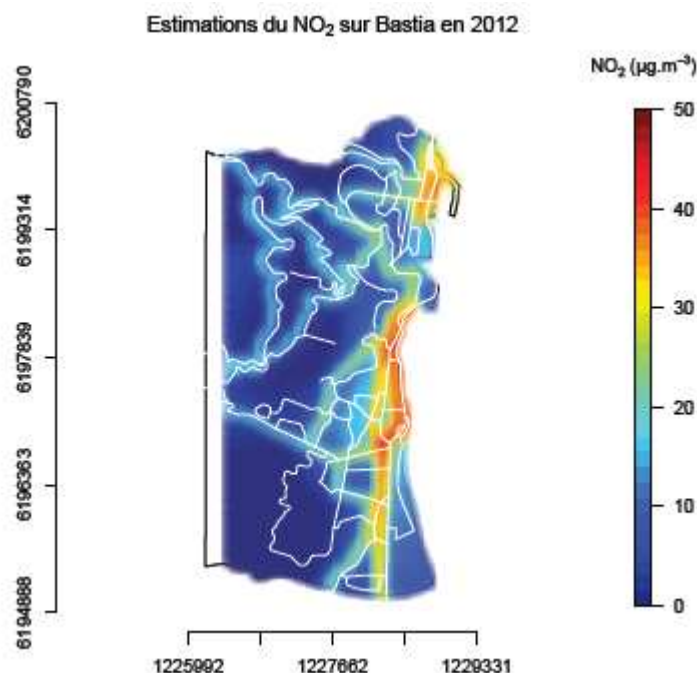


Figure 6 : Cartographie de la pollution du NO2 en moyenne annuelle intégrant fond et proximité automobile sur Bastia

Le nuage de points modèle-mesure en validation croisée n'est pas très bon mais c'était attendu à la vue des lacunes du cadastre d'émissions. Néanmoins, cette carte identifie les zones présentant des forts niveaux de concentrations en NO₂ et est cohérente avec les mesures faites par tubes passifs.

5. PERSPECTIVES

Afin de construire des cartographies plus fiables sur l'agglomération, plusieurs options sont à envisager. Dans l'idéal, la topographie de la ville étant assez spécifique (bord de mer et relief prononcé), un modèle urbain serait plus adapté. Il faudra pour cela développer et consolider l'inventaire des émissions. L'échelle de temps sera néanmoins longue.

Cette session de travail a aussi été l'occasion de réfléchir à de nouvelles approches pour la spatialisation des données. On pourrait imaginer spatialiser l'incrément de pollution par krigeage et sommer les cartes de fond et de l'incrément de pollution lié au trafic routier. Il faudrait pour cela y intégrer le phénomène de décroissance des concentrations, ce qui requiert des développements mathématiques conséquents.