



Assistance à l'utilisation des outils de modélisation

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

Convention n°115/03

*Laurence ROUÏL
Unité Modélisation et Analyse Economique
pour la gestion des Risques (MECO)*

Direction des Risques Chroniques (DRC)

Décembre 2003



Assistance à l'utilisation des outils de modélisation

Laboratoire Central de Surveillance de la Qualité de l'Air

Convention n°115/03

financée par la Direction de la Prévention des Pollutions
et des Risques (DPPR)

Décembre 2003

Laure MALHERBE
Laurence ROUÏL

Ce document comporte 11 pages (hors couverture et annexes).

	Rédaction	Vérification	Approbation
NOM	Laure MALHERBE	Laurence ROUÏL	Martine RAMEL
Qualité	Ingénieur Etudes et Recherches Direction des Risques Chroniques	Ingénieur Etudes et Recherches Direction des Risques Chroniques	Coordination LCSQA Direction des Risques Chroniques
Visa			

TABLE DES MATIÈRES

1. Résumé	Erreur! Signet non défini.
2. Assistance en modélisation déterministe	5
3. Assistance en prévision	6
3.1 Travail réalisé	6
3.2 Bilan sur AASQARIUM	6
3.2.1 Bilan sur le logiciel	7
3.2.2 Bilan sur les algorithmes statistiques	7
3.3 Perspectives	8
4. Assistance en interpolation	9
5. Assistance au GT Moyens Mobiles	10
6. Liste des annexes	11

1. RESUME

1- Modélisation déterministe

Comme pour l'année passée et suivant le cadrage formulé par le MEDD en matière de modélisation déterministe, les demandes d'assistance des AASQA ne concernent pas de nouveaux projets, mais plutôt le suivi de programmes déjà engagés. Notre contribution s'exprime par une participation active aux Comités de Suivi des projets, avec une attente exprimée en terme d'analyse et d'expertise des choix techniques effectués par l'Association. L'avis du LCSQA est parfois requis en terme d'orientation des études et d'interprétation des résultats obtenus.

2- Prévision statistique

Aucune demande précise n'a été adressée à l'INERIS dans ce domaine. Durant l'année 2003, l'INERIS a continué à suivre le projet AASQARIUM, entrepris en 2001, par l'IRISA de l'INSA de Rennes a entrepris, sous l'égide de l'association ECRIN et avec la participation d'AASQA volontaires (COPARLY-ASCOPARG, AIR BREIZH, AIR COM). Le but est de développer et de mettre à la disposition du public une interface logicielle consacrée à la prévision statistique locale à court terme de la pollution atmosphérique.

Un tel projet est destiné à permettre à toute AASQA l'accès à des outils numériques de prévision. Le projet AASQARIUM, dont l'ADEME et l'INERIS suivent le déroulement, a abouti en 2003 à 1) la mise au point d'un outil générique opérationnel et convivial permettant la récupération, l'apprentissage et l'utilisation de modèles statistiques de prévision ; 2) la validation cet outil sur un algorithme statistique et des jeux de données tests. L'INERIS s'assure de la cohérence de ces développements par rapport aux besoins des AASQA et de leur valorisation auprès des AASQA.

3- Interpolation

Le travail de l'INERIS sur ce thème a essentiellement consisté à 1) finaliser un guide d'utilisation des méthodes de la géostatistique ; 2) assurer le suivi et la valorisation d'une étude conduite par le Centre de Géostatistique de l'Ecole des Mines de Paris, dans le cadre d'une convention INERIS/ARMINES.

4- Traitement statistique des données d'échantillonnage

Concernant son activité d'animation auprès du GT « Moyens Mobiles », l'INERIS avait mis en évidence en 2002 plusieurs méthodes susceptibles de répondre à la question de l'échantillonnage temporel et de l'estimation d'une moyenne annuelle à partir des données de moyens mobiles (Rapport LCSQA 2002, Assistance en modélisation de la qualité de l'air).

L'INERIS, a évalué en 2003 la méthode des plans de sondage mise au point par ATMO Poitou-Charentes. Cette évaluation a été réalisée en collaboration avec ATMO Poitou-Charentes. Les données exploitées à cette fin ont été fournies par plusieurs AASQA membres du GT Moyens Mobiles.

2. ASSISTANCE EN MODELISATION DETERMINISTE

Comme pour l'année passée et suivant le cadrage formulé par le MEDD en matière de modélisation déterministe, les demandes d'assistance des AASQA ne concernent pas de nouveaux projets, mais plutôt le suivi de programmes déjà engagés. Notre contribution s'exprime par une participation active aux Comités de Suivi des projets, avec une attente exprimée en terme d'analyse et d'expertise des choix techniques effectués par l'Association. L'avis du LCSQA est parfois requis en terme d'orientation des études et d'interprétation des résultats obtenus. Les projets pour lesquels le LCSQA a été sollicité sont les suivants :

- Air de l'Ain et des Pays de Savoie, programme POVA (Pollution dans les Vallées Alpines) : Constitution et présidence du Groupe de Suivi Scientifique du projet. Ce groupe réunit des représentants des DRIRE et services de l'Equipement, du MEDD, de l'ADEME, des AASQA voisines, du LCSQA, et d'experts appartenant au Conseil Scientifique de PRIMEQUAL. L'ampleur du programme et son possible impact en terme de communication et d'appui aux pouvoirs publics obligent à une certaine rigueur sur son suivi scientifique. Il comprend une large part de modélisation, qui étant donnée la nature des problèmes traités (pollution en zone montagneuse) requiert l'usage de modèles numériques sophistiqués, de type tridimensionnels. Le Comité a un rôle d'évaluation et de conseil dans les choix techniques qui sont faits. L'INERIS assure l'animation de ce groupe.
- AIRMARAIX : Participation au Comité d'évaluation du projet de mise en place de la plate-forme régionale de modélisation déterministe. Le consortium Air-Alpes-Méditerranée a mandaté l'association AIRMARAIX afin de porter un programme de modélisation déterministe à l'échelle régionale afin de réaliser des cartographies de surveillance et des prévisions de la qualité de l'air sur la région PACA. Un comité d'évaluation consulté en 2002 sur l'élaboration du cahier des charges de ce projet a été constitué et reconduit en 2003, afin de procéder à l'évaluation et au choix du prestataire. L'INERIS a été consulté dans ce cadre et sollicité afin de remettre un avis d'expert concernant la faisabilité technique et la pertinence de ce programme.
- ASPA : programme Interreg 3 d'évaluation des émissions et des immissions de polluants atmosphériques dans le Fossé Rhéna. L'INERIS participe aux réunions de suivi de ce projet. L'activité de l'année 2003 a été assez dense, puisque outre la préparation de l'exercice de modélisation associé au projet, une campagne de mesure de grande envergure a été organisée en juin 2003. La contribution du LCSQA/INERIS se résume à une activité de conseil.

De manière ponctuelle plusieurs AASQA ont sollicité l'avis de l'INERIS sur la mise en place de modèles de rue, en vue de l'évaluation de l'exposition des populations citadines. Plusieurs AASQA s'interrogent sur la pertinence des résultats fournis par des modèles empiriques plus ou moins sophistiqués et souhaitent une implication du LCSQA sur ce thème (Air Pays de Loire par exemple). Quelques échanges, qui seront certainement plus importants en 2004 ont eu lieu sur ce thème.

3. ASSISTANCE EN PREVISION

Aucune demande précise n'a été adressée à l'INERIS dans ce domaine. Durant l'année 2003, l'INERIS a continué à suivre le projet AASQARIUM dont sont retracées ci-après les principales étapes.

3.1 TRAVAIL REALISE

En 2001, l'IRISA de l'INSA de Rennes a entrepris, sous l'égide de l'association ECRIN et avec la participation d'AASQA volontaires (COPARLY-ASCOPARG, AIR BREIZH, AIR COM), de développer et de mettre à la disposition du public une interface logicielle consacrée à la prévision statistique locale à court terme de la pollution atmosphérique. Cette interface, appelée AASQARIUM, a pour fonction de faciliter l'acquisition et la mise en œuvre de modèles de prévision statistiques élaborés par des chercheurs.

Un tel projet est destiné à permettre à toute AASQA l'accès à des outils numériques de prévision. Le projet AASQARIUM, dont l'ADEME et l'INERIS suivent le déroulement, a donné lieu en septembre 2002 à une convention entre l'IRISA et l'INERIS. Par cette convention, l'IRISA s'est engagé à :

- mettre au point un outil générique opérationnel et convivial permettant la récupération, l'apprentissage et l'utilisation de modèles statistiques de prévision ;
- valider cet outil sur un algorithme statistique et des jeux de données tests.

Ce travail a été réalisé à l'aide de données fournies par COPARLY- ASCOPARG et par AIR BREIZH. En 2002, plusieurs réunions de cadrage ainsi que des tests préliminaires effectués par COPARLY ont permis d'adapter les fonctionnalités du logiciel aux besoins réels des AASQA. Une première version d'AASQARIUM est librement accessible sur Internet depuis juin 2003, sur un site conçu par l'IRISA à cet effet (www.insa-rennes.fr/d-info/aasqarium). Elle a fait l'objet d'une démonstration lors du colloque « STIC¹ & Environnement » (Rouen, 19 et 20 juin 2003).

Pendant l'été 2003, AASQARIUM a été installé et évalué au sein d'AIR COM par un étudiant de l'IRISA. L'algorithme statistique de prévision dont disposait déjà AIR COM (prévision par arbre) a été adapté aux spécifications d'AASQARIUM à cette occasion.

Parallèlement à ce travail, des algorithmes statistiques ont été développés et testés par l'IRISA. Ils sont désormais disponibles sur le site Internet mentionné plus haut, et peuvent être chargés et utilisés grâce à AASQARIUM.

3.2 BILAN SUR AASQARIUM

Un réunion de fin de convention a eu lieu le 20 novembre 2003 en présence de l'IRISA, de l'INERIS et d'ECRIN. Le compte-rendu de cette réunion est joint en annexe. Nous en restituons ici les principales conclusions.

¹ Sciences et Techniques de l'Information et de la Communication

3.2.1 Bilan sur le logiciel

Le logiciel AASQARIUM, c'est-à-dire la structure informatique permettant l'exploitation des modèles statistiques, est désormais opérationnel. Il peut être librement téléchargé à l'adresse suivante :

www.insa-rennes.fr/d-info/aasqarium

Il fonctionne aussi bien dans l'environnement Windows que sous Unix ou Linux. Développé sous licence GPL (general public license), il est libre d'utilisation, de modification et de redistribution.

Il permet plus précisément:

- de charger un ou plusieurs algorithmes statistiques ;
- de réaliser l'apprentissage de ces algorithmes à l'aide d'un historique de données propre à la zone étudiée (ces données doivent se trouver au format ISO de l'ADEME), créant ainsi des modèles locaux de prévision ;
- de mettre en œuvre ces modèles pour effectuer des prévisions
- de visualiser graphiquement les résultats.

De manipulation aisée, il est doté d'une rubrique d'aide. Un manuel d'utilisation est également disponible à l'adresse Internet mentionnée ci-dessus. AASQARIUM prévoit en outre la possibilité de rédiger des scripts, afin de lancer la récupération des données et le calcul des prévisions de façon automatique. Cette fonctionnalité n'a toutefois pas été testée.

3.2.2 Bilan sur les algorithmes statistiques

Quatre algorithmes statistiques sont actuellement proposés sur le site Internet d'AASQARIUM :

- Squelette (persistance) : ce programme, qui prédit du jour pour le lendemain la concentration maximale observée le jour même, sert de référence pour l'évaluation des modèles. Il définit en outre la structure de programmation que doit respecter tout algorithme destiné à fonctionner dans AASQARIUM.
- PersistPlus : il s'agit d'un modèle de persistance amélioré, qui apporte au modèle précédent une correction fonction de la température. (algorithme conçu par l'IRISA).
- ArbreO3 : cet algorithme de prévision par arbre, utilisé à AIR COM, a été rendu compatible avec les spécifications d'AASQARIUM.
- SifEnLigne : cet algorithme fondé sur la logique floue a été élaboré par l'IRISA. Il permet de réaliser une prévision à 8hTU du pic d'ozone du jour et une prévision à 14hTU du pic d'ozone du lendemain. De plus, il associe à chaque estimation un indice de confiance. Il a été évalué sur trois sites au cours de l'été 2003 : Rennes, Lorient et Brest.

Chaque algorithme est spécifique à un polluant donné et à un ensemble donné de prédicteurs. Tout utilisateur qui le souhaite peut l'adapter à un autre polluant, remplacer un prédicteur ou introduire d'autres variables explicatives. Il convient à cette fin de modifier le fichier .info qui

établit le lien entre l'algorithme et AASQARIUM ainsi que le fichier de configuration associé à l'algorithme. Une notice contenant la procédure à suivre sera rédigée par l'IRISA.

3.3 PERSPECTIVES

La mise en œuvre opérationnelle d'AASQARIUM au sein des AASQA intéressées par cet outil est désormais indispensable à la survie, à l'amélioration et à la valorisation du logiciel. L'implication des AASQA et de la communauté scientifique pour alimenter le site Internet en nouveaux algorithmes de prévision est également nécessaire. L'association ECRIN et l'IRISA se sont rapprochées dans ce but d'une action de recherche en qualité de l'air conduite par le CNRS (Département *Sciences et Techniques de l'Information et de la Communication et Environnement*). Des rapprochements avec des partenaires européens seraient aussi utiles.

Afin de favoriser la création d'un groupe d'utilisateurs, l'association ECRIN est prête à héberger sur son site Internet un forum de discussion. Celui-ci permettrait aux utilisateurs potentiels d'échanger sur les difficultés rencontrées et de transmettre leurs réflexions et commentaires sur l'outil. Un lien pourrait être fait de ce site vers le site AASQARIUM de l'IRISA. L'organisation d'une journée de démonstration et de formation sur AASQARIUM a été également proposée par ECRIN.

D'autre part, l'IRISA est disposé à apporter une fois par an un certain nombre de corrections mineures à AASQARIUM et à adapter des algorithmes statistiques existants aux spécifications du logiciel. En revanche, la question de la réalisation de modifications majeures reste en suspens.

4. ASSISTANCE EN INTERPOLATION

Aucune demande précise pour l'année 2003 n'a été formulée par les AASQA sur ce thème. A la suite d'un courrier envoyé par l'INERIS à l'ensemble des associations, plusieurs AASQA (AIR BREIZH, ATMO Champagne-Ardenne, ATMO Picardie, AIRLOR, AIRAQ, ORA, ARPAM, AIR DE L'AIN ET DES PAYS DE SAVOIE, ASQAB, AIRPARIF) ont toutefois manifesté leur intérêt pour l'activité d'assistance proposée par l'INERIS ou pour les documents produits dans le cadre de cette activité.

En 2003, le travail de l'INERIS a consisté à :

- finaliser un guide d'utilisation des méthodes de la géostatistique.

Une version préliminaire de ce guide était parue en 2002 sous le titre *Représentation de la qualité de l'air dans les zones peu couvertes par les stations de mesure fixes : partie I, stratégie et préconisations*.

A la demande du MEDD et de l'ADEME, ce guide a été recentré exclusivement sur les méthodes de la géostatistique.

- assurer le suivi et la valorisation d'une étude conduite par le Centre de Géostatistique de l'Ecole des Mines de Paris, dans le cadre d'une convention INERIS/ARMINES.

En raison de l'importance prise par l'analyse préliminaire des données et du travail de modélisation qu'il est nécessaire de réaliser au préalable, l'objectif initial de l'étude (extrapolation dans le temps de cartes de concentration à l'aide de données de tubes et de stations fixes, et étude de stratégies d'échantillonnage temporel) a été révisé au cours de son développement. L'étude a été ainsi orientée sur l'apport des variables auxiliaires et sur la façon d'introduire ces variables dans le modèle géostatistique. La question de l'échantillonnage et de l'extrapolation temporelle a été cependant abordée d'un point de vue qualitatif et des perspectives intéressantes se sont ouvertes sur ce point. Ce problème sera considéré plus en détail dans une fiche LCSQA 2004.

Les données exploitées, fournies par l'ASPA et par AIR Languedoc-Roussillon, mesuraient le dioxyde d'azote dans deux agglomérations (Mulhouse et Montpellier) et au voisinage d'un axe routier (Vallée de la Thur).

Des conclusions synthétiques de cette étude et les recommandations méthodologiques qui en résultent ont été formulées par l'INERIS et sont données en annexe de ce document (Cartographie du dioxyde d'azote dans une agglomération ou le long d'un axe routier à l'aide des méthodes géostatistiques, synthèse des travaux du Centre de Géostatistique de l'Ecole des Mines de Paris). Les AASQA pourront également consulter cette synthèse ainsi que les rapports détaillés du Centre de Géostatistique sur le site Internet du LCSQA.

Le travail accompli par l'Ecole des Mines de Paris a fait l'objet de trois communications, la première lors des Journées de Géostatistique organisée par le Centre de Géostatistique (Ecole des Mines de Paris, Fontainebleau, 18-19 septembre 2003), la seconde lors de la journée technique « cartographie » du LCSQA (LNE, Paris, 23 octobre 2003), la troisième lors d'une journée organisée par Géovariances sur la géostatistique et ses applications en qualité de l'air (Paris, 17 novembre 2003).

5. ASSISTANCE AU GT MOYENS MOBILES

En 2002, l'INERIS avait mis en évidence plusieurs méthodes susceptibles de répondre à la question de l'échantillonnage temporel et de l'estimation d'une moyenne annuelle à partir des données de moyens mobiles (Rapport LCSQA 2002, Assistance en modélisation de la qualité de l'air).

L'INERIS, ainsi qu'il s'y était engagé en 2002, a évalué durant l'année suivante la méthode des plans de sondage mise au point par ATMO Poitou-Charentes. Cette évaluation a été réalisée en collaboration avec ATMO Poitou-Charentes. Les données exploitées à cette fin ont été fournies par plusieurs AASQA membres du GT Moyens Mobiles. Le principe de la méthode ainsi que l'ensemble des résultats et des préconisations qui découlent du travail réalisé seront présentés au GT Moyens Mobiles à l'occasion d'une réunion plénière le 21 janvier 2004.

La méthode des plans de sondage offre une réponse intéressante et simple d'usage aux questions de l'optimisation de l'échantillonnage, de l'estimation de moyennes annuelles et d'intervalles de confiance autour de celles-ci, et de l'évaluation de la qualité de l'air en comparaison avec des valeurs seuils. Elle n'exige aucun travail de modélisation ni de données de variables externes (du type météorologie)

La garantie de son efficacité ne va cependant pas sans quelques contraintes. Au vu des simulations réalisées, la précision des estimations et la fiabilité des intervalles de confiance estimés sont d'autant meilleures

- que le nombre de passages du moyen mobile en un point est plus important (de multiples mais brefs passages sont préférables à des passages plus longs mais peu nombreux) ;
- que ces passages se répartissent sur l'ensemble de l'année préalablement découpée en périodes homogènes (ces périodes, par exemple les saisons, sont appelées strates) ;
- qu'il est possible de disposer des données d'une station fixe proche géographiquement du site à sonder et de caractéristiques similaires.

De plus, la méthode n'est rigoureusement applicable que si, au sein de chaque strate, l'échantillonnage est réalisé de façon aléatoire.

Dans une gamme relativement large de situations, elle se révèle néanmoins un moyen utile pour concevoir des stratégies d'échantillonnage et estimer des indicateurs annuels de la qualité de l'air. ATMO Poitou-Charentes a ainsi élaboré une dizaine de fiches exemples illustrant son application dans différents contextes.

Les résultats de ce travail seront consignés dans un document qui sera diffusé au premier semestre 2004. Les programmes informatiques permettant la mise en œuvre de la méthode ont été développés avec R, logiciel de statistique en libre accès sur Internet. Ils pourront être transmis à l'ensemble des AASQA.

La comparaison avec les résultats de la méthode développée par l'Ecole des Mines de Douai (étude de l'influence des paramètres météorologiques et mise en œuvre d'un échantillonnage stratifié), également présentée lors de la prochaine réunion, aidera à formuler des recommandations précises en terme de stratégie d'échantillonnage et d'exploitation des mesures.

6. LISTE DES ANNEXES

Repère	Désignation précise	Nb/N°pages
1	Projet AASQARIUM - Evaluation d'un modèle statistique de prévision locale des pics de pollution (IRISA, 2003)	20
2	Cartographie du dioxyde d'azote dans une agglomération ou le long d'un axe routier à l'aide des méthodes géostatistiques. Synthèse des travaux réalisés par le Centre de Géostatistique de l'Ecole des Mines de Paris	13

Annexe 1

Projet AASQARIUM - Evaluation d'un modèle statistique de prévision locale des pics de pollution (IRISA, 2003)

Annexe 2

Cartographie du dioxyde d'azote dans une agglomération ou le long d'un axe routier à l'aide des méthodes géostatistiques

Synthèse des travaux réalisés par le Centre de Géostatistique de l'Ecole des Mines de Paris

TABLE DES MATIERES

1. Introduction 2

2. Cartographie du dioxyde d’azote dans une agglomération. Exemples de Mulhouse et de Montpellier 3

 2.1 Modélisation avec variables auxiliaires 3

 2.1.1 Méthodologie..... 3

 2.1.2 Bilan 7

 2.2 Évaluation des modèles variographiques 8

 2.2.1 Méthodologie..... 8

 2.2.2 Bilan 8

 2.3 Représentativité temporelle des campagnes de mesure et réalisation de cartes moyennes 9

 2.3.1 Comparaison des quinzaines de mesure 9

 2.3.1.1 Méthodologie..... 9

 2.3.1.2 Bilan 9

 2.3.2 Réalisation de cartes moyennes..... 10

 2.3.2.1 Méthodologie..... 10

 2.3.2.2 Bilan 10

3. Cartographie du dioxyde d’azote au voisinage d’un axe routier 11

 3.1 Méthodologie..... 11

 3.2 Bilan 12

4. Conclusion..... 13

1. INTRODUCTION

Les études confiées en 2002 par l'INERIS au Centre de Géostatistique de l'Ecole des Mines de Paris sont une contribution aux travaux conduits par le LCSQA sur l'usage des différents outils de cartographie de la qualité de l'air. Face à l'emploi de plus en plus fréquent dans les AASQA des techniques de la géostatistique, elles ont pour but d'approfondir divers points de méthode relatifs à l'application de ces techniques.

Elles ont été réalisées en collaboration avec l'ASPA et Air Languedoc-Roussillon qui ont fourni l'ensemble des données. Le polluant considéré est le dioxyde d'azote (NO₂). Deux types de situations ont été examinés :

- les études du NO₂ à Mulhouse et Montpellier portent sur l'estimation de la pollution urbaine de fond dans une agglomération ;
- l'analyse des données de la Vallée de la Thur (Haut-Rhin) traite de la question plus spécifique de la modélisation de la pollution de part et d'autre d'un axe routier.

Les résultats des deux premières études nous permettent de tirer des enseignements sur :

- l'apport réel des variables auxiliaires et les difficultés liées à l'introduction de ces variables. **La sensibilité des relations entre concentrations et variables auxiliaires à l'implantation des données et au type d'environnement** est mise en évidence, ce qui pose le **problème de l'extrapolation spatiale de ces relations**.
- **l'évaluation des modèles par validation croisée ;**
- **la représentativité temporelle des campagnes de mesure et l'exploitation de ces campagnes pour établir des moyennes annuelles.**

La troisième étude montre comment on peut aborder un problème de pollution non stationnaire autour d'une source polluante et modéliser l'évolution spatiale des concentrations alentour.

Tous ces points, pour lesquels seront supposées acquises les principales notions de géostatistique (on se reportera à la version préliminaire du guide *Représentation de la qualité de l'air dans les zones peu ou pas couvertes par les stations de mesure fixes*), sont développés dans les paragraphes suivants. Chaque paragraphe comprend des indications méthodologiques suivies de remarques et de mises en garde plus générales sur l'application des méthodes et l'interprétation des résultats. Pour des informations plus détaillées, il convient de se reporter aux rapports suivants :

- Chauvet P., Examen préliminaire des données de Mulhouse, Rapport N-15/02/G, octobre 2002
- Fouquet C.(de), Méthodologie de cartographie de la concentration annuelle de NO₂ sur l'agglomération de Mulhouse, Rapport d'avancement N-6/03/G, avril 2003
- Le Loc'h G., Etude exploratoire du dioxyde d'azote sur l'agglomération de Montpellier, Rapport N-8/03/G, juin 2003
- Fouquet C.(de), Etude sur la réalisation de cartographies de la qualité de l'air dans les zones peu/pas couvertes par les réseaux de stations fixes à l'aide de méthodes géostatistiques (complément d'étude et synthèse), Rapport N-9/03/G, juin 2003
- Séguret S., Estimation du dioxyde d'azote routier dans la vallée de la Thur, Rapport N-1/03/G, février 2003

Ceux-ci seront disponibles dès janvier 2004 sur le site Internet du LCSQA.

2. CARTOGRAPHIE DU DIOXYDE D'AZOTE DANS UNE AGGLOMERATION. Exemples de Mulhouse et de Montpellier

2.1 MODELISATION AVEC VARIABLES AUXILIAIRES

L'information apportée par des variables complémentaires, liées directement ou indirectement aux concentrations et dont les valeurs sont connues en de nombreux points de l'espace, est susceptible d'améliorer la précision des cartes de pollution.

La sélection des variables auxiliaires les plus pertinentes, qui expliquent le mieux les concentrations, ne peut se faire sans une analyse approfondie :

- des relations entre les différentes variables auxiliaires ;
- des relations entre les variables auxiliaires et les concentrations.

Les variables auxiliaires disponibles étaient :

- pour Mulhouse :
 - l'altitude (grille de données avec un pas de 250 m ou 1000 m) ;
 - les émissions (grille de données avec un pas de 1000 m) ;
 - la densité de population (grille de données avec un pas de 1000 m) ;
 - l'occupation du sol (grille de données avec un pas de 1000 m).
- pour Montpellier :
 - les émissions (grille de données avec un pas de 1000 m ou 200 m) ;
 - la densité de population dans un rayon de 100 m, 1000 m ou 1500 m autour des points de mesure.

2.1.1 Méthodologie

Les tableaux 1 et 2 synthétisent la méthodologie adoptée pour cette analyse. L'ensemble des calculs a été réalisé avec le logiciel Isatis.

Tableau 1 – Analyse des corrélations entre variables

Calculs réalisés	Objectifs
<p>Calcul des statistiques des variables auxiliaires : statistiques élémentaires, histogrammes, variogrammes, corrélations entre variables. Étude de la stabilité de ces statistiques lorsque l'on modifie l'ensemble de points sélectionnés pour le calcul. Dans certains cas, une transformation logarithmique des variables initiales peut se révéler utile car elle a un effet stabilisateur. Exemple de transformation : $Y = \log(1 + Z / \bar{Z})$ ou $Y = \log(1 + Z / k)$ <i>Z</i> : variable d'origine \bar{Z} : moyenne de Z <i>k</i> : facteur à déterminer</p>	<p>Ces calculs permettent d'apprécier la cohérence de l'ensemble des variables auxiliaires, en vue d'en réduire le nombre et d'étudier les relations de ces variables avec les concentrations.</p>
<p>Analyse en composantes principales (ACP) Cette technique mathématique d'analyse multivariable permet de réduire un système complexe de corrélations en un plus petit nombre de dimensions. Elle remplace l'ensemble initial de variables par de nouvelles variables de variance maximale (les facteurs de l'ACP), non corrélées deux à deux et qui sont des combinaisons linéaires des variables d'origine. Réalisée sur la totalité des variables auxiliaires, l'ACP permet d'observer les corrélations, similarités ou oppositions entre ces variables et de mettre en évidence des familles de variables auxiliaires. Pour une analyse plus fine des corrélations, une ACP peut être ensuite effectuée au sein de chaque famille.</p>	<p>Ce travail a pour fin de réduire le nombre initial de variables auxiliaires. Il permet d'éliminer les informations redondantes pour ne conserver que certaines variables représentatives de l'ensemble. Il fournit en outre des variables synthétiques (les facteurs) qui peuvent être également employées comme variables auxiliaires.</p>

Remarque technique :

Lorsque les variables auxiliaires se présentent sous la forme d'une grille de valeurs, l'utilisateur a deux possibilités pour conduire l'ACP :

1. L'ACP est effectuée aux nœuds de la grille. Ainsi obtient-on les valeurs des facteurs en tout point de cette grille, en vue de l'utilisation éventuelle de ces variables synthétiques comme dérives externes.

Pour disposer des valeurs des facteurs aux points de mesure on procèdera à une interpolation classique, à un krigeage ou à une migration. Cette dernière solution,

d'usage courant en géostatistique, consiste à affecter aux points expérimentaux les valeurs des facteurs calculées aux nœuds les plus proches.

2. Une valeur des variables auxiliaires est attribuée préalablement à chaque point d'échantillonnage (par interpolation linéaire classique, krigeage ou migration) puis l'ACP est réalisée sur cet ensemble plus restreint de points.

On suppose que les valeurs des facteurs fournies aux points de mesure par l'ACP se calculent aux nœuds de la grille par la même combinaison linéaire des variables initiales. Cette hypothèse est admissible **si les points de mesure sont représentatifs de l'ensemble de la zone à estimer.**

Cette seconde méthode a l'avantage de préparer l'étape suivante c'est-à-dire l'étude des corrélations entre variables auxiliaires et concentrations aux points de mesure.

Une comparaison aux points de mesure des facteurs d'ACP obtenus par ces deux méthodes permet de vérifier si les relations entre variables auxiliaires diffèrent sensiblement ou non entre les tubes et l'ensemble de la zone à estimer.

Tableau 2 – Analyse des corrélations entre variables (suite) et modélisation

Calculs réalisés (suite)	Objectifs
<p>Relations entre variables auxiliaires et concentrations</p> <p>Tracé du nuage de corrélation et calcul du coefficient de corrélation entre les variables auxiliaires (brutes, transformées ou synthétiques) précédemment sélectionnées et les concentrations</p> <p>Etude de la correspondance entre les classes de valeurs des variables auxiliaires et des classes de valeurs des variables auxiliaires</p> <p>Analyse en composantes principales effectuée sur les variables auxiliaires présélectionnées et les concentrations.</p> <p>Il convient là encore de vérifier si les corrélations et relations entre variables sont sensibles ou non à une modification de l'ensemble des points étudiés.</p> <p>La relation entre les données d'une variable auxiliaire et les données de concentration peut être rendue plus linéaire, soit en remplaçant la variable auxiliaire par sa régression expérimentale sur la variable de pollution $E(Z_{aux}/Z)$, soit, si cela n'a pas été fait, en transformant la variable auxiliaire (ex : $\log(1+Z_{aux}/k)$)</p>	<p>Sélection des variables auxiliaires qui expliquent le mieux les concentrations.</p>
<p>Définition d'un modèle géostatistique intégrant ces variables auxiliaires</p> <p><i>La méthode d'estimation ici préconisée et pour laquelle doit être défini le modèle est le krigeage avec dérive externe.</i></p> <p><i>Cette méthode suppose qu'en moyenne, Z est une fonction linéaire de la ou des variables auxiliaires mises en dérive, ce qui impose une condition supplémentaire dans le krigeage :</i></p> $E[Z(x)] = \sum_1^N a_i \Phi_i(x)$ <p>Φ_i : variables explicatives</p> <p><i>Le modèle de variogramme est ajusté dans Isatis en utilisant la fonction de calage automatique (option d'ajustement non stationnaire).</i></p>	<p>Construire un modèle géostatistique en accord avec les données expérimentales et tenant compte du pouvoir explicatif des variables auxiliaires.</p>

<p><i>Le choix de cette méthode s'explique par :</i></p> <ul style="list-style-type: none"> - <i>le fait que la plupart des variables auxiliaires sont données avec une densité d'information très supérieure à celle des tubes ;</i> - <i>la propriété qu'a le krigeage avec dérive externe d'optimiser localement (dans le voisinage) la combinaison linéaire des variables explicatives.</i> 	
---	--

2.1.2 Bilan

Les variables auxiliaires qui se sont révélées les plus explicatives de la pollution sont :

- pour Mulhouse :
 - les émissions de NOx
 - la densité de population
 - le bâti dense.

(Cette dernière variable, peu corrélées aux précédentes, apporte une information différente de celle qui est contenue dans les deux premières variables.)

- pour Montpellier :
 - le cadastre des émissions à 1000 m
 - la densité de population à 1000 ou 1500 m

Ces études multivariées très approfondies mettent en évidence plusieurs points importants qui exigent la plus grande vigilance.

- Le problème de la représentativité spatiale des points de mesure et de l'extrapolation dans l'espace du modèle multivariable calé sur ces points est ainsi soulevé. Les relations entre variables de concentration et variables auxiliaires sont-elles représentatives de la zone à estimer ? La question se pose particulièrement lorsque les échantillons sont implantés de façon préférentielle. L'étude de Montpellier montre clairement que le type d'environnement dans lequel ont été exposés les tubes influence significativement les relations entre concentrations et variables explicatives. **Dans ce cas, il faut être prudent sur l'utilisation du modèle hors des zones dans lesquelles il a été construit.**
- Le pouvoir explicatif des variables auxiliaires dépend aussi :
 - de leur **résolution spatiale**.
Ce paramètre peut avoir un effet déterminant dans les relations avec les concentrations. Dans l'étude sur Montpellier par exemple, les émissions de NOx fournies suivant une maille kilométrique sont plus corrélées aux concentrations de NO₂ que les émissions de NOx fournies avec un pas de 200 m.
 - de la **période de l'année**.
Ainsi à Mulhouse, la corrélation entre densité de population et concentration est plus forte en hiver qu'en été.

Notons que dans les deux zones étudiées les variables auxiliaires améliorent ostensiblement la qualité du modèle en hiver mais n'apportent qu'un gain très limité en été. En été, les concentrations de NO₂ sont moins structurées dans l'espace et l'implantation des fortes valeurs se modifie, de là les différences de liens entre moyennes saisonnières et variables auxiliaires.

Il semble dès lors impossible d'établir une règle systématique sur le type de variables auxiliaires à employer en toutes circonstances. Des variables considérées comme explicatives de la pollution ne rendent pas toujours compte de la variabilité des concentrations. **Avant toute modélisation, on ne peut en aucun cas s'affranchir d'une étude approfondie des corrélations entre variables.**

2.2 ÉVALUATION DES MODELES VARIOGRAPHIQUES

Comment choisir un modèle adéquat et quelle précision attendre des modèles ?

2.2.1 Méthodologie

En réponse à ces questions, une **validation croisée** est mise en œuvre pour chaque modèle ajusté. Elle consiste à supprimer temporairement un point de mesure et à estimer par krigeage la valeur en ce point à l'aide des autres données expérimentales. Cette opération est répétée pour chaque donnée.

Si Z désigne la concentration observée et Z^* son estimation par validation croisée, les erreurs, erreurs standardisées et erreurs relatives de validation croisée sont respectivement $(Z-Z^*)$, $(Z-Z^*)/\sigma_K$ et $|Z-Z^*|/Z$.

Les statistiques de ces erreurs ainsi que les corrélations entre observations et estimations sont alors calculées. La qualité d'un modèle se juge sur plusieurs critères :

- le nombre de données robustes, c'est-à-dire de données dont l'erreur standardisée est inférieure à 2,5 en moyenne absolue ;
- la moyenne de l'erreur ou de l'erreur standardisée, qui doit s'approcher de 0 ;
- la variance de l'erreur, qui doit être la plus faible possible ;
- la variance de l'erreur standardisée, qui doit s'approcher de 1 ;
- la corrélation entre estimations et observations (critère généralement concordant avec celui de la variance de l'erreur).

Le calcul de l'erreur relative apporte en outre une information sur la précision des estimations à l'intérieur de la zone d'échantillonnage.

2.2.2 Bilan

Dans l'étude de Mulhouse comme dans celle de Montpellier, la validation croisée est mise en œuvre pour toutes les variantes de modèles considérées (modèles avec ou sans dérive externe, avec des variables différentes en dérive ou ajustés avec des fonctions mathématiques). Les écarts entre variantes sont souvent faibles. En revanche, la précision des estimations fondées sur ces variantes n'est pas toujours satisfaisante (proportion importante d'erreurs relatives supérieures à 30%). Comme il est mentionné en conclusion du rapport sur Mulhouse, **il semble plus important de chercher à améliorer la précision des estimations en augmentant le nombre de tubes que de vouloir optimiser à tout prix le modèle de variogramme.**

Par ailleurs, les statistiques de la validation croisée dépendent de l'implantation des données. **Elles reflètent principalement la qualité du modèle dans les zones les plus riches en données**, là où les points de mesure sont proportionnellement plus nombreux.

Avant de conclure sur la qualité du modèle, il convient de s'interroger sur les zones dans lesquelles on souhaite que ce modèle soit le plus fiable : centre de l'agglomération ? périphérie de l'agglomération ? Si la ou les zones d'intérêt ne sont pourvues que d'un petit nombre de points de mesure, la validation croisée est insuffisante pour y évaluer le modèle géostatistique.

2.3 REPRESENTATIVITE TEMPORELLE DES CAMPAGNES DE MESURE ET REALISATION DE CARTES MOYENNES

Dans les deux zones étudiées, les données de tubes ont été collectées pendant plusieurs quinzaines consécutives et à des saisons contrastées :

- Mulhouse :
 - Hiver 2001 : 3 quinzaines successives
 - Été 2001 : 3 quinzaines successives
- Montpellier :
 - Hiver 2001 : 4 quinzaines successives
 - Été 2001 : 4 quinzaines successives
 - Hiver 2002 : 4 quinzaines successives

Quelle information les données de ces quinzaines de mesure apportent-elles d'un point de vue temporel?

2.3.1 Comparaison des quinzaines de mesure

2.3.1.1 Méthodologie

Une analyse détaillée des corrélations entre quinzaines et entre saisons complète l'étude des relations avec les variables auxiliaires.

Elle a pour objet d'observer les ressemblances et disparités entre quinzaines et de voir si ces dernières semblent ou non correspondre à des situations de pollution similaires. Cet examen permet d'évaluer en première approche la représentativité temporelle d'une quinzaine par rapport à la saison et de la saison par rapport à l'année.

2.3.1.2 Bilan

Pour Mulhouse comme pour Montpellier, on constate *a posteriori* **l'intérêt d'effectuer des mesures pendant plusieurs quinzaines afin de couvrir diverses situations de pollution**. Dans chaque cas en effet, et aussi bien en hiver qu'en été, une quinzaine de mesure présente des caractéristiques différentes des autres quinzaines.

D'autre part, les zones de fortes et de faibles concentrations de l'hiver ne coïncident pas nécessairement avec les zones de fortes et de faibles concentrations de l'été et des différences de structure spatiale apparaissent d'une saison à l'autre (Le phénomène de pollution est moins structuré l'été que l'hiver). Le modèle d'une saison ne peut donc pas être déduit du modèle de l'autre saison par un simple recalage.

2.3.2 Réalisation de cartes moyennes

Les deux études se limitent à l'exploitation des données de campagnes, soit à trois ou quatre quinzaines en hiver et à trois ou quatre quinzaines en été. Mais la représentativité saisonnière de ces campagnes n'a pas été démontrée. Comme à Mulhouse la plupart des tubes ont été implantés aux mêmes endroits en été et en hiver, une moyenne de toutes les quinzaines, dite « annuelle », a été calculée aux points où les données des deux saisons étaient disponibles. Mais là encore la représentativité annuelle de cette moyenne n'a pas été établie. Par ailleurs une nette différence de variabilité apparaît suivant les saisons.

Réaliser des cartes de concentrations moyennes annuelles par une modélisation spatio-temporelle, en évaluant la précision de ces concentrations, se révèle une opération complexe. Des données de stations fixes suffisamment nombreuses dans l'espace et le temps sont nécessaires à cette fin. Ce problème n'a pas été abordé. Il constitue à lui seul un sujet d'étude.

2.3.2.1 Méthodologie

Une démarche intéressante et novatrice est toutefois proposée pour établir une carte dite « annuelle » à partir des campagnes saisonnières et en exploitant au mieux l'ensemble des mesures et des variables auxiliaires. Cette méthode consiste à cokriquer la moyenne annuelle² par les moyennes saisonnières, ce qui permet :

- de conserver tous les points de mesure dans l'estimation, y compris ceux pour lesquels seule la valeur estivale ou hivernale est connue ;
- d'assurer une bonne cohérence des estimations saisonnières et annuelles.

Cela suppose de construire un modèle multivariable qui respecte l'égalité :

$$C_{NO2\text{annuelle}} = 1/2 * (C_{NO2\text{hiver}} + C_{NO2\text{été}}).$$

Cette contrainte implique que les variogrammes simples et croisés satisfassent aux relations suivantes :

$$\gamma_{\text{annuel}} = \frac{1}{4}(\gamma_{\text{hiver}} + \gamma_{\text{été}}) + \frac{1}{2}\gamma_{\text{hiver,été}}$$

$$\gamma_{\text{annuel,hiver}} = \frac{1}{2}(\gamma_{\text{hiver}} + \gamma_{\text{hiver,été}})$$

$$\gamma_{\text{annuel,été}} = \frac{1}{2}(\gamma_{\text{été}} + \gamma_{\text{hiver,été}})$$

Des variables auxiliaires peuvent être de plus utilisées en dérive externe dans les modèles saisonniers.

2.3.2.2 Bilan

La version actuelle d'ISATIS ne permet pas de mettre en œuvre la méthode de cokrigeage, telle qu'elle est ici proposée.

Sur les données de Mulhouse, on constate cependant qu'en validation croisée, la moyenne des estimations estivale et hivernale obtenues par krigeage avec dérive externe fournit de meilleurs résultats qu'une estimation réalisée à partir des moyennes « annuelles ». Ce résultat montre tout l'intérêt qu'il y aurait à effectuer un cokrigeage et ouvre des perspectives intéressantes pour établir des cartes annuelles plus précises et plus cohérentes.

² Le terme annuel correspond ici à la moyenne des concentrations mesurées en été et en hiver comme indiqué en introduction de ce chapitre.

La signification et la pertinence de telles cartes, qui font la moyenne de situations saisonnières très contrastées, appellent cependant quelques réserves et demeurent étroitement liées à la stratégie d'échantillonnage adoptée.

3. CARTOGRAPHIE DU DIOXYDE D'AZOTE AU VOISINAGE D'UN AXE ROUTIER

Les données exploitées résultent d'un échantillonnage réparti suivant des sections perpendiculaires à la route. Elles montrent une diminution du NO₂ dans le fond de la vallée et une croissance rapide des concentrations à proximité de la route. Elles révèlent une double non stationnarité qui concerne à la fois la moyenne et la variance.

L'étude réalisée par le Centre de Géostatistique illustre une méthode de cartographie adaptée à ce type de contexte.

3.1 METHODOLOGIE

La modélisation retenue est de la forme dérive+résidu, soit :

$$NO_2(x,y)=m(x,y)+R(x,y)$$

x : distance à la route

y : distance le long de la route

m(x,y) : dérive

R(x,y) : résidu

La cartographie se décompose ainsi en trois grandes étapes :

1. Modélisation de la dérive

Le modèle de dérive est de la forme :

$$m(x,y) = C_0 + \sum_i C_i f_i(x,y)$$

Le modèle de dérive repose sur des fonctions de base qui sont choisies en examinant l'évolution des concentrations sur des sections parallèles et perpendiculaires à la route. Les coefficients multiplicatif et additifs sont calculés par moindres carrés sur l'ensemble des données exploitables. Le modèle est calibré séparément pour chaque saison.

2. Analyse structurale des résidus

Aux points de mesure, la dérive permet de construire des résidus R(x,y), que l'on considère comme la réalisation de variables aléatoires de moyenne stationnaire nulle :

$$R(x_i, y_i) = NO_2(x_i, y_i) - m(x_i, y_i)$$

NO₂(x_i, y_i) : concentration au point de coordonnées (x_i, y_i)

m(x_i, y_i) : dérive calculée au point de coordonnées (x_i, y_i)

Sous réserve qu'ils soient stationnaires d'ordre 2, c'est-à-dire que leur espérance et leur covariance existent et soient stationnaires, ces résidus peuvent faire l'objet d'une analyse géostatistique classique.

Dans le cas particulier étudié, il s'avère que la variance des résidus $\sigma^2(x,y)$ n'est pas stationnaire (la dispersion des résidus augmente avec la dérive). L'analyse se reporte donc sur les résidus normés $R(x,y)/\sigma(x,y)$ où $\sigma(x,y)$ est exprimé comme une fonction de m(x,y).

3. Estimation

Les résidus normés sont estimés dans la zone d'étude par krigeage simple.

En un point (x_0, y_0) , la concentration estimée de NO_2 s'écrit donc :

$$\text{NO}_2^*(x_0, y_0) = m(x_0, y_0) + \sigma(x_0, y_0) \cdot \text{RN}^*(x_0, y_0)$$

$\text{RN}^*(x_0, y_0)$: résidu normé, estimé au point (x_0, y_0)

3.2 BILAN

La méthode proposée pour cartographier la pollution le long d'un axe routier pourrait s'appliquer à d'autres types de sources dont on souhaite évaluer l'impact atmosphérique : aéroport, zone industrielle ou toute autre source bien circonscrite autour de laquelle les concentrations risquent de présenter une non stationnarité spatiale marquée.

Toutefois, du fait de la spécificité de la zone étudiée, elle ne saurait se généraliser sans vérification préalable à des situations autres que celle pour laquelle elle a été mise en œuvre.

4. CONCLUSION

Les études réalisées par le Centre de Géostatistique se sont principalement attachées à la modélisation spatiale du NO₂ à partir de données d'échantillonnage passif. Vu que l'usage de la géostatistique se répand dans les AASQA, il s'est en effet révélé utile de mettre au clair différentes questions relatives à l'introduction de variables auxiliaires et à l'apport de ces variables dans une estimation.

D'un point de vue pratique, il ressort que pour être efficace, une étude avec variables auxiliaires requiert un travail long et approfondi : examen des corrélations entre variables, analyse de la sensibilité de ces corrélations à une modification des points sélectionnés, à l'implantation des données, au type d'environnement ou au support des variables auxiliaires. Sans cette étape préalable on ne peut présumer du gain de précision apporté par les variables supplémentaires.

D'autre part, l'évaluation des incertitudes d'un modèle géostatistique est une exigence majeure auquel la validation croisée, à défaut de disposer d'un jeu de données de validation, offre une réponse partielle. Les résultats de cette validation dépendent toutefois de la répartition des tubes et sont influencés par les zones d'échantillonnage préférentiel. Aussi faut-il s'assurer que les statistiques de validation croisée utilisées pour juger la qualité du modèle soient bien représentatives des zones d'intérêt (où il importe d'avoir les estimations les plus fiables).

La réalisation à partir de campagnes de mesure de cartes moyennes saisonnières ou annuelles compte également parmi les préoccupations des AASQA. L'attention portée à la modélisation avec variables auxiliaires n'a pas permis d'étudier le problème temporel en profondeur. En particulier, aucune modélisation spatio-temporelle faisant usage de données de stations fixes n'a été effectuée. Le travail accompli a néanmoins débouché sur des recommandations méthodologiques intéressantes pour exploiter l'ensemble des mesures disponibles et établir de façon plus précise et plus cohérente des cartes moyennes sur l'année à partir de campagnes estivales et hivernales. (Des différences sensibles observées entre quinzaines de mesure d'une même saison montrent de plus l'intérêt d'étendre ces campagnes sur plusieurs quinzaines).

Enfin, les démarches différentes adoptées pour estimer la pollution d'une agglomération (krigeage ordinaire, krigeage avec dérive externe, cokrigeage) et l'évolution des concentrations de part et d'autre d'une route (modélisation non stationnaire de type dérive + résidu) font voir la nécessité de bien cerner le contexte et le but de la cartographie avant d'engager une étude géostatistique.