O P T E M I

OPTimisation des **É**missions par **M**odélisation **I**nverse

Rapport Final

VERSION PROVISOIRE

2 janvier 2006

Participants au projet :

- Laurent MENUT, coordinateur, IPSL/LMD Palaiseau
- Isabelle PISON, LISA Créteil
- Laurent DEGUILLAUME, LISA Créteil
- Matthias BEEKMANN, LISA Créteil
- Nadège BLOND, LIV Strasbourg

Table des matières

1 Introduction **9**
- 1.1 Motivations . . . 9
- 1.2 Principe général du projet . . . 11
- 1.3 Présentation des résultats dans ce rapport . . . 12
- 1.4 Types de résultats attendus . . . 13
  - 1.4.1 Points communs entre les méthodes . . . 13
  - 1.4.2 Différences entre les méthodes . . . 14

2 Cadre de travail et outils **15**
- 2.1 Le modèle CHIMERE . . . 15
  - 2.1.1 Choix de ce modèle particulier . . . 15
  - 2.1.2 Domaine et météorologie . . . 16
  - 2.1.3 Les émissions . . . 17
  - 2.1.4 Le mécanisme chimique . . . 18
- 2.2 Les outils pilotant le modèle . . . 18
  - 2.2.1 La méthode de krigeage . . . 18
  - 2.2.2 Les méthodes adjointes . . . 19
  - 2.2.3 L'optimiseur non linéaire . . . 20
  - 2.2.4 Le modèle inverse . . . 21
  - 2.2.5 L'analyse de Monte Carlo Bayésienne . . . 21
- 2.3 Les cas d'étude choisis . . . 23
  - 2.3.1 Choix de la région . . . 23
  - 2.3.2 Choix des périodes . . . 23

3 Modélisation inverse par approche adjointe **25**
- 3.1 Généralités . . . 25
- 3.2 Les limites de la modélisation inverse à l'échelle régionale . . . 26
  - 3.2.1 Type de problème étudié . . . 26
  - 3.2.2 Limites dues aux mesures . . . 27
  - 3.2.3 Limites dues au modèle . . . 28
- 3.3 Mise en place de la méthode avec cas académiques . . . 29
  - 3.3.1 Principe de base d'un cas académique . . . 29
  - 3.3.2 Deux sources et une ou deux stations de mesure . . . 30
- 3.4 Ajout de contraintes par krigeage . . . 32
  - 3.4.1 Comparaisons mesures et krigeage . . . 32

3.4.2 Mesures, background et krigeage . . . 34
3.4.3 Krigeage itératif . . . 35
3.4.4 Applications du krigeage avec des mesures d'ozone . . . 36
3.5 Agrégation spatiale . . . 36
3.5.1 Principe général . . . 36
3.5.2 Choix méthodologique . . . 37
3.5.3 Construction d'un modèle dynamique de zonage . . . 37
3.6 Cas réels : ESQUIF POI2, le 7 Août 1998 en Ile de France . . . 40
3.6.1 Les émissions à optimiser . . . 40
3.6.2 Tests de sensibilité aux paramètres météorologiques . . . 41
3.6.3 Résultats moyens sur le domaine . . . 42
3.6.4 Résultats sur quelques sites . . . 43
3.6.5 Impact de l'optimisation sur le panache d'ozone . . . 45
3.7 Cas réels : ESQUIF POI6, du 16 au 18 Juillet 1999 en Ile de France . . . 46
3.7.1 Simulations directes . . . 46
3.7.2 Corrections apportées aux paramètres d'entrée . . . 47
3.7.3 Emissions optimisées . . . 48
3.7.4 Conclusion sur les POIs ESQUIF . . . 49
3.8 Inversions climatologiques . . . 49
3.8.1 Variabilité entre les deux mois d'été . . . 50
3.8.2 Variabilité pour les jours de semaine . . . 50
3.8.3 Variabilité pour les jours de week-end . . . 50
3.8.4 Variabilité intra-type de jour . . . 52
3.8.5 Variabilité spatiale . . . 53
3.8.6 Variabilité jour à jour . . . 54
3.8.7 Synthèse des résultats climatologiques . . . 55

4 Modélisation Monte-Carlo sous contrainte **57**
4.1 Adaptation et mise en oeuvre de la méthode . . . 57
4.1.1 Attribution des distributions de probabilité à priori . . . 57
4.1.2 Estimation des incertitudes : . . . 60
4.2 Résultats . . . 61
4.2.1 Distribution à posteriori des paramètres d'entrée du modèle : Études des POIs 62
4.2.2 Interprétation des résultats : confrontation modèle/observations . . . 65
4.2.3 Etude de la période climatologique : juillet et août 1999 . . . 67
4.2.4 Conclusion sur l'étude climatologique . . . 69

5 Conclusions et perspectives **71**
5.1 Conclusion . . . 71
5.2 Perspectives . . . 72

Annexes **73**

A Principes théoriques des outils utilisés **75**
A.1 Principes théoriques de l'adjoint 75
A.2 Principes théoriques du krigeage 77
A.3 Définitions des fonctions liées à l'inverse par adjoint 78
A.3.1 Principe : imbrication des modules de calcul 78
A.3.2 La fonction coût 79
A.3.3 Le vecteur d'état 80
A.3.4 La matrice $\boldsymbol{R}$ 80
A.3.5 La matrice $\boldsymbol{B}$ 80
A.3.6 Le critère d'arrêt de l'optimisation 81
A.3.7 Les indicateurs statistiques 82

B Définition du cas académique **83**

Résumé court :

Le projet OPTEMI visait **à mettre en place et évaluer des techniques de modélisation inverse afin de mieux quantifier les flux d'émissions surfaciques employés par les modèles de chimie-transport**. Pour cela, deux approches différentes mais complémentaires ont été baties : (i) une méthode d'optimisation basée sur le modèle adjoint, (ii) une méthode Monte-Carlo basée sur des perturbations aléatoires du modèle direct.
Afin d'évaluer au mieux la qualité des résultats obtenus, une stratégie de développement pas à pas a été réalisée : les optimisations ont donc été mises en place depuis des cas académiques simples jusqu'à des cas réels complexes. Les limites de ce type de méthode, directement liées à la quantité et qualité d'information disponible comme contrainte, ont ainsi été quantifiées. Ainsi, il a été montré que des résultats réalistes pouvcaient être obtenus en utilisant l'information sous forme de krigeage des mesures et de zonage dynamique des flux à estimer.
Le domaine d'étude choisi pour les cas réels est la région parisienne. deux types de péridoes ont é'e étudiées : (i) des cas "extrèmes" de pollution atmosphérique les POIs 2 et 6 de la campagne ESQUIF, (ii) une climatologie régionale avec l'inversion de l'ensemble des étés 1998 et 1999.
Pour la région parisienne étudiée, nous avons montré que les émissions disponibles devaient avoir un pic de trafic matinal plus réduit en temps dans Paris, mais des flux plus intenses en banlieue. Enfin, des informations supplémentaires ont été apportées sur la différenciation des jours de semaine et de week-end dansles cadastres. L'ensemble des modifications à apporter, calculé par l'inversion adjointe, est dans une gamme de $\pm 15\%$ du cadastres actuel, valeur confirmée par l'approche Monte-Carlo.

Résume long :

Le projet OPTEMI visait **à mettre en place et évaluer des techniques de modélisation inverse afin de mieux quantifier les flux d'émissions surfaciques employés par les modèles de chimie-transport** (tel que CHIMERE). Pour cela, deux approches différentes mais complémentaires ont été baties : (i) une méthode d'optimisation basée sur le modèle adjoint, (ii) une méthode Monte-Carlo basée sur des perturbations aléatoires du modèle direct.
Dans les deux cas, des méthodologies originales ont été élaborées :

- Pour l'**approche adjointe**, l'approche est basée sur le krigeage des contraintes (les mesures) et un zonage dynamique des flux à inverser (les émissions surfaciques). Cette nouvelle méthode a été validée sur des cas académiques simples, à même de quantifier les limites de la méthode.
- Pour l'**approche Monte-Carlo**, une approche originale basée sur le calcul de probabilités de réalisation, permet d'obtenir une correction mais aussi l'incertitude associée. Les simulation sont aussi porté sur les étés 1998 et 1999.

Une recherche bibliographique a mis en avant le manque de validation des travaux précédemment réalisés dans le domaine. La plupart des auteurs présentent des études directement sur des cas réels complexes, sans estimation au préalable de la validité de leur approche. Nous avons donc choisi de développer graduellement notre méthodologie en mettant en place des cas académiques à complexité croissante pour l'adjoint. Cela nous a permis de quantifier des limites réelles de l'approche. Les méthodologies développées ont ensuite été appliquées sur des cas réels : tout d'abord des cas "extrèmes" correspondants à des périodes intensives du projet ESQUIF (POI2, Août 1998 et POI6, Juillet 1999), puis une climatologie sur les étés 1998 et 1999 en région parisienne.
Avec ces deux approches, l'information obtenue est donc des flux d'émissions corrigés en temps et en espace à la résolution du modèle d'accueil, ainsi que l'incertitude associé à ces flux. Sachant qu'au départ on ne dispose que de masses moyennes émises, par secteurs d'activité, cet ajout d'information reste précieux pour la modélisation de chimie-transport régionale.
Nous avons d'abord cherché à **calculer de manière précise les limitations de nos méthodes** :

- Afin de pouvoir réellement jauger des flux d'émissions, il fallait s'affranchir un minimum des erreurs liées aux autres paramètres. Des études préliminaires ont donc été menées afin de quantifier des biais sur les concentrations liés à la météorologie (dont hauteur de la couche limite et mélange vertical) et aux conditions aux limites. Cela a permis d'assurer une meilleure précision sur les flux inversés ainsi que de meilleures contraintes sur la méthode Monte-Carlo. De plus, nous avons envisagé la possibilité de ne pas prendre en compte des processus existants : nous avons donc développé la prise en compte des émissions tri-dimensionnelles liées aux décollages et aterrissages des avions de ligne en région parisienne. L'impact modéré ne pouvant expliquer les différences modèle/mesures, nous avons procédé aux inversions sans cet ajout.
- Pour les deux méthodes, les limites de leur application ont été clairement quantifiées : la méthode adjointe ne peut inverser des émissions sans un minimum de renseignements sur les sources étudiées. Ainsi, il est plus bénéfique d'utiliser le NO comme contrainte plutot que l'ozone si l'on veut estimer des flux de NO émis. De même pour la méthode Monte-Carlo, un minimum d'information étant nécessaire pour construire des distributions, il faut beaucoup de contraintes pour retrouver un facteur correctif fiable.

Connaissant ces limitations, nous avons pu dégager, de l'ensemble des simulations réalisées, des **résultats augmentant notre connaissance de la variabilité spatio-temporelle des émissions**, tels

que :

- En région parisienne et sur les étés considérés, les émissions utilisées par le modèle correspondent à un cadastre de bonne qualité. Si on pouvait le soupconner au vu des résultats de prévision, on peut aujourd'hui le quantifier. Hormis une variabilité intra-semaine pas assez bien prise en compte (notamment les Mercredi et Vendredi pour le trafic), les valeurs moyennes des flux de NO (principale source) sont correctes, et la variabilité spatiale adaptée à la région. la méthode Monte-Carlo a de plus, montré que le rapport COVs/NOx était lui aussi correct que ce soit à Paris même que dans le panache d'ozone.
- La tendance générale est à une réduction des émissions liées au trafic dans Paris. Cela se traduit surtout par un "pic" temporel d'émission moins étalé dans le temps le matin. A l'opposé, on note une légère tendance à augmenter les émissions en périphérie.
- Temporelllement, l'inversion montre que l'on sous-estime les émissions les matins des jours de semaine et Dimanche. Ils sont, par contre, surestimés le Samedi. Ces variations sont dans une gamme de $\pm 15\%$, valeur confirmée par l'approche Monte-Carlo.

Chapitre 1

Introduction

1.1 Motivations

Ce document est le rapport final du projet **OPTEMI** (pour **OPT**imisation des **E**missions par **M**odélisation **I**nverse) qui a été accepté en Décembre 2002. Nous présentons tout d'abord les principaux buts du projet, en rappelant le contexte et cadre de travail de l'époque. Puis, chapitre par chapitre, nous présentons la méthodologie développée et les principaux résultats obtenus. Une dernière partie, de conclusions et perspectives, présente une analyse critique de ce qu'il reste à réaliser dans ce domaine de recherche qu'est l'optimisation de paramètres.
Ces dernières années, différents projets visant à la mise au point d'outils de modélisation pour la pollution photochimique ont été développés, dont le modèle de chimie-transport CHIMERE. L'ensemble de ces outils avait pour objectif le développement d'une meilleure description des épisodes de pollution atmosphérique. Dans ce cadre, deux buts principaux ont été identifiés : (1) réaliser des scénarii réalistes pour estimer comment réduire au mieux la pollution, (2) disposer d'outils permettant d'effectuer des prévisions.
Ces outils nécessitent de nombreux paramètres d'entrée. Or, que ce soit en modélisation directe ou adjointe, que ce soit via les études de scénarii Monte-Carlo, les résultats convergent vers un point important : au sein d'un modèle de chimie-transport, les **émissions surfaciques** sont à la fois un **paramètre d'entrée indispensable** mais **très incertain**, auxquelles les concentrations finalement simulées sont **très sensibles**.
Il faut donc, pour espérer mieux simuler ces champs de concentrations, améliorer notre connaissance sur ces émissions. Ce constat n'est pas un cas particulier des études réalisées dans le cadre de la pollution au sein de la couche limite à l'échelle régionale. Il s'applique aussi à tous types de modélisation impliquant ce type de sources et la prise en compte d'une chimie atmosphérique complexe : de l'échelle régionale à l'échelle globale, que ce soit pour la troposphère ou la stratosphère.
Les émissions peuvent être décomposées en deux familles principales :

- Les **émissions anthropiques** : En milieu urbain, elles sont principalement issues des transports et des industries. Ce sont les sources d'émissions majoritaires (en intensité) même si elles sont relativement limitées spatialement. Actuellement, d'importants efforts sont réalisés pour mieux quantifier ces émissions, par définition hautement variables en temps (le trafic, par exemple, n'est pas un processus régulier) et en espace (de nombreuses sources sont très localisées)
- Les **émissions biogéniques** (ou naturelles) : En dehors de milieux fortement urbanisés, les émissions naturelles deviennent la principale source à prendre en compte si l'on veut estimer correc-

tement des champs de concentrations de polluants comme l'ozone et les oxydes d'azote.

Ces sources sont relativement mal connues, et ce, pour des raisons différentes :

- **Les émissions anthropiques :**
  - Sources majoritaires de polluants primaires, ces émissions sont particulièrement intenses mais relativement limitées spatialement : en milieu urbain, il faut recenser de nombreuses sources très localisées.
  - Ces émissions sont hautement variables en temps : le trafic, par exemple, ne peut être considéré comme une source constante d'un jour à l'autre, même si l'on sait estimer des profils diurnes types.
  - Certaines de ces sources n'ont pas encore fait l'objet d'études spécifiques. Le projet OPTEMI initialement déposé, citait comme exemple le trafic aérien à proximité des aéroports. Nous avons entre-temps étudié cet aspect et les résultats ont fait l'objet d'une publication reprise en **Annexe ??, p.??** et référencée [*Pison and Menut* (2004)].
- **Les émissions naturelles :**
  - L'identification du rôle non négligeable de ces émissions au sein des processus de pollution photo-oxydante est récente. L'incertitude sur les totaux annuels d'une base de données d'émissions d'isoprène et de terpènes à l'échelle de l'Europe est estimée à un facteur 2 à 3, [*Simpson* (1999)]. Cette grande incertitude est dûe à la grande variabilité des facteurs d'émissions d'une espèce à l'autre voire même d'une plante à l'autre. De plus, la variation des ces facteurs d'émissions en fonction de facteurs climatiques (température moyenne, pluviométrie, ...) reste encore mal comprise. En ce qui concerne les émissions des composés azotés par les sols, les processus fondamentaux microbiologiques et leur dépendance à l'état du sol ne sont pas encore bien comprises. Il n'existe donc pas à ce jour de base de données d'émissions naturelles, prévue pour une modélisation fine à l'échelle régionale ; La prise en compte de ces émissions, en modélisation, repose sur des travaux théoriques eux aussi très récents.
  - Les espèces chimiques émises ont souvent des temps de vie assez courts ce qui rend les mesures spatialement peu représentatives et l'évaluation de bases de données d'émissions difficiles. D'autres espèces, tels que les carbonylés sont difficiles à mesurer. De plus, isoler dans la concentration mesurée, la part issue de l'émission reste délicat. Il n'existe pas encore de structure, à plusieurs niveaux de réflexion, basée à la fois sur l'instrumentation et la modélisation pour quantifier la part due aux émissions.

Pour mieux comprendre la mise en place de la pollution, il est indispensable de bien connaitre à la fois les sources anthropiques et biogéniques. Il faut aussi réduire l'incertitude que l'on a sur leur détermination. Pour cela, on peut utiliser deux moyens, en étudiant les **méthodes de constitution** des cadastres, ou le **cadastre résultant en lui-même**. Revoir les méthodes de construction à la base semble irréaliste, car :

- La composition des "cadastres" d'émissions est complexe : les données "brutes" permettant de les constituer sont souvent incomplètes et insuffisantes. Hétérogènes en temps et en espace, il est difficile de regrouper ces nombreuses informations en un ensemble cohérent.
- En milieu urbain, les émissions seront en grande partie issues d'activités anthropiques comme la circulation automobile, les petites industries. Très variables en temps et en espace, il est difficile de les estimer et, surtout, d'en faire le suivi.

- Les émissions anthropiques issues de l'industrie sont estimées statistiquement à partir des activités industrielles déclarées. Ces déclarations ne sont pas de manière sûre exactes et/ou à jour.

Il s'avère donc quasiment impossible de disposer de données d'émissions à la résolution spatiale et temporelle voulue et à jour par ce biais. L'étude va donc porter sur la **modification du cadastre déjà constitué**.
Pour affiner ce cadastre, il est nécessaire d'ajouter de l'information, des contraintes au système de modélisation. Ces contraintes existent : les mesures chimiques de surface effectuées soit par les réseaux de surveillance de la qualité de l'air (comme AIRPARIF) soit, ponctuellement, lors de campagnes de mesure. Il faut donc **assimiler ces données**. Cette technique algorithmique est déjà largement employée dans de nombreux domaines, et dans le cadre de l'atmosphère, en dynamique. Utiliser une donnée et la comparer à un résultat de simulation pour optimiser un paramètre d'entrée du système est appelée **modélisation inverse**.
Ce projet se place donc dans un cadre très spécifique : celui de la modélisation. Dans la suite de ce rapport, nous n'aborderons pas l'aspect "réel" des émissions. Ce que nous cherchons à savoir ici, c'est si les outils de modélisation que nous avons développés sont capables de nous dire vers quel secteur d'activité réel il est le plus habile de s'orienter si l'on veut réduire la pollution. Mais la réponse restera toujours dépendante de la qualité du modèle employé, c'est à dire de sa capacité à se rapprocher de ce que l'on cherche finalement à décrire.

1.2 Principe général du projet

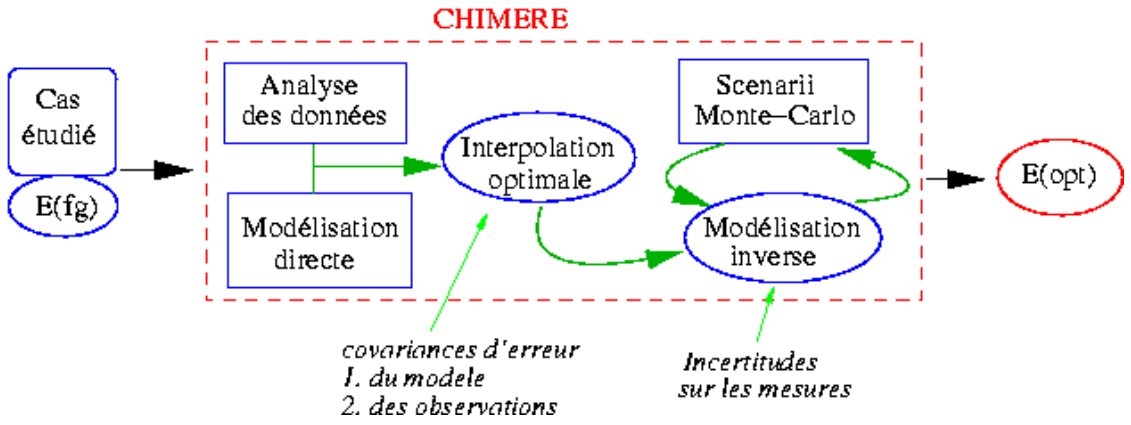


FIG. 1.1 – *Schéma de principe du passage des émissions dite de première ébauche $E_{(fg)}$ "first guess" aux émissions optimisées $E_{(opt)}$*

Le principe général du projet est relativement simple. Pour avoir une meilleure précision sur les émissions, il faut **ajouter de l'information sur leur valeur, variabilité et incertitude**. Sachant qu'il est impossible d'en obtenir directement sur les émissions elles-mêmes, il faut utiliser d'autres formes d'informations. D'autre part, nous voulons une information qui permette, avant tout, d'**affiner la variabilité temporelle des flux** : les seules informations de ce type qui sont disponibles sont les mesures de concentrations de polluants à la surface. Tout le principe du projet repose donc sur la question :

Comment utiliser au mieux toute l'information disponible dans les concentrations mesurées pour mieux contraindre les émissions utilisées dans les modèles ?
Le principe général du projet est synthétisé sur la Figure 1.1. Ce schéma représente la méthodologie telle qu'elle a été initialement proposée dans le projet et telle qu'elle a été développée ces trois dernières années. Nous reprenons ici les principales articulations méthodologiques. Ces points seront développés bien plus en détail dans les chapitres suivants.
Le point de départ est un "cas étudié". Cette dénomination n'est restrictive ni en terme de lieu (taille de la région ou de la ville), ni de temps (durée du cas), ni de composé étudié (une ou plusieurs espèces). L'unique information est que l'on dispose d'émissions connues notées E(fg) (avec "fg" pour *first-guess*, la première ébauche). Cette information est "habituellement" utilisé comme forcage d'un modèle de chimie-transport. On obtient alors un champs de concentrations. Si l'on combine ce champs de concentrations modèle avec des mesures, afin de minimiser cette différence (le contenu de la "boite" notée CHIMERE), on peut arriver à déterminer une émission plus proche de ce qu'il aurait fallu avoir dès le départ et alors notée E(opt) (avec "opt" pour optimisé). La Figure 1.1 indique principalement deux grandes phases :

- La combinaison de mesures avec des simulations directes permet d'obtenir des cartes par interpolation optimale. Ces cartes peuvent servir de contraintes à la modélisation inverse (i.e ici au sens de minimisation avec un adjoint)
- L'utilisation alternée de modélisation inverse et de scénario Monte-Carlo peut permettre d'optimiser les émissions tout en réduisant leur incertitude.

Il est à noter que quelques ajustements méthodologiques ont été réalisés pendant la réalisation du projet, ce dernier restant de la recherche et donc évoluant en fonction des résultats obtenus. On notera que le changement le plus important reste le remplacement de l'interpolation optimale par le krigeage. Ce choix particulier sera explicité en détail dans le chapitre 3.

1.3 Présentation des résultats dans ce rapport

Ce rapport est divisé en plusieurs grandes parties :

- une présentation des résultats liés à la modélisation inverse. Ce chapitre reprend en grande partie le travail de thèse réalisé par Isabelle PISON sur ce thème (soutenue le 7 Décembre 2005 à l'Université Paris12), encadrée par Laurent MENUT.
- Une présentation des résultats liés aux approches Monte-Carlo. Ce travail reprend les recherches de post-doctorat réalisées par Laurent DEGUILLAUME, encadré par Matthias BEEKMANN. Ce post-doctorat correspond au principal financement de ce projet.
- En fin de document, les Annexes reprennent certains des travaux publiés dans le cadre de cette étude :
  - ☐ Deguillaume, L. and Beekmann, M. : *Inversion of pollutant emissions via Bayesian Monte Carlo Analysis*, (en préparation)
  - ☐ Beekmann M. and Derognat, C., *Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign*, JGR, 108-D17, doi : 10.1029/2003JD003391, 2003.
  - ☐ Menut L., I. Coll and S. Cautenet, *Impact of meteorological data resolution on the forecasted ozone concentrations during the ESCOMPTE IOP 2a and 2b*, Atmospheric Research - ESCOMPTE Special Issue, 74, 139-159, 2005

- ☐ Menut L., *Adjoint modelling for atmospheric pollution processes sensitivity at regional scale during the ESQUIF IOP2*, Journal of Geophysical Research - Atmospheres, 108, D17, 2003 (repris en **Annexe ??, p.??**).
- ☐ Pison I. and L.Menut, *Quantification of the impact of aircraft traffic emissions on tropospheric ozone over Paris area*, Atmospheric Env., 38-7, 971-983, 2004 (repris en **Annexe ??, p.??**).
- ☐ Pison I., L.Menut and N.Blond, *Inverse modeling of emissions for local photo-oxidant pollution : Testing a new methodology with kriging constraints*, Annales Geophysicae, sub-judice, (repris en **Annexe ??, p.??**).

1.4 Types de résultats attendus

L'utilisation en parallèle de la modélisation inverse par adjoint et par Monte-Carlo va apporter des réponses différentes et complémentaires. Dans cette section, nous décrivons la nature de ce que nous allons obtenir de manière concrète comme résultats. Nous commencons tout d'abord par aborder les *points communs* entre les deux méthodes, pour terminer par les *aspects différents*. L'ensmeble de ces informations sont synthétisées dans le Tableau 1.1.

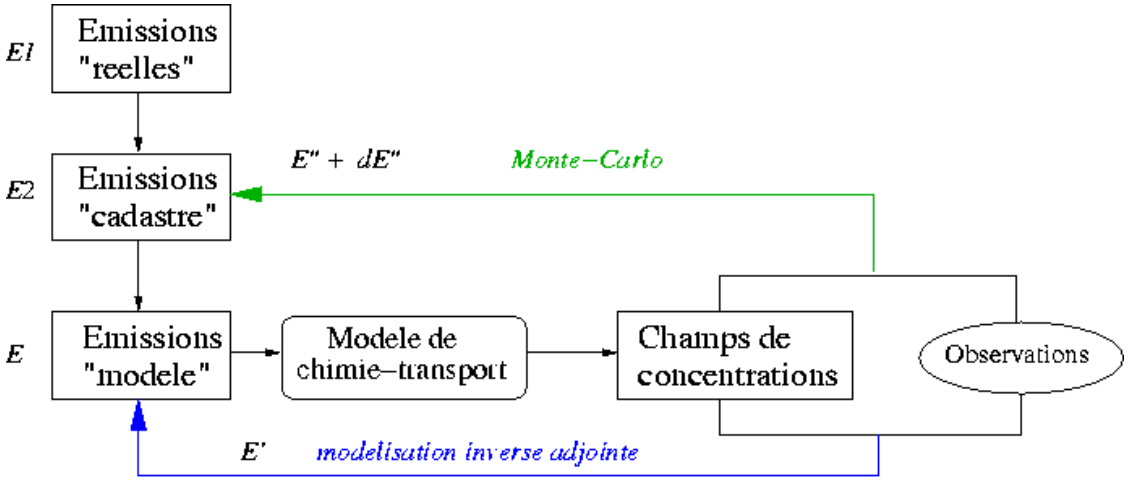


FIG. 1.2 – *Schéma des différents degrés d'informations sur les émissions et des apports respectifs des deux méthodes de modélisation : inverse para djoint et Monte-Carlo.*

1.4.1 Points communs entre les méthodes

Le premier aspect tout à fait commun est la définition que l'on peut donner de la variable que l'on recherche : E la valeur à un instant donné, en un lieu donné et pour une espèce gazeuse chimique modèle donnée d'un flux surfacique d'émissions (Figure 1.2). C'est une valuer de flux liée au modèle utilisé et au mécanisme chimique associé. On connait donc cette grandeur car on l'a construite pour nos besoins de modélisation : c'est donc une grandeur uniquement virtuelle, celle qui correspond à ce qu'il y'a **de plus adapté pour simuler les champs de concentrations que l'on cherche à représenter**. Elle a été élaborée à partir des émissions "cadastre", E_2 (Figure 1.2) qui correspondent, en général à des "tonnes/an" pour un secteur d'activité (comme le trafic, les industries) ou pour des

familles d'espèces (les NOx ou les COVs, par exemple). E_2 est lui-même approximé à partir de comptages, de statistiques, de "vraies" mesures de terrains compilées sous forme de bases de données : c'est l'approche dite "bottom-up" : lourde et imprécise mais correspondante au seul lien avec la réalité de ces grandeurs physiques E_1.
Lorsqu'on utilise E avec un modèle de chimie-transport, on obtient des champs de concentrations. On peut comparer ces champs à des observations et jauger de la différence entre les deux. Pour réduire cette différence, on peut tenter d'améliorer les valeurs que l'on a construit de ces émissions.
C'est le principal point commun entre les deux méthodes : à partir des émissions modèles, utiles au calcul de champs de concentrations, on va disposer d'informations sur E en proposant des corrections sur cette valeur en elle-même.

1.4.2 Différences entre les méthodes

La modification proposée sur E ne sera pas de même nature. La modélisation inverse par adjoint proposera des "corrections" pour chacun des flux utilisés par le modèle : on aura ainsi des facteurs correctifs pour chaque espèce modèle optimisée (en l'occurence NO), à chaque heure et dans chaque maille horizontale du modèle ou ces flux sont représentés. Ce changement correspondra à un cas particulier, sans renseignement sur l'incertitude associée au flux initial. La modélisation Monte-Carlo va, par contre, apport cette information sur l'incertitude liée au flux d'émission étudié (noté dE'' Figure 1.2), ainsi qu'une modification de la valeur du flux elle-même (noté E''). Par contre, et afin d'avoir une representativité statistique suffisante, cette information devra être intégrée en temps et en espace (sur tout le domaine d'étude).

Information	Méthode adjointe	Méthode Monte-Carlo
Spatiale	Chaque maille	moyenne sur le domaine
Temporelle	Chaque heure	moyenne temporelle
Espèce chimique	Espèce modèle	Espèce modèle
Incertitude	non	oui

TAB. 1.1 – *Synthèse des informations obtenues avec les deux approches*

Chapitre 2

Cadre de travail et outils

Dans ce chapitre, nous présentons succinctement les outils que nous avons choisi pour le projet, qu'ils soient déjà disponibles ou qu'ils aient été développés à cette occasion. Puis nous décrivons dans quelle configuration ils ont été utilisés. Les chapitres suivants ne seront alors consacrés qu'aux résultats du projet. De manière résumée, ces différents outils se déclinent selon qu'ils ont été développés en dehors du groupe de travail "OPTEMI", ou par le groupe (mais hors projet) ou spécifiquement pour le projet. On a donc :

1. Le modèle de chimie-transport CHIMERE ... existant et développé par OPTEMI (et d'autres)
2. La modélisation inverse par adjoint, avec :
   - Le modèle adjoint ..... développé par et pour OPTEMI
   - le krigeage ..... développé par OPTEMI
   - le minimiseur non-linéaire ..... développé par l'INRIA
   - Le modèle inverse ..... développé par et pour OPTEMI
3. L'analyse de Monte-Carlo Bayésienne ..... développé par et pour OPTEMI

Le modèle de chimie-transport CHIMERE est le "moteur" central de l'étude. Les autres composantes "outils" sont principalement des outils de pilotage de ce noyau, développés précisement pour obtenir les résultats proposés pour le projet.

2.1 Le modèle CHIMERE

2.1.1 Choix de ce modèle particulier

Afin d'optimiser les émissions surfaciques sur une région, à des fins de modélisation, le choix du modèle est un élément important des résultats attendus. Le modèle CHIMERE a été choisi pour différentes raisons :

- Il reproduit tous les phénomènes physiques et chimiques d'intérêt aux échelle spatiales (quelques kilomètres) et temporelles (l'heure) considérées ;
- Il est assez optimisé pour permettre, avec une bonne précision sur les concentrations calculées, de réaliser de nombreuses simulations en un temps réaliste : point vital lorsque l'on veut mettre en place des processus convergents à base de modèle direct et adjoint ou réaliser des centaines de simulations Monte-Carlo.

- Il a fait ses preuves sur le type de problème étudié. De nombreux travaux publiés ont pu montrer sa pertinence ; notamment en région parisienne avec la campagne ESQUIF, la prévision quotidienne mise en place avec AIRPARIF (la première prévision déterministe de la pollution à l'échelle régionale en France).
- Enfin, les proposants de ce projet font partie de l'équipe de développement du modèle, et donc à ce titre, connaissent assez bien l'outil pour que son emploi et ses particularités ne soient pas des facteurs limitants.

2.1.2 Domaine et météorologie

Le modèle CHIMERE utilisé dans le cadre de ce projet est la version V200211, mise à disposition en Novembre 2002. C'est cette version qui a été utilisée pour la prévision notamment dans le cadre de PREVAIR. Les versions suivantes voient principalement l'ajout du calcul des aérosols : la version utilisée ici est donc celle calculant encore aujourd'hui les processus gazeux. Cette version correspond aussi aux derniers développements de l'adjoint complet du modèle. Une description complète de cet outil peut-être trouvée sur son site internet : **http ://euler.lmd.polytechnique.fr/chimere/**.

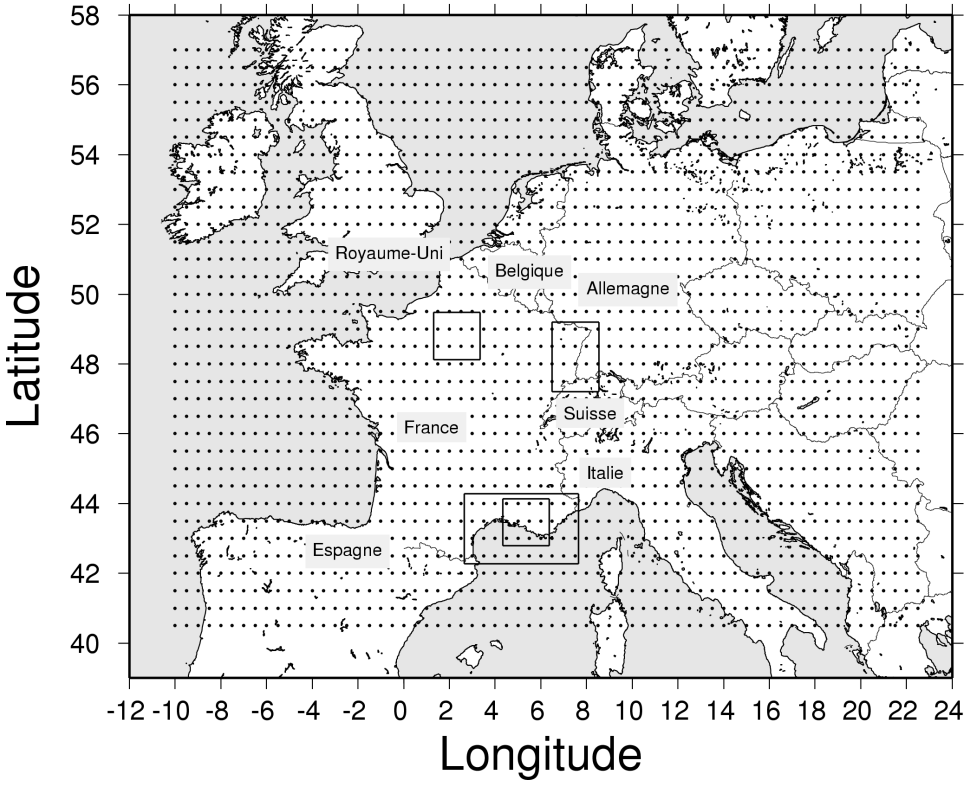


FIG. 2.1 – *Domaines de simulation définis pour les différentes simulations réalisées avec CHIMERE. Le grand domaine Européen permet des calculs avec une résolution spatiale de 0.5×0.5 degrés. Ils permettent de fournir des conditions aux limites chimiques horaires aux domaines de plus petite taille, avec une résolution de 6km, comme les domaines ESQUIF et ESCOMPTE.*

Le principe de base de CHIMERE-gaz est de transporter et de faire réagir chimiquement des espèces de gaz trace dans l'atmosphère. La météorologie doit être prescrite et donc calculée en amont. Pour ce

travail, elle est issue des données du Centre Européen (ECMWF), réinterpolées sur la grille régionale. Les conditions aux limites chimiques sont horaires et précalculées avec le même modèle mais utilisé à l'échelle continentale, comme cela est illustré sur la Figure 2.1. La grille de calcul est celle utilisée en prévision composée de huit niveaux verticaux, de la surface à 750hPa (afin de couvrir largement toute l'épaisseur de la couche limite).

2.1.3 Les émissions

Lors de l'écriture du projet, les deux campagnes de mesures ESQUIF et ESCOMPTE voyaient leurs phases expérimentales terminées. Dans le cas d'ESQUIF, de nombreux travaux de modélisation avaient aussi été menés à bien. Un cadastre était donc à la disposition de la communauté scientifique, proposé par AIRPARIF, représentant les flux d'émissions au mieux des connaissances en 1998. Dans le cas d'ESCOMPTE, une partie non négligeable du projet était de créer un inventaire d'émissions à jour, servant de référence. Les délais de réalisation de ce cadastre, ainsi que des difficultés de validation n'ont permis de le rendre utilisable que très récemment. Cela explique que, pour cette étude, nous nous soyons en priorité focalisés sur la région parisienne.
Ces émissions ont été élaborées par AIRPARIF pour la campagne ESQUIF sur la base du mois de Juillet 1998 à l'aide des modules "émissions" de SIMPAR (Système Informatique de Modélisation de la Pollution Atmosphérique à l'échelle Régionale). Plus de détails sont donnés dans [*Vautard et al.* (2003a)].
Le cadastre est constitué de flux surfaciques d'émissions de monoxyde de carbone, dioxyde de carbone, dioxyde de soufre, oxydes d'azote, COV non méthaniques et méthane pour trois types de jours correspondant aux jours de semaine, au samedi et au dimanche (ou jour férié) à une résolution spatiale de 3 km (sur le domaine de 150 km $\times$ 150 km représentant l'Ile de France) avec un pas horaire d'une heure.
Les masses totales annuelles sont calculées en prenant en compte les onze secteurs d'activité définis par la nomenclature SNAP (Selected Nomenclature for sources of Air Pollution). En pratique, les sources sont divisées en trois catégories pour chacune desquelles un mode de calcul adpaté est appliqué :

1. les grandes sources ponctuelles telles que cheminées industrielles, centrales thermiques, incinérateurs, soit en tout deux cent trois en Ile de France. Les émissions sont calculées à partir des données collectées par la DRIRE (Direction Régionale de l'Industrie, de la Recherche et de l'Environnement) d'Ile de France en 1998. Les profils temporels permettant de passer des masses totales annuelles à la résolution horaire sont tirés du programme GENEMIS (comme à l'échelle continentale) et des déclarations des industries.
2. les sources linéiques (grands axes routiers, autoroutes). Elles sont calculées en combinant quatre types d'information :
   - les données sur le réseau routier et le trafic, fournies par la DREIF (Direction Régionale de l'Équipement en Ile de France) et qui concernent les types de voies (voies urbaines, routes et autoroutes), leur débit, la vitesse du trafic, le pourcentage de véhicules roulant "à froid" pour les trente cinq mille brins du réseau.
   - la composition du trafic c'est-à-dire les pourcentages de chaque catégorie de véhicules (véhicules particuliers, deux roues, véhicules utilitaires légers, poids lourds, bus et cars) circulant sur les trois types de voies selon l'heure de la journée.

- les facteurs d'émissions pour ces véhicules, obtenus d'après la méthode COPERT III (Computer Program to Calculate Emissions from Road Transport)
- la température extérieure, qui influence les émissions à froid et les pertes par évaporation.

Les profils temporels diurnes sont calculés par le modèle de trafic mais une modulation, obtenue à partir des mesures d'une station "trafic" du réseau AIRPARIF, est appliquée pour différencier les trois types de jours.

3. les sources surfaciques c'est-à-dire celles qui ne font pas partie des deux premières catégories comme le chauffage domestique, le trafic diffus, les petites sources industrielles. Les données du CITEPA (Centre Interprofessionnel Technique d'Études de la Pollution Atmophérique) pour 1994, fournies pour des arrondissements ou départements, sont combinées à l'occupation des sols indiquée par l'IFEN (Institut Français de l'ENvironnement) afin d'obtenir la résolution spatiale voulue. Les profils temporels proviennent également de GENEMIS.

La spéciation des COV en espèces-modèle est effectuée pour chaque secteur SNAP d'après [*Theloke et al.* (2000)],[*Theloke et al.* (2001)],[*Schmitz et al.* (2000)],[*Rudd and Marlowe* (1998)],[*Hassel et al.* (2000)]. Les oxydes d'azote sont répartis en 10% de dioxyde d'azote et 90% de monoxyde d'azote. Dans le cadastre obtenu, le trafic routier est la principale source de NO_x, avec 53% des émissions totales. Les solvants et le trafic sont les principales sources de COV (36 et 35% respectivement).

2.1.4 Le mécanisme chimique

Le mécanisme chimique est tout aussi important comme choix de calcul que la facon dont les émissions ont été élaborées. CHIMERE peut travailler avec divers mécanismes chimiques. En prévision, une version réduite du mécanisme chimique MELCHIOR (Modèle d'Étude Lagrangienne de la CHImie de l'Ozone à l'échelle Régionale) est utilisée. Elle comporte 40 espèces et 100 réactions (80 espèces chimiques et 200 réactions pour le mécanisme complet) [*Lattuati* (1997)]. Le rayonnement est pris en compte grâce à des taux de photolyse tabulés, valables par ciel clair [*Madronich and Flocke* (1998)]. Un modèle statistique prenant en compte la nébulosité et le profil d'humidité dans les couches basses permet de paramétrer l'atténuation du rayonnement due aux nuages [*ESQUIF* (2001)].
L'ensemble des autres caractéristiques de CHIMERE sont décrites de manière exhaustive dans les différentés publications citées au début de ce rapport ; ainsi que dans la documentation du modèle disponible à l'adresse **http ://euler.lmd.polytechnique.fr/chimere/**.

2.2 Les outils pilotant le modèle

Nous présentons dans ce paragraphe les différents outils que nous avons développés et utilisés pour cette étude. Ce ne sont pas en eux-mêmes des modèles mais plutôt des outils mathématiques qui vont piloter les différents calculs du modèle dans un cadre particulier.

2.2.1 La méthode de krigeage

la méthode de krigeage employée dans ce projet est une dérivation d'une technique appelée 'interpolation optimale'. Le but est de tirer le maximum des informations disponibles sous forme de mesures afin de compenser l'aspect sous-contraint du problème. On ne peut donc pas véritablement créer de l'information, mais on va par contre mieux la répartir. Pour cela, on ne va pas uniquement interpoler

les données disponibles mais les redistribuer en tenant compte de la physique du problème via l'utilisation du modèle lui-même. Le krigeage s'applique donc sur des mesures ponctuellement distribuées en espace et sur des champs de concentrations simulés. Le résultat est alors une carte dite "analysée". L'hypothèse sous-jacente à cette approche est que le modèle utilisé sait déjà assez bien reproduire les grandes tendances des structures des champs de polluants sur la zone d'étude.
Dans le cadre de ce travail, on peut faire cette hypothèse avec une certaine confiance, le modèle CHIMERE ayant été déjà largement validé sur des cas d'études particuliers ; et étant validé quotidiennement dans le cadre de la prévision PREVAIR **http ://www.prevair.org/fr/**. Un rappel théorique du krigeage est reporté en **Annexe A.2, p.77**.

2.2.2 Les méthodes adjointes

Le principe de base des méthodes adjointes est illustré sur la Figure 2.2. Le **modèle direct** calcule à chaque pas de temps une nouvelle concentration, dépendante de sa valeur au pas de temps précédent et de tous les processus l'affectant au pas de temps de calcul. Le **modèle adjoint** est le gradient du modèle direct : plus précisement, il permet de calculer la dérivée d'une quantité simulée à tous les paramètres d'entrée. C'est donc un second modèle, différent du premier, mais qui doit reprendre en tout point les paramétrisations et l'architecture de calcul du premier.

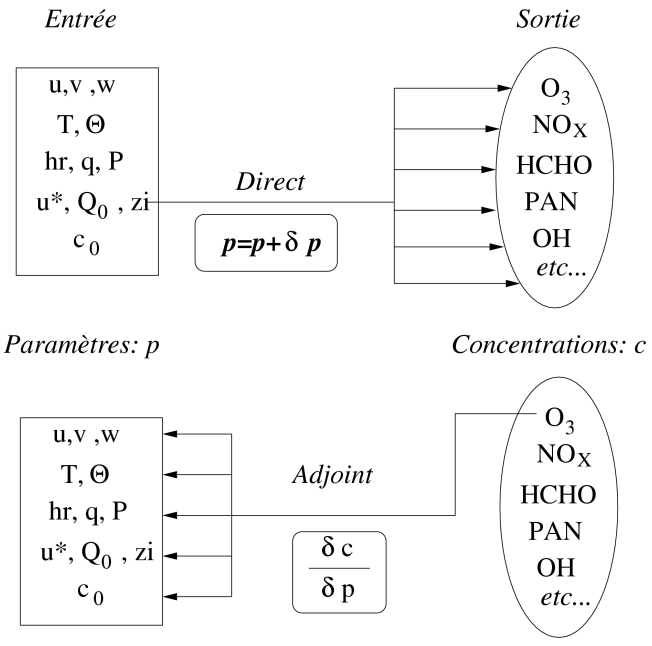


FIG. 2.2 – *Principe schématique de l'adjoint d'un modèle de chimie-transport.*

Comme cela est présenté sur la Figure 2.2 :

1. Le modèle direct considère les paramètres de météorologie, émissions et autres comme des données d'entrée : on calcule via le module de chimie-transport, la concentration chimique (O_3, NO_x et autres) d'un pas de temps au pas de temps suivant. Si l'on veut réaliser une étude de scénario, on peut modifier un ou plusieurs paramètres d'entrée : cette modification va cependant affecter toutes les concentrations en sortie.

2. Au contraire, le modèle adjoint va calculer la dépendance d'une unique espèce en sortie à tous les paramètres d'entrée. Cela se fait en une seule intégration du modèle. On est donc capable de quantifier, pour une espèce donnée, quel a été le paramètre le plus important, où, quand et pourquoi.

Le modèle adjoint permet donc d'estimer pour une concentration donnée, l'impact de la variation d'un des paramètres. Ce contrôle permet alors de contraindre la trajectoire du modèle dans cet espace des paramètres.

Pour la modélisation inverse, l'avantage est donc de pouvoir connaitre au même instant, tous les paramètres auxquels la concentration étudié est sensible. Par contre, la limitation est de ne pas pouvoir "trop" s'écarter de la valeur initiale. "trop" est un concept flou, assez dépendant du problème étudié. Dans le cas de la pollution photo-oxydante à l'échelle régionale, [*Menut* (2003)] a montré que l'on pouvait raisonnablement modifier les paramètres d'entrée de l'ordre de leur incertitude (de 10 à 30%) sans modifier la "direction" de variation de la concentration étudiée.

D'un point de vue plus mathématique, l'adjoint a pour but de calculer le gradient d'une fonction scalaire des paramètres (voir **Annexe A,p.75** pour le cadre mathématique). Le modèle adjoint est donc le gradient du modèle direct. Les équations adjointes sont obtenues à partir du modèle linéaire tangent qui est la linéarisation du modèle direct (non-linéaire) autour des variables directes d'entrée. L'opérateur d'adjonction est défini de telle sorte qu'il est équivalent à une transposition si on utilise le produit scalaire canonique. À partir du modèle linéaire tangent, une simple transposition des lignes en colonne (ainsi que l'inversion de l'ordre d'actualisation des variables) permet d'obtenir les équations adjointes, les variables directes d'entrée devenant les variables adjointes de sortie.

En pratique, pour construire l'adjoint d'un modèle, il faut donc d'abord écrire le modèle linéaire-tangent puis l'adjointiser. Il existe des programmes qui effectuent cette tâche automatiquement comme Odyssée à l'INRIA, FastOpt (Allemagne). L'adjoint de CHIMERE v200211 a cependant été réalisé "à la main" afin d'obtenir une programmation permettant d'optimiser le temps de calcul (la structure même de calcul du modèle ne permet pas, en effet, une adjointisation "au fil du programme"). De plus, une intégration adjointe étant environ trois fois plus longue qu'une intégration directe, une attention toute particulière a été apportée à ce développement et à son optimisation. Enfin, il est à noter que ce développement a été réalisé spécifiquement afin de mener à bien ce projet particulier de modélisation inverse OPTEMI.

2.2.3 L'optimiseur non linéaire

Le minimiseur utilisé est N1QN3 [*Gilbert and Lemaréchal* (1989)], dans la version décrite plus récemment par [*Gilbert and Lemaréchal* (1995)]. Il utilise une méthode de type quasi-Newton à mémoire limitée pour résoudre les problèmes de minimisation sans contrainte. Son principe consiste en la détermination à chaque itération d'une direction de descente et d'un pas adapté le long de cette direction. Celle-ci est calculée en utilisant une approximation du hessien et le gradient de la fonction coût au point courant. Il est à noter que pour cette étude, cet outil a été utilisé tel qu'il avait été développé par ses concepteurs. Les détails plus techniques sont décrits de manière exhaustive dans [*Pison* (2005)].

2.2.4 Le modèle inverse

Finalement, l'ensemble constitué par le modèle direct CHIMERE, son adjoint, le minimiseur non-linéaire est encapsulé dans le modèle dit **modèle inverse**. Ce programme va piloter un ensemble complexe composé de modèles mais aussi de gestion de lecture des mesures disponibles. Cette structure est schématisée sur la Figure 2.3.

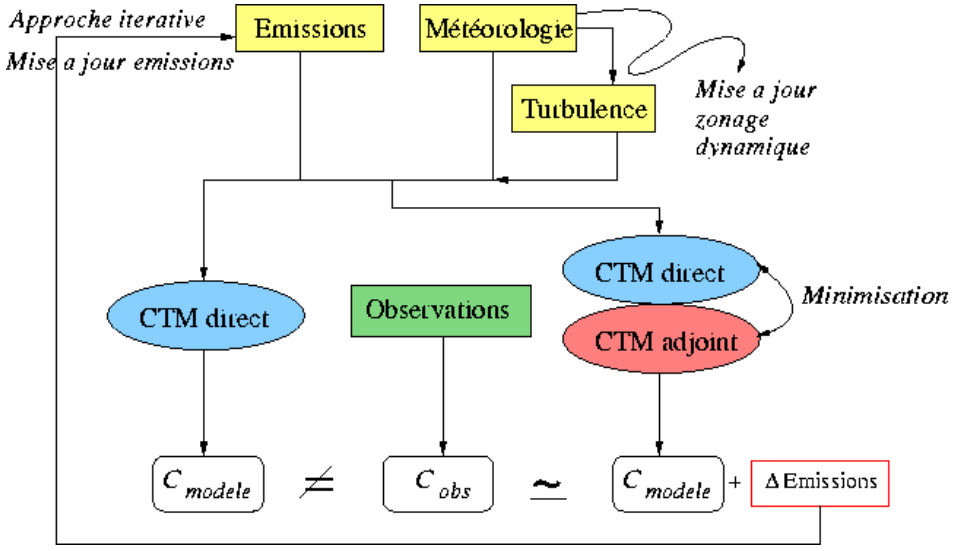


FIG. 2.3 – *Modélisation inverse pour l'estimation des sources - Principe*

Par principe, on réalise tout d'abord une simulation directe afin de jauger de la différence entre mesure (si disponible) et modèle. Le schéma représente ce processus à chaque heure et en chaque point de grille. La simulation adjointe donne la sensibilité des espèces calculées aux émissions, et permet d'estimer précisément une fonction coût. Cette fonction sert au minimiseur non-linéaire à apprécier dans quelle "direction" il faut modifier les émissions pour réduire la différence modèle/mesures. L'ajout de la technique de krigeage modifie la structure des mesures, le zonage modifie la structure des émissions et l'approche itérative crée un effet de boucle sur l'ensemble du processus.

2.2.5 L'analyse de Monte Carlo Bayésienne

L'analyse de Monte Carlo Bayésienne est différente de ce qui a été précédemment décrit et est complémentaire. On va dans ce cas, obtenir une information plus moyenne (en temps et en espace) mais avec le gain de l'incertitude associée.
Plusieurs auteurs [*Gao et al.* (1996)], [*Hanna et al.* (1998)], [*Bergin and Millford* (2000)] ont mis en évidence de larges incertitudes des modèles de chimie et de transport par rapport à leur données d'entrées (émissions, variables météorologiques, paramètres cinétiques des réactions chimiques et photolytiques). Les méthodes de Monte Carlo permettent une évaluation globale de l'incertitude des modèles par rapport à l'ensemble des variables d'entrées ou de paramétrisations d'un modèle (e.g. [*Hanna et al.* (2001)]). Ces méthodes consistent à répéter plusieurs centaines de simulations, mais à chaque fois avec des ensembles de données d'entrée différentes, choisies aléatoirement dans leur domaine d'incertitude. Une extension de cette méthode, appelée "analyse de Monte Carlo Bayésienne",

introduit un forcage par les observations permettant d'estimer le réalisme de chaque simulation de Monte Carlo en fonction de l'accord entre sorties de modèles et observations [*Dilks et al.* (1992)] ; [*Brand and Small* (1995)]. En d'autres termes, la méthode consiste à déterminer des termes de probabilité $p(O|Y_k)$ qui expriment la chance de réaliser un vecteur d'observations O étant donné un vecteur de simulations Y_k, k désignant une simulation de Monte Carlo individuelle. En faisant l'hypothèse d'une distribution normale de l'erreur associée aux observations et d'une indépendance entre les différents éléments O_i du vecteur O, la probabilité $p(O|Y_k)$ peut être évaluée comme suit :

$$p(O|Y_k) = \Pi_{j=1,N} \left(\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_{e,j}} \times exp \left(-0.5 \left[\frac{O_j - Y_{k,j}}{\sigma_{e,j}} \right]^2 \right) \right) \tag{2.1}$$

Dans l'équation 2.1, l'écart entre simulations et observations est en effet pondéré par les incertitudes $\sigma_{e,j}$ associées aux observations, qu'il faut savoir estimer. Par ailleurs, l'équation 2.1 s'applique indépendamment de la durée de la simulation. Il faut simplement veiller à une mise en oeuvre cohérente de la méthode, i.e. fixer les erreurs sur les observations et les variables d'entrée représentatives pour l'échelle de temps choisie (i.e. une heure, un jour, quelques mois).
L'équation 2.1 permet d'obtenir des distributions de probabilité à posteriori de variables prédictives du modèle, en pondérant chaque valeur C par la probabilité $p(O|Y_k)$. De même, pour chaque variable d'entrée e du modèle, telle que les émissions, une distribution de leur probabilité à posteriori peut être obtenue en pondérant chaque valeur issue de la distribution à priori par $p(O|Y_k)$:

$$p(e1 \leq e \leq e2) = \frac{1}{C} \int_{e_1}^{e_2} p(O|Y_k) de \tag{2.2}$$

Dans l'équation 2.2, la probabilité p (e1 $\leq$ e $\leq$ e2) de trouver une valeur e entre e_1 et e_2 est donnée par l'intégrale sur les probabilités $p(O|Y_k)$, associées aux simulations de Monte Carlo ayant des émissions entre e_1 et e_2. C représente un facteur de normalisation pour assurer que l'intégrale totale des probabilités donne l'unité.

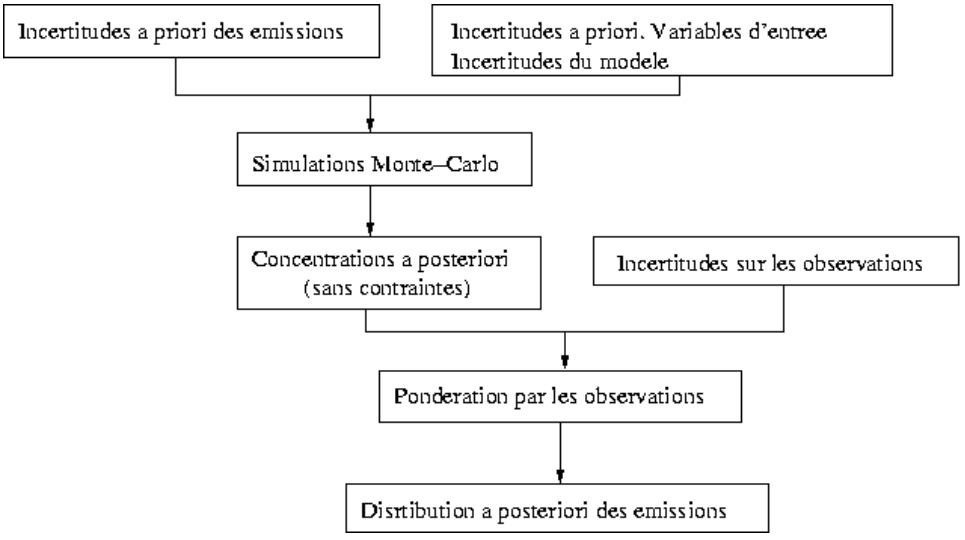


FIG. 2.4 – *Principe général de la méthode Monte-Carlo*

Dans ce sens, la méthode Monte Carlo Bayésienne constitue une méthode de modélisation inverse des émissions, complémentaire à la méthode adjointe présentée en amont. Notamment, elle fournit l'incertitude l'estimation à posteriori des émissions, comme décrite ci-dessus, Figure 2.4.
Par rapport à la méthode de modélisation inverse des émissions avec l'adjoint de CHIMERE, cette méthode permet donc, en plus, de fournir l'incertitude, en tenant compte de leur incertitude à priori, de l'incertitude des observations et de l'incertitude des autres variables d'entrée du modèle. Par exemple, l'incertitude sur la dispersion verticale et horizontale des polluants est prise en compte en fixant une distribution de probabilité sur la hauteur de la couche limite convective, sur le coefficient de mélange vertical et sur la vitesse et la direction du vent. Par contre, le coût de calcul des simulations de Monte Carlo à effectuer limite le nombre de variables d'entrée à perturber. Ainsi, on doit se limiter dans ce travail à dériver un facteur correctif moyenné sur la durée et le domaine de la simulation.

2.3 Les cas d'étude choisis

2.3.1 Choix de la région

Pour réaliser cette étude, deux régions particulières étaient envisagées : la région Parisienne et la région de Marseille Fos-Berre. D'une part car ce sont les deux plus grandes agglomérations urbaines Francaises : elles sont donc pourvues d'un réseau de surveillance de la qualité de l'air conséquent. D'autre part car elles avaient été le siège de deux récentes campagnes de mesures sur la pollution atmosphérique régionale, ESQUIF en 1998-1999 et ESCOMPTE en 2001.
Mais entre ces deux campagnes, des différences d'avancement des travaux scientifiques sont à noter :

- Alors que le cadastre des émissions sur Paris existait avant la campagne ESQUIF, celui d'ESCOMPTE a été crée après la période de mesures (des retards ont vu ce cadastre n'être livré à la communauté scientifique qu'en 2003). Il est encore aujourd'hui en cours d'évaluation, notamment dans le cadre de l'exercice d'inter-comparaison des modèles. Des erreurs importantes n'en ont fait que très récemment un outil stable (donc largement après le début du projet OPTEMI). Au contraire, le cadastre sur la région Parisienne était déjà existant et utilisé en 1998. Il a été revu en 2000, puis encore affiné par le réseau AIRPARIF dans le cadre du projet HEAVEN. On peut donc directement valider l'intérêt de notre méthodologie par rapport aux améliorations bottom-up faites en parallèle.
- En région parisienne, la topographie est à faible relief. Cela permet de pouvoir considérer que parmi tous les processus incertains à modéliser, l'écoulement atmosphérique n'est pas le point le plus délicat sur cette région. Ce n'est pas le cas de la région de Marseille, ou les effets de relief, mais aussi de bord de mer (avec des brises) font de la météorologie une première étape de compréhension importante et délicate, avant de pouvoir s'attaquer à d'autres processus comme les émissions.

Pour ces deux raisons, il a été décidé de commencer le travail par la région Parisienne.

2.3.2 Choix des périodes

Pour les périodes plus précises d'études, il a alors été logiquement choisi les périodes d'observations intensives de la campagne ESQUIF. Les POI 2 et 6 ont été retenues pour leurs configurations météorologiques très marquées et représentatives de l'ensmeble de ce que l'on obsreve en général sur cette région. Ces caractéristiques ont naturellement fait que ce sont les périodes qui ont été les plus

étudiées. On dispose donc d'une bonne connaissance des processus de pollution atmosphérique qui ont été relevés pendant la campagne.
On notera que toutes les informations concernant la campagne ESQUIF sont disponibles soit dans le rapport final de la campagne (en francais) à cette adresse : **http ://clim-serv.ipsl.polytechnique.fr/esquif/**, soient dans les articles consacrés à ce projet dans la section spéciale de *Journal of Geophysical Research*[1].
Parmi les 12 périodes d'observations intensives réalisées de 1998 à 2000, nous nous focaliserons sur les POI2 et POI6. Leurs principales caratéristiques sont détaillées dans le Tableau 2.1.

POI	**Dates**	**Origine de la masse d'air**	**Remarques** (sur toute la POI)
Été 1998			
1	25 → 26/07/1998	Continentale 1-2 jours de stagnation	Période nuageuse (moyens et hauts), vent faible
2	07 → 09/08/1998	Continentale 2-5 jours de stagnation	Ciel clair, fortes températures et vent faible
Été 1999			
3	16 → 17/06/1999	Continentale 3-4 jours de stagnation	Vent dominant de Nord-Est < 3 m.s^{-1}
4	25 → 26/06/1999	Continentale 3-4 jours de stagnation	Ciel clair, vent venant souvent de l'Est ≈ 5 m.s^{-1}
5	01 → 02/07/1999	Maritime	Ciel clair, vent de Sud-Est≈ 4 m.s^{-1}
6	16 → 18/07/1999	Maritime puis stagnant	Ciel clair
7	24 → 25/07/1999	Continentale 2-3 jours de stagnation	Expérience de suivi du panache parisien au-dessus de la Bretagne. Période ventée.
8	28 → 31/07/1999	Continentale 3-4 jours de stagnation	Vent très faible < 2 m.s^{-1}
Hiver 1999-2000			
9	07 → 08/10/1999	Continentale 2-3 jours de stagnation	Conditions anticycloniques. Expéreinces pour la qualification de la conversion NO→NO_2
10	26 → 28/01/2000	Continentale	Expérience pour l'observation de situations à très fortes [NO_x]. Panache parisien vers l'Ouest.
Été 2000			
11	19 → 20/07/2000	Épisode photochimique modéré sans stagnation continentale	Expérience sur les aérosols
12	30 → 31/07/2000	Conditions sèches, pas de stagnation	Expérience sur les aérosols

TAB. 2.1 – *Récapitulatif des douze POI d'ESQUIF. La colonne "origine de la masse d'air" indique l'origine continentale ou maritime de celle-ci. Les "remarques" mettent en relief la spécificité de chacune des situations météorologiques.*

[1]Liste complète des articles sur http ://euler.lmd.polytechnique.fr/menut/esquif/esquif2jgr.html

Chapitre 3

Modélisation inverse par approche adjointe

Nous avons vu au chapitre précédent les différents buts recherchés ainsi que les outils à mettre en oeuvre afin de les atteindre. Dans ce chapitre, nous décrivons plus précisément l'adaptation de la méthode générale telle qu'elle a été faite avec CHIMERE sur la région parisienne et avec le modèle adjoint.

3.1 Généralités

Le but des méthodes inverses est d'estimer, à partir d'observations, les valeurs des paramètres du modèle permettant de retrouver les prévisions souhaitées En pratique, les méthodes inverses consistent à minimiser la différence entre l'observation et la prévision pour une variable (ici les concentrations de polluants) afin de mieux estimer certains paramètres (les émissions par exemple) dont dépend celle-ci. La méthode d'inversion modifie les valeurs de départ des paramètres, dites *a priori* (ou "first-guess") pour obtenir les valeurs **optimisées**. On appelle **dimension du problème** (inverse) le nombre de paramètres à inverser. Dans le cas des émissions par exemple, chaque flux de chaque espèce émise constitue un paramètre d'entrée. Enfin, on appelle **contraintes** les observations c'est-à-dire les valeurs que les prévisions doivent approcher. Des mesures de concentration sont classiquement utilisées comme contraintes pour l'inversion des émissions.
La modélisation inverse fait donc partie de la famille des techniques d'hybridation mesures/modèle ou assimilation de données. Le but commun de ces techniques est de définir au mieux l'état d'un système en utilisant le maximum d'informations. Il faut noter cependant que le but de l'inversion est différent de celui de l'assimilation de données qui consiste à intégrer des informations telles que des observations pour améliorer l'évolution du modèle (voir [*Talagrand* (1997)],[*Daley* (1997)] pour un exposé général et [*Elbern et al.* (1997)] pour l'assimilation de données chimiques dans la troposphère).
Les premières applications des techniques d'hybridation mesures/modèle avec des modèles déterministes eulériens ont consisté en le développement de l'assimilation de données pour l'océanologie et la météorologie (voir [*Daley* (1996)] pour une revue historique). Il s'agissait donc de traiter de gandes échelles spatiales sur de longues périodes (de type "climatologies"). L'application de ces techniques à des problèmes concernant les gaz-trace de l'atmosphère est plus récente et porte la trace de cette histoire.
Les méthodes inverses ont principalement été utilisées à l'échelle globale (couvrant toute la planète) ou continentale. Les espèces d'intérêt à cette échelle spatiale sont celles qui ont des temps de vie longs

comme le méthane [*Hein et al.* (1997); *Houweling et al.* (1999); *Wang and Bentley* (2002); *Bergamaschi et al.* (2000)], le dioxyde de carbone [*Bousquet et al.* (1999a); *Rodenbeck et al.* (2003); *Kaminski and Heimann* (2001)], les CFC [*Hartley and Prinn* (1993); *Mahowald et al.* (1997)] . En plus de l'inversion des sources d'émissions, certaines études se préoccupent des puits comme [*Bousquet et al.* (1999a)] pour le dioxyde de carbone, voire du bilan génŕal de l'espèce comme [*Jacob et al.* (2002)] pour l'acétone. Des cas-tests permettent d'aborder le problème de la qualité et de l'éventuelle amélioration du réseau de mesure pour l'inversion [*Hartley and Prinn* (1993); *Mahowald et al.* (1997)]. La question de la variabilité temporelle des sources est particulièrement détaillée par [*Haas-Laursen et al.* (1996)] et [*Hein et al.* (1997)], [*Bousquet et al.* (1999a)] traitant une année représentative de la climatologie de toute une décennie. Les résultats obtenus consistent en de nouvelles valeurs pour les sources bien sûr mais aussi en une réduction de l'incertitude sur celles-ci qui peut atteindre 75% pour un hémisphère entier [*Houweling et al.* (1999)] ou de 25 à 80% pour un continent [*Wang and Bentley* (2002)]. Les références citées ici ne forment pas une liste exhaustive des études effectuées à l'échelle globale. De nombreuses équipes travaillent actuellement sur l'inversion des émissions des gaz à effet de serre comme le dioxyde de carbone.

A l'échelle régionale, les études inverses sont moins nombreuses. On peut citer principalement des études de acs précises sur Atlanta, [*Chang et al.* (1996)] et [*Chang et al.* (1997)], avec des inversions d'isoprène et de monoxyde de carbone. L'inversion des émissions de monoxyde de carbone par [*Mulholland and Seinfeld* (1995)] à Los Angeles conduit les auteurs à relever la prépondérance des corrections dues aux sources mobiles, notamment le trafic routier. Plus récemment, [*Mendoza-Dominguez and Russell* (2001)] se sont intéressés aux précurseurs de l'ozone que sont les COV et les NO_x. À partir de la campagne menée à Atlanta, les émissions ont été ajustées en distinguant les contributions des sources anthropiques mobiles, ponctuelles, ou surfaciques et biogéniques. Le principal objectif de [*Elbern and Schmidt* (1999)] était l'assimilation de mesures d'ozone. Cependant, comme le cadastre disponible n'était pas à jour, la méthode a été adaptée pour optimiser les émissions des précurseurs de l'ozone. Bien que le problème soit mal posé, les résultats sur les concentrations d'ozone sont améliorés de façon significative avec les émissions optimisées. Les études de [*Elbern et al.* (2000)] sur des cas académiques ont montré qu'il est possible d'optimiser les émissions de précurseurs de l'ozone dans des sites urbains complexes à condition de disposer d'un grand nombre de contraintes. Par ailleurs, l'étude conclut à une importante sous-estimation des COV tant en milieu urbain qu'autour de la ville, ce qui rejoint les résultats de [*Chang et al.* (1996)] pour l'isoprène.

3.2 Les limites de la modélisation inverse à l'échelle régionale

3.2.1 Type de problème étudié

Les techniques d'assimilation de données furent originellement développées pour traiter des problèmes météorologiques, qui sont des problèmes de **conditions initiales**. En effet, les équations régissant l'écoulement atmosphérique étant fortement non-linéaires, la qualité des conditions initiales à une influence déterminante sur celle des résultats de l'intégration complète du modèle.

Les problèmes de chimie-transport sont beaucoup moins non-linéaires. Ils sont du type **conditions aux limites**. Les principaux paramètres dont la qualité détermine celle des sorties du modèle sont en effet les sources surfaciques et, dans le cas de la pollution à l'échelle régionale, les concentrations aux limites du domaine étudié, qui peuvent jouer un rôle prépondérant lors de certains épisodes (voir

par exemple [*Menut et al.* (2000b)] pour le cas de l'Île-de-France). Les théories développées pour la météorologie et l'étude d'espèces chimiques à long temps de vie à l'échelle globale ne peuvent être appliquées directement au cas de la pollution à l'échelle régionale.
Nous nous proposons donc de répertorier les limitations auxquelles nous allons nous heurter lors de l'élaboration d'une méthode d'inversion des émissions à l'échelle régionale. Elles sont principalement dues aux incertitudes et aux erreurs des mesures, du modèle et de ses données d'entrée.

3.2.2 Limites dues aux mesures

Toutes les informations nécessaires ne sont jamais disponibles dans un cas réel, ceci pour des raisons variées. Les informations fournies par les mesures ne sont pas assez nombreuses, ne sont pas adaptées à tous les problèmes, intègrent des effets différents, sont bruitées et enfin, ne sont pas indépendantes les unes des autres.

- **Quantité :** Lorsque l'on souhaite utiliser les mesures de routine de réseaux, la **quantité de mesures disponibles** est bien inférieure à la dimension du problème qui est alors dit **sous-contraint**. Par exemple en Île-de-France, il y a environ 40 stations Airparif mesurant les concentrations d'ozone. Le cadastre *a priori* est constitué de 625 flux surfaciques par espèce émise. Le rapport entre le nombre de contraintes et la dimension du problème pour l'inversion d'une seule espèce est donc d'environ 5%. Il faut aussi noter que les mesures de surface ne fournissent aucune information sur la verticale.
- **Pertinence :** La **pertinence de l'information disponible** constitue un deuxième obstacle. Il n'est pas possible d'inverser des émissions qui n'ont pas d'influence sur les mesures disponibles puisqu'on ne dispose alors d'aucune information. Par exemple, inverser des émissions de HONO avec des mesures de NO_y peut ne rien donner parce que le HONO peut varier du simple au double sans que son influence dépasse l'ordre du pourcent sur les concentrations mesurées. Ceci est aussi valable si l'on considère l'évolution temporelle des émissions et des concentrations. En effet, les émissions auxquelles les concentrations sont les plus sensibles peuvent être celles d'heures bien antérieures. C'est le cas par exemple des concentrations d'ozone lors du pic de l'après-midi qui sont très sensibles aux émissions de NO_x du pic de trafic le matin et peu sensibles à celles de l'heure précédant le pic.
- **Qualité :** La **qualité de l'information** disponible peut n'être pas assez bonne. L'information fournie par une mesure intègre de nombreux effets différents. La concentration mesurée est un bilan de l'influence de plusieurs sources. Il n'est pas possible a priori de séparer les différentes contributions. Lorsque les effets de plusieurs sources se superposent en un même point de mesure, il n'est pas possible de retrouver les informations concernant chaque source sans d'autres mesures pour lesquelles les effets des sources sont découplés. Même pour une mesure située idéalement dans le panache d'une unique source, il peut être impossible de distinguer entre un mauvais placement de la source et une amplitude mal estimée.
- **Indépendance :** La question de l'**indépendance des mesures** est d'autant plus importante que l'échelle est petite. Pour des études à l'échelle globale [*Bousquet et al.* (1999a,b)], on peut considérer qu'une mesure de CO_2 sur un continent est indépendante d'une mesure effectuée sur un autre continent. Mais plus l'échelle est petite, plus les mesures peuvent être liées. À l'échelle régionale, cette information est contenue dans des covariances. Or, si les variances sont (habituellement) fournies avec les mesures, les covariances ne le sont pas.

- **Erreurs instrumentales :** Enfin, les **instruments de mesure** eux-mêmes ne sont pas parfaits, les mesures comportent incertitudes et erreurs. L'optimisation peut alors aboutir à des corrections qui transfèrent les erreurs de mesures sur les émissions.

3.2.3 Limites dues au modèle

Les erreurs dues au modèle peuvent provenir des données d'entrée ou du modèle direct lui-même.

- **Paramètres d'entrée :** La méthode d'inversion choisie suppose que tout l'écart entre les mesures et les valeurs simulées est dû aux émissions à inverser. Or, **les valeurs d' autres paramètres d'entrée** du modèle direct contribuent bien sûr à cet écart, notamment les données météorologiques. Une erreur sur la direction du vent, par exemple, peut décaler le panache d'une source par rapport aux mesures ; la température a aussi un rôle important dans la chimie photo-oxydante [*Menut* (2003)]. Les émissions inversées risquent d'être sur-corrigées pour compenser les erreurs sur les autres paramètres d'entrée en même temps. **Les paramètres météorologiques pris en compte dans le modèle** par le biais de paramétrisations peuvent poser problème. Même si la direction et la vitesse du vent sont correctes, le schéma de transport qui les utilise peut avoir une précision insuffisante. À petite échelle, la diffusion numérique joue un rôle crucial dans l'obtention d'un panache d'ozone de taille et de forme réaliste par exemple. Ce type de problème est particulièrement difficile à diagnostiquer puisque les paramétrisations utilisées dans les modèles sont en général les meilleures disponibles.
- **Qualité des paramétrisations :** Pour la modélisation chimie-transport, les deux paramètres les plus importants sont (i) la vitesse de friction et (ii) la hauteur de couche-limite (HCL) [*Menut* (2003)], [*Hanna et al.* (1998)], [*Hanna et al.* (2001)]. La vitesse de friction u_* est utilisée directement dans le calcul du dépôt sec des espèces. À l'échelle régionale, la sensibilité des polluants à u_* n'est pas négligeable sans être la principale. La hauteur de couche-limite, $\overline{h}$, est également un paramètre diagnostiqué à partir de certains paramètres : sa valeur, déduite des profils verticaux de vent et de température, comporte une grande incertitude. Malheureusement, elle est directement liée à l'impact des émissions surfaciques sur les concentrations de polluants : une mauvaise estimation de la hauteur de couche-limite peut amener à une trop grande concentration ou dilution des émissions. Par exemple, une surestimation des concentrations d'ozone peut être interprétée comme due à des émissions de COV ou de NO_x (selon le régime chimique [*Honoré and Vautard* (2000)]) trop importantes ou à une faible sous-estimation de $\overline{h}$.
- **Exhaustivité des processus représentés :** Il est par définition difficile de se rendre compte qu'un **processus important manque** dans un modèle. Des simulations à long terme et des comparaisons avec de nombreuses données permettent en général de détecter un tel manque.
- **Représentativité modèle/mesure :**
  À l'échelle régionale, au même endroit, des mesures d'O_3 et de NO_x n'ont pas la même représentativité spatiale. L'ozone est un polluant secondaire, produit loin des sources après plusieurs heures. La diffusion naturelle fait donc qu'une mesure d'ozone a une représentativité qui la rend comparable à une maille de modèle à cette échelle, c'est-à-dire 3 à 6 km. Par contre, les NO_x sont plutôt représentatifs de l'environnement de leur source, c'est-à-dire quelques centaines de mètres : une échelle inférieure à une taille de maille de modèle régional.

3.3 Mise en place de la méthode avec cas académiques

3.3.1 Principe de base d'un cas académique

Les paragraphes précédents ont montré que de nombreuses incertitudes et méconnaisssances pouvaient interférer et nuire aux résultats que l'on espère obtenir. La première étape consiste donc à jauger des poids relatifs de l'ensemble de ces incertitudes en environnement complètement contrôlé. Il a donc **été décidé de commencer l'étude par des simulations sur des cas académiques**. Cela a aussi l'avantage de tester si l'ensemble de la méthode qui a été mise en place fonctionne d'un point de vue informatique.

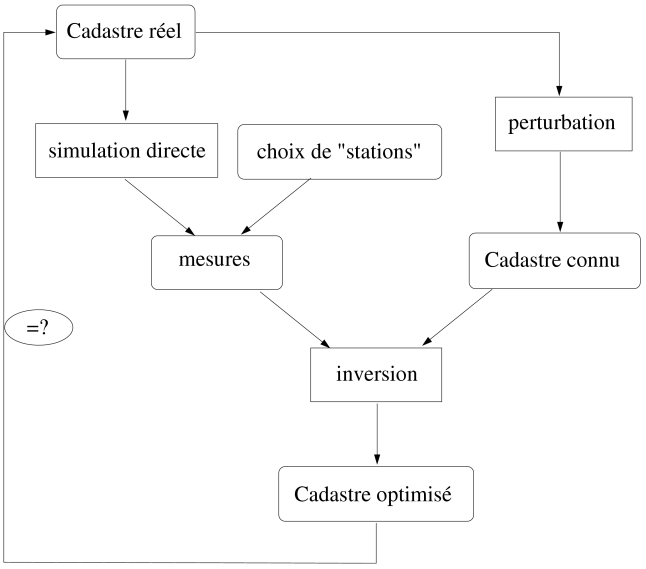


FIG. 3.1 – *Principe d'élaboration d'un cas académique*

On crée donc un système dont on maitrise tous les élements pouvant générer une erreur. On est alors certain que la différence entre les émissions de "première ébauche" et celles optimisées correspond vraiment à la contrainte apportée au modèle entre la mesure et le calcul d'un polluant. On liste dans la suite les paramètres, autres que les émissions, pouvant perturber le système et les choix faits pour éviter des erreurs :

- La météorologie (vent, température, humidité, mélange vertical turbulent, hauteur de la couche limite). On l'impose, la plus simple possible tout en restant réaliste. On crée donc des cas de vent ne changeant pas de direction, avec des gradients de température imposés permettant de cerner au mieux le mélange vertical et la stabilité atmosphérique
- Le mécanisme chimique, notamment les taux de réactions, mais aussi l'atténuation dûe aux nuages. On peut supprimer cette notion d'espèces réactives dans un premier temps et chercher à optimiser une espèce trace émise.
- Le dépôt : ce processus n'étant pas un puits majoritaire, il peut simplement ne pas être considéré en première approximation.
- L'erreur modèle : Même si les paramètres sont imposés, les paramétrisations peuvent fausser leur prise en compte. Pour éviter cette erreur modèle, il faut travailler en "expériences jumelles", tel

que :

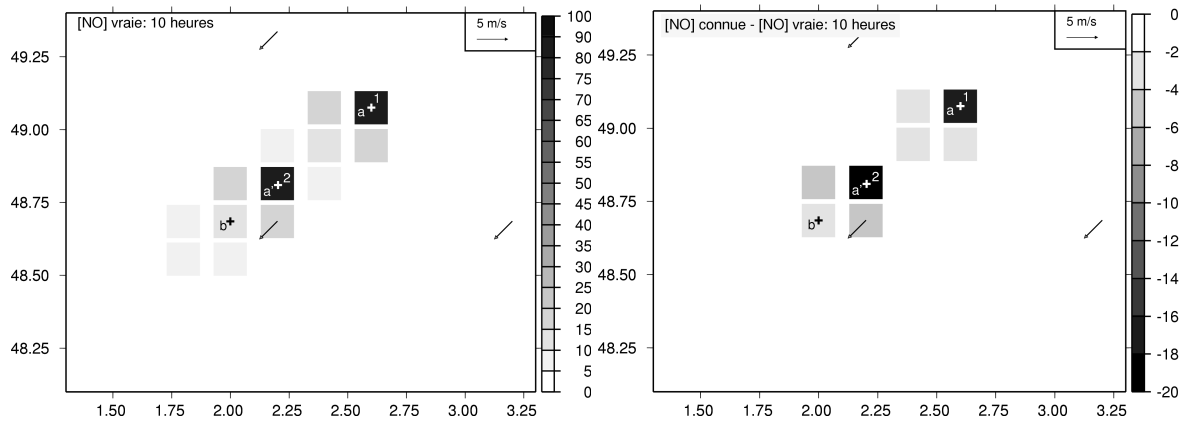
1. On réalise une première simulation où l'on considère que les émissions utilisées sont celles finalement recherchées. Les polluants simulés sont dès lors considérés comme les observations. On fait donc l'hypothèse qu'il n'y a pas d'erreur modèle.
2. On perturbe les émissions en entrée. Et on considère que ce nouveau jeu d'émissions est l'unique information dont on dispose. On obtient donc un nouveau jeu de concentrations, différent du premier.
3. On effectue la procédure de minimisation : si le modèle est à même de retrouver les émissions de départ, on peut considérer que la méthodologie est fiable.

Même si cette approche ne permet pas d'apporter toutes les réponses, elle est cependant un passage obligé pour une bonne connaissance du modèle et du système d'optimisation. ce doit donc être la première étape de travail avant de passer à un cas réel. L'ensemble de la mise en oeuvre de cette technique est résumée sur la Figure 3.1.
L'élaboration d'un cas académique permet en effet de disposer d'une "réalité" à laquelle il suffit de comparer les résultats de l'optimisation pour déterminer le degré de précision de la méthode. La simulation directe effectuée avec le cadastre "réel" (qui donne donc les valeurs de concentrations considérées comme "réelles" pour tous les polluants dans toutes les mailles du modèle) et le choix des stations de mesure (emplacement et espèces mesurées) permettent de générer les mesures, plus ou moins entachées d'erreurs selon l'incertitude choisie. Il faut noter que les problèmes dus à la diffusion numérique sont en partie évités puisque mesures et "réalité" sont générées avec le même maillage. Les paramètres *a priori* sont créés en perturbant les paramètres "réels". Dans le cas du cadastre d'émission indiqué sur la Figure 3.1, les émissions de certaines espèces sont multipliées ou divisées. La simulation inverse utilise le cadastre *a priori* (qui fournit donc les concentrations simulées *a priori*) et les mesures pour produire un cadastre optimisé. La comparaison entre cadastre "réel" et cadastre optimisé est possible, contrairement à ce qui se passe dans la réalité.

3.3.2 Deux sources et une ou deux stations de mesure

Le cas créé est très simple : deux sources sont placées au centre d'un domaine et sont notées 1 et 2. Elles sont chacun equipées d'une station, respectivement notées, a et a'. Le vent est constant et dans l'alignement des sources (d'autres cas ont été réalisés ou la direction du vent changeait mais ne sont pas présentés ici car les conclusions variaient peu). Les caractéristiques complètes de ce cas sont décrites en détail dans l'**Annexe B, p.83**.
Dans les deux cas d'inversion, les émissions de NO_x sont inversées pendant une fenêtre temporelle de six heures, de 7 heures à midi. Le coefficient de correction qui devrait être obtenu est donc de 1,25 pour les cinq premières heures et 1 pour la dernière. Les différences entre cadastre *a priori* et cadastre "réel" amènent aux écarts entre concentrations de NO simulées représentées sur la Figure 3.2.
Pour jauger de l'impact de la disponibilité des informations en terme de mesures, on inverse tout d'abord les deux sources avec les deux stations puis avec une seule. Les résultats sont présentés globalement sur la Figure 3.3.

- Dans le premier cas d'inversion, les deux stations de mesure a et a' sont disponibles (indiquées sur la Figure 3.2). Ces stations fournissent des mesures "parfaites" de NO, l'inverse doit donc en principe permettre de retrouver les valeurs "réelles" des émissions. Il s'agit en effet d'un cas parfait



(a) Concentrations de NO simulées avec le cadastre réel

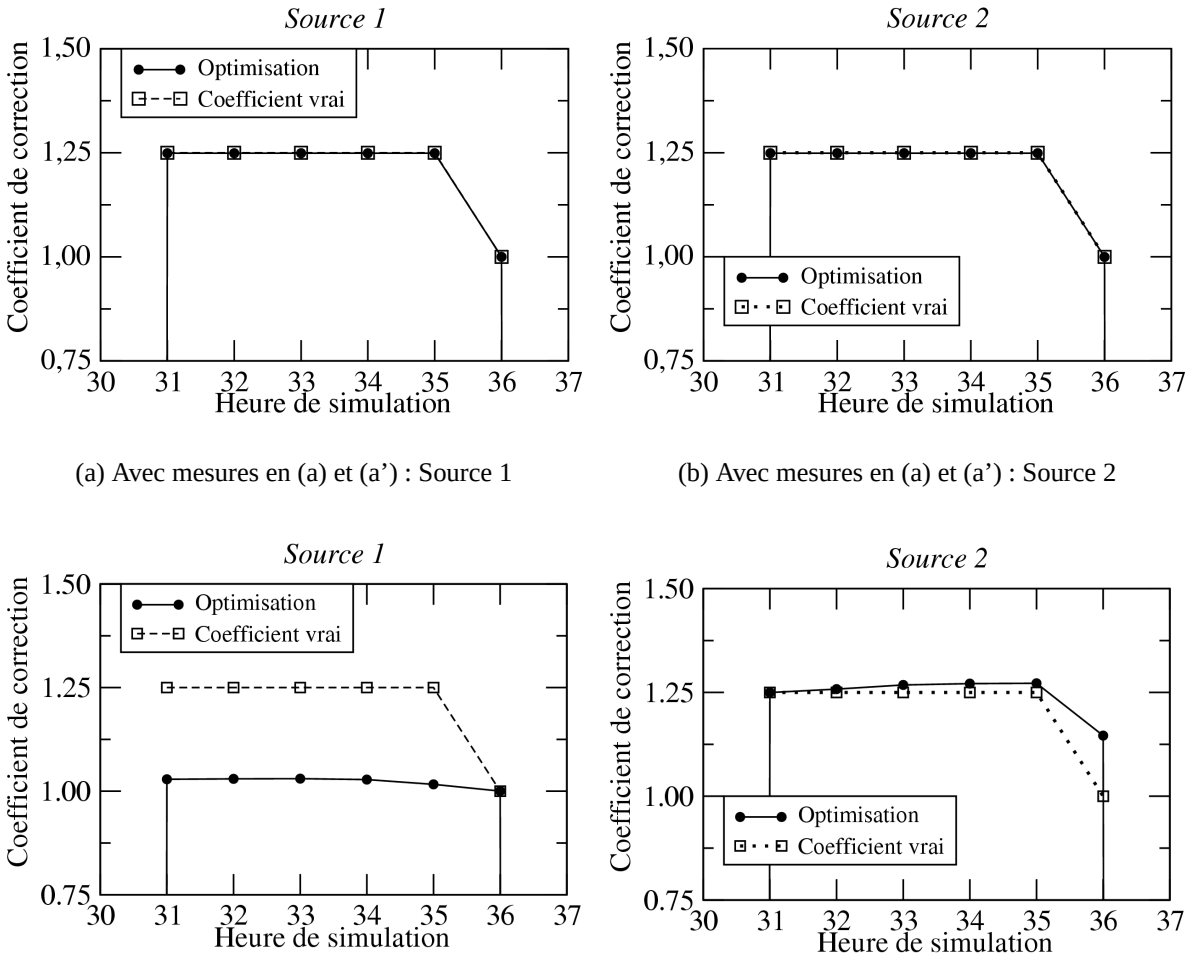
(b) NO(a priori) - NO(réel)

FIG. 3.2 – *Concentrations surfaciques de NO et écarts (ppb) entre concentrations a priori et "réelles" à 10 heures.*

pour l'inversion : toute l'information nécessaire est fournie sans erreur ni incertitude. Seules les erreurs dues au modèle direct peuvent dégrader les résultats. L'optimisation converge et le coût s'annule puisque les contraintes sont parfaites.

- Pour le second cas, seule la station [a'] est disponible et fournit des mesures parfaites. La seule information disponible intègre donc les effets des deux sources. La source [2], dans laquelle est placée la station [a'] est parfaitement corrigée par l'inversion. L'intensité de la source 1 n'est quasiment pas modifiée par l'inversion : bien que la station [a'] soit située en aval de la source [1], l'influence de celle-ci sur les concentrations mesurées est faible à cause de la distance séparant la source et la station. On note aussi que la source [2] a un impact très fort sur cette mesure. L'information fournie par la station [a'] n'est donc pas pertinente pour la source [1] qui n'est donc presque pas modifiée par l'optimisation. A la dernière heure, la source [1] n'est plus sous-estimée mais la source [2] reste surestimée car la réponse des concentrations au changement d'émissions est décalée par le délai lié au transport.

On relève tout d'abord que dans le cas ou les mesures ne sont pas entachées d'erreur, et ou le nombre de contraintes (mesures) équilibre le nombre de paramètres à inverser (les sources) le résultat est parfait : on retrouve les émissions que l'on avait au départ avant perturbation. On a donc tout d'abord une **validation complète du système de calcul mis en place**.
On note donc que si le nombre de contraintes est trop faible (une seule mesure pour deux sources distinctes), le modèle va avoir tendance à donner le plus de poids à la source qui lui semble la plus efficace, au détriment de celle qui semble avoir le moins de poids sur les concentrations. Cette tendance s'aggrave avec le transport qui fait perdurer artificiellement cet effet.
Ces premiers résultats nous ont donc conduit à imaginer un système dans lequel nous pouvons **ajouter de l'information non pas sur les émissions directement mais sur les mesures en elles-mêmes**. Ceci afin de réduire le problème, difficilement contournable du sous-conditionnement du problème de minimisation. Une technique basée sur **le krigeage** a donc été ajoutée à notre modèle d'inversion et



(a) Avec mesures en (a) et (a') : Source 1

(b) Avec mesures en (a) et (a') : Source 2

(c) Avec mesures uniquement (a') : Source 1

(d) Avec mesures uniquement (a') : Source 2

FIG. 3.3 – *Coefficients correctifs pour les sources 1 et 2 à chaque heure de la fenêtre temporelle pour le cas avec les deux stations (en haut) et avec uniquement une station (2) en bas.*

les résultats sont présentés dans la suite.

3.4 Ajout de contraintes par krigeage

3.4.1 Comparaisons mesures et krigeage

Nous ajoutons donc maintenant des contraintes, concues à partir des mesures et de krigeage. Le principe théorique est repris succinctement en **annexe A.2, p.77**. L'intégralité de la méthode (développement et validation) sont explicités dans [*Blond* (2002)].
Pour jauger de l'impact de cette approche, on va réaliser des calculs sur un cas un peu plus réaliste que précédemment : on se place donc dans le contexte de la région parisienne et, plus particulièrement, de la POI2 du projet ESQUIF, c'est à dire le 7 Août 1998. Ce cas a l'avantage d'avoir une météorologie assez simple, bien contrainte ([*Menut et al.* (2000a)]) et d'être un épisode de pollution

dite "locale" (i.e notre approche ne sera pas perturbée par des conditions aux limites variables). On fait donc l'hypothèse (très simplificatrice ici) que le modèle, la météorologie sont corrects et que les émissions n'ont pas la bonne amplitude mais que les sources sont bien placées (hypothèse réaliste).
On cherche dans ce cas à inverser les émissions de NO_x de 8 et 9 heures le matin et uniquement pour le centre de Paris (Figure 3.4). Pour cela, on utilise comme contrainte les concentrations mesurées de NO entre 9h et 17h. Pour plus de réalisme on utilise comme mesuers les emplacements des vraies stations AIRPARIF en région parisienne.

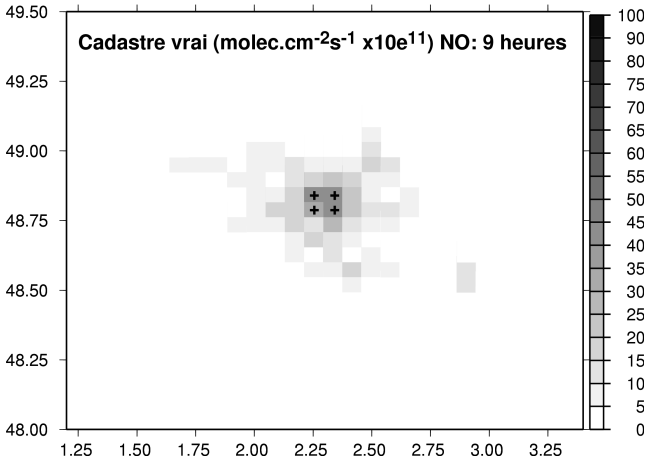


FIG. 3.4 – *Cadastre d'émission "réel" : flux de NO à 9 heures le 7 août 1998. + = flux perturbé pour créer le cadastre a priori.*

Pour différentes perturbations des émissions à priori (de -70% à +100%), et pour différentes qualités de mesures (0, 10 et 15%), on peut donc refaire un cas *quais-académique* et comparer de l'impact du krgigeage par rapport à l'emploi des mesures seules. Les résultats sont synthétisés dans le Tableau 3.1.
Les résultats sont présentés sous forme de "score" défini tel que :

$$score_x = (1 - \frac{x_{opt} - x_{réel}}{x_{apriori} - x_{réel}}) \times 100$$

avec x_{opt} la valeur optimisée de x, $x_{réel}$ la valeur "réelle" et $x_{apriori}$ la valeur *a priori*, x étant un flux d'émission. La valeur du score correspond donc au changement en pourcentage de la différence entre le cadastre *a priori* et la "réalité" après l'optimisation. Un score positif indique que la différence avec la "réalité" a été réduite alors qu'un score négatif indique qu'elle a été augmentée par l'optimisation. Une optimisation parfaite correspondrait donc à un score de 100% indiquant que la différence entre *a priori* et "réalité" a été annulée après optimisation.
Ces résultats amènent différentes conclusions :

- **Quantité de cas convergents :** Sans aucun doute, la méthode de krigeage permet de satisfaire les critères de convergence bien mieux qu'en utilisant les mesures seules ;
- **Type de contrainte utilisée :** Autant l'inversion fonctionne avec les mesures de NO, autant les résultats ne semblent pas corrects avec l'ozone.
- **Qualité du résultat :** Pour uniquement les cas ou la convergence est atteinte, il semble que la méthode avec les mesures donne de meilleurs scores. Cependant le résultat est délivré sous la forme d'une moyenne des huit flux inversés. Or, un examen plus approfondi des résultats montre que le

Espèce utilisée	Type de contraintes	Qualité mesures ϵ_m	qualité du cadastre *a priori* ϵ_E						
			-70%	-50%	-20%	+20%	+50%	+70%	+100%
NO	mesures	0	53.8	35.4	x	x	x	x	x
		10	49.3	29.	x	-77.4	x	x	x
		15	57.	x	x	x	x	x	x
	krigeage	0	29.8	29.1	30.3	28.8	x	x	21.9
		10	31.6	31.4	32.7	26.8	25.7	25.3	21.
		15	29.7	30.2	42.8	17.4	26.2	22.2	20.9
ozone	mesures	0	x	x	x	x	x	x	x
		10	x	x	x	x	x	x	x
		15	x	x	x	x	x	x	x
	krigeage	0	x	8.3	7.3	5.7	0.6	x	x
		10	x	10.9	13.8	x	-4.1	x	x
		15	3.5	3.1	x	10.9	12.4	8.5	6.8

TAB. 3.1 – *Scores moyens obtenus pour l'optimisation des émissions de NO_x avec des concentrations d'ozone ou de NO comme contraintes. x = convergence non atteinte.*

score est artificiellement plus haut pour les mauvaises raisons. Ce point est discuté maintenant plus en détail.

Si l'on reprend les valeurs du Tableau 3.1 mais en se focalisant uniquement sur les cas qui ont convergé et en ajoutant la variabilité autour du score trouvé, on montre que l'écart moyen avec les émissions "réelles" est réduit de 45,1 à 59,3% en utilisant les mesures et de 26,8 à 31,6% en utilisant le krigeage. Cependant, des variances beaucoup plus élevées sont obtenues avec les mesures seules. Cela signifie que les scores semblent meilleurs mais que, pour cela, l'optimisation a redistribué des modifications de flux sans respecter la **logique de la répartition des émissions** prescrites au départ.

3.4.2 Mesures, background et krigeage

Pour mieux contraindre cette répartition, des tests sont alors réalisés en ajoutant la prise en compte d'une matrice de background dans le cas de l'inversion avec les mesures. Cette matrice a pour but d'ajouter une limitation aux flux que l'on cherche à inverser en leur imposant des limites.
On note tout d'abord, avec le Tableau 3.2 que les problèmes de mauvaise "équilibre" de l'optimisation pour la méthode avec les mesures uniquement est un probleme à la fois spatial et temporel. On retrouve ici des problèmes numériques de type oscillatoires tels que calculés dans les études de [*Chang et al.* (1997)] avec le CO. De manière assez logique, le minimiseur fait porter le poids à une première heure sur ce qui lui semble "instantanément" le plus efficace. A l'heure suivante, ayant été un peu "trop fort" dans sa correction, il fait machine arrière. Le bilan des corrections est finalement une fonction oscillante autour de la moyenne des flux au départ. Cela devrait être moins le cas avec la matrice de background et le krigeage qui apportent une cohérence spatiale, respectivement dans l'espace des émissions et des mesures.
Les résultats du Tableau 3.2 montrent que l'ajout de cette matrice permet de réduire le problème de déséquilibre d'optimisation des flux mais pas totalement. En tout cas, semble moins efficace que la méthode de krigeage.

Exp.	08 :00		09 :00	
Mesures seules	-75.8	-56.4	-33.5	-33.2
	-14.1	-26.8	+37.0	+7.2
Mesures avec **B**	-69.4	-51.5	-35.3	-38.3
	-16.7	-27.9	+12.3	-14.6
Krigeage	-34.9	-35.6	-35.8	-34.
	-26.6	-33.4	-39.8	-35.2

TAB. 3.2 – *Differences en % entre les flux réels et optimisés pour $OPT_{NOx}measCONC_{NO}$, $OPT_{NOx}krigCONC_{NO}$ et $OPT_{NOx}measBCONC_{NO}$ avec ϵ_E=-50% et ϵ_m=10%. Chaque bloc de 4 valeurs représente les 4 mailles urbaines optimisées. Chaque colonne représente l'heure optimisée.*

3.4.3 Krigeage itératif

Les résultats cohérents et homogènes nous montrent que cette approche est certainement un bon compromis méthodologique. Cependant les scores, relativement faibles, ne sont pas au niveau de précision que l'on souhaite atteindre. On va donc chercher à les améliorer en mettant en place une procédure itérative d'optimisation.

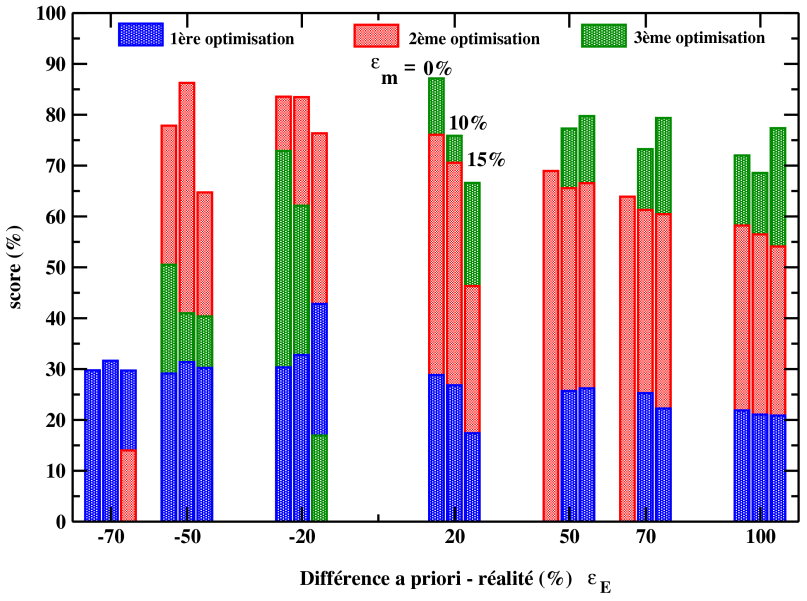


FIG. 3.5 – *Scores moyens (%) après un à trois cycles krigeage-optimisation.*

La Figure 3.5 présente les résultats pour l'optmisation du NO et toujours avec les mêmes configurations du modèle que précédemment. On montre qu'à partir de deux itérations, le score est grandement amélioré. On montre aussi que faire un troisième cycle d'optimisation n'améliore pas forcément les résultats : cela est du à un manque d'optimalité du système à optimiser. A partir d'un certain niveau

de modification des flux et en l'absence d'information nouvelles supplémentaires, le système n'a plus de contraintes claires pour se rapprocher de la "bonne solution". Dans la suite des calculs, et sachant que ces itérations ont un coût numérique non négligable, nous resterons sur ce compromis de deux cycles d'optmisation-krigeage.

3.4.4 Applications du krigeage avec des mesures d'ozone

Nous reprenons les méthodes exposées précédemment et basées sur l'inversion de flux d'émissions de NO à partir de mesures de NO. Nous cherchons toujours à inverser des flux de NO mais en utilisant cette fois des mesures d'ozone. La logique derrière ces calculs est que ces mesures sont généralement plus nombreuses et disponibles que celle du NO. De plus, la représentativité spatiale de l'ozone est plus proche des échelles de transport que l'on étudie avec le modèle.
Les résultats montrent, pour le domaine étudié et la situation météorologique étudiée, que les scores restent inférieurs à 15%. Cela s'explique simplement par le fait que là ou sont les informations importantes sur l'ozone (dans le panache), peu d'informations sont disponibles sur ce cas (panache de sud-ouest). De plus, l'impact des modifications des émissions de NO se traduit par une variation (relativement) moins importante des concentrations d'ozone dans le panache. En résumé, les mesures d'ozone n'ont pas un poids assez fort pour contraindre efficacement les émissions urbaines de NO, Figure 3.6, et l'inversion conduit à des flux optimisés proches des flux *a priori*.

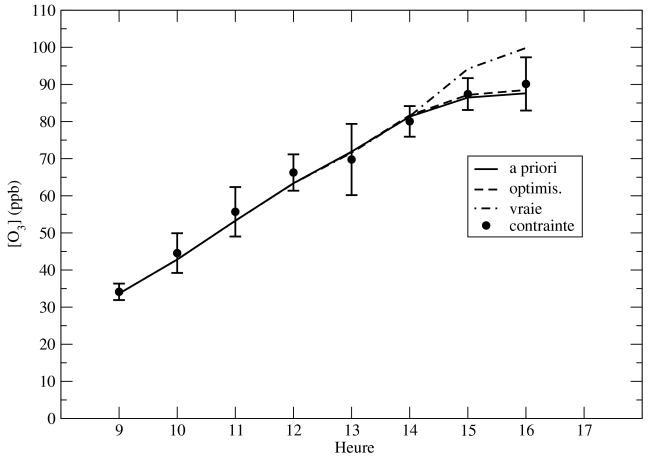


FIG. 3.6 – *Évolution des concentrations d'ozone "réelles", a priori et optimisées pour E_{NO} =+100% avec les contraintes obtenues à partir de mesures à ±=15%, au coeur du panache.*

3.5 Agrégation spatiale

3.5.1 Principe général

L'utilisation du krigeage permet d'exploiter au mieux toutes les informations disponibles. La quantité totale d'informations reste cependant limitée et le problème de la grande dimension du problème par rapport à la quantité d'informations disponible demeure. Dans le cas de la pollution à l'échelle régionale, il est cependant particulièrement difficile d'ajouter de l'information dans l'espace des émissions.

Nous avons donc développé une **approche originale d'agrégation spatiale**. Il s'agit de regrouper les informations dans l'espace des paramètres à optimiser. Par principe, si l'on cherche à inverser des émissions de NO liées au trafic, on sait dès le départ que le principe de construction de ces émissions est le même pour deux mailles différentes (i.e une voiture émet autant en roulant 1h au nord ou au sud de Paris). On peut donc imaginer réduire la dimension du problème à inverser en regroupant des flux d'émissions à inverser, dans la mesure ou on fait l'hypothèse de comportements analogues.
On va donc chercher à créer des zones spatiales. Le principe a déjà été appliqué pour des études à l'échelle globale, [*Heimann and Kaminski* (1999)], [*Bousquet et al.* (1999a)], [*Prinn and Hartley* (1995)] : les zones définies étaient des continents ou des parties d'océans. A cette échelle, la création de zones étaient plus simples à gérer car on pouvait considérer que la variabilité des émissions et concentrations étaient "à l'intérieur" d'une maille. Ce n'est pas notre cas, en pollution régionale, ou les zones qui pourraient être définies en terme d'émissions ne sont pas les mêmes que celles que l'on peut définir en terme de concentrations. De plus, il n'est pas possible de conserver de manière statique une zone pendant toute la durée d'une simulation alors que les conditions météorologiques peuvent complètement changer en quelques heures.

3.5.2 Choix méthodologique

Le regroupement en "zones" des flux d'émissions vise avant tout à réduire la dimension du problème. Notre problème étant spatial et temporel, nous vons dû faire un choix sur le type de dimension dans laquelle nous voulions apporter de l'information. Il nous a semblé pertinent de considérer que :

- Les informations que nous possédons aujourd'hui sur les sources permettent d'avoir des certitudes sur la localisation des principales sources. Cela est aussi vrai pour le trafic puisque les principaux "brins" autoroutiers sont bien connus.
- Par contre, la ventilation temporelle est mal connue. Les émissions sont fournies en masse intégrée en temps. De plus, on peut bien imaginer qu'en fonction des types de jours (semaine, mais avec Mercredi, Samedi, Dimanche, jour férié ou de retours de vacances), cette ventilation temporelle montre une large variabilité.

Pour ces raisons, nous avons choisi de faire l'hypothèse de fixer nos zones de travail spatialement. Nous cherchons alors à faire varier, et donc à apporter de l'information, en temps.

3.5.3 Construction d'un modèle dynamique de zonage

Connaissant la complexité du problème, sa variabilité en temps, et voulant créer une méthode adaptable ensuite à d'autres sites, nous nous sommes attachés à définir des critères souples pour le choix des zones. Nous avons donc construit un **modèle de construction des zones** : ce modèle est basé sur des règles simples et définit un **nouveau domaine de calcul à chaque heure de simulation**.
Les hypothèses sont les suivantes :

- **Des flux à modifier pour les mêmes raisons :** comme par exemple le trafic, qui s'il est sous-estimé en intensité pour un flux de véhicule donné, le sera partout de la même manière.
- **Prendre en compte les sensibilités des concentrations aux émissions :** il a été montré, [*Menut* (2003)], que l'on pouvait avoir une sensibilité de l'ozone au NO négative dans l'heure précédente mais positive plusieurs heures avant. Comme notre problème est basé sur une évolution de plusieurs heures, nous devons prendre en compte la sensibilité exprimée telle que :

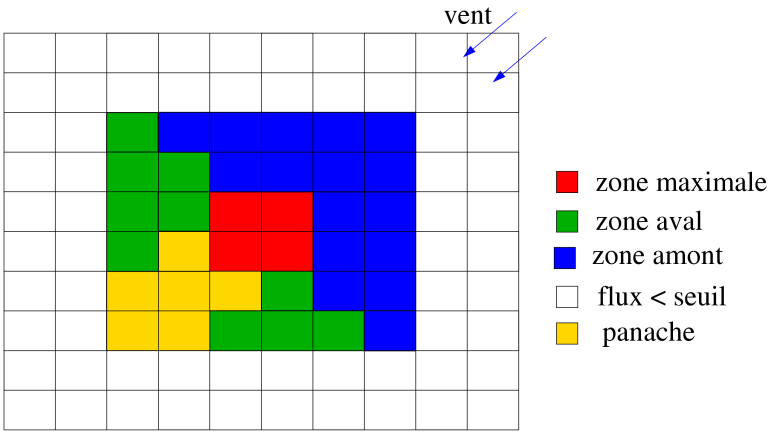


FIG. 3.7 – *Schéma d'un découpage en grandes régions selon la direction du vent.*

$$\frac{\partial J}{\partial x_z} = \sum_{i=1}^{n} \frac{\partial J}{\partial p_i} \times \frac{\partial p_i}{\partial x_z} = \sum_{i=1}^{n} \frac{\partial J}{\partial p_i} \times p_i^0 \tag{3.1}$$

avec p_i^0 la valeur *a priori* de p_i. Il faut donc regrouper les paramètres p_i de telle sorte que les termes en $\frac{\partial J}{\partial p_i}$ ne se compensent pas.

- **La physique du cas étudié :** Dans les cas où une zone de sources très intense existe, on peut distinguer, en plus de cette zone, trois régions, schématisées sur la Figure 3.7 :
  - ☐ une région en amont de la zone source, particulièrement sensible aux conditions aux limites et aux émissions locales
  - ☐ une région située directement sous le vent de la source, dans le panache, qui est donc particulièrement sensible à l'intensité de la source
  - ☐ une région en aval de la source mais qui n'est pas située directement dans son panache.
  - ☐ **Les conditions météorologiques :** les zones "amont", "aval" et "de panache" (de la zone d'émission maximale) sont déterminées à partir de la direction du vent dans la zone maximale.

Pour jauger de la pertinence de cette approche, on recalcule les cas académiques déjà présentés et on définit une tolérance : τ_E, le pourcentage maximum d'écart souhaité entre le flux maximal et les flux regroupés dans la même zone : celle-ci englobe donc le flux maximal E_{max} et tous ceux qui sont au moins égaux à $(1 - \tau_E) \times E_{max}$. cela permet, en pratique, de choisir le nombre de zones en fonctions des gradients d'émissions sur la région étudiée. Plus la tolérance est grande, plus le nombre de zones va être réduit, puisque plus on tolère les agrégations de flux à intensités différentes.
La Figure 3.8 présente un exemple de résultat sur un cas académique pour trois tolérances : 0, 10 et 20% conduisant à la création de 57, 33 et 21 zones distinctes respectivement (Tableau dans la partie gauche). On voit alors que dans le cas à 21 zones, et la où les méthodes précédentes (sans krigeage) ne convergeait pas, on obtient directement en un cycle d'optimisation, une réuction de l'écart de 54,64
A la suite de ce résultat, la question se pose naturellement de savoir s'il existe une configuration optimale. C'est à dire : *connaissant la physique du problème, le nombre de flux, peut-on choisir une tolérance qui permet d'avoir les meilleurs résultats d'optimisation ?*. De nombreux cas ont donc été testés : il a alors été montré que les résultats ne dépendaient pas que de la tolérance, mais aussi de la

Tolérance (%)	Dimension	Score (%)
0	57	x
10	33	45,66
20	21	54,64

Scores moyens obtenus pour l'optimisation des émissions de NO_x avec des concentrations de NO comme contraintes, selon la dimension du problème. x = convergence non atteinte.

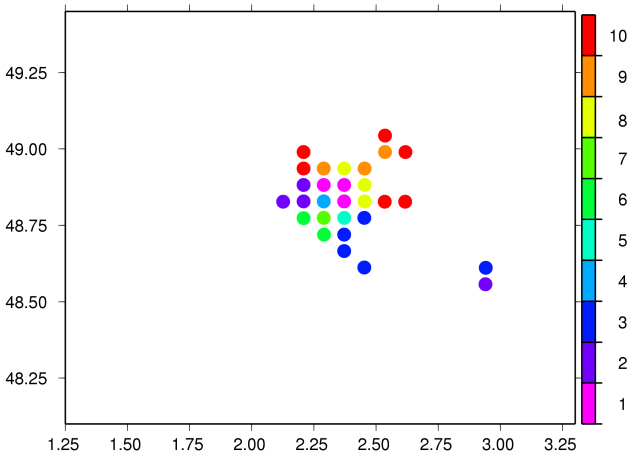
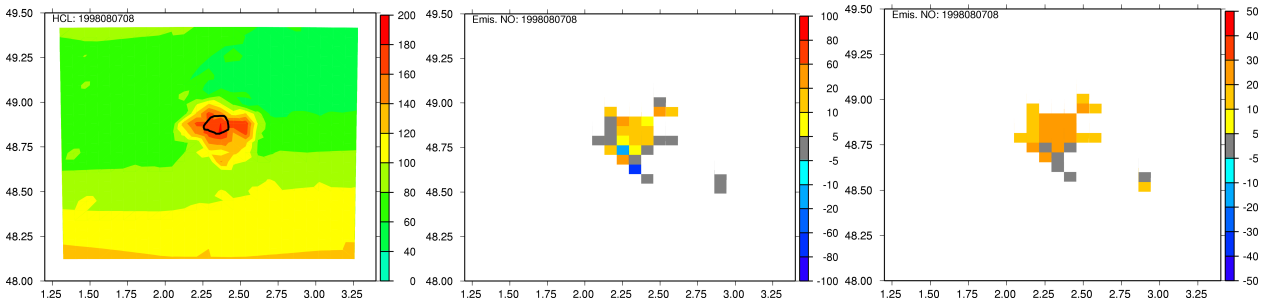


FIG. 3.8 – *Résultats pour différentes tolérances (à gauche). Zones obtenues à 8 heures (et leur numéro pour le calcul), selon la tolérance τ_E choisie : 21 zones (à droite)*

qualité du cadastre *a priori* et des contraintes. cela amène donc à la conclusion qu'il n'y a pas encore, à notre disposition, de critère "universel" permettant d'établir l'optimalité de la procédure.



(a) HCL "réelle" (b) Inversion avec HCL parfaite (c) Inversion avec HCL +10%

FIG. 3.9 – *Différence (en %) avec le cadastre "réel" de NO à 8 heures avec une HCL parfaite ou surestimée de 10% dans tout le domaine.*

Avant de passer à l'application de la méthode sur des cas réels, [*Pison* (2005)] a montré que l'adjonction d'autres paramètres tels que la température au cours de l'inversion, était difficile à gérer. En effet, des paramètres additionnels donnaient des degrés de liberté réalistes au problème mais, en contrepartie, rendait le problème difficile à gérer en terme de direction de l'optimiseur. Il a donc été conclu qu'il était plus adapté de corriger en amont les paramètres météorologiques sur des cas particuliers, afin de réduire l'incertitude de ces autres paramètres. Par exemple, dans le cas d'analyse de cas, il est assez efficace de débiaiser la température, la direction du vent et la hauteur de la couche limite avant de procéder à l'inversion. Si cette correction ne peut être faite (par manque de données disponibles en général), la prise en compte de l'incertitude peut permettre de corriger ce manque.
Pour quantifier l'impact d'une méconnaissance de la hauteur de la couche limite sur les résultats de NO_x inversés, on reprend ici le même cas académique avec la hauteur de couche limite supposée "parfaite" et une hauteur surestimée de 10%. [*Pison* (2005)] montre, Figure 3.9, que le score obtenu est

de 40,6% au lieu de 45,7% lorsque la HCL est "réelle" : La différence moyenne entre les corrections obtenues avec la HCL "réelle" et avec la HCL surestimée de 10% est de 10,2 points, ce qui correspond à la relation prévisible quasiment linéaire entre l'écart sur les corrections des émissions de NO et la variation de la HCL. On peut donc considérer, en première approximation, que débiaiser la hautreur de la couche ,imite avant inversion apporte autant qu'ajouter ce paramètre incertain pendant l'inversion.

3.6 Cas réels : ESQUIF POI2, le 7 Août 1998 en Ile de France

3.6.1 Les émissions à optimiser

- **hypothèses préliminaires**
  Avant toute application à des cas réels, nous avons montré qu'il était nécessaire de jauger des biais systématiques des simulations par rapport aux mesures. Un travail présenté dans [*Pison* (2005)] permet d'évaluer les biais et la variabilité du modèle pour les mois d'été de 1998 et 1999, choisis pour réaliser les études en cas réel d'inversion des émissions. La qualité des résultats obtenus avec CHIMERE permet de considérer qu'aucun processus important n'est manquant ou trop mal représenté. De plus, en région Parisienne, on peut aussi faire l'hypothèse que les données disponibles pour la météorologie représente bien les processus réellement observés.

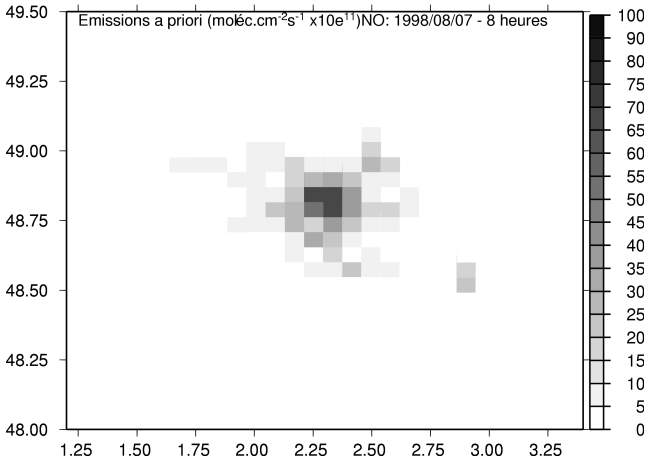


FIG. 3.10 – *Cadastre d'émissions a priori : flux de NO le 7 août 1998 à 8 heures TU. En été, TU = heure légale - 2 heures.*

- **conservation des rapports à l'émission**
  La situation décrite ici est bien plus complexe que les simples cas académiques. On conserve donc l'hypothèse que les **sources sont bien placées** (une étude précédente de [*Vautard et al.* (2003a)] a déjà permis de corriger les sources mal placées pour le cadastre que l'on utilise). On cherchera donc uniquement à **modifier leur intensité**. De plus, on cherchera à inverser des familles d'espèces (NO_x, COV) en considérant que l'agrégation des espèces émises au sein de chaque famille est suffisamment réaliste. Cela signifie que les rapports fournis par le cadastre *a priori* entre les différentes espèces composant la famille, par exemple le NO et le NO_2, sont conservés lors de l'inversion ; c'est donc l'intensité du flux de toute la famille qui est inversée.

- **prise en compte des biais**
  Les résultats statistiques obtenus pour les concentrations suggèrent que les émissions du pic matinal de NO_x en Île-de-France pourraient être mal évaluées dans le cadastre *a priori*. Le rôle du mélange vertical dans les écarts entre concentrations mesurées et simulées doit cependant être étudié avant de procéder éventuellement à l'inversion de ces émissions. Dans ce cas, il apparaît donc nécessaire d'inverser les émissions de NO_x afin d'obtenir un profil temporel optimisé pour le pic de concentration du matin, de 4 à 11 heures d'après les évolutions du biais.
- **focus sur les flux urbains**
  Le problème présent concerne principalement les flux en milieu urbain. On s'attachera donc à optimiser les flux les plus intenses c'est-à-dire ceux qui recouvrent les zones de trafic routier. Cela correspond aux flux de NO_x supérieurs à 5×10^{11}moléc.cm^{-2}.s^{-1}, soit environ 30 à 60 flux pour chacune des huit heures de la fenêtre temporelle définie précédemment. La Figure 3.10 représente les émissions de NO (soit 90% des émissions de NO_x) *a priori* à 8 heures : les flux non inversés sont en blanc et correspondent aux environnements ruraux ou péri-urbains peu denses.
- **choix des mesures considérées**
  Dans ce cadre, l'agrégation spatiale est effectuée sur les flux sélectionnés comme décrit précédemment : notre technique étant adaptative, le nombre et la taille des zones définies varient à chaque heure. Enfin, les contraintes utilisées pour l'inversion des émissions de NO_x sont les analyses de NO effectuées de 4 à 16 heures. Le krigeage des concentrations est effectué grâce aux mesures du réseau AIRPARIF décrites. On notera enfin que les stations "observation", dont la représentativité n'est pas contrôlée, et les stations "trafic", dont la représentativité n'est pas adaptée au problème (trop petite par rapport à la maille du modèle), ne sont pas utilisées pour l'analyse.

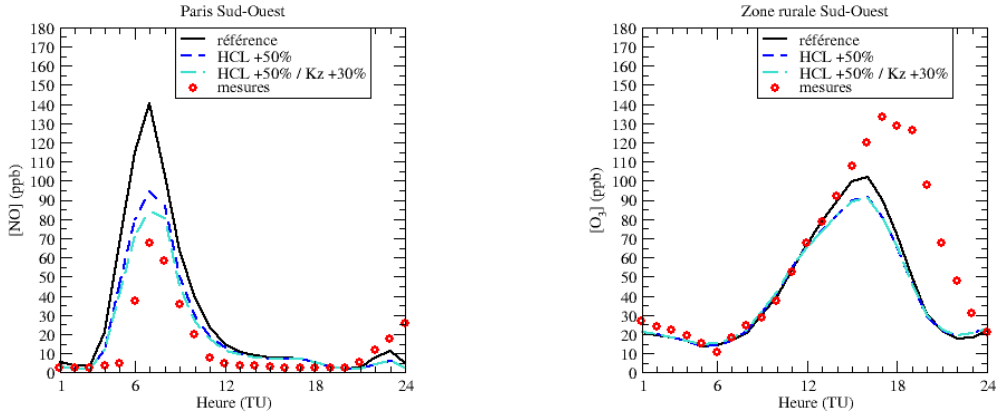
3.6.2 Tests de sensibilité aux paramètres météorologiques

L'examen des concentrations simulées pour la journée du 7 Août 1998 (cas étudié et décrit en détail dans [*Menut* (2003)], [*Vautard et al.* (2003a)]) a montré que la surestimation des concentrations de monoxyde d'azote en ville pourrait expliquer la sous-estimation des concentrations d'ozone l'après-midi dans le panache. Il est aussi probable que la surestimation des concentrations simulées de NO (par rapport aux mesures) peut être due à une sous-estimation du mélange vertical ou à une surestimation des émissions de NO_x. L'influence du mélange vertical sur les concentrations simulées le 7 août 1998 sera donc estimée en examinant les rôles du coefficient de mélange turbulent et de la hauteur de couche-limite avant l'optimisation des émissions de NO_x.

Le principal résultat montre que la hauteur de la couche limite est sous-estimée par le modèle de l'ordre de 50% par rapport aux mesures. On effectue donc cette correction et on peut alors refaire une simulation directe et jauger de l'impact. les résultats sont présenté sur la Figure 3.11

On note que les concentrations de NO simulées le matin en ville avec une HCL augmentée de moitié sont de 15 à 39,5% inférieures aux concentrations obtenues avec la HCL de référence, l'impact du K_z restant faible (Figure 3.11(a)) et sont nettement plus proches des valeurs mesurées. De plus, la sous-estimation des concentrations d'ozone simulées par rapport aux mesures est peu accentuée par l'augmentation de la HCL (Figure 3.11(b)) .

Dans la suite de cette étude, on utilisera donc le modèle avec une hauteur de couche limite corrigée de l'ordre de +50% et sans modification particulière du coefficient de mélange.



(a) NO (b) Ozone

FIG. 3.11 – *Concentrations de monoxyde d'azote et d'ozone (ppb) obtenues avec différentes HCL et valeurs de K_z. En été, TU = heure légale - 2 heures.*

3.6.3 Résultats moyens sur le domaine

Les flux de NO_x vont être optimisés, en laissant libre une variation possible du rapport COV/NO_x. On choisit une tolérance sur le zonage permettant de regrouper les flux de type urbains les plus importants ($\tau_E \geq$25%) et on réalise deux cycles consécutifs "krigeage-optimisation".
Un premier résultat global (c'est à dire moyenné sur le domaine) est présenté sur la Figure 3.12. Par rapport au cadastre *a priori*, le pic du matin du cadastre optimisé est intensifié entre 4 et 5 heures et réduit les heures suivantes. L'augmentation la plus forte est obtenue à 5 heures avec un flux optimisé moyen plus intense de 11,7% que le flux *a priori* ; la réduction la plus forte est obtenue à 6 heures avec un flux optimisé moyen plus faible de 16,1% que le flux *a priori*. La masse moyenne émise au cours de la journée dans tout le domaine n'est donc diminuée que de 1% par l'optimisation. Ces corrections font apparaître sur le profil temporel moyen deux pics nettement séparés le matin, entre 4 et 5 heures et 8 et 9 heures, au lieu d'un plateau de 4 à 9 heures.
De manière complémentaire, la Figure 3.13 présente les mêmes résultats mais spatialisés sur le domaine et pour des heures particulières : 05 :00 et 08 :00h. On note que :

- à 05 :00h : Les flux les plus intenses (en ville) sont multipliés par 0,5 à 0,75 alors que les flux intenses extérieurs à Paris sont multipliés par 1,25 à 2,75. Les flux les plus faibles sont peu modifiés, le coefficient de l'optimisation restant compris entre 0,95 et 1,25. Ainsi le premier pic des émissions optimisées présentés sur la Figure 3.12 est dû à une augmentation des flux *a priori* en amont de Paris.
- à 08 :00h : les flux les plus intenses sont modifiés de moins de 5% avec des coefficients de correction compris entre 0,95 et 1,05. Les flux moins intenses de l'extérieur de la ville sont réduits, multipliés

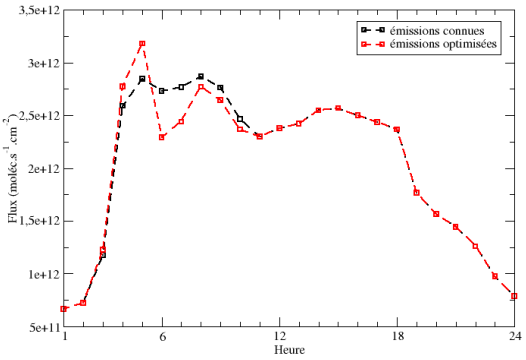
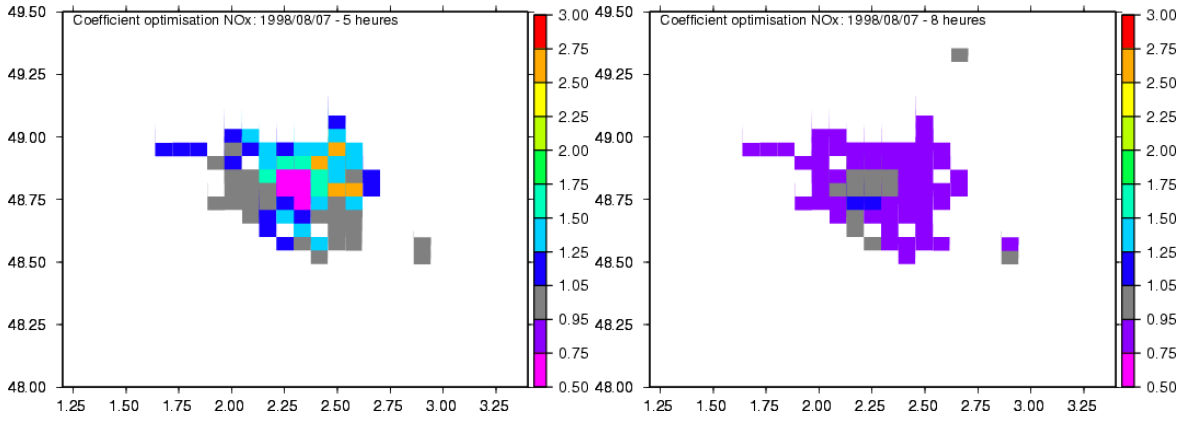


FIG. 3.12 – *Profil temporel moyen sur tout le domaine du flux d'émission de NO le 7 août 1998. En été, TU = heure légale - 2 heures. Les zones grisées correspondent aux heures non optimisées.*

par 0,75 à 0,95. Les corrections sont beaucoup plus homogènes qu'à 5 heures : le deuxième pic du profil temporel des émissions optimisées moyennes sur tout le domaine correspond donc à un abaissement général des flux d'émissions inversés.



(a) à 5 heures TU (b) à 8 heures TU

FIG. 3.13 – *Répartition spatiale des corrections pour les NO_x le 7 Août 1998, à 05 :00 et 08 :00.*

3.6.4 Résultats sur quelques sites

Pour regarder plus en détail les résultats, on extrait des séries chronologiques de flux d'émissions optimisés pour quelques points particuliers (choisis comme ayant des mesures disponibles).
La Figure 3.14 présente pour deux sites différents (Paris sud-ouest et Tremblay) les concentrations mesurées et simulées, ainsi que les flux d'émissions optimisés.

Les modifications apportées par l'optimisation sont modérées et réalistes à Paris (réduction de la "largeur" temporelle du pic de trafic). Cette adaptation du pic peut tout à fait se corréler avec le principe de construction même du cadastre : les profils moyens mensuels sont similaires sur les mois d'été et seule une réduction globale de 20% est appliquée en Août par rapport à Juillet. La réalité peut donc amener vers une réduction de la période temporelle en Août et pas seulement une réduction de l'intensité du pic.
Au contraire, à Tremblay, les modifications apportées semblent plus importantes que ce qui peut être envisagé sur ce type de site. La question reste donc posée puisque les mesures présentent bien des concentrations très élevées pour ce site. On est vraisemblablement confronté à un problème de représentativité de la mesure par rapport à la maille du modèle censé la représenter.

3.6.5 Impact de l'optimisation sur le panache d'ozone

Nous disposons maintenant de valeurs d'émissions optimisées pour la journée complète. L'un des buts de nos travaux étant d'améliorer notre représentation des concentrations d'ozone au pic (en milieu d'après midi), nous comparons les simulations avec les émissions "a priori" et "optimisées". La carte de différences présentée sur la Figure 3.15(a) montre que le changement des émissions du matin a un impact important sur les concentrations dans le panache de l'après-midi : au maximum on note une augmentation de 5ppb au coeur du panache, améliorant ainsi considérablement la simulation initiale. Mais le changement des concentrations d'ozone dépend aussi de l'abondance relative des COVs par rapport aux NO_x. Ce concept de régime chimique ne peut être négligé en région parisienne, car, comme l'a récemment montré [*Sillman et al.* (2003)] et en Île-de-France, le régime chimique dans le panache de la ville est le plus souvent proche de la transition entre NO_x-limité et COV-limité. De plus, des régimes différents sont diagnostiqués
On a donc cherché à faire la même optimisation mais en fixant ce rapport COVs/NO_x afin de conserver le régime chimique prescrit dans la simulation de référence. Le résultat est présenté sur la Figure 3.15(b). On voit alors que l'impact du changement des émissions est moindre : on modifie nettement moins l'ozone dans le panache et de plus, la modification a un signe opposé à ce qu'il aurait fallu pour se rapprocher des mesures. On ne peut cependant pas en conclure qu'il ne faut pas modifier ce rapport. Cependant, en l'absence de contraintes supplémentaires sur les COVs, il apparait plus approprié de conserver l'approche qui permet de se rapprocher à la fois des mesures de NO le matin et d'ozone l'après-midi, c'est à dire de fixer le rapport COVs/NO_x.

3.7 Cas réels : ESQUIF POI6, du 16 au 18 Juillet 1999 en Ile de France

3.7.1 Simulations directes

Après avoir étudié une seule journée, nous allons procéder à l'inversion des émissions pour trois journées consécutives : 16, 17 et 18 Juillet 1999, constituant la POI6 du projet ESQUIF. L'intérêt est triple :

- une météorologie bien étudiée et bien documentée dans de précédentes études (comme [*Vautard et al.* (2003b)], [*Hodzic* (2005)]), donc bien contraintes ;
- des conditions anticycloniques propices à un épisode de pollution locale et

- des journées qui correspondent aux types de jours définis pour caractériser les émissions : semaine (avec un vendredi), Samedi et Dimanche.

Une première simulation directe permet de jauger des performances du modèle sur ce cas complexe. Des panaches d'ozone très concentrés (plus de 120 ppb le 17) se sont développés à des endroits différents pendant les trois journées de la POI 6. On note principalement :

- Le 16 juillet, le panache se situe au Sud-Est de Paris et les concentrations n'y dépassent pas 80 ppb pour des concentrations au vent d'environ 50 ppb (Figure 3.16 (a)). Mesures et modèle sont en bon accord pour les valeurs des concentrations malgré un léger décalage du panache simulé vers le Sud.
- Le 17 juillet, le panache se développe à l'Ouest de Paris et les concentrations y dépassent 100 ppb pour des concentrations en amont supérieures à 60 ppb (Figure 3.16 (b)). Pour cette journée, les mesures et le modèle sont en bon accord tant pour les valeurs des concentrations que pour la position du panache.
- Enfin, le 18 juillet, le panache est situé près de la ville, au Nord-Ouest. Les concentrations d'ozone y sont supérieures à 100 ppb pour des concentrations au vent supérieures à 65 ppb (Figure 3.16 (c)). Les valeurs des concentrations mesurées et simulées sont en bon accord bien que le modèle situe le panache légèrement plus à l'Ouest que les mesures (cf. [*Hodzic* (2005)]).

Une comparaison plus détaillée des concentrations de monoxyde d'azote mesurées et simulées indique cependant que les concentrations de NO simulées le matin en zone urbaine (Paris) ou péri-urbaine (Garches) sont très supérieures aux concentrations mesurées le samedi et le dimanche (cf. [*Pison* (2005)]). Le pic matinal de NO simulé le samedi est plus de deux fois supérieur au pic mesuré et le pic simulé le dimanche matin n'est pas observé. La surestimation par le modèle des concentrations de NO_x explique la sous-estimation des concentrations d'ozone simulées en ville (écart par exemple de 10 ppb pour le pic du samedi à Paris Sud-Ouest) et dans le panache (écart de plus de 10 ppb pour le pic du vendredi).

3.7.2 Corrections apportées aux paramètres d'entrée

Les différences entre concentrations de NO_x mesurées et simulées sont particulièrement grandes le samedi et le dimanche. Or, les variations diurnes des émissions utilisées pour ces deux types de jour sont très incertaines [*ESQUIF* (2001)]. De plus, les faibles concentrations d'ozone observées en amont de la ville sont bien reproduites. On peut donc supposer que le rapport COV sur NO_x à l'émission utilisé par le modèle joue un rôle prépondérant dans les écarts entre concentrations mesurées et simulées, en plus d'éventuelles erreurs sur la hauteur de couche-limite ou la chimie.
On va donc :

- corriger la hauteur de la couche limite employée par le modèle en utilisant les mesures dont on dispose ;
- inverser les flux d'émissions de NO_x en portant une attention particulière à la réponse du modèle en fonction des trois types de jour : semaine, Samedi et Dimanche

Les résultats sont présentés sur la Figure 3.17 pour la modulation temporelle de la hauteur de couche limite en région Parisienne. Des mesures réalisées par sondages (à Trappes) et par lidar rétrodiffusion (à Palaiseau) permettent d'avoir des renseignements à haute fréquence temporelle sur la variabilité de la HCL. Cela indique principalement une sous-estimation de la hauteur utilisée par le modèle par rapport aux mesures. Les facteurs correctifs ont donc été appliqué avec +10% le 16 et le 17 et +20% le 18 (en période diurne uniquement).

3.7.3 Emissions optimisées

Sous forme de cartes de corrections, les résultats sont présentés sur la Figure 3.18. On peut voir clairement que les corrections obtenues pour les trois types de jours sont très différentes. L'allure du profil temporel du vendredi est conservée, l'intensité des émissions étant réduite de moins de 5% pour la majorité des flux inversés à 50% pour quelques flux plus intenses. Le samedi, les résultats obtenus diffèrent selon les lieux : les émissions optimisées forment en ville un pic plus intense de 25% que le pic *a priori* mais d'une durée plus courte (deux heures au lieu de cinq) ; à l'extérieur de Paris, les flux d'émission sont réduits de telle sorte que le pic matinal *a priori* disparaît complètement du cadastre optimisé. Les corrections obtenues sont importantes mais ne comportent aucun élément irréaliste à ce stade. Enfin, le dimanche, le pic de concentration de NO qui était simulé en ville avec le cadastre *a priori* mais non observé disparaît avec le cadastre optimisé. Cependant, les corrections obtenues pour les mailles péri-urbaines, qui amènent un pic d'émission isolé à 7 heures, paraissent peu réalistes. Ce manque de réalisme peut être dû à la réduction d'un très grand écart entre concentrations mesurées et simulées (7,5 fois plus élevées à 8 heures lors du pic simulé mais non observé).
Enfin, la comparaison des deux vendredis, le 7 août 1998 et le 16 juillet 1999, indique des tendances différentes : dans le premier cas, le profil temporel des émissions est modifié (durée du pic réduite) alors que dans le deuxième cas, il est conservé. Cependant, dans tous les cas, le réalisme des corrections obtenues par l'optimisation est plus satisfaisant pour les flux intenses, en ville, c'est-à-dire là où sont concentrées de nombreuses mesures, que pour les flux périphériques, moins intenses et situés dans des zones où le réseau de mesure est moins dense. On retrouve donc les limites de la méthodologie malgré l'utilisation du krigeage qui permet de répartir au mieux l'information disponible dans les concentrations mesurées.
Les deux études réalisées sur des cas de pollution photo-oxydante bien documentés ont permis de montrer que **la méthodologie d'inversion était applicable à des cas réels**. Les hypothèses posées lors de la validation sur des cas académiques sont donc raisonnablement réalistes. Cependant, les résultats obtenus pour ces deux études ne sont pas généralisables à d'autres périodes. En effet, les cadastres optimisés comportent des caractéristiques particulières à chacun des cas, ce qui ne permet pas de dégager des tendances valables pour une semaine ou un mois entiers. Il s'avère donc nécessaire de mener une étude d'inversion à plus long terme.

3.7.4 Conclusion sur les POIs ESQUIF

L'étude de deux cas particulièrement bien documentés a d'abord montré que la méthodologie d'inversion était applicable à des cas réels.
Deux cas de pollution dite locale (dans lesquels les émissions du domaine jouent donc un rôle prépondérant) ont été partiuclièrement étudiés : les POI 2 et 6 de la campagne ESQUIF. Les émissions matinales de NO_x du vendredi 7 août 1998 et des vendredi 16, samedi 17 et dimanche 18 juillet 1999 ont été inversées après que la hauteur de couche-limite a été corrigée en fonction des mesures obtenues par les instruments supplémentaires.
Les résultats obtenus pour les deux vendredis sont très différents. En effet, le profil temporel des émissions du 7 août 1998 est modifié de telle sorte que la durée du pic matinal en ville est réduite de cinq heures *a priori* à deux heures après optimisation alors que le profil temporel du 16 juillet 1999 est conservé. Ces modifications peuvent être reliées au fait que le cadastre *a priori* a été élaboré sur la base du mois de juillet, les émissions du mois d'août étant obtenues en diminuant tous les flux de

20% mais en suivant les mêmes variations diurnes. Les modifications apportées par l'optimisation le 7 août peuvent donc correspondre à une différence réelle dans le trafic entre les mois de juillet et août.

Les profils temporels de type "samedi" et "dimanche et jour férié" utilisés pour l'élaboration du cadastre *a priori* comportent une grande incertitude. Le samedi 17 juillet 1999 présente finalement des résultats très contrastés selon les lieux. En ville, le pic matinal d'émission de NO_x est réduit en durée (de cinq heures *a priori* à deux heures après optimisation) mais augmenté de 25% en intensité. À l'extérieur de la ville, le pic matinal d'émission disparaît du cadastre optimisé. Le dimanche 18 juillet, le cadastre optimisé permet de rectifier une erreur importante en éliminant un pic d'émission simulé en ville mais non observé. Il comporte néanmoins, pour les flux moins intenses, des modifications peu réalistes

La répartition spatiale des corrections a montré que les résultats obtenus pour les flux de la ville, c'est-à-dire là où sont concentrées de nombreuses mesures, étaient plus réalistes que ceux obtenus pour les flux périphériques, situés dans des zones où le réseau de mesure est moins dense. L'utilisation du krigeage pour répartir au mieux l'information disponible dans les concentrations mesurées ne permet donc pas de surmonter toutes les limitations dues à la quantité limitée d'information.

3.8 Inversions climatologiques

Après ces études de cas, permettant de jauger du comportement de la méthode, nous proposons ici une étude complète sur l'ensemble des étés 1998 et 1999 (c'est à dire sur 124 jours simulés). La période englobe les POIs 2 et 6 d'ESQUIF étudiées précédemment et va permettre de jauger du comportement général de l'optimisation. L'intérêt de cette approche climatologique va être d'estimer si la modélisation inverse permet d'améliorer statistiquement le comportement du modèle.

Le premier but de ce travail est donc d'estimer la variabilité que l'on peut estimer entre les "types de jours" que l'on utilise déjà dans la cadastre a priori : *le jour de semaine moyen est-il représentatif de ce que l'on veut représenter ? le samedi et le dimanche aussi ?* Le second but est de donner plus de finesse au cadastre dont on dispose et de quantifier si notre méthode permet de dégager des tendances, principalement sur le trafic, en terme de "Lundi", "Mardi" types etc.

Avant de quantifier ces grandeurs, une étude préliminaire a été menée sur les paramètres (autres que les émissions) et qui pourraient modifier nos résultats : les conditions aux limites chimiques et notre représentation de la hauteur de la couche limite. Comme nous étudions ici les émissions d'un point de vue climatologique, nous devons estimer le réalisme de ces grandeurs d'un point de vue statistique et moyen. Nous ne sommes plus, en effet, confrontés à des études de cas, comme les POI 2 et 6 d'ESQUIF, qui sont, par définition, des extrêmes. Différents tests de sensibilités, présentés dans [*Pison* (2005)], montrent qu'en moyenne ces grandeurs sont bien reconstituées par le modèle et représentent bien ce qui est nécessaire à ces simulations (la POI2 et la hauteur de couche limite trop basse est en fait une exception sur les deux étés).

Nous allons donc nous focaliser plus particulièrement sur les flux d'émissions. Pour représenter au mieux nos résultats, nous utiliserons la médiane (moins sensible aux valeurs extrèmes, quelques fois aberrantes, que la moyenne).

3.8.1 Variabilité entre les deux mois d'été

Les émissions *a priori* du mois d'août sont, à toute heure, inférieures de 20% aux émissions du mois de juillet : c'est un choix de modélisation (pas de construction du cadastre) fait spécifiquement pour Paris et basé sur de longues séries de mesures AIRPARIF à partir de données "trafic". Les profils temporels médians des flux optimisés aux mois de juillet et août, représentés sur la Figure 3.19, montrent que cette différence est presque conservée après optimisation : elle varie entre -19,5% à 8 heures et -22% à 6 heures. La première conclusion est donc une confirmation de l'intérêt du choix modèle qui avait été fait.

3.8.2 Variabilité pour les jours de semaine

Par rapport au profil temporel *a priori*, l'intensité des émissions optimisées est augmentée de 8,6 à 12,8% entre 4 et 10 heures (Figure 3.19). Cependant, l'allure du profil est conservée, avec deux pics, à 5 et 8 heures. Au total, la masse émise est augmentée de 11,5%.

L'intensité plus élevée des flux d'émission optimisés est un résultat attendu. En effet, la comparaison des concentrations de monoxyde d'azote simulées *a priori* et mesurées avait mis en évidence la sous-estimation des concentrations matinales de NO, au moins en ville. La conservation de l'allure du profil temporel est cohérente avec les résultats de l'étude des corrélations qui avait indiqué que la variabilité et le phasage temporel des concentrations de monoxyde d'azote étaient corrects, au moins en milieu urbain.

3.8.3 Variabilité pour les jours de week-end

Le samedi (Figure 3.19(c)), l'intensité des émissions optimisées est réduite, entre 4 et 10 heures, de 6,4 à 14,4% par rapport au cadastre *a priori*. Les réductions les plus importantes sont obtenues à 5 et 6 heures (13,2 et 14,5%). Au total, la masse émise est inférieure de 9,5% pour le profil optimisé par rapport au profil *a priori*. Le profil optimisé comporte un pic d'émissions entre 7 et 9 heures (flux supérieurs à $2{,}5.10^{12}$ moléc.cm^{-2}.s^{-1}) alors que les flux *a priori* sont quasiment constants entre 5 et 9 heures (flux compris entre 2,7 et 2,9 10^{12} moléc.cm^{-2}.s^{-1}). On peut évoquer la possibilité que ce pic corresponde au trafic de "loisir" du samedi matin, plus tardif et moins intense que le trafic dû aux déplacements de travail en semaine.

Le dimanche (Figure 3.19(d)), l'impact de l'optimisation est plus faible que le samedi : les flux optimisés sont augmentés de 12 et 10,5% à 8 et 9 heures par rapport aux flux *a priori* et différent de ceux-ci de moins de 5% de 4 à 7 heures. La masse émise entre 3 et 10 heures est augmentée de moins de 6% après optimisation. Le profil médian du type "dimanche ou jour férié" a été élaboré à partir de 18 dimanches plus les 14 juillet 1999 et 1998 et le samedi 15 août 1998. Un sixième des flux optimisés concernent donc des jours très particuliers pour le trafic. Par exemple, le mardi 14 juillet 1998 a pu donner lieu à des retours de "pont" (depuis le samedi 11), bien que ceux-ci influencent en principe les émissions de la soirée.

Finalement, l'allure des profils optimisés diffère peu de celle des profils *a priori* qui étaient cependant supposés comporter une grande incertitude.

3.8.4 Variabilité intra-type de jour

Nous cherchons tout d'abord à caractériser la variabilité des émissions que nous avons optimisées par rapport à la variabilité des émissions *a priori*. Cette dernière est loin d'être négligeable puisque notre connaissance des émissions nous permet de distinguer les jours de semaine des Samedis et Dimanche, ainsi qu'une différence notable entre Juillet et Août (sur le trafic principalement). Cette variabilité est représentée pour un profil de flux d'émissions de NO sur la Figure 3.20.
A gauche sur la Figure 3.20, les six profils typiques utilisés sont représentés. On peut noter la variabilité déjà prise en compte dans les simulations. Cette variabilité s'applique aussi aux autres composés émis. A droite, les résultats de la climatologie sont reportés et comparés aux profils "extrèmes" : "semaine en Juillet", les plus fortes émissions et "dimanche en août" les plus faibles (pour le trafic). La courbe tiretée rouge représente la médiane des flux de NO émis et optimisés. Les rectangles rouges représentent le percentile 50 et leur continuation le percentile 80. On note tout d'abord que l'optimisation n'a pas ajouté de variabilité vers les valeurs faibles des flux d'émissions. Par contre, une variabilité non négligeable a été ajoutée vers les valeurs plus fortes. Un examen des paramètres liés à cette forte majoration de la variabilité correspond à une prise en compte pas assez fine du taux de croissance de la hauteur de la couche limite à 05 :00 et 06 :00 TU : la dilution brutale des concentrations tend à provoquer une réaction exagérée du processus d'optimisation. Hormis ces deux heures particulières, le percentile 50 du profil médian optimisé correspond à des flux d'émission variant de -28 à +23% autour de la médiane. Cette variabilité est donc du même ordre de grandeur que l'écart d'intensité *a priori* de 20% entre le mois d'août et le mois de juillet d'une part et que la différence de 35% dans la masse émise de 3 à 10 heures entre les jours de la semaine et le type "dimanche ou jour férié" d'autre part. La variabilité de l'ensemble des flux optimisés paraît donc réaliste.

3.8.5 Variabilité spatiale

La Figure 3.21 indique, pour les heures optimisées et dans chaque maille, la correction la plus fréquente sur l'ensemble des jours.
On remarquera tout d'abord que, quelle que soit l'heure considérée, les corrections les plus souvent effectuées sont des réductions des émissions de NO_x. En fonction de leur intensité : les flux les plus forts (urbains) sont les plus souvent modifiés (jusqu'à 65%), les flux périphériques toujours dans une moindre mesure (toujours moins de 55%). Le "pic" de modification est calculé à 07 :00TU, pic de trafic, et correspond à une réduction quasi systématique sur le domaine.

3.8.6 Variabilité jour à jour

L'ensemble des résultats précédents nous a permis de "valider" les émissions utilisées en Ile de France. L'inversion a permis de modifier les profils par rapport à des mesures et pour les différents types de profils prescrits.
Mais la méthodologie mise en place permet plus : nous allons maintenant chercher à affiner les résultats et à apporter une information dont on ne dispose pas habituellement : **la caractérisation des profils diurnes d'émissions jour à jour**.
En effet, si l'on étudie la réalité par rapport à l'outil employé, on peut imaginer qu'un profil de type "semaine" est loin d'être exhaustif lorsque l'on observe la variabilité du trafic quotidien en région parisienne. Le trafic comptant pour la majorité des émissions de NO_x (53% des émissions totales

dans le cadastre *a priori*), on peut supposer que les profils temporels des différents jours de la semaine présentent des caractéristiques particulières.
Pour différencier chacun des cinq jours de la semaine dans le cadastre optimisé, nous comparons le profil médian de chaque jour au profil médian de toute la semaine. Afin de séparer la variabilité due aux différences d'intensité entre les mois de la variabilité éventuelle due aux diférences entre les jours, les médianes ont été calculées séparément pour juillet et août. Au total, on compte dix-huit lundis (8 en juillet, 10 en août), jeudis (10 en juillet, 8 en août) et vendredis (10 vendredis en juillet, 8 en août), seize mardis (7 en juillet, le 14 juillet 1998 étant un mardi, 9 en août) et seize mercredis (8 en juillet, le 14 juillet 1999 étant un mercredi, 8 en août).
Pour le mois de juillet, les profils temporels médians des lundi, mardi, jeudi et vendredi sont très semblables au profil temporel médian de la semaine : les écarts sont inférieurs à 3,2% à toute heure, l'écart maximal étant de 3,2% le jeudi à 3 heures (Figure 3.22). Le profil temporel du mercredi est également très proche du profil typique de la semaine (différences inférieures à 1,6%) sauf à 6 heures où l'écart atteint +6,3% (Figure 3.22). Les flux d'émission du mercredi ne sont donc pas quasiment constants de 5 à 8 heures (à environ 3,6 moléc.cm^{-2}.s^{-1}), comme avec le profil hebdomadaire, mais forment un pic à près de 3,9 moléc.cm^{-2}.s^{-1} à 6 heures.
Au mois d'août, les profils temporels médians des mardi, jeudi et vendredi sont très semblables au profil temporel médian de la semaine : les écarts sont inférieurs à 4,5% à toute heure, l'écart maximal étant de 4,4% le vendredi à 6 heures (Figure 3.22). Le profil temporel du mercredi conserve l'allure du profil hebdomadaire mais avec une intensité, entre 4 et 8 heures, supérieure de 5,4 à 8,7%. Le profil temporel du lundi est également très proche du profil typique de la semaine (différences inférieures à 4,3%) sauf entre 4 et 5 heures où l'écart atteint +6 et +11,7% (Figure 3.22). Les flux d'émission du lundi présentent donc, à 5 heures, un pic à près de 3,4 moléc.cm^{-2}.s^{-1}. On peut en proposer une explication en terme de trafic de retour de longs week-end. En effet, dans l'ensemble des 10 lundis du mois d'août se trouvent quatre lundis susceptibles d'être des jours de trafic dû aux vacanciers (les lundis 2 et 30 août 1999 et les lundis 3 et 31 août 1999) ainsi que les deux lundis 16 août 1999 et 17 août 1998 clôturant les week-ends du 15 août. Finalement, les cinq jours de la semaine ne présentent pas de caractéristiques moyennes très marquées par rapport aux profils hebdomadaires.

3.8.7 Synthèse des résultats climatologiques

L'optimisation des émissions de NO_x matinales sur les mois de juillet et août 1998 et 1999 a permis d'obtenir des concentrations d'ozone en meilleur accord avec les mesures. Le biais et l'erreur moyenne ont été diminués de 2 et 1,8% et la corrélation a été améliorée de 1,2%. Après avoir comparé les 7 août 1998 et 16 juillet 1999, on considère que la correction de la hauteur de couche-limite, importante pour l'étude de journées spécifiques (c'est-à-dire comportant des situations météorologiques extrêmes), n'est pas essentielle pour l'inversion d'une longue période comportant peu de situations exceptionnelles. L'examen de la médiane de l'ensemble des flux optimisés montre que la variabilité du percentile 50, regroupant donc la moitié des jours inversés, est comparable à celle qui a été introduite dans le cadastre *a priori* (différence entre les mois et les types de jour). Les corrections les plus fréquemment effectuées sont réparties spatialement selon l'intensité des flux d'émission. Après optimisation, les flux *a priori* les plus intenses du centre du domaine sont le plus souvent diminués de 65% ou plus alors que l'intensité de la majorité des flux moins intenses en périphérie est modifiés de moins de 55%. Pour tous les types de jour (semaine, samedi et "dimanche ou jour férié"), les écarts entre profils optimisés et *a priori* sont inférieurs à 15%. Cette valeur rejoint les résultats obtenus lors

de l'évaluation du cadastre d'émission effectuée grâce aux mesures aéroportées au cours de la campagne [*ESQUIF* (2001)] : l'erreur globale sur les émissions de NO_x avait alors été estimée à 15% au maximum. Enfin, la comparaison entre le profil obtenu pour toute la semaine et celui de chacun des cinq jours ne fait apparaître de caractéristique particulière que pour les lundis et les mercredis du mois d'août.

Émissions a priori et optimisées	[NO] a priori, optimisé et mesuré

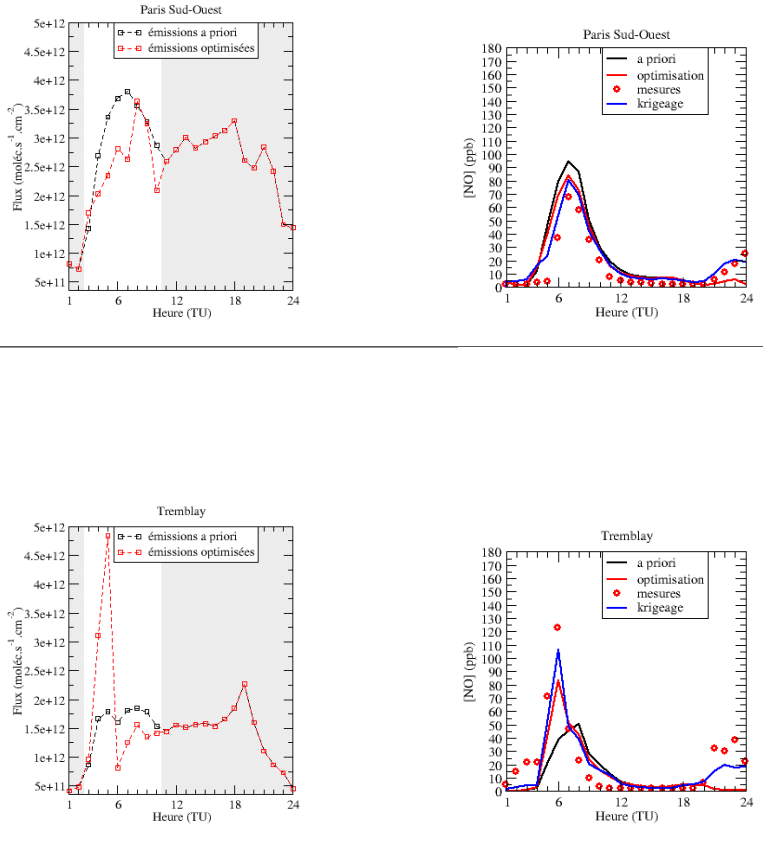
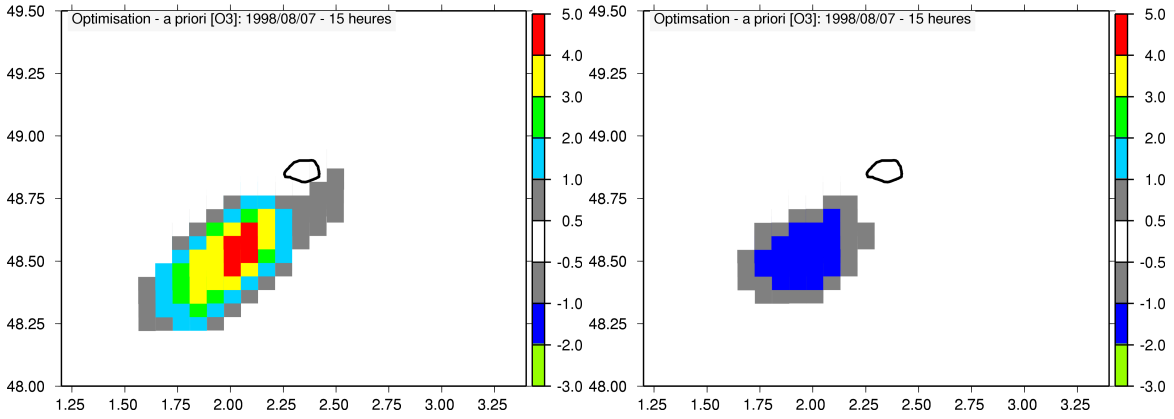
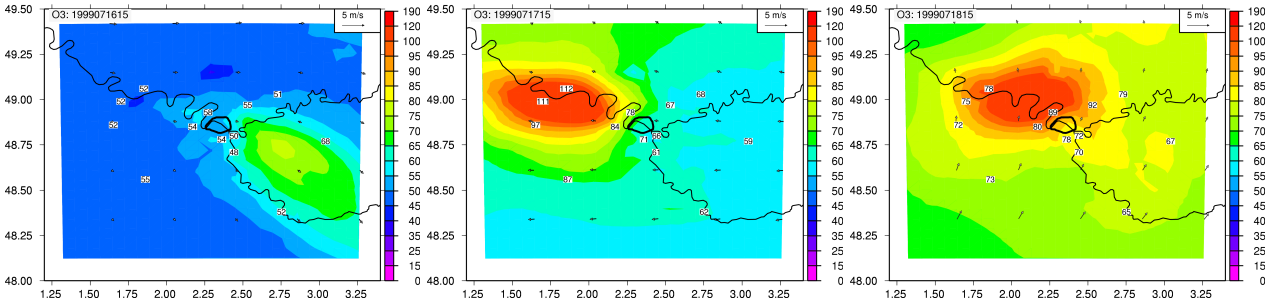


FIG. 3.14 – *Évolutions temporelles des émissions et des concentrations de monoxyde d'azote le 7 août 1998 dans les maille Paris Sud-Ouest et Tremblay. Les zones grisées correspondent aux heures non optimisées.*



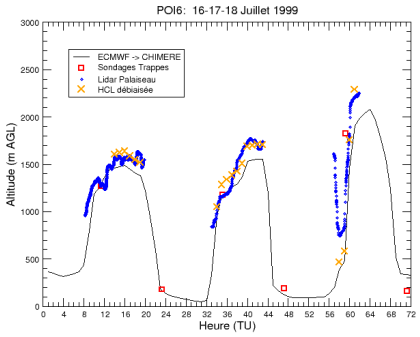
(a) Rapport COV/NOx variable (b) Rapport COV/NOx fixé

FIG. 3.15 – $[O_3]$*(optimisé) -* $[O_3]$*(a priori)- 7 août 1998 à 15h TU.*

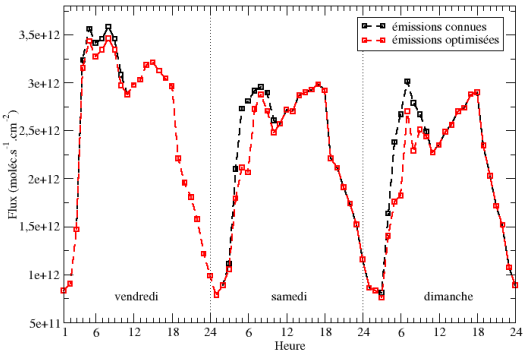


(a) vendredi 16 (b) samedi 17 (c) dimanche 18

FIG. 3.16 – *Concentrations d'O_3 (en ppb) en surface les 16, 17 et 18 juillet 1999 à 15 heures TU. Les vecteurs indiquent le sens et l'intensité du vent, les étiquettes correspondent aux valeurs relevées par des stations de mesure AIRPARIF. En été, TU = heure légale - 2 heures.*

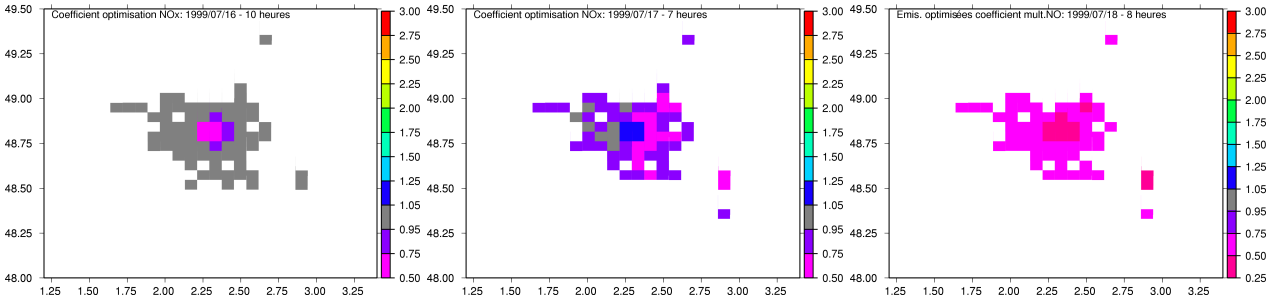


(a) *Hauteur de couche-limite pendant la POI 6 d'ESQUIF : comparaison entre les mesures et les valeurs fournies au modèle par l'ECMWF.*



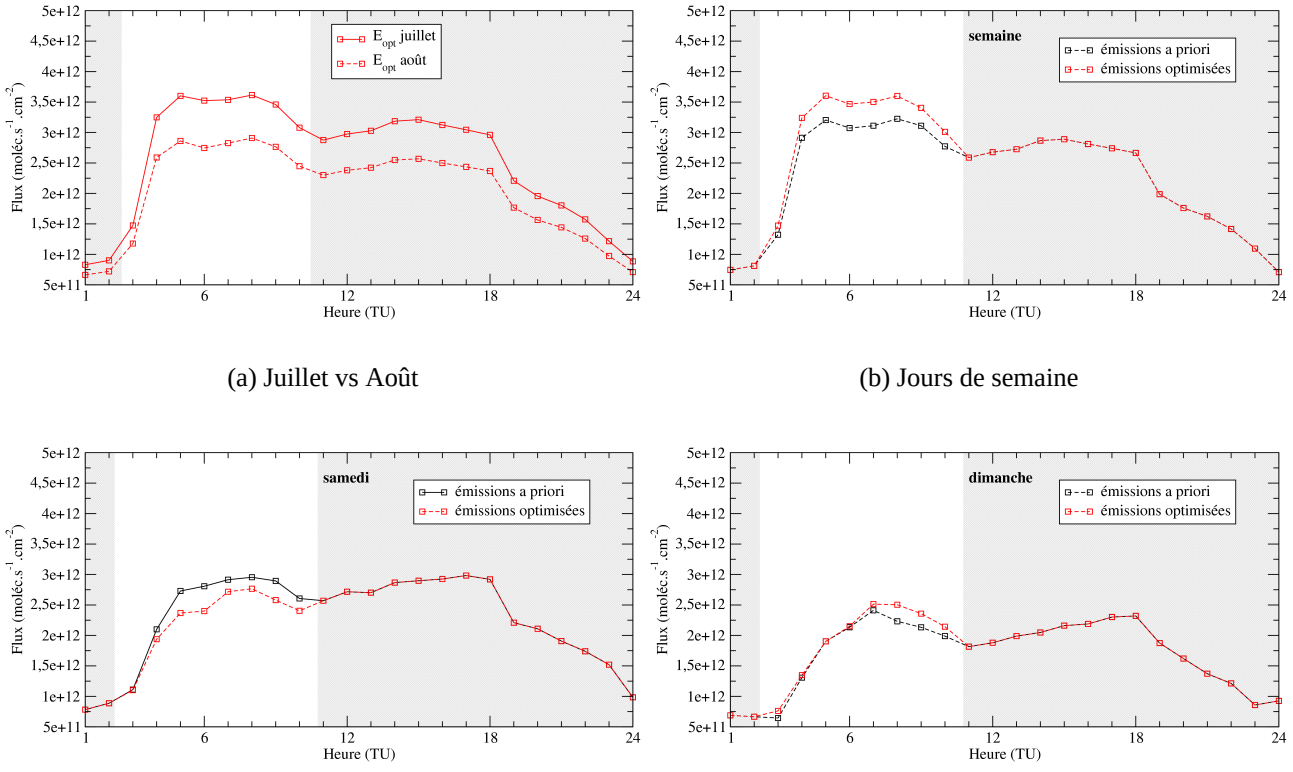
(b) *Profil temporel moyen sur tout le domaine du flux d'émission de NO du 16 au 18 juillet 1999. Les zones grisées correspondent aux heures non optimisées.*

FIG. 3.17 – *Correction de la hauteur de la couche limite et flux de No optimisés pour la POI6 : 16, 17 et 18 Juillet 1999*



(a) vendredi 16 à 10 heures TU (b) samedi 17 à 7 heures TU (c) dimanche 18 à 8 heures TU

FIG. 3.18 – *Répartition spatiale des corrections pour les NO_x et intensité des émissions de NO (représentant 90% des flux de NO_x) les 16, 17 et 18 juillet 1999.*



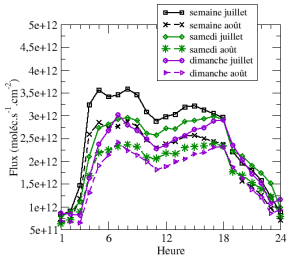
(a) Juillet vs Août

(b) Jours de semaine

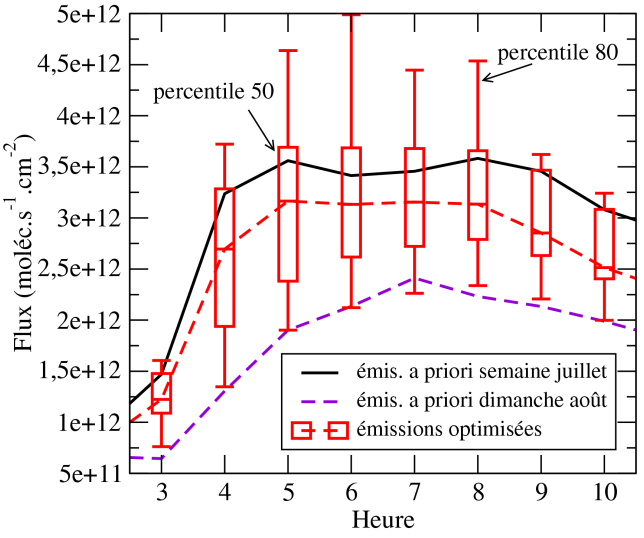
(c) Samedi

(d) Dimanche

FIG. 3.19 – *Profils temporels médians des flux d'émission de NO optimisés moyennés sur tout le domaine, obtenu pour différentes périodes. Les zones grisées correspondent aux heures non optimisées.*

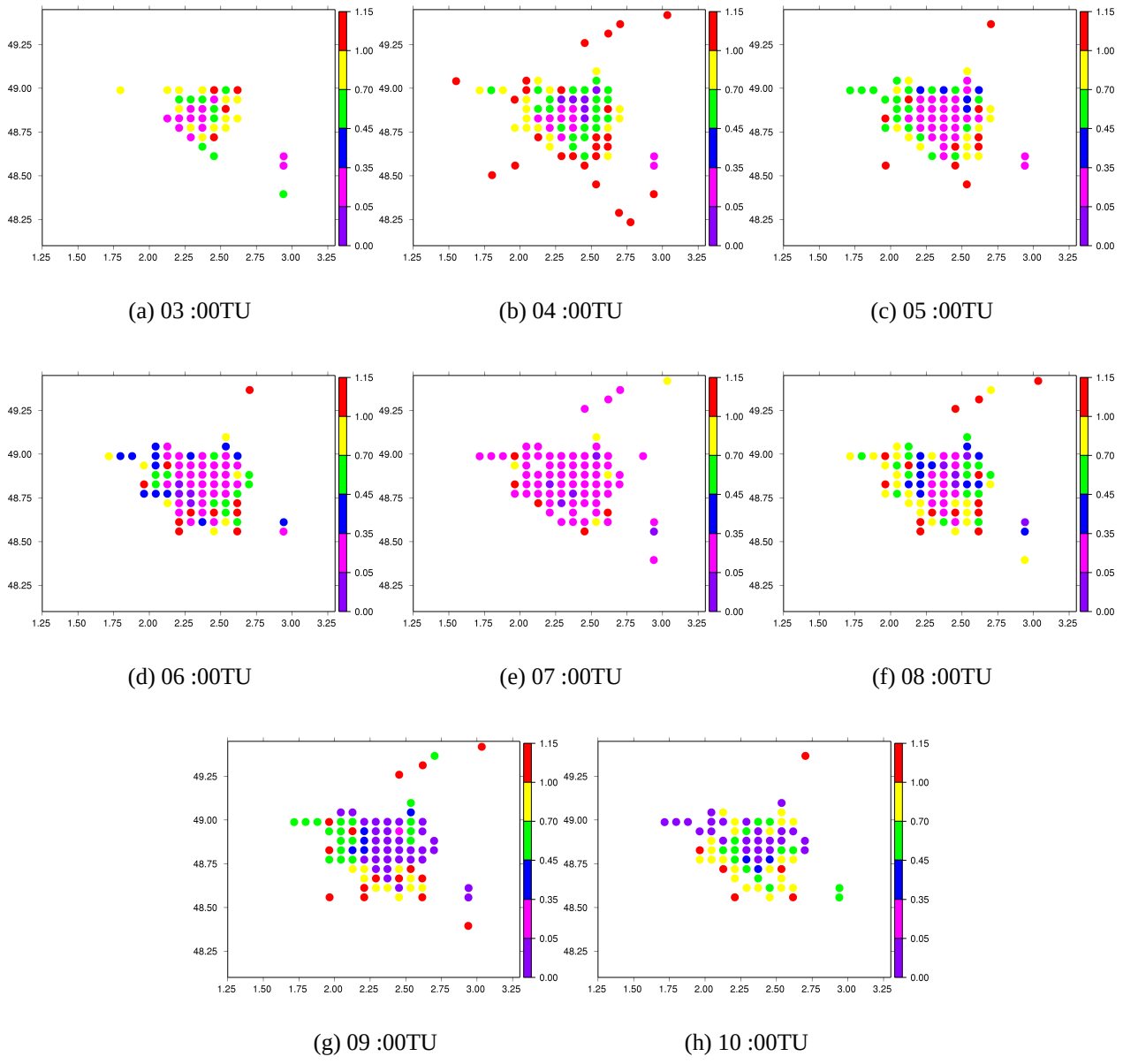


(a) Profils temporels a priori des flux d'émission de NO moyennés sur tout le domaine, selon les différents types de jour pour les mois de juillet et août 1998 et 1999.



(b) Profil temporel médian des flux d'émission de NO moyennés sur tout le domaine, obtenu sur les périodes du 1er juillet au 31 août 1998 et 1999.

FIG. 3.20 – *Variabilité des émissions a priori et optimisées*



(a) 03 :00TU (b) 04 :00TU (c) 05 :00TU

(d) 06 :00TU (e) 07 :00TU (f) 08 :00TU

(g) 09 :00TU (h) 10 :00TU

FIG. 3.21 – *Représentation spatiale des corrections les plus fréquemment appliquées aux flux d'émission de NO_x sur les périodes du 1er juillet au 31 août 1998 et 1999. En été, TU = heure locale - 2 heures.*

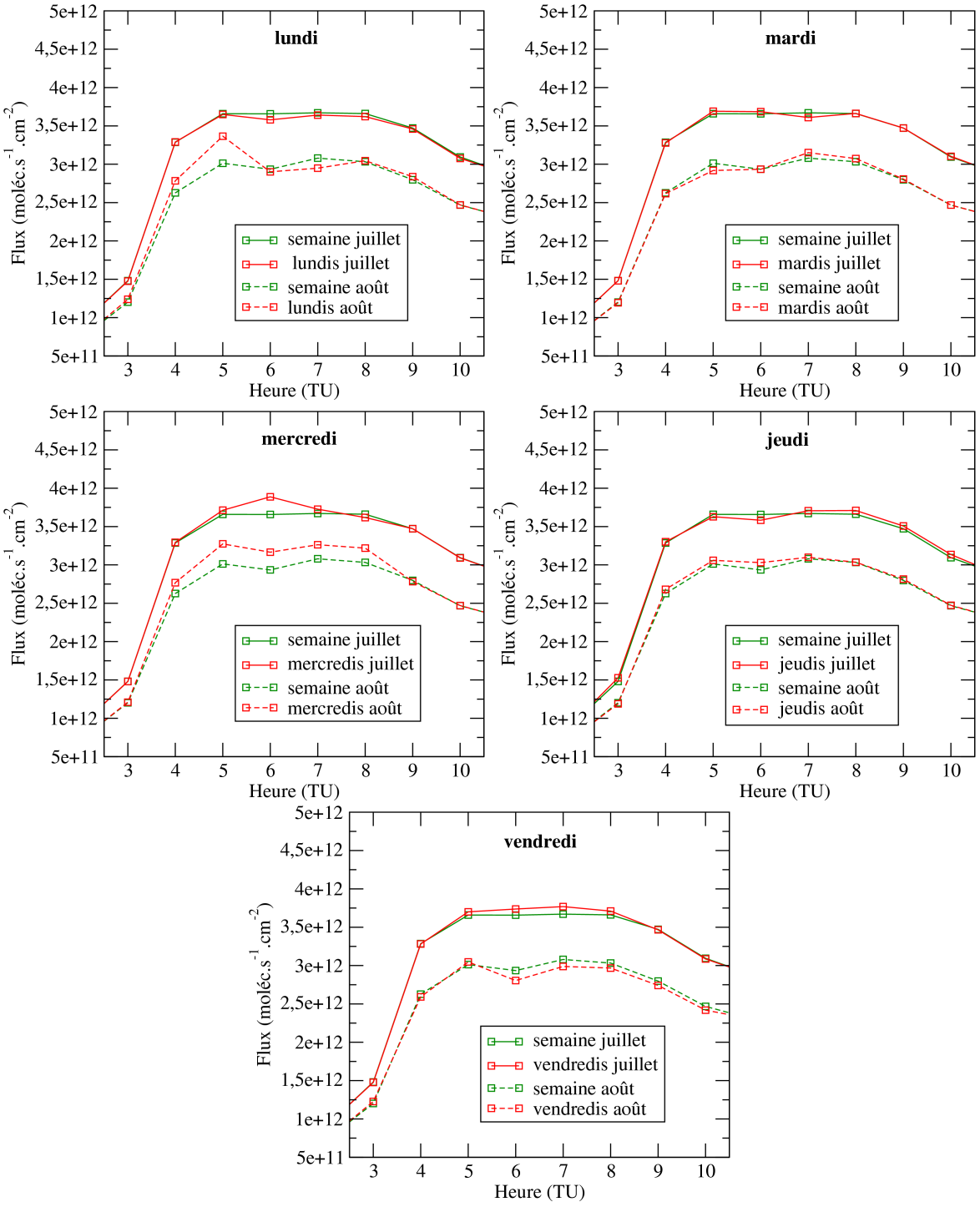


FIG. 3.22 – *Profils temporels médian sur tout le domaine des flux d'émission de NO optimisés pour les jours de type "semaine" obtenus sur les mois de juillet et août 1998 et 1999. En été, TU = heure légale - 2 heures.*

Chapitre 4

Modélisation Monte-Carlo sous contrainte

4.1 Adaptation et mise en oeuvre de la méthode

Dans un premier temps, l'étude a porté sur trois jours d'observations intensives (POI's) au cours de la campagne d'ESQUIF (7 août 1998, POI2 ; 16 et 17 juillet 1999, POI6), pour lesquels le panache urbain était bien développé et pour lesquels des mesures aéroportées d'ozone et des composés azotés (NOx, NOy) autour de Paris étaient disponibles, Figure 4.1.
Le deuxième volet de cette étude a porté sur une période plus longue de deux mois, juillet et août 1999, pour laquelle seules les mesures issues du réseau d'AIRPARIF étaient disponibles. La période choisie est identique à celle considérée dans l'étude de modélisation inverse utilisant l'adjoint.
La mise en oeuvre de la méthode de Monte Carlo Bayésienne pour ce deux cas d'étude est décrite ci-dessous. Le modèle utilisé est CHIMERE, dans la même version que celle utilisée pour l'étude utilisant l'adjoint.

4.1.1 Attribution des distributions de probabilité à priori

Un problème délicat dans la mise en place de la méthode de Monte Carlo réside dans la nécessité de fixer les distributions de probabilité à priori pour les paramètres d'entrée de CHIMERE, i.e. pour les émissions, les coefficients de vitesse des réactions et pour les fréquences de photolyse. La forme la plus naturelle, pour une distribution de probabilité contenant uniquement des valeurs positives, est la distribution log-normale qui a été retenue pour toutes les grandeurs à l'exception du vent et de la température. Concernant les émissions anthropiques, des études dédiées concernant la propagation d'incertitude dans des modèles d'émissions [*Kuhlwein and Friedrich* (2000)], des expériences spécifiques d'évaluation d'émissions à partir d'observations en Ile de France [*Vautard et al.* (2003a)] et dans le Sud de l'Allemagne [*Kuhlwein et al.* (2002)], [*Mannschreck et al.* (2002)] ont été recoupées afin d'estimer leur incertitude à 40% (avec un niveau de confiance de 1 sigma). Pour les vitesses de réactions, les incertitudes indiquées dans les compilations de [*Atkinson et al.* (1997)] et [*DeMore et al.* (1997)] ont été utilisées indiquant des valeurs pour les incertitudes entre 10 et 30%. Les incertitudes sur les paramètres météorologiques sont déduites des différences aléatoires entre observations et analyses météorologiques pour les deux POI's d'ESQUIF. Par exemple, l'incertitude sur la hauteur de la couche limite est de 40%, correspondant aux différences importantes constatées entre hauteurs de la couche limite utilisées dans CHIMERE et dérivées à partir de profils lidar (aérosols) et de radiosondages. Les valeurs retenues sont indiquées dans le Tableau 4.1, ainsi que la forme de la distribution

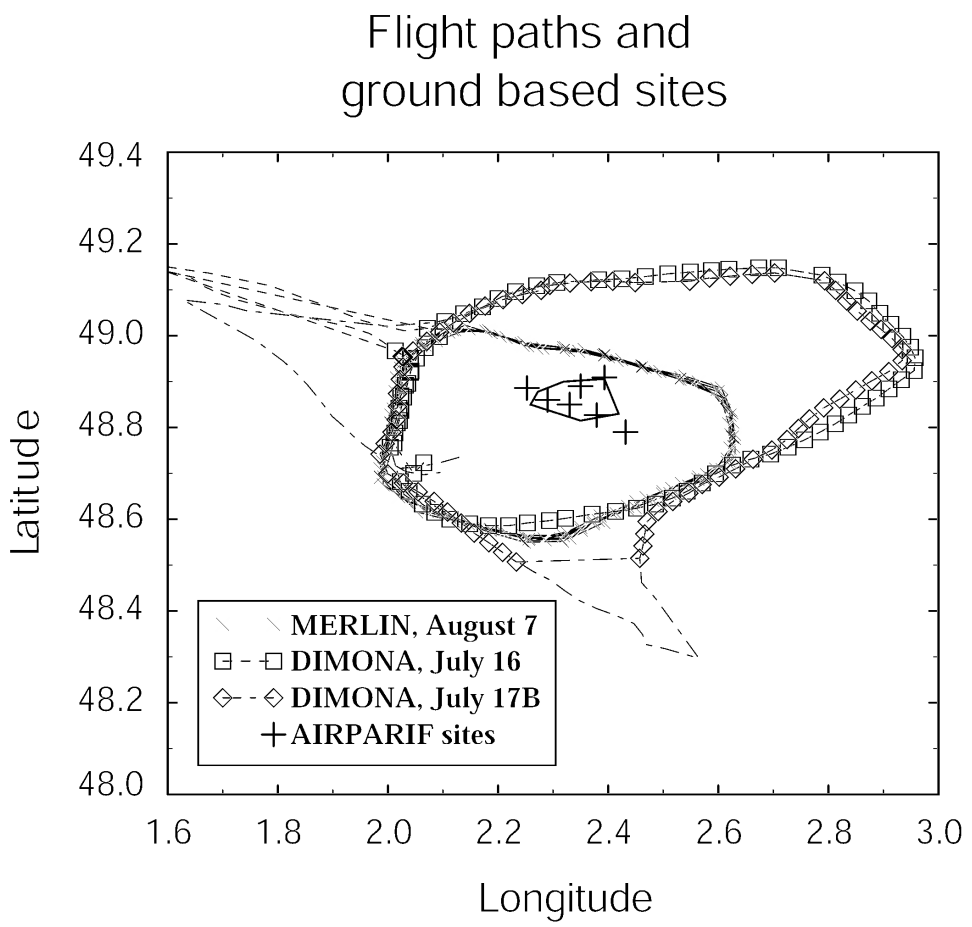


FIG. 4.1 – *Carte de la région Parisienne et des différentes vols utilisés pour cette étude.*

des incertitudes (normale ou log-normale).
Il est important de souligner que ces distributions à priori ont été dérivées pour représenter la totalité de l'incertitude d'une variable d'entrée, donc à la fois les erreurs systématiques et aléatoires. Par ailleurs, elles représentent des erreurs moyennes sur tout le domaine, puisque pour une variable d'entrée et une simulation de Monte Carlo donnée, une seule perturbation est appliquée pour tout le domaine et toute la durée de la simulation.

4.1.1.1 Choix des observations pour les POI's :

Les observations utilisées pour contraindre les simulations de Monte Carlo doivent contenir un maximum d'informations sur les émissions des NOx et COVs émis dans l'agglomération Parisienne. Dans cette optique, les mesures de composés azotés (NO, NOx, NOy), contentant de l'information directe sur les émissions des NOx et des mesures d'ozone, contenant une information indirecte notamment sur les émissions des COVs, via les processus de production photochimique de l'ozone, ont été utilisées. De plus, les observations doivent idéalement avoir lieu à une certaine distance des sources d'émissions, afin de s'affranchir de leur hétérogénéité spatiale. Enfin, les mesures doivent permettre d'estimer les concentrations de fond des polluants afin de pouvoir remonter aux sources.

Paramètres	Incertitudes
Emissions :	
COV anthropiques	± 40 %
NOx anthropiques	± 40 %
COV biogéniques	± 50 %
Vitesses de réactions :	
NO + O3	± 10 %
NO2 + OH	± 10 %
NO + HO2	± 10 %
NO + RO2	± 30 %
HO2 + HO2	± 10 %
RO2 + HO2	± 30 %
RH + OH	± 10 %
CH3COO2 + NO	± 20 %
CH3COO2 + NO2	± 20 %
PAN + M	± 30 %
Fréquences de photolyse et rayonnement :	
Flux actinique	± 10 %
J(O3 → → 2 OH)	± 30 %
J(NO2 → NO+O3)	± 20 %
J(CH2O → CO+2 HO2)	± 40 %
J(CH3COCO → ....)	± 50 %
J(carbonyle issu d'aromatiques)	± 40 %
Paramètres météorologiques :	
Vent zonal	± 1 m/s (normale)
Vent vertical	± 1 m/s (normale)
Hauteur de la couche de mélange	± 40 %
Température	± 1.5 K (normale)
Humidité relative	± 20 %
Coefficient vertical de mélange	± 50 %
Vitesses de dépôt	± 25 %

TAB. 4.1 – *Incertitudes fixées pour les paramètres d'entrées du modèle (à 1 sigma) ; la distribution de probabilité des paramètres est supposée log-normale, excepté pour le vent et la température, où la distribution est normale.*

Ainsi, cette étude s'est appuyée sur les mesures des vols circulaires autour de l'agglomération Parisienne lors des POI's 2 et 6 de la campagne d'ESQUIF, qui donnent une information à la fois sur les concentrations dans le panache et dans le fond. En pratique, pour chaque vol, à la fois pour les concentrations observées et simulées, 30% (70%) des valeurs les plus élevées (plus faibles) ont été moyennées afin de fournir une valeur représentative du panache (du fond). Cette technique permet d'éviter que des erreurs dans la simulation de la direction du panache et des conditions de fond n'affectent les résultats d'inversion.

Ainsi, les différences entre la moyenne des concentrations du panache et du fond, pour OX (= O3 + NO2), NOy et NOx ont été retenues comme contraintes pour le modèle. Le choix des OX plutôt que O3 s'explique par le fait que cette grandeur représente mieux la production photochimique des photo-oxydants que l'ozone seul, car on s'affranchit des transformations rapides entre O3 et NO2.

De plus, une valeur moyenne pour les concentrations des OX au sein de l'agglomération (OX urbain), a été rajoutée dans le calcul de la fonction de coût, car elle contient également de l'information sur les

émissions via le taux de production photochimique locale. La moyenne est déterminée sur six stations urbaines de fond du réseau AIRPARIF situées à Paris et dans la Petite Couronne et entre 14 et 18 h TU, i.e lorsque la formation d'ozone accumulée est maximale. La Figure 4.1 montre l'emplacement des sites AIRPARIF et les trajectoires de vol pour les trois jours de POI's. Ces quatre contraintes constituent la contrainte appelée "contrainte standard".
Afin de rendre nos résultats comparables avec ceux obtenus par la méthode adjointe, une contrainte supplémentaire, basée sur les mesures de NO à Paris et en Petite Couronne a été rajoutée. Il s'agit de la moyenne pour six stations d'AIRPARIF entre 4 et 10h TU, correspondant à la durée habituelle en moyenne du pic de NO du matin.

4.1.1.2 Choix des observations pour la période climatologique :

Dans le cas des simulations de période climatologique (deux mois), seules les observations AIRPARIF ont été utilisées. De la même facon que pour les deux POI's traitées dans cette étude, les mesures de composés azotés (NO) et des mesures d'ozone contenant des informations directes et indirectes sur les émissions de NOx et COVs ont été utilisées. Le premier set d'observations pris en compte concerne les mesures d'ozone obtenues par les stations AIRPARIF situées en zone rurale. La technique utilisée dans cette étude consiste à déterminer une concentration journalière d'ozone dans le panache et de fond. Pour cela, lors de chaque journée, deux maximums d'ozone observés sur deux stations distinctes sont moyennés (ozone dans le panache) ; les trois stations présentant les valeurs d'ozone journalières les plus faibles permettent de calculer une valeur de l'ozone de fond. Deux conditions sont imposées en amont dans l'étude des 62 journées de la période climatologique. Tout d'abord, une différence entre les valeurs calculées pour l'ozone dans le panache et de fond de plus de 10 ppb est imposée. L'objectif est de ne considérer que les jours avec présence d'un panache d'ozone. Ensuite, la sélection s'effectue sur les journées où le maximum d'ozone observé sera localisé sur la même station où une de ces deux stations consécutives que le maximum d'ozone simulé. Il s'agit ici de ne considérer que les jours où le modèle est capable de représenter spatialement le panache. Par conséquent, les journées d'observation considérées et confrontées aux données simulées pourront différer en fonction des simulations perturbées. Le deuxième jeu de données AIRPARIF pris en compte dans notre étude concerne les concentrations de NO à Paris et en Petite Couronne (6 stations). Les valeurs de NO sont moyennées pour ces 6 stations entre 4 et 10H TU, de la même facon que pour l'étude des deux POI's.

4.1.2 Estimation des incertitudes :

La solution de l'équation 2.1 requiert la connaissance de l'incertitude associée aux observations. De facon générale, cette incertitude est liée à la fois à une erreur instrumentale et à une incertitude de représentativité de la mesure ponctuelle pour la maille du modèle. Le cas échéant, on doit tenir compte de l'incertitude des mesures dans le panache et du fond. Les incertitudes relatives pour les différentes observations (ou plutôt agrégats d'observations) sont indiquées dans le Tableau 4.2, pour les différents jours des deux POI's et pour l'étude climatologique.
Pour les mesures aéroportées, l'incertitude liée à la représentativité verticale des observations est le terme prépondérant de l'incertitude totale. Elle provient du fait que les mesures utilisées ont été obtenues seulement à une altitude spécifique ($\approx$300 m au-dessus du sol), tandis que les simulations donnent une image plus moyenne (en altitude) de la structure verticale de la couche de mélange, due à la difficulté de résoudre finement les processus de mélange vertical. Une estimation de l'incertitude

Terme fonction coût	Plate-forme d'observation	Incertitudes relatives
POI6 :		
Ox panache - Ox fond	METAIR/DIMONA	0.29 / 0.25
NOy panache - NOy fond	METAIR/DIMONA	0.23 / 0.36
NOx panache - NOx fond	METAIR/DIMONA	0.34 / 0.67
Ox Urbain - Ox fond	AIRPARIF	0.22 / 0.58
NO urbain	AIRPARIF	0.35 / 0.24
POI2 :		
O3 panache - O3 fond	MF/MERLIN	0.30
NO panache - NO fond	MF/MERLIN	0.68
Ox Urbain - Ox fond	AIRPARIF	1.043
NO urbain	AIRPARIF	0.2
Etude climatologique		
O3 panache - O3 fond	AIRPARIF	0.29
NO urbain	AIRPARIF	0.32

TAB. 4.2 – *Observations et incertitudes (1 sigma) utilisées pour la construction des fonctions de coût*

verticale des mesures a alors été obtenue à partir de plusieurs profils verticaux effectués lors des vols. Pour les observations issues du réseau AIRPARIF, les incertitudes instrumentales sont prépondérantes. Elles sont de l'ordre de 3 à 4 ppb [rapport ESQUIF, 2002] pour les mesures de NO et de l'ozone, et relativement indépendantes des concentrations mesurées. Concernant les observations d'Ox en ville, une incertitude forte est attribuée aux mesures de NO2, qui montrent une interférence avec d'autres espèces azotées (PAN, HNO3,...). L'étendue de l'erreur a été estimée à partir de simulations de ces composés avec CHIMERE (en absence d'observations). Pour les observations de NO en ville, une correction a été appliquée, et son incertitude estimée, afin de palier au fait que les concentrations mesurées à environ 3 m du sol soient comparées à des concentrations simulées représentatives des premiers 50 m. Cette correction a été effectuée à partir des observations au premier étage de la tour Eiffel qui se situe à environ 50 m au dessus du sol.
Pour l'étude climatologique, les incertitudes indiquées par AIRPARIF ont été interprétées comme incertitudes systématiques. C'est un choix conservateur, puisque ces termes contiennent à la fois une composante systématique et aléatoire. En revanche, il n'a pas été possible d'estimer un terme réaliste pour l'incertitude liée à la représentativité de l'ozone observé dans le panache.

4.2 Résultats

Dans cette partie, nous allons montrer les résultats sur l'inversion des émissions en Ile de France obtenue avec la méthode de Monte Carlo Bayésienne. Nous allons insister sur des éventuelles modifications des moyennes des émissions à posteriori et de leur incertitude, par rapport à l'estimation à priori. Il s'agira donc d'analyser les fonctions de densité de probabilité (FDP) des paramètres d'entrée, à priori et à posteriori, les dernières étant contraintes par les observations via un facteur de pondération $p(Y_k|O)$ définie dans l'équation 2.1.
Nous allons traiter, dans un premier temps, les deux POIs d'ESQUIF afin de déterminer de quelles facons les observations contraignent le modèle et l'améliorent. Nous allons considérer séparément les effets des contraintes standard (production de OX (= O3 + NO2), NOy et NOx dans le panache), de la contrainte par le NO urbain et enfin, de ces deux contraintes cumulées. Dans un deuxième temps,

nous allons traiter la période "climatologique" de Juillet et Août 1999, et considérer les contraintes par les observations de routine du réseau AIRPARIF.
Il est important de rappeler que nous dérivons des facteurs correctifs pour les différentes données d'entrée qui sont constants sur le domaine de CHIMERE et pendant la durée de la simulation (la journée ou la période de deux mois, dans notre cas).

4.2.1 Distribution à posteriori des paramètres d'entrée du modèle : Études des POIs

La Figure 4.2 représente les FDPs à priori et à posteriori des deux POIs pour les émissions de NOx, de COVs et pour le rapport de ces deux émissions. Le Tableau 4.3 détaille les déplacements de la moyenne des FDPs après modifications par les observations, la déviation standard de la fonction à posteriori, et la diminution relative de la variance pour les trois types de contraintes considérées. Les données relatives au vent et à la hauteur de la couche limite sont aussi indiquées dans cette table.

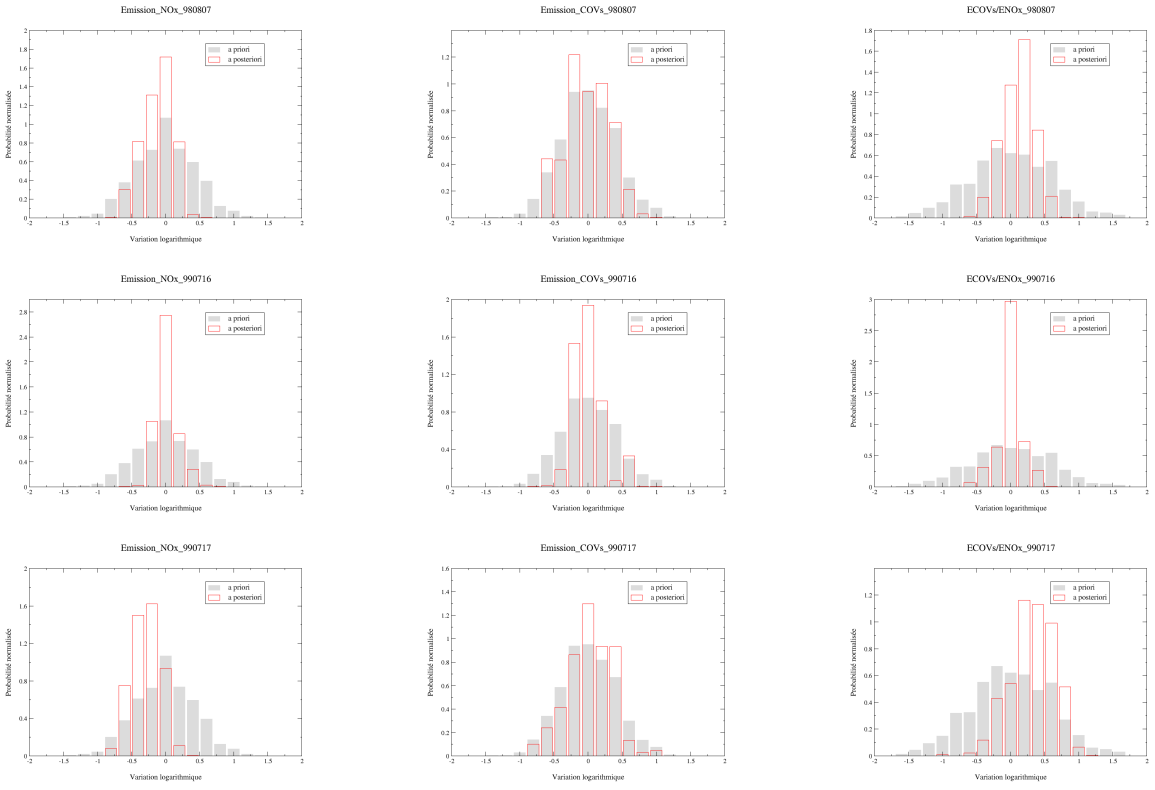


FIG. 4.2 – *Fonctions de densité de probabilité pour les deux POIs des paramètres d'entrée : émissions de NOx, de COVs et rapport de ces deux émissions. La distribution standard à priori est indiquée en bleu ; et la distribution alternative à posteriori est représentée en rouge. L'axe x représente la variation relative logarithmique du paramètre d'entrée étudié.*

Les distributions à priori des incertitudes des paramètres d'entrée sont représentées par des distributions normales (bleu), Figure 4.2. Les distributions à posteriori (en rouge) semblent plus "perturbées",

		Contraintes standard		Contraintes NO urbain		Contraintes standard + NO urbain	
POIs	Paramètres d'entrée	$\overline{x}$ et $\sigma(x)$ (%)	$\Delta(var)$ (%)	$\overline{x}$ et $\sigma(x)$ (%)	$\Delta(var)$ (%)	$\overline{x}$ et $\sigma(x)$ (%)	$\Delta(var)$ (%)
POI2	Émission NOx	- 1.8 $\pm$ 27	- 60	- 18 $\pm$ 24	- 69	- 12 $\pm$ 22	- 73
7 août	Émission COVs	3.1 $\pm$ 31	- 39	2 $\pm$ 39	- 2	0 $\pm$ 31	- 36
	ECOVs/ENOx	4 $\pm$ 26	- 80	18 $\pm$ 40	- 53	10 $\pm$ 23	- 85
	Vitesse du vent	0.19 $\pm$ 0.66	- 46	0.13 $\pm$ 0.96	+15	0.18 $\pm$ 0.71	- 38
	Hauteur couche limite	10 $\pm$ 32	- 33	1.6 $\pm$ 38	- 8	11 $\pm$ 33	- 31
POI6	Émission NOx	21 $\pm$ 26	- 61	- 26 $\pm$ 29	-53	0 $\pm$ 17	- 85
16 juillet	Émission COVs	6 $\pm$ 27	- 53	0 $\pm$ 39	-1	0 $\pm$ 23	- 67
	ECOVs/ENOx	- 16 $\pm$ 22	- 86	25 $\pm$ 48	- 33	0 $\pm$ 18	- 90
	Vitesse du vent	- 0.14 $\pm$ 0.66	- 44	0.14 $\pm$ 0.83	- 12	- 0.5 $\pm$ 0.58	- 57
	Hauteur couche limite	- 33 $\pm$ 37	- 11	8 $\pm$ 36	- 18	- 44 $\pm$ 35	- 23
POI6	Émission NOx	- 20 $\pm$ 26	- 64	- 26 $\pm$ 25	- 68	- 29 $\pm$ 21	- 76
17 juillet	Émission COVs	1 $\pm$ 33	- 28	3 $\pm$ 39	- 3	2 $\pm$ 32	- 34
	ECOVs/ENOx	20 $\pm$ 22	- 54	28 $\pm$ 42	- 52	30 $\pm$ 32	- 70
	Vitesse du vent	0.15 $\pm$ 0.6	- 56	0.18 $\pm$ 0.90	- 2.5	0.17 $\pm$ 0.6	- 56
	Hauteur couche limite	3.9 $\pm$ 33	- 28	8 $\pm$ 36	- 18	4 $\pm$ 31	- 40

TAB. 4.3 – *Modifications des distributions à posteriori des émissions, du vent et de la hauteur de la couche limite pour différentes contraintes et pour les deux POIs considérés.* $\overline{x}$ *et* $\sigma(x)$ *sontr respectivement la moyenne et la déviation standard.* $\Delta(var)$ *est la modification de la variance.*

moins symétrique mais leur distribution se rapproche d'une distribution log-normale ce qui permet de caractériser leur déviation standard (logarithmique) et la valeur moyenne de leur logarithme.

4.2.1.1 7 août 1998 (POI2)

Pour la journée du 7 août 1998 (POI2), les déviations standard des distributions des émissions de NOx et de COVs se réduisent fortement lorsque les contraintes standard sont appliquées (Tableau 4.3), alors que les valeurs de leurs émissions ne sont que très faiblement modifiées (- 1.8% pour les émissions de NOx et +3.1% pour les émissions de COVs). Par ailleurs, on observe une vitesse de vent et une hauteur de couche limite plus élevée. Lorsque l'on applique les contraintes des stations urbaines de NO, la distribution des émissions de COVs n'est que très légèrement modifiées alors que les émissions de NOx sont baissées de 18% avec une déviation standard beaucoup plus faible (baisse de variance de -69%). Le vent est légèrement augmenté et la hauteur de couche limite n'est que très peu modifiée. L'information apportée influe donc directement sur les valeurs des émissions de NOx et très faiblement sur les émissions de COVs. Enfin, en cumulant les deux types de contraintes, les émissions de NOx à posteriori sont diminuées de 12% par rapport aux émissions à priori, mais cette diminution n'est pas significative, compte tenue de l'incertitude (déviation standard) sur cette estimation (à un sigma) de 22%. Par rapport à l'incertitude à priori fixée à 40%, ceci correspond à une réduction de 73%, en terme de variance. Les émissions de COVs à priori ne sont pas modifiées, leur déviation standard diminue à 31%, ce qui correspond à une diminution de variance de 36%, plus faible que celle pour les émissions de NOx. Ceci s'explique par le fait que les contraintes sur les émissions de NOx par les observations de NO sont plus "directes" que la contrainte des OX sur les émissions des COV. Le rapport des émissions de COVs sur les émissions de NOx n'est pas modifié de facon significative,

mais son incertitude (en terme de variance) est très fortement réduite (de 85%).

4.2.1.2 16 juillet 1999 (POI 6)

Dans le cas d'une application des contraintes standard à la journée du 16 juillet 1999 de la POI 6, les modifications des fonctions à posteriori se caractérisent par une augmentation des émissions des NOx (21%) et COVs (6%) et par une baisse de la vitesse du vent et de la hauteur de la couche limite. Toutes ces modifications des fonctions à posteriori des paramètres d'entrée s'accompagnent de diminutions importantes de leurs déviations standard. Lorsque l'on applique les contraintes des stations de mesures urbaines de NO, le modèle a tendance, à l'inverse, à diminuer les émissions de NOx, et à augmenter les valeurs de la HPBL et du vent. Cette contrainte impose aux émissions de NOx un comportement à l'opposé de celui des contraintes standard. En cumulant les deux types de contraintes, l'effet antagoniste des deux contraintes aboutit à une modification quasiment nulle des moyennes des émissions de NOx et COVs par rapport à la distribution à priori. Par contre, la diminution de la variance atteint des valeurs de 85% pour les émissions de NOx, de 67% pour les COVs, et de 90% pour le rapport des émissions des COVs sur les émissions de NOx.

4.2.1.3 17 juillet 1999 (POI 6)

Après application des contraintes standard, le 17 juillet se caractérise par une baisse des émissions de NOx de 20%, une stagnation des émissions de COVs et par une augmentation faible de la HPBL et plus importante du vent. Dans le cas avec les contraintes de NO urbain, le modèle diminue encore les émissions de NOx, augmente le vent et la HPBL, montrant un comportement similaire au cas avec contraintes standard. Par conséquent, le cumul des deux types de contraintes aboutit à une diminution des émissions de NOx (-29%), qui n'est cependant pas significative à un niveau de confiance de 95%, malgré une diminution de la variance de -76%. Les émissions de COVs restent presque inchangées, avec une diminution de leur variance plus faible, de 34%.

4.2.1.4 Trois journées étudiées

En conclusion, pour aucune des trois journées de POIs d'ESQUIF, les émissions de NOx à posteriori ne sont modifiées de facon significative (à deux sigmas), même si une diminution notable est obtenue pour un jour (le 17 juillet). Pour les émissions des COV, les modifications sont assez faibles pour les trois jours, en dessous de 2%. L'incertitude sur les émissions de NOx diminue très nettement (de 73 à 85%), celle sur les émissions moins fortement (de 34 à 67%). Pour les trois jours, l'adjonction d'une contrainte sur les concentrations de NO en milieu urbain agit de facon synergique ou antagoniste par rapport à la contrainte standard, selon les jours. En revanche, pour les trois jours, elle permet de diminuer l'incertitude sur les émissions de NO. L'information apportée par les observations semble moins modifier les émissions de COVs que les émissions de NOx excepté dans le cas du 16 juillet 1999 caractérisé par un régime chimique plus COVs limité que les deux autres jours étudiés. L'information apportée par les observations et son rôle dans la modification des émissions vont être discutés plus en détail dans la suite.

Espèces chimiques	P50 : Simulation de référence	P50 : Sans contrainte	P50 : Contraintes standard	P50 : Contrainte NO urbain	P50 : Contraintes standard + NO urbain	Obs.
POI2 : 7 août 1998						
Concentration moyenne dans le panache - concentration fond						
ΔOX, ppb	36.18	24.33	27	30.1	26.5	29.8
ΔNO, ppb	1.3	0.63	0.75	0.28	0.59	0.88
Concentration moyenne en milieu urbain ; OX urbain normalisé par OX fond						
NO, ppb	72.48	53.53	51.10	38.77	39.1	37.53
ΔOX, ppb	16.75	16.25	13.25	16.79	12.74	10.72

TAB. 4.4 –

4.2.2 Interprétation des résultats : confrontation modèle/observations

Les Tableau 4.4 (POI2) et Tableau 4.5 (POI6) permettent de confronter les observations aux simulations et d'analyser comment la modification des distributions des données d'entrée du modèle (à posteriori versus à priori) permet d'améliorer l'accord entre simulations et observations. On analysera ensuite, comment les modifications des données d'entrée du modèle affectent les différents types d'observations, ce qui permet d'obtenir des renseignements sur la nature de la contrainte appliquée aux distributions *à priori*.
Pour cela, les fonctions de densité de probabilité cumulative (FDPCs) des paramètres étudiés de sortie du modèle sont tracées pour des simulations de Monte Carlo avec et sans contraintes. De ces figures (non montrées ici) sont extraites les valeurs des paramètres de sortie pour une probabilité cumulée de 50% (P50 dans la table, correspondant à la médiane) qui sont ensuite confrontées aux observations.
A partir du Tableau 4.4 pour la POI2, on constate d'abord, sans surprise, que l'application des contraintes standard permet un meilleur accord avec les observations d'OX et de NO dans le panache, tandis que la contrainte de NO urbain induit un meilleur accord pour le NO urbain, toujours par rapport au système sans contraintes. Pour le cas des contraintes cumulées, des différences un peu plus fortes subsistent entre les valeurs P50 et les observations, de l'ordre de 10 à 30%. Ces différences sont nettement inférieures aux incertitudes (à un sigma) inhérentes aux observations indiquées dans le Tableau 4.2. On conclut alors, que les distributions à posteriori des données d'entrée du modèle (émissions, mais aussi vitesse de vent, hauteur de la couche limite, etc.) permettent un accord avec toutes les observations, à leur incertitude près. Comme nous le verrons, le même constat s'applique pour le 17 Juillet de la POI6 mais pas pour le 16 Juillet.

4.2.2.1 Accord pour les observations des composés azotés

A partir des Tableau 4.4 et Tableau 4.5, nous constatons, que les concentrations du NO urbain réagissent à peu près de facon linéaire par rapport aux modifications des émissions des NOx. En revanche, les simulations du NO dans le panache varient très fortement pour de faibles modifications des émissions des NOx. Deux explications qualitatives pour ce comportement peuvent être avancées, dont il faudrait préciser la portée quantitative par des tests supplémentaires. Premièrement, les concentrations dans le panache sont presque deux ordres de grandeur plus faibles que celles obtenues en milieu urbain. Elles seront donc très fortement influencées par les variations des processus de perte chimique et de dispersion dans les différentes simulations de Monte Carlo. Deuxièmement, pour des conditions de fortes concentrations de NOx (au-delà de quelques ppb de NOx), il existe une rétroac-

tion positive permettant d'expliquer une réponse plus que linéaire des changements de concentrations des NOx par rapport aux changements de leurs émissions. Plus le niveau des concentrations de NOx augmentent, plus on diminue les concentrations du radical OH via la réaction entre NO2 et OH et plus le temps de vie et le niveau des concentrations de NOx augmentent. Il est clair que cette relation plus que linéaire permet d'exercer une contrainte assez forte sur les émissions des NOx.

Concernant la POI6, pour le 16 juillet, le modèle a tendance à sur estimer le NO en milieu urbain (18.37 ppb pour la référence / 9.6 ppb pour les observations) mais aussi à sous estimer les concentrations de NOx dans le panache (5.18 ppb pour la référence / 10.5 ppb pour les observations). Cette différence explique l'effet antagoniste des contraintes appliquées (standard et NO urbain) sur les émissions de NOx, avec une moyenne inchangée pour le cumul des contraintes. Il est intéressant de noter la diminution de la hauteur de la couche limite. Elle permet de mieux réconcilier les deux contraintes antagonistes, car elle a pour effet d'augmenter les concentrations simulées des NOx et des NOy dans le panache (cas bien mélangé), sans pour autant augmenter aussi fortement le NO urbain à la surface. Néanmoins, les observations de lidar Raleigh de la hauteur de la couche limite à Palaiseau font plutôt état d'une sous-estimation de la hauteur de la couche limite par le modèle. Ainsi, il apparaît pour ce jour une certaine incohérence entres mesures dans le panache et à Paris et les simulations obtenues avec CHIMERE.

La journée du 17 juillet se caractérise par une sur estimation des NOx simulés à la fois dans panache et du NO urbain par rapport aux observations. Dans ce cas, les observations vont alors un effet "cumulatif" permettant, lorsque l'on somme ces contraintes, d'aboutir à de bons accords entre les concentrations simulées et mesurées (Tableau 4.5).

4.2.2.2 Accord pour les observations d'OX dans le panache :

Il est ensuite intéressant de savoir comment les observations d'ozone dans le panache et dans la ville contraignent les émissions des COVs. En effet, dans les conditions relativement polluées de l'Ile de France, il existe une relation clairement positive entre les émissions de COVs et production photochimique de l'ozone (POX), tandis que le signe de la relation entre émission de NOx et POX est variable d'un jour à l'autre.

Pour schématiser, les contraintes par les observations des composés azotés induisent un ajustement sur les émissions des NOx. Cette modification agira aussi sur POX. Il reste un degré de liberté pour rapproché le POX simulé au POX observé : les émissions des COVs. En réalité, les modifications d'autres variables d'entrée interviennent également.

Pour la POI2, la diminution des émissions des NOx cause une augmentation de POX, en accord avec les observations. En conséquence, il n'y a pas besoin d'ajustement des émissions des COVs. Pour le 16 Juillet de la POI6, une diminution de la hauteur de la couche limite mène à une accumulation plus forte des polluants et augmente POX, en le rapprochant des observations. Egalement pour ce jour, il n'y a pas besoin d'ajustement des émissions des COVs. Enfin pour le 17 Juillet, la diminution des émissions des NOx de presque 30% mène à une faible augmentation de POX dans le panache qui s'approche des observations. Comme pour les autres jours, le modèle n'ajuste pas les émissions de COVs.

Espèces chimiques	P50 : Simulation de référence	P50 : Sans contrainte	P50 : Contraintes standard	P50 : Contrainte NO urbain	P50 : Contraintes standard + NO urbain	Obs.
POI6 : 16 juillet 1999						
Concentration moyenne dans le panache - concentration fond						
ΔOX, ppb	14.88	14.98	17.78	16.65	18.11	20.7
ΔNOy, ppb	9.29	8.69	15	5.83	13.72	18.1
ΔNOx, ppb	5.18	4.37	9.93	1.9	9.37	10.5
Concentration moyenne en milieu urbain ; OX urbain normalisé par OX fond						
NO, ppb	18.37	18.38	33.55	10.71	19.75	9.60
ΔOX, ppb	32.89	25.77	31.04	23.9	30.23	29.08
POI6 : 17 juillet 1999						
Concentration moyenne dans le panache - concentration fond						
ΔOX, ppb	30.95	25.9	27.1	25.88	28.07	32.3
ΔNOy, ppb	11.81	11.02	8.02	7.46	7.45	8.1
ΔNOx, ppb	4.85	4.25	2.82	2.52	2.55	2.8
Concentration moyenne en milieu urbain ; OX urbain normalisé par OX fond						
NO, ppb	36.83	31.98	22.18	17.95	18.08	17.14
ΔOX, ppb	20.61	20.5	20.34	20.53	19.64	14.43

TAB. 4.5 –

4.2.2.3 Considérations sur l'incertitude des émissions à posteriori :

Même si la contrainte par les observations ne change pas les moyennes des émissions de facon significative, les incertitudes des émissions à posteriori sont fortement diminuées. Ceci correspond au fait que les valeurs extrêmes dans les distributions à priori recoivent un poids relativement faible. De facon générale, les incertitudes des distributions à posteriori dépendent de facon complexe des incertitudes à priori des émissions, mais aussi des autres variables, des incertitudes sur les observations et du dégrée de contrainte d'une observation sur les émissions. Sans rentrer dans une analyse détaillée de ces différents facteurs, notons que la réduction de l'incertitude sur les émissions est la plus importante pour le 16 Juillet quand à la fois :

- les incertitudes sur les observations des composés azotés dans le panache sont les plus faibles ;
- les contraintes des observations sur les émissions sont les plus fortes en raison du caractère non linéaire de la relation entre concentrations et émissions ;
- les contraintes sur les émissions sont les plus fortes en raison des écarts élevés entre simulations et observations.

4.2.3 Etude de la période climatologique : juillet et août 1999

L'objectif de l'étude sur une période climatologique est de tenter de généraliser les résultats obtenus pour des journées spécifiques.
La Figure 4.3 représente de la même facon que précédemment les FDPs à priori et à posteriori pour les émissions de NOx, de COVs et pour le rapport de ces deux émissions. Là encore, il est important de souligner que les facteurs correctifs pour les différentes données d'entrée dérivées de cette étude sont constants sur le domaine de CHIMERE et pendant la durée de la simulation, à savoir, dans ce cas, 62 jours.

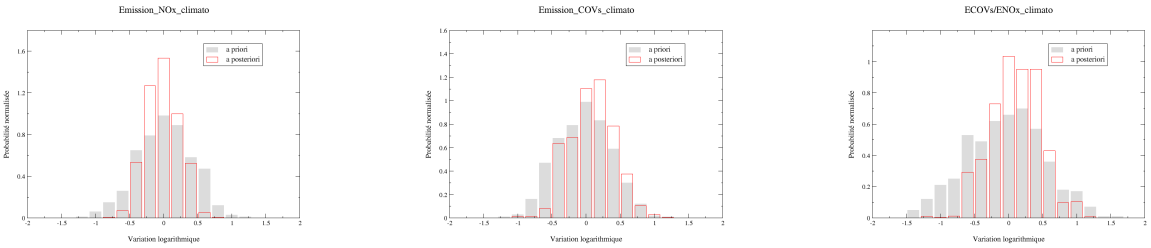


FIG. 4.3 – *Fonctions de densité de probabilité pour la période climatologique des paramètres d'entrée : émissions de NOx, de COVs et rapport de ces deux émissions. La distribution standard à priori est indiquée en bleu ; et la distribution alternative à posteriori est représentée en rouge. L'axe x représente la variation relative logarithmique du paramètre d'entrée étudié.*

Étude climatologique Juillet-Août 1999	Contraintes O3 panache		Contraintes NO urbain		Contraintes O3 panache + NO urbain	
Paramètres d'entrée	$\overline{x}$ et $\sigma(x)$ (%)	$\Delta(var)$ (%)	$\overline{x}$ et $\sigma(x)$ (%)	$\Delta(var)$ (%)	$\overline{x}$ et $\sigma(x)$ (%)	$\Delta(var)$ (%)
Émission NOx	- 4.4 $\pm$ 37	-14	1.5 $\pm$ 26	-57	- 3 $\pm$ 24	-64
Émission COVs	17 $\pm$ 34	-26	- 1.7 $\pm$ 37	-11	12 $\pm$ 32	-32
ECOVs/ENOx	21 $\pm$ 49	- 24	- 3 $\pm$ 45	- 35	15 $\pm$ 37	- 57
Vitesse du vent						
Hauteur couche limite	- 5.2 $\pm$ 40	$\pm$ 0	1.2 $\pm$ 39	-7	- 1.2 $\pm$ 38	-14

TAB. 4.6 – *Modifications des distributions à posteriori des émissions, du vent et de la hauteur de la couche limite pour différentes contraintes et pour la période climatologique de juillet-août 1999. $\overline{x}$ et $\sigma(x)$ sont respectivement la moyenne et la déviation standard. $\Delta(var)$ est la modification de la variance.*

Le Tableau 4.6 détaille les déplacements de la moyenne des FDPs après modifications par les observations, la déviation standard de la fonction à posteriori, et la diminution relative de la variance pour les trois types de contraintes considérées (O3 panache, NO urbain et cumul de ces deux contraintes). Lorsque l'on applique la contrainte sur l'ozone dans le panache, on observe une diminution de la variance des émissions, notamment pour les émissions de COVs (-26%) et une modification sensible des valeurs d'émissions pour les COVs (+17%) et faible pour les NOx. Lorsque les contraintes de NO urbain sont utilisées, ce sont surtout les émissions de NOx qui sont alors modifiées à travers une baisse notable de leur variance de -57%. Enfin, lorsque ces deux contraintes sont appliquées en simultané, on observe une augmentation de 12% des émissions de COVs et une légère baisse des émissions des NOx de 3%, toutes deux non significatives, puisque leur déviation standard (à 1 sigma) est respectivement 24% et 32%. Ceci correspond à une baisse de la variance des émissions des NOx de 64%, et de celle COVs de 32%. La plus faible diminution de la variance des émissions des COV est à reliée au fait que les émissions de NOx sont directement contraintes par le NO urbain et les émissions de COVs seulement de facon indirecte par la contrainte sur l'ozone dans le panache. Il faut aussi remarquer que les distributions à posteriori des données météorologiques et des autres variables sont beaucoup moins modifiées que celle des émissions. Le tableau XXX.X permet de confronter les observations aux simulations afin d'obtenir des renseignements sur la nature de la contrainte appliquée

Espèces chimiques	P50 : Simulation de référence	P50 : Sans contrainte	P50 : Contraintes NO urbain	P50 : Contrainte O3 panache	P50 : Contraintes NO urbain + O3 panache	Obs.
ΔO3, panache	13.85	12.12	12.84	16.02	16.16	18.06
NO, urbain	12.21	12.49	12.43	12.18	12.57	12.06

TAB. 4.7 –

aux distributions à priori.
Le Tableau 4.7 met en avant le fait, que le modèle est déjà, dans le cadre de données moyennes, assez représentatif des observations notamment concernant le NO urbain moyen. Par ailleurs, la contrainte sur les observations d'ozone dans le panache a permis de mieux représenter les concentrations d'ozone dans le panache puisque ces contraintes ont permis de fournir des concentrations moyennes de l'ordre de 16 ppb, proches des observations moyennes d'environ 18 ppb. Le modèle a corrigé la sous évaluation des concentrations d'ozone dans le panache en augmentant le rapport des émissions de COVs sur NOx de 15%. Par la suite, il serait intéressant de pouvoir effectuer le même type d'étude avec des contraintes complémentaires à celles utilisées dans cette étude à savoir l'ozone urbain et la production d'OX dans le panache.

4.2.4 Conclusion sur l'étude climatologique

En conclusion, l'étude climatologique a permis de confirmer globalement les résultats obtenus pour trois jours, mais pour une période plus longue de deux mois : pas de modification significative des émissions de NOx et de COVs en Ile de France. La diminution de la variance des émissions est plus faible que pour les jours des POI's. Ceci est dû au fait que seules les observations du réseau d'AIRPARIF ont pu être utilisées comme contraintes. Les informations des mesures aéroportées permettant une mesure dans le panache n'ont donc pas été prises en compte. En comparant les résultats pour les deux expériences, il est à noter que les distributions à priori n'ont pas exactement la même signification. Les incertitudes à priori pour les POI's peuvent ître interprétées comme la somme de l'erreur systématique et aléatoire (changeant d'un jour à l'autre), tandis que les incertitudes pour la période climatologique représentent l'erreur systématique seule. Comme nous ne connaissions pas la répartition entre erreurs systématiques et aléatoires, nous avons fait le choix conservateur de reporter toute l'erreur relative à une journée sur l'erreur systématique relative à la période de deux mois. Dans ce sens, les incertitudes à priori et à posteriori sur les émissions et les autres variables d'entrée représentent des bornes supérieures. Par ailleurs, il serait intéressant de stratifier les résultats selon le type du jour considéré (journée de semaine, week-end, différence entre le mois de juillet et le mois d'août etc.).

Chapitre 5

Conclusions et perspectives

5.1 Conclusion

Le projet OPTEMI visait **à mettre en place et évaluer des techniques de modélisation inverse afin de mieux quantifier les flux d'émissions surfaciques employés par les modèles de chimie-transport** (tel que CHIMERE). Pour cela, deux approches différentes mais complémentaires ont été baties : (i) une méthode d'optimisation basée sur le modèle adjoint, (ii) une méthode Monte-Carlo basée sur des perturbations aléatoires du modèle direct.
Dans les deux cas, des méthodologies originales ont été élaborées :

- Pour l'**approche adjointe**, l'approche est basée sur le krigeage des contraintes (les mesures) et un zonage dynamique des flux à inverser (les émissions surfaciques). Cette nouvelle méthode a été validée sur des cas académiques simples, à même de quantifier les limites de la méthode.
- Pour l'**approche Monte-Carlo**, une approche originale basée sur le calcul de probabilités de réalisation, permet d'obtenir une correction mais aussi l'incertitude associée. Les simulation sont aussi porté sur les étés 1998 et 1999.

Une recherche bibliographique a mis en avant le manque de validation des travaux précédemment réalisés dans le domaine. La plupart des auteurs présentent des études directement sur des cas réels complexes, sans estimation au préalable de la validité de leur approche. Nous avons donc choisi de développer graduellement notre méthodologie en mettant en place des cas académiques à complexité croissante pour l'adjoint. Cela nous a permis de quantifier des limites réelles de l'approche. Les méthodologies développées ont ensuite été appliquées sur des cas réels : tout d'abord des cas "extrèmes" correspondants à des périodes intensives du projet ESQUIF (POI2, Août 1998 et POI6, Juillet 1999), puis une climatologie sur les étés 1998 et 1999 en région parisienne.
Avec ces deux approches, l'information obtenue est donc des flux d'émissions corrigés en temps et en espace à la résolution du modèle d'accueil, ainsi que l'incertitude associé à ces flux. Sachant qu'au départ on ne dispose que de masses moyennes émises, par secteurs d'activité, cet ajout d'information reste précieux pour la modélisation de chimie-transport régionale.
Nous avons d'abord cherché à **calculer de manière précise les limitations de nos méthodes** :

- Afin de pouvoir réellement jauger des flux d'émissions, il fallait s'affranchir un minimum des erreurs liées aux autres paramètres. Des études préliminaires ont donc été menées afin de quantifier des biais sur les concentrations liés à la météorologie (dont hauteur de la couche limite et mélange vertical) et aux conditions aux limites. Cela a permis d'assurer une meilleure précision sur les

flux inversés ainsi que de meilleures contraintes sur la méthode Monte-Carlo. De plus, nous avons envisagé la possibilité de ne pas prendre en compte des processus existants : nous avons donc développé la prise en compte des émissions tri-dimensionnelles liées aux décollages et aterrissages des avions de ligne en région parisienne. L'impact modéré ne pouvant expliquer les différences modèle/mesures, nous avons procédé aux inversions sans cet ajout.

- Pour les deux méthodes, les limites de leur application ont été clairement quantifiées : la méthode adjointe ne peut inverser des émissions sans un minimum de renseignements sur les sources étudiées. Ainsi, il est plus bénéfique d'utiliser le NO comme contrainte plutot que l'ozone si l'on veut estimer des flux de NO émis. De même pour la méthode Monte-Carlo, un minimum d'information étant nécessaire pour construire des distributions, il faut beaucoup de contraintes pour retrouver un facteur correctif fiable.

Connaissant ces limitations, nous avons pu dégager, de l'ensemble des simulations réalisées, des **résultats augmentant notre connaissance de la variabilité spatio-temporelle des émissions**, tels que :

- En région parisienne et sur les étés considérés, les émissions utilisées par le modèle correspondent à un cadastre de bonne qualité. Si on pouvait le soupconner au vu des résultats de prévision, on peut aujourd'hui le quantifier. Hormis une variabilité intra-semaine pas assez bien prise en compte (notamment les Mercredi et Vendredi pour le trafic), les valeurs moyennes des flux de NO (principale source) sont correctes, et la variabilité spatiale adaptée à la région. la méthode Monte-Carlo a de plus, montré que le rapport COVs/NOx était lui aussi correct que ce soit à Paris même que dans le panache d'ozone.
- La tendance générale est à une réduction des émissions liées au trafic dans Paris. Cela se traduit surtout par un "pic" temporel d'émission moins étalé dans le temps le matin. A l'opposé, on note une légère tendance à augmenter les émissions en périphérie.
- Temporelllement, l'inversion montre que l'on sous-estime les émissions les matins des jours de semaine et Dimanche. Ils sont, par contre, surestimés le Samedi. Ces variations sont dans une gamme de $\pm 15\%$, valeur confirmée par l'approche Monte-Carlo.

5.2 Perspectives

Les résultats ont montré que les deux méthodes mises en oeuvre permettait d'obtenir de nombreuses informations sur des flux d'émissions à partir de mesures de routines.
Ces résultats ont néanmoins été obtenus avec un seul modèle : c'est à dire un outil particulier, avec son équilibre particulier entre les différentes données d'entrée et paramétrisations. Mais le but de notre méthode est d'obtenir des coefficients correctifs les plus proches possibles de la réalité : il faut donc qu'ils apportent un bénéfice quel que soit le modèle. Une application, servant de validation supplémentaire, sera donc de mettre à disposition nos coefficients d'optimisation pour d'autres modèles : si leur propres scores sont améliorés, le bien-fondé de la méthode sera d'autant plus confirmé.
D'un point de vue applications, il existe de nombreux autres sites subissant une pollution atmosphérique, et disposant de mesures de routines en surface. D'autres perspectives portent donc sur le portage des méthodes sur d'autres sites et d'autre périodes. On peut évidemment cité le projet ESCOMPTE et l'environnement de la région de Fos-Berre Marseille. A topographie plus complexe et donc météorologie moins facilement contrainte, des études visant à réduire ces incertitudes devront être conduites au préalable.

Enfin, des perspectives supplémentaires sont liées aux améliorations des méthodes. Ainsi, il pourra être envisagé d'ajouter des flux à inverser comme ceux liés aux émissions d'aérosols, PM2.5 et PM10 en milieu urbain.

Annexe A

Principes théoriques des outils utilisés

A.1 Principes théoriques de l'adjoint

Le principe du modèle adjoint est de calculer les gradients de fonctions scalaires, ces fonctions étant dépendantes de paramètres et variables du modèle. Dans notre cas, on cherche à estimer le gradient de concentrations chimiques calculées par rapport aux valeurs des paramètres du modèle.
Ces gradients ont plusieurs types d'utilisation : d'abord la sensibilité, puisqu'ils quantifient directement la dépendance au paramètre considéré, ensuite la modélisation inverse puisqu'ils permettent d'utiliser une méthode d'optimisation non-linaire.
Le principal objectif du modèle adjoint est donc le calcul du gradient $\nabla x(p)$ de la fonction scalaire $x(p)$, avec p l'ensemble des paramètres du modèle M (p est un vecteur).
On considère un ensemble de variables x évoluant spatialement et temporellement suivant le système d'équations suivant :

$$\frac{dx}{dt} = \mathbf{M}(x, p), \tag{A.1}$$

x représente donc les concentrations, et est le "vecteur d'état" du modèle ; et **M** est le modèle. Puisque le temps est discret dans le modèle, on peut noter ce vecteur d'état x_n à l'instant $t = ndt$. De plus, le modèle M peut être exprimé sous la forme d'une fonction non linéaire *M*, donnant la tendance instantanée de x_n, c'est à dire contenant tous les processus du modèle, mais aussi ses approximations numériques. L'équation discrétisée d'évolution des variables du modèle s'écrit :

$$x_{n+1} = M(x_n, p) \tag{A.2}$$

On a donc deux grandes familles de cas à traiter :

- **Sensibilité** : On cherche la sensibilité de x_n par rapport à p
- **Modélisation inverse** : On cherche à déterminer p en comparant x_n avec y_n(observations) et en minimisant leur différence.

Dasn le cas du développement théorique de l'adjoint, nous nous placons dans le cas le plus simple : la sensibilité. Le problème de la modélisation inverse reprend exactement les mêmes notions en y ajoutant d'autres plus complexes.
La sensibilité S de la fonction scalaire $M(x)$ de x à un instant donné t sera le gradient tel que :

$$S = \nabla_p(M(x(t))). \tag{A.3}$$

En se donnant un produit scalaire ($< . >$) dans l'espace des paramètres, le gradient S de $M(x_n)$ par rapport à p est tel que toute perturbation infinitésimale δp de p va créer une perturbation infinitésimale $\delta M(x_n)$ de $M(x_n)$, tel que :

$$\delta M(x_n) =< S, \delta p > . \tag{A.4}$$

A chaque instant t, la perturbation de x s'exprimera avec le système suivant :

$$\delta x_{n+1} = \left(\frac{\partial M}{\partial x}(x_n, p)\delta x_n\right) + \left(\frac{\partial M}{\partial p}(x_n, p)\delta p\right). \tag{A.5}$$

En faisant l'hypothèse que l'on a pas de perturbation initiale sur δx_0, on obtient, par remplacement successif dans (A.5), une expression pour la perturbation finale δx_n :

$$\delta x_n = \mathbf{A}\delta p, \tag{A.6}$$

où **A** est une matrice composée des sommes successives des itérations des dérivées partielles de la fonction M, par rapport à x et p. On a donc :

$$\delta M(x_n) = \frac{\partial M}{\partial x}\mathbf{A}\delta p, \tag{A.7}$$

ce qui peut, aussi, s'écrire, en utilisant l'opérateur d'adjonction (noté par une *) :

$$\delta M(x_n) = \left\langle \mathbf{A}^* \frac{\partial M^*}{\partial x}, \delta p \right\rangle. \tag{A.8}$$

Dans ce cas, on notera que l'opérateur d'adjonction est identique à une transposition car le produit scalaire choisi est canonique. Cette dernière relation, eq. (A.8) montre en particulier (si on la compare à l'équation (A.4)) que :

$$S = \mathbf{A}^* \left(\frac{\partial M}{\partial x}\right)^* \tag{A.9}$$

L'intégration "à rebours" dans le temps des équations adjointes du système (A.2) peut résoudre numériquement le système (A.9) tel que : avec

$$\delta^* x_n = \left(\frac{\partial M}{\partial x}(x_n)\right)^*$$

et

$$\delta^* x_{n-1} = \left(\frac{\partial M}{\partial x}(x_{n-1}, p)\right)^* \delta^* x_n; \tag{A.10}$$

la valeur finale du calcul $\delta^* x_0$ représentera la sensibilité de M aux conditions initiales et la sensibilité S aux paramètres sera alors donnée par la condition finale du système, c'est à dire :

$$\left\{\begin{array}{lcl} \delta^* p_n & = & 0 \\ \delta^* p_{n-1} & = & \delta^* p_n + \left(\dfrac{\partial M}{\partial p}(x_{n-1}, p)\right)^* \delta^* x_n \\ S & = & \delta^* p_0 \end{array}\right. \tag{A.11}$$

A.2 Principes théoriques du krigeage

Le but du krigeage est de calculer les analyses $y^a_{t,h}(\mathbf{s})$ d'un champ de concentrations $y_{t,h}(\mathbf{s})$ pour tout point $\mathbf{s}$, le jour t ($t = 1, \ldots, T$) à l'heure h ($h = 1 \ldots, 24$). Un ensemble de K mesures de concentrations $y^o_{t,h}(\mathbf{s_k})$ avec $\mathbf{k} = 1, ..., K$ et $\mathbf{s_k}$ l'emplacement de la $k^{ième}$ station de mesure et un champ de concentrations *a priori* (appelé aussi "first-guess") $y^b_{t,h}(\mathbf{s})$ sont disponibles. Les concentrations analysées $y^a_{t,h}(\mathbf{s})$ sont alors données par :

$$y^a_{t,h}(\mathbf{s}) = y^b_{t,h}(\mathbf{s}) + \sum_{k=1}^{K} \lambda^k_{t,h}(\mathbf{s})(y^o_{t,h}(\mathbf{s_k}) - y^b_{t,h}(\mathbf{s_k})) \tag{A.12}$$

L'équation A.12 montre donc que le champ analysé consiste en une correction du champ *a priori* par une combinaison linéaire des *innovations*, $y^o_{t,h}(\mathbf{s_k}) - y^b_{t,h}(\mathbf{s_k})$. Les coefficients $\lambda^k_{t,h}(\mathbf{s})$ sont des poids à déterminer.
Les poids $\lambda^k_{t,h}(\mathbf{s})$ sont solutions du système :

$$\begin{cases} \forall k = 1, \ldots, K, \\ \sum_{l=1}^{K} \widehat{\lambda}^l_{t,h}(\mathbf{s})\gamma_{t,h}(\mathbf{s_k}, \mathbf{s_l}) - \mu = \gamma_{t,h}(\mathbf{s_k}, \mathbf{s}), \\ \sum_{l=1}^{K} \widehat{\lambda}^l_{t,h}(\mathbf{s}) = 1 \end{cases} \tag{A.13}$$

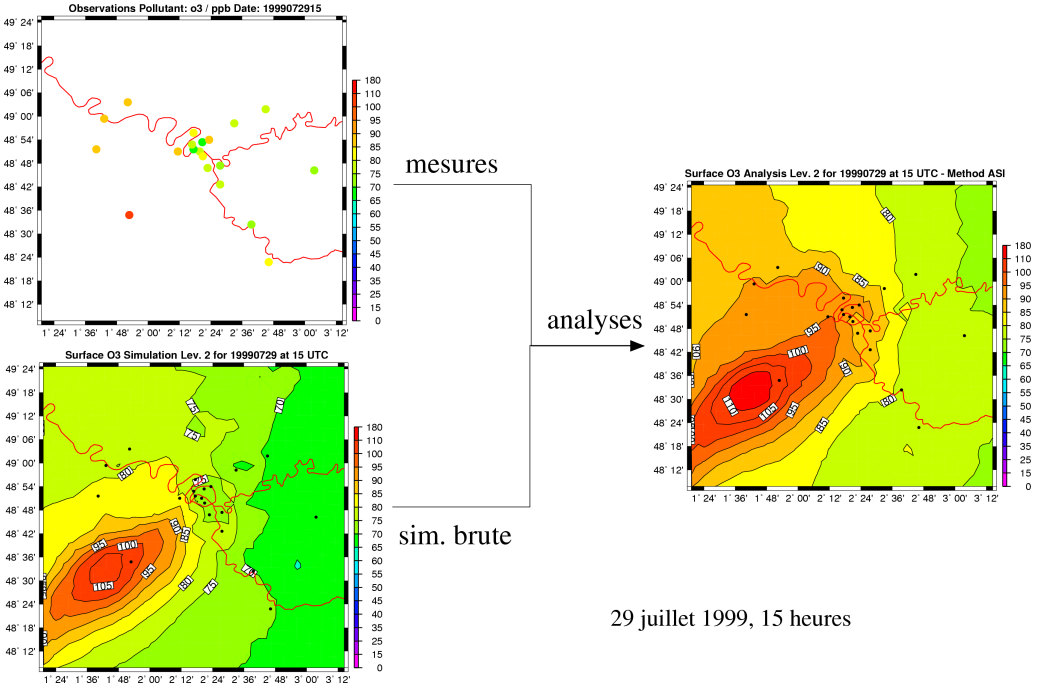


FIG. A.1 – *Exemple de résultats obtenus en Ile de France avec le réseau AIRPARIF, le modèle direct CHIMERE et la méthode de krigeage INK.*

où $\gamma_{t,h}(\mathbf{s_k}, \mathbf{s_l}) = \frac{1}{2}E(X(\mathbf{s_k}) - X(\mathbf{s_l}))^2 = \gamma_{t,h}(\|\mathbf{s_k} - \mathbf{s_l}\|) = \gamma_{t,h}(r)$ est la fonction de covariance avec $r = \|\mathbf{s_k} - \mathbf{s_l}\|$. $E(.)$ représente la moyenne. La variance de l'erreur d'observation $\sigma^{2o}_{t,h}(\mathbf{s_k})$ est supposée homogène et constante dans le temps, soit $\sigma^{2o}_{t,h}(\mathbf{s_k}) = \sigma^{2o}_h$.

La semi-variance $\gamma_{t,h}(r)$ est modélisée grâce à un variogramme expérimental $\gamma_{t,h}^{exp}$. Celui-ci est calculé à partir de toutes les mesures disponibles à l'heure h, le jour t. Tout d'abord, les classes de distance sont construites avec un intervalle L et une tolérance τ. Pour chaque classe, la variance $\frac{1}{2}E(X^o(\mathbf{s_k}) - X^o(\mathbf{s_l}))^2$ est ensuite calculée. Enfin, le graphe des variances en fonction de la distance moyenne dans la classe est tracé et ajusté au semi-variogramme exprérimental par un modèle exponentiel avec un "effet pépite" (noté $NUG_h = \sigma_h^{2o}$).
La technique et le code de krigeage utilisés pour ce travail ont été élaborés par Nadège BLOND au cours de sa thèse, sous le nom de "méthode INK". C'est un outil implanté dans la version de CHIMERE utilisée quotidiennement en Ile de France par AIRPARIF pour réaliser les cartes de prévision.
La Figure A.1 illustre un exemple d'analyse en Ile de France, avec le réseau AIRPARIF, le modèle CHIMERE et la méthode de krigeage INK.
En Ile de France, avec le réseau AIRPARIF, on utilise un intervalle L =5 km et une tolérance τ=4 km pour les classes de distance. De plus, nous avons choisi pour σ_h^{2o} la variance maximale de toutes les mesures disponibles à l'heure h.

A.3 Définitions des fonctions liées à l'inverse par adjoint

Dans cette section, nous définissons les différentes termes que nous allons utilisé pour l'optimisation. Cela inclut particulièrement :

- La fonction coût J calculée au coeur du modèle, à chaque pas de temps et pour chaque maille ou des mesures sont disponibles. Cela permet de quantifier la "distance" qui sépare une concentration mesurée d'une concentration simulée.
- Le vecteur d'état qui gère les différentes variables modèles que l'on cherche à optimiser
- Les matrices de covariance d'erreur notées **B** et **R**.
- Le critère d'arrêt de l'optmisation.

Tout d'abord nous présentons la chaine algorithmique générale qui utilise ces différents outils.

A.3.1 Principe : imbrication des modules de calcul

L'inverse ne représente en principe qu'un gestionnaire permettant de calculer le coût et son gradient puis de procéder à leur minimisation. Il s'agit donc de faire appel ou modèle direct, à l'adjoint et au minimiseur.
La Figure A.2 indique les différentes entrées et étapes d'une simulation inverse. La procédure d'inversion comprend cinq étapes :

1. le modèle est initialisé avec les paramètres *a priori* x_f et les concentrations y_{sim} sont calculées
2. une première valeur du coût $J(x(t_0)) = J(x_f)$ est évaluée avec les contraintes y_{obs} disponibles
3. une simulation adjointe est faite et les composantes du gradient de la fonction coût $\frac{\partial J}{\partial x}(x_f)$ sont calculées
4. le minimiseur fournit le meilleur ajustement $x(t_1)$ de x pour minimiser J et $\partial J/\partial x$
5. le modèle est ré-initialisé avec les nouveaux paramètres $x(t_1)$

Les itérations se poursuivent jusqu'à ce que le critère d'arrêt (p.81) soit atteint.

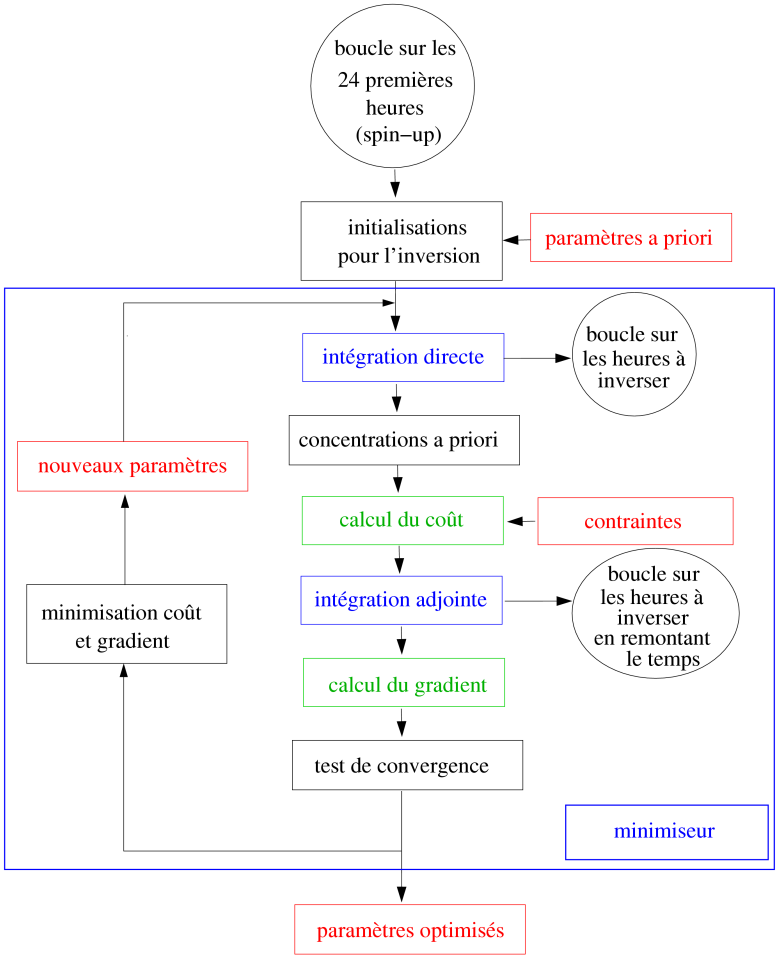


FIG. A.2 – *Principe de l'inversion d'une simulation.*

A.3.2 La fonction coût

La fonction coût d'un porblème de minimisation est typiquement définie selon [*Talagrand* (1997)] sous la forme suivante :

$$J(x(t)) = (y_{obs}(t) - \mathbf{H}y_{sim}(t))^T \boldsymbol{R} \underbrace{(y_{obs}(t) - \mathbf{H}y_{sim}(t))}_{y(t)} + (x_f(t) - x(t))^T \boldsymbol{B} \underbrace{(x_f(t) - x(t))}_{X(t)} \quad \text{(A.14)}$$

avec :

- $x(t)$ est le vecteur d'état ; t n'indique pas une heure mais un état de l'optimisation
- $y(t)$ est le vecteur qui contient les différences entre les concentrations simulées ($y_{obs}(t)$) et les contraintes correspondant ($y_{sim}(t)$, avec $\mathbf{H}$ l'opérateur d'observation)
- $X(t)$ est le vecteur qui contient les différences entre les paramètres optimisés ($x(t)$) et *a priori* ($x_f(t)$)
- $\boldsymbol{R}$ est la matrice des covariances d'erreur d'observation
- $\boldsymbol{B}$ est la matrice des covariances d'erreur du modèle

Le premier terme indique donc à quel point on souhaite se rapprocher des contraintes et le deuxième, à quel point on peut s'éloigner des valeurs *a priori* des paramètres.

A.3.3 Le vecteur d'état

Le vecteur d'état contient en principe les paramètres à inverser. Comme ceux-ci pourront être de natures différentes (flux d'émissions ou paramètres météorologiques) et couvrir plusieurs ordres de grandeur (pour les flux d'émissions), notre vecteur d'état contient en fait les coefficients correctifs à appliquer aux paramètres correspondant. On a alors $X(t) = (1 - x(t))$.
Les paramètres de plusieurs heures peuvent constituer différentes composantes de $x(t)$. En effet, les espèces émises n'ont pas une influence immédiate sur les espèces mesurées. Le pic d'ozone de l'après-midi par exemple est principalement sensible aux émissions de NO_2 du pic de trafic du matin. Il apparaît donc nécessaire d'effectuer l'inversion dans des "fenêtres temporelles" de taille pertinente par rapport à la physique et à la chimie de la pollution photo-oxydante, par exemple, de 9 heures à 15 heures pour l'inversion du NO_2 avec des mesures d'ozone. La fonction coût prend donc en compte les mesures disponibles pendant toute la "fenêtre" et une seule exécution de l'adjoint (en remontant le temps pour toute la période) permet d'obtenir toutes les composantes du gradient. Il faut remarquer que l'utilisation de fenêtres temporelles augmente rapidement la dimension du problème ; leur durée ne peut donc pas dépasser quelques heures et dépend des liens entre espèces mesurées et inversées.

A.3.4 La matrice $\boldsymbol{R}$

Un des objectifs de ce travail est d'élaborer une méthode utilisant les mesures de routine des réseaux. Or, pour un réseau de mesures, les covariances des erreurs d'observation c'est-à-dire les termes extra-diagonaux de $\boldsymbol{R}$, sont supposés nuls car tous les instruments sont indépendants. Il reste donc à estimer les termes diagonaux de $\boldsymbol{R} = diag(w_1, \cdots, w_N)$ (avec N le nombre de contraintes disponibles).
Les poids des termes de la fonction coût comportant une différence entre concentrations simulée et mesurée sont proportionnels à la confiance que l'on accorde à la contrainte. Le poids w_{ii} doit donc être proportionnel à la confiance accordée à $y_{obs}(i)$. On pourrait utiliser $w_i \alpha \frac{1}{var(i)}$ avec $var(i)$ la variance associée à $y_{obs}(i)$. Mais si l'on utilise la variance absolue, de petites concentrations auront une petite variance et donc un grand poids dans l'optimisation alors que les concentrations élevées auront une variance absolue beaucoup plus grande et donc une influence moindre dans l'optimisation. Or, c'est notamment la reproduction correcte des pics de concentrations qui est recherchée. Au lieu d'utiliser directement la variance des contraintes, on utlise donc l'incertitude relative transformée en "variance relative" : $var(i) = I(i)^2$ et $I_{rel}(i) = I(i)/val(i)$ donnent $var_{rel}(i) = I_{rel}(i)^2 = (I(i)/val(i))^2$. Pour plus de facilité de lecture, tout est exprimé en pourcentage : $I_{rel}(i)(pourcent) = I(i)/val(i) \times 100$ d'où $var_{rel}(i)(pourcent) = var_{rel}(i) \times 10000$

A.3.5 La matrice $\boldsymbol{B}$

À l'échelle régionale, les émissions surfaciques ne sont pas un phénomène continu, ni spatialement, ni temporellement. De plus, un flux estimé peut être faux à une heure donnée mais pas à la suivante. Il devient donc particulièrement difficile d'ajouter de l'information *a priori* dans les termes extra-diagonaux de la matrice "de background".

Même avec une matrice diagonale, la valeur des termes est difficile à estimer et actualiser. L'étude récente menée par [*Hoelzemann et al.* (2001)], basée sur des informations statistiques (voir page **??**) est tout à fait correcte pour des simulations à long terme (par exemple des scénarios de variation d'émissions d'année en année) mais ne peut être utilisée à l'échelle régionale et horaire à cause de la forte variabilité des sources et des puits.
De plus, dans le cas de notre étude, aucune information sur l'incertitude du cadastre a priori n'est disponible. Celle-ci pourrait être de 35% [*Kuhlwein and Friedrich* (2000)] pour les espèces anthropiques et d'un facteur 2 à 3 pour les espèces biogéniques [*Beekmann and Derognat* (2003); *Sillman et al.* (2003)].
Finalement, l'espace des émissions ne peut pas être enrichi par des informations *a priori* permettant de construire une matrice $\boldsymbol{B}$ **pour les émissions**. Dans le cas de l'inversion de paramètres météorologiques, une matrice "de background" au moins diagonale devra être élaborée. En effet, la météorologie étant calculée "off-line", il est nécessaire d'éviter que les valeurs optimisées de certains paramètres ne s'éloignent trop des valeurs *a priori*, sous peine d'obtenir des résultats incohérents. Dans le cas de la hauteur de couche-limite par exemple, la valeur optimisée ne doit pas s'éloigner de la valeur *a priori* au-delà de l'incertitude de la fonction de diagnostic qui dépend de la température et de la pression fournies "off-line". Dans le cas contraire, la hauteur de couche-limite optimisée n'est plus en accord avec les profils verticaux de température et de pression.

A.3.6 Le critère d'arrêt de l'optimisation

Nous avons vu que l'optimisation était un processus itératif basé sur al convergence (ou non) de la valeur de la fonction coût vers une précision pré-définie. Nous précision ici le critère que nous avons défini pour notre étude.
Le critère de convergence sur la norme du gradient est noté $epsg$. L'optimisation s'arrête quand le nouveau rapport ϵ_k calculé est inférieur à $epsg$ choisi au départ :

$$\epsilon_k = \frac{\|g_k\|}{\|g_1\|} < epsg.$$

On veut s'arrêter quand $\|g_k\|$ est presque nulle (en supposant qu'elle peut l'être).

$$\|g_k\| = \sqrt{\sum_{n=1}^{dim} \frac{\partial J}{\partial X_n}^2}$$

$$\frac{\partial J}{\partial X_n} = \lim_{h \to 0} \frac{J(\ldots, X_n + h, \ldots) - J(\ldots, X_n, \ldots)}{h}$$

Mais comme le vecteur x contient des coefficients correctifs, on ne cherche pas à faire tendre h vers 0 pour ne pas tenir compte de corrections inférieures à 1% par exemple. On cherche plutôt une situation où $\frac{\partial J}{\partial X_n}$ est quasiment nulle avec un $h = 10^{-2}$ fini. Il ne reste plus alors qu'à estimer une valeur pour laquelle la différence $\Delta J = J(\ldots, X_n + h, \ldots) - J(\ldots, X_n, \ldots)$ est à peu près nulle. ΔJ est de la forme :

$$\Delta J = \sum_{i=1}^{N} w_i[(y_{obs}(i) - y_{sim}(i, X_n + h))^2 - (y_{obs}(i) - y_{sim}(i, X_n))^2]$$

La différence entre deux termes en ppb est non significative quand elle est strictement inférieure à 10^{-2}. ΔJ est donc non significative quand elle est strictement inférieure à $10^{-4} \times \sum_{i=1}^{N} w_i$. Pour une meilleure approximation, on considère que ΔJ «nulle» quand elle est égale à $10^{-5} \times \sum_{i=1}^{N} w_i$
Finalement, on obtient une valeur $\|g_k\|$ satisfaisante quand on a :

$$\begin{aligned}\|g_k\| &\approx \sqrt{dim \times (\frac{10^{-5} \times \sum_{i=1}^{N} w_i}{10^{-2}})^2} \\ &\approx \sqrt{dim} \times \frac{10^{-5} \times \|\sum_{i=1}^{N} w_i\|}{10^{-2}} \\ &\approx \sqrt{dim} \times 10^{-3} \times \sum_{i=1}^{N} w_i\end{aligned}$$

On peut donc choisir

$$epsg = \frac{\sqrt{dim} \times 10^{-3} \times \sum_{i=1}^{N} w_i}{\|g_1\|}.$$

A.3.7 Les indicateurs statistiques

Dans tout ce qui suit, on note O_n la n-ième valeur observée de l'échantillon considéré de taille N et S_n la n-ième valeur simulée. La moyenne de chacune de ces variables est :

$$\overline{X_N} = \frac{1}{N} \sum_{i=1}^{N} X_n \tag{A.15}$$

On étudie les performances du modèle avec les trois indicateurs statistiques suivants :

- **le biais** : c'est l'écart moyen entre les valeurs simulées et observées. Il indique les erreurs systématiques. On a :

$$biais = \frac{1}{N} \sum_{i=1}^{N} (S_n - O_n)$$

- **la moyenne quadratique** : elle donne l'erreur moyenne en valeur absolue. On a :

$$rms = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (S_n - O_n)^2}$$

- **la corrélation** : elle permet de savoir si la variabilité est bien reproduite, indépendamment de l'amplitude. On a :

$$cor = \frac{\frac{1}{N} \sum_{i=1}^{N} (S_n - \overline{S_N}) \times (O_n - \overline{O_N})}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (S_n - \overline{S_N})^2 \times \frac{1}{N} \sum_{i=1}^{N} (O_n - \overline{O_N})^2}}$$

Annexe B

Définition du cas académique

Nous présentons ici la configuration du cas académique dont les résultats sont présentés p.29.

- Le domaine représente l'Île-de-France, de 1,3 à 3,3°E et de 48,1 à 49,4°N. Le maillage est constitué de 10×10 mailles pour huit niveaux verticaux dont l'épaisseur ne varie pas dans le temps. L'épaisseur du premier niveau est fixée à 20 mètres, les niveaux supérieurs sont de plus en plus épais et englobent jusqu'à 300 mètres au-dessus de la hauteur de couche-limite maximale (Figure B.2).
- Les conditions météorologiques reproduisent celles d'un épisode de pollution estival en Île-de-France : températures élevées, des vitesses de vent faibles et aucun nuage. Le vent et la température sont uniformes dans le domaine.
- La direction du vent est fixée à 45° c'est-à-dire soufflant du Nord-Est à une vitesse en surface de $4\ \mathrm{m.s^{-1}}$. Cette vitesse correspond à une valeur réaliste en Île-de-France assez élevée pour permettre une bonne visualisation des résultats dans le maillage choisi. Le profil vertical du module du vent est déterminé à partir de cette vitesse de surface et une loi logarithmique dans la couche-limite. Au-dessus de la couche limite, le module du vent est constant.
- La valeur moyenne de la température à deux mètres pour la journée est fixée à 23°C. La température à 2 mètres suit un profil temporel sinusoïdal (Figure B.1) qui représente un cycle diurne typique avec un maximum à midi et un minimum une heure avant le lever du Soleil. Un profil vertical réaliste est obtenu en appliquant une décroissance linéaire de 6K par kilomètre avec une inversion de 2K à la hauteur de la couche limite.

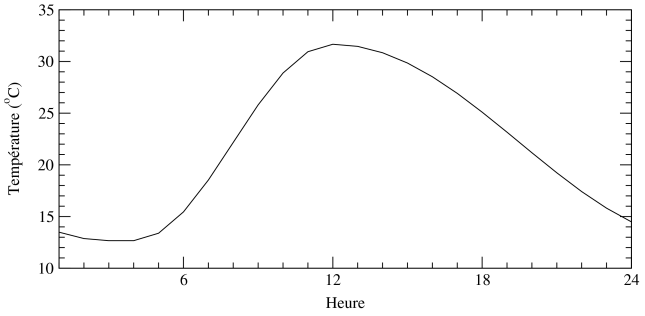


FIG. B.1 – *Evolution de la température à 2 mètres sur 24 heures.*

- La hauteur de couche-limite minimale est fixée à 300 mètres, la hauteur maximale est fixée à 2200 mètres. Le Soleil se levant à 5h et se couchant à 19h (il est au zénith à 12h), le maximum

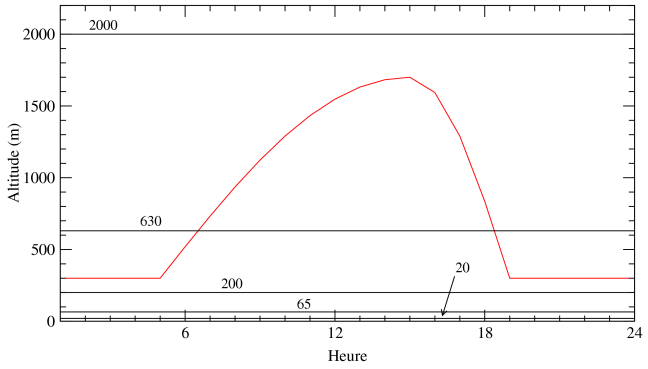


FIG. B.2 – *Niveaux verticaux et évolution de la hauteur de couche-limite en 24 heures.*

est atteint à 15h (trois heures après le passage du Soleil au zénith). L'évolution de la hauteur de couche-limite est sinusoïdale pendant les deux parties du jour alors que la hauteur de la couche-limite nocturne est constante et minimale.

- Les concentrations aux limites (latérales et au sommet) sont nulles et le temps de mise à l'équilibre est de 24 heures (durée typique pour une simulation en Île-de-France).
- Les émissions :
  - ☐ Les émissions biogéniques sont quasiment nulles afin de mettre en évidence les relations entre émissions anthropiques et concentrations pour le monoxyde d'azote (NO est en effet une espèce émise anthropique et biogénique). Elles ne sont pas exactement égales à zéro pour éviter tout problème de calcul dans certaines routines traitant des émissions.
  - ☐ Le cadastre d'émissions comporte deux sources anthropiques. Elles émettent les 16 espèces-modèle anthropiques. Les proportions des flux des différentes espèces sont réalistes. L'intensité maximale des flux d'émission des principales espèces sont de 10^{14}moléc.cm^{-2}.s^{-1} pour CO, 10^{13}moléc.cm^{-2}.s^{-1} pour NC_4H_{10} et NO, 10^{12}moléc.cm^{-2}.s^{-1} pour C_2H_4, C_2H_6, C_3H_6, HCHO, NO_2, OXYL, SO_2. Pour chaque source, les émissions sont constantes et égales à E_{max} entre 7 et 11 heures et égales à $E_{max}/10$ le reste du temps.
    Le cadastre *a priori* est créé en sous-estimant les émissions "réelles" de NO_x de 20% pendant les cinq heures d'intensité maximale. La proportion de 20% est considérée comme une estimation réaliste de l'incertitude du cadastre pour les émissions dues au trafic. La durée choisie correspond au nombre d'heures de pic de trafic pendant une matinée ou un après-midi. Deux inversions des émissions de NO_x perturbées sont effectuées, en conservant un rapport de 90% de NO pour 10% de NO_2.

Bibliographie

Atkinson, R., D. L. Baulch, R. A Cox., R. F. Hampson, J. A. Kerr Jr., M. J. Rossi, and J. Troe (1997), Evaluated kinetic, photochemical and heterogeneous data for atmospheric chemistry, *J. Phys. Chem. Ref. Data*, *26*, 521–1012.

Beekmann, M., and C. Derognat (2003), Monte-Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign, *J. Geophys. Res.*, *108*(D17), 8559, doi :10.1029/2003JD003391.

Bergamaschi, P., R. Hein, M. Heinmann, and P. Crutzen (2000), Inverse modeling of the global CO cycle : 1. Inversion of CO mixing ratio, *Journal of Geophysical Research*, *105*(D2), 1909–1927.

Bergin, M., and J. Millford (2000), Application of Bayesian Monte Carlo analysis to Lagrangian photochemical air quality model, *Atmos. Environ.*, *34*, 781–792.

Blond, N. (2002), Assimilation de données photochimiques et prévision de la pollution troposphérique, Physique et Chimie de l'atmosphère, Ecole Polytechnique, France, 202 p.

Bousquet, P., P. Ciais, P. Peylin, and P. Monfray (1999a), Inverse modeling and annual atmospheric co_2 sources and sinks : 1. method and control inversion, *J. Geophys. Res.*, *104*(D21), 26,161–26,178.

Bousquet, P., P. Ciais, P. Peylin, and P. Monfray (1999b), Inverse modeling and annual atmospheric co_2 sources and sinks : 2. sensitivity study, *J. Geophys. Res.*, *104*(D21), 26,179–26,193.

Brand, K., and M. Small (1995), Updating uncertainty in an integrated risk assessment : Conceptual framework and methods, *Risk Anal.*, *15*, 719–731.

Chang, M., D. Hartley, and W.-L. Chang (1996), Inverse modeling of biogenic isoprene emissions, *Geophysical research letters*, *23*(21), 3007.

Chang, M., D. Hartley, C. Cardelino, D. Haas-Laursen, and W. Chang (1997), On using inverse methods for resolving emissions with large spatial inhomogeneities, *J. Geophys. Res.*, *102*(D13), 16,023–16,036.

Daley, R. (1996), *Atmospheric Data Assimilation*, Cambridge University Press.

Daley, R. (1997), Atmospheric data assimilation, *Journal of the meteorological society of Japan*, *75*(1B), 319–329.

DeMore, W. B., S. P. Sander, D. M. Golden, R. F. Hampson, M. J. Kurylo, C. J. Howard, A. R. Ravishankara, C. E. Kolb, and Molina (1997), Chemical kinetics and photochemical data for use in stratospheric modeling - evaluation 12, *JPL Publ. 94, Jet Propulsion Laboratory, Pasadena, Calif.*

Dilks, D., R. Canal, and P. Meier (1992), Development of Bayesian Monte Carlo techniques for water quality modelling uncertainty, *Ecol. Modell.*, *62*, 149–162.

Elbern, H., and H. Schmidt (1999), A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling, *J. Geophys. Res.*, *104*, 18,583–18,598.

Elbern, H., H. Schmidt, and A. Ebel (1997), Variational data assimilation for tropospheric chemistry modeling, *J. Geophys. Res.*, *102*(D13), 15,967–15,985.

Elbern, H., H. Schmidt, O. Talagrand, and A. Ebel (2000), 4D-variational data assimilation with an adjoint air quality model for emissions analysis, *Environmental modelling & software*, *15*, 539–548.

ESQUIF (2001), ESQUIF Etude et Simulation de la QUalité de l'air en Ile-de-France, rapport final, *Tech. rep.*, Institut Pierre-Simon Laplace, Laboratoire Interuniversitaire des Systèmes Atmosphériques, Météo-France, Laboratoire d'Aérologie, Forschungszentrum Jülich, AIRPARIF.

Gao, D., W. Stockwell, and J. Milford (1996), Global uncertainty analysis of a regional scale gas phase chemical mechanism, *J. Geophys. Res.*, *101*, 9107–9120.

Gilbert, J.-C., and C. Lemaréchal (1989), Some numerical experiments with variable-storage quasi-Newton algorithms, *Mathematical programming*, *45*, 407–435.

Gilbert, J.-C., and C. Lemaréchal (1995), The modules M1QN3 and N1QN3, version 2.0c, *Tech. rep.*, INRIA.

Haas-Laursen, D., D. Hartley, and R. Prinn (1996), Optimizing an inverse method to deduce time-varying emissions of trace gases, *Journal of Geophysical Research*, *101*(D17), 22,823–22,831.

Hanna, S., J. Chang, and M. Fernau (1998), Monte-carlo estimates of uncertainties in predictions by a photochemical grid model (uam-iv) due to uncertainties in input variables, *Atmos. Env.*, *32*, 3609–3618.

Hanna, S., Z. Lu, H. Frey, N. Wheeler, J. Vukovitch, S. Arunachalam, M. Fernau, and D. Hansen (2001), Uncertainties in predicted ozone concentrations due to input uncertainties for the uam-v photochemical grid model applied to the july 1995 otag domain, *Atmos. Env.*, *35*, 891–903.

Hartley, D., and R. Prinn (1993), On the feasibility of determining surface emissions of trace gases using an inverse method in a three-dimensional chemical transport model, *J. Geophys. Res*, *98*, 5183– 5198.

Hassel, D., F. Weber, and T. Schmitz (2000), Zusammensetzung der Kohlenwasserstoffe im Abgas unterschiedlicher Fahrzeugkonzepte, JĺJlich, Germany.

Heimann, M., and T. Kaminski (1999), *Inverse modeling approaches to infer surface trace gas fluxes from observed atmospheric mixing ratios*, *Approaches to scaling of trace gas fluxes in ecosystems*, vol. 14, 275-295 pp., Elsevier, Amsterdam.

Hein, R., P.J.Crutzen, and M.Heimann (1997), An inverse modeling approach to investigate the global atmospheric methane cycle, *Global Biogeochem. Cycles*, (11), 43–76.

Hodzic, A. (2005), Modélisation des aérosols de pollution en Ile de France, Thèse de sciences, Ecole Polytechnique, France.

Hoelzemann, J., H. Elbern, and A. Ebel (2001), PSAS and 4D-var data assimilation for chemical state analysis by urban and rural observation sites, *Phys. Chem. Earth. B*, *26*(10), 807–812.

Honoré, C., and R. Vautard (2000), Photochemical regimes in urban atmospheres : The influence of dispersion, *Geophys. Res. Letters*, *27*, 1895–1898.

Houweling, S., T. Kaminski, F. Dentener, J. Lelieveld, and M. Heinmann (1999), Inverse modeling of methane sources and sinks using the adjoint of a global transport model, *J. Geophys. Res.*, *104*(D21), 26,137–26,160.

Jacob, D., B. Field, E. Jin, I. Bey, Q. Li, J. Logan, and R. Yantosca (2002), Atmospheric budget of acetone, *Journal of Geophysical Research*, *107*(D10), ACH5–1–19.

Kaminski, T., and M. Heimann (2001), Inverse modeling of atmospheric carbon dioxide fluxes, *Science*, *294*, 259.

Kuhlwein, J., and R. Friedrich (2000), Uncertainties of modelling emissions from road transport, *Atmos. Environ.*, *34*, 4603–4610.

Kuhlwein, J., R. Friedrich, U. Kalthoff, C. U., F. Slemr, M. Habram, and M. Möllmann-Coers (2002), Comparison of modeled and measured total co and nox emissions rates, *Atmos. Environ.*, *36*(suppl.1), S53–S60.

Lattuati, M. (1997), Contribution à l'étude du bilan de l'ozone troposphérique à l'interface de l'Europe et de l'Atlantique Nord : modélisation lagrangienne et mesures en altitude, Thèse de sciences, Université Paris 6, France.

Madronich, S., and S. Flocke (1998), *Handbook of Environment Chemistry*, chap. The role of solar radiation in atmospheric chemistry, pp. 1–26, P. Boule, Springer & Verlag, Heidelberg.

Mahowald, N., R. Prinn, and P. Rasch (1997), Deducing ccl3 f emissions using an inverse method and chemical transport models with assimilated winds, *J. Geophys. Res.*, *102*, 28,153–28,168.

Mannschreck, K., D. Klemp, D. Kley, R. Friedrich, J. Kuhlwein, B. Wickert, P. Matuska, M. Habram, and F. Slemr (2002), Evaluation of an emission inventory by comparison of modeled and measured emission ratios of individual hc's, co, and nox, *Atmos. Environ.*, *36*(suppl. 1), S81–S94.

Mendoza-Dominguez, A., and A. Russell (2001), Estimation of emission adjustments from the application of four-dimensional data assimilation to photochemical air quality modeling, *Atm. Env.*, *35*, 2879–2894.

Menut, L. (2003), Adjoint modelling for atmospheric pollution processes sensitivity at regional scale during the ESQUIF IOP2, *J. Geophys. Res.*, *108*(D17), 8562, doi :10.1029/2002JD002549.

Menut, L., R. Vautard, M. Beekmann, and C. Honoré (2000a), Sensitivity of photochemical pollution using the adjoint of a simplified chemistry-transport model, *J. Geophys. Res.*, *105*(D12), 15,379–15,402.

Menut, L., et al. (2000b), Measurements and modelling of atmospheric pollution over the Paris area : an overview of the ESQUIF Project, *Annales Geophysicae*, *18*(11), 1467–1481.

Mulholland, M., and J. Seinfeld (1995), Inverse air pollution modelling of urban-scale carbon monoxide emissions, *Atm. Env.*, *29*, 497–516.

Pison, I. (2005), Modélisation inverse des émissions à l'échelle régionale, Physique et Chimie de l'atmosphère, Université Paris 12, France, 150 p.

Pison, I., and L. Menut (2004), Quantification of the impact of aircraft traffic emissions on tropospheric ozone over Paris area, *Atmospheric Environment*, *38*(7), 971–983.

Prinn, R., and D. Hartley (1995), *Progress and Problems in Atmospheric Chemistry*, chap. chapter 6 - inverse methods in atmospheric chemistry, pp. 172–197, XXX.

Rodenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003), CO2 flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmospheric Chemistry And Physics*, *3*, 1919–1964.

Rudd, H., and I. Marlowe (1998), Improvements to the VOC inventory through speciation of white spirits : a report produced for the department of the environment, transport and the regions, *Tech. rep.*, AEA Technology, Oxfrodshire, England.

Schmitz, T., D. Hassel, and F. Weber (2000), Determination of VOC components in the exhaust of gasoline and diesel passenger cars, *Atmospheric Environment*, (34), 4639–4647.

Sillman, S., R.Vautard, L.Menut, and D.Kley (2003), O3-nox-voc sensitivity indicators in paris : results from models and esquif measurements, *J. of Geophys. Res.*, *108*(D17).

Simpson, D. (1999), Inventorying emissions from nature in Europe, *Journal of Geophysical Research*, *104*, 8113–8152.

Talagrand, O. (1997), Assimilation of observations : an introduction, *Journal of the meteorological society of Japan*, *75*(1B), 191–209.

Theloke, J., A. Obermeier, and R. Friedrich (2000), Ermittlung des Lösemittelemissionen 1994 in Deutschland und Methoden zur Fortschreibung, in *Auftrag des Umweltbundesamtes*, Universität Stuttgart, Germany.

Theloke, J., A. Obermeier, and R. Friedrich (2001), Abschätzung des Lösemittelemissionen in Deutschland, *Gefahrstoffe Reinhaltung Luft*, (61), 105–111.

Vautard, R., et al. (2003a), Paris emission inventory diagnostics from esquif airborne measurements and a chemistry transport model, *J. Geophys. Res.*, *108*(D17), 8564, doi :10.1029/2002JD002797.

Vautard, R., et al. (2003b), A synthesis of the Air Pollution Over the Paris Region (ESQUIF) field campaign, *Journal of Geophysical Research*, *108*(D17), 8558, doi :10.1029/2003JD003380.

Wang, Y., and S. Bentley (2002), Development of a spatially explicit inventory of methane emissions from Australia and its verification using atmospheric concentration data, *Atmospheric Environment*, *36*, 4965–4975.