Simulation numérique et assimilation de données variationnelle pour la dispersion atmosphérique de polluants

Denis Quélo

▶ **To cite this version:**

Denis Quélo. Simulation numérique et assimilation de données variationnelle pour la dispersion atmosphérique de polluants. Modélisation et simulation. Ecole des Ponts ParisTech, 2004. Français. NNT : 2004ENPC0441 . pastel-00002101

HAL Id: pastel-00002101
https://pastel.hal.science/pastel-00002101

Submitted on 3 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE L'ECOLE NATIONALE DES PONTS ET CHAUSSEES

Spécialité

Mathématiques et Informatique

Présentée et soutenue publiquement par

Denis Quélo

pour obtenir le titre de Docteur de l'École Nationale des Ponts et Chaussées

Simulation numérique et assimilation de données variationnelle pour la dispersion atmosphérique de polluants

Soutenue le 8 Décembre 2004 devant le jury composé de :

Olivier Thual	Rapporteur
Philippe Ciais	Rapporteur
Laurent Menut	Examinateur
Olivier Isnard	Examinateur
François-Xavier Le Dimet	Directeur de thèse
Bruno Sportisse	Co-Directeur de thèse

Remerciements

Tout d'abord merci Bruno pour m'avoir accordé ta confiance en me proposant ce sujet, puis pour m'avoir guidé et conseillé, et aussi pour ta contribution importante dans ce travail. J'ai beaucoup appris à ton contact. Merci ensuite à Vivien dont j'ai grandement apprécié l'aide.

Je remercie messieurs Olivier Thual et Philippe Ciais d'avoir accepté d'être rapporteur de ce mémoire. Je remercie également les autres membres du jury : François-Xavier Le Dimet, Laurent Menut et Olivier Isnard. Merci pour le temps consacré à la lecture de ce document et les améliorations suggérées.

Je profite de cette page pour remercier l'ensemble des membres du CEREA. Une pensée pour tous ceux avec qui j'ai partagé des moments agréables au labo ou en dehors.

J'aimerais aussi remercier tout ceux qui par leur compétence, leur amitié, leur soutien, leurs encouragements ou tout simplement leur présence ont contribué à la réalisation de ce travail. Qu'ils trouvent ici l'expression de ma reconnaissance.

Résumé

L'assimilation de données consiste à réduire les incertitudes sur les paramètres d'entrée du modèle à l'aide de mesures. L'objectif peut être d'améliorer les prévisions ou encore de mieux connaître des paramètres incertains comme par exemple les inventaires d'émissions. Cette thèse s'est effectuée dans ce contexte, appliqué aux modèles de Chimie-Transport atmosphériques.

Dans une première partie, un modèle de chimie-transport tridimensionnel (Polair3D) a été développé pour simuler de façon réaliste les concentrations de polluants responsables de la pollution photochimique. Des comparaisons aux mesures de surface d'ozone et d'oxydes d'azote sur la région de Lille pendant l'année 1998 ont validé ce modèle à l'échelle régionale.

Dans une deuxième partie, la méthode d'assimilation de données variationnelle 4D-var a été implémentée. Elle repose sur la minimisation d'une fonctionnelle (fonction coût) mesurant l'écart entre des concentrations et des mesures par un algorithme itératif. Un des ingrédients essentiels est le calcul du gradient de la fonction coût. Il est basé sur le modèle adjoint de Polair3D obtenu par différenciation automatique. Une application à la modélisation inverse des émissions sur Lille avec des observations réelles a montré la faisabilité de la méthode. En particulier, l'inversion de paramètres temporels d'émissions d'oxydes d'azote permet d'améliorer notablement les prévisions d'ozone et d'oxydes d'azote au-delà de la période considérée.

La sensibilité des résultats de l'inversion aux paramètres du système (paramètres physiques, erreur d'observations, positionnement des capteurs par exemple) est reliée au conditionnement de ce système et peut être étudiée par la sensibilité dites de "second-ordre". Une dernière partie à caractère plus méthodologique étudie cette problématique. Une synthèse bibliographique dresse un état des lieux théoriques de cette thématique, qui est également illustrée par une application à la dispersion de radionucléides à petite échelle. Dans le cas de la cinétique chimique, la grande dispersion des échelles temporelles est un exemple typique de mauvais conditionnement. Son impact sur l'assimilation de données est étudié de manière théorique et est illustré numériquement sur une chimie simple.

Table des matières

Introduction **9**

1 Le modèle de Chimie–Transport POLAIR3D **17**

1.1 Modélisation de la pollution photochimique troposphérique 18

1.2 La plate-forme POLAIR3D 28

2 Modélisation de la qualité de l'air sur la région de Lille **43**

2.1 Les observations 45

2.2 Conditions de la simulation 49

2.3 Résultats pour l'année de référence 1998 et pour l'exercice 2015 57

2.4 Conclusion 65

3 Assimilation de données **69**

3.1 Quelques notations et définitions 73

3.2 Modèles stationnaires 81

3.3 Modèles d'évolution 84

3.4 Conditionnement et qualité du résultat 95

4 Modèles linéaire tangent et adjoint de POLAIR3D **103**

4.1 Equations dérivées continues pour les modèles de dispersion réactive 106

4.2 Approche continue versus approche discrète 108

4.3 Algorithmes mis en œuvre dans POLAIR3D 113

4.4 Application à des tests de sensibilité 118

5 Modélisation inverse des émissions pour un CTM 123

5.1 Cadre de l'étude 126

5.2 Etude de sensibilité 127

5.3 Modélisation inverse des émissions : expériences synthétiques 135

5.4 Résultats sur le cas réel 142

Conclusion 151

A Observational network design, adaptive observations and ... 155

A.1 Optimal design theory 158

A.2 A direct application to data assimilation : the targeting problem 163

A.3 Singular Vector techniques 167

A.4 Sensitivity techniques 171

A.5 Ensemble approaches 176

A.6 Time scheduling 178

B Assimilation de données pour les modèles lent/rapide 183

B.1 Un modèle académique 186

B.2 Assimilation de données pour les modèles de chimie "lent/rapide" 188

B.3 Tests numériques 196

C Assimilation de données pour la dispersion à petite échelle 203

C.1 Problématique 205

C.2 Analyse de sensibilité 215

C.3 Expérience de modélisation inverse 221

C.4 Analyse du second ordre et expériences de positionnement de capteurs . . . 227

C.5 Conclusion 230

Introduction

Modélisation de la pollution atmosphérique

Les modèles de Chimie-Transport (comme CMAQ [Byun et Ching, 1999], EURAD [Ebel *et al.*, 1989], CHIMERE [Vautard *et al.*, 2001], POLAIR3D [Sportisse *et al.*, 2002], ...) décrivent l'évolution en temps et en espace des espèces chimiques considérées que ce soit dans la phase gazeuse ou la matière condensée de l'atmosphère (gouttes de nuage, aérosols). Leur utilisation va de la compréhension des processus à la prévision opérationnelle en passant par les études d'impact.

Au niveau de la phase gazeuse, les modèles tridimensionnels actuels commencent à bien représenter les processus physiques et se comparent de manière satisfaisante aux mesures disponibles (figure 1), la donne restant largement différente pour les aérosols ([Seigneur *et al.*, 1982]).

De nombreuses incertitudes demeurent néanmoins. On peut citer par exemple :

- les incertitudes sur les données d'émission (les inventaires d'émission restent largement incertains) ;
- les incertitudes sur les modèles (notamment à travers la problématique de la paramétrisation sous-maille et des effets de ségrégation).

Dans ce contexte, il peut s'avérer utile d'avoir recours à d'autre sources d'informations supplémentaires aux modèles, en l'occurrence les observations délivrées par les réseaux de mesures atmosphériques (voir la figure 2 par exemple pour le réseau EMEP sur l'Europe).

Le couplage des résultats de modèles numériques et d'observations délivrées par des réseaux de mesure se fait dans le cadre classique de ce que l'on appelle l'*assimilation de données*, comme elle est pratiquée par ailleurs dans de nombreux autres domaines, notamment en météorologie. L'intérêt peut être double :

- affiner notre connaissance de l'état chimique de l'atmosphère en réduisant les incertitudes,

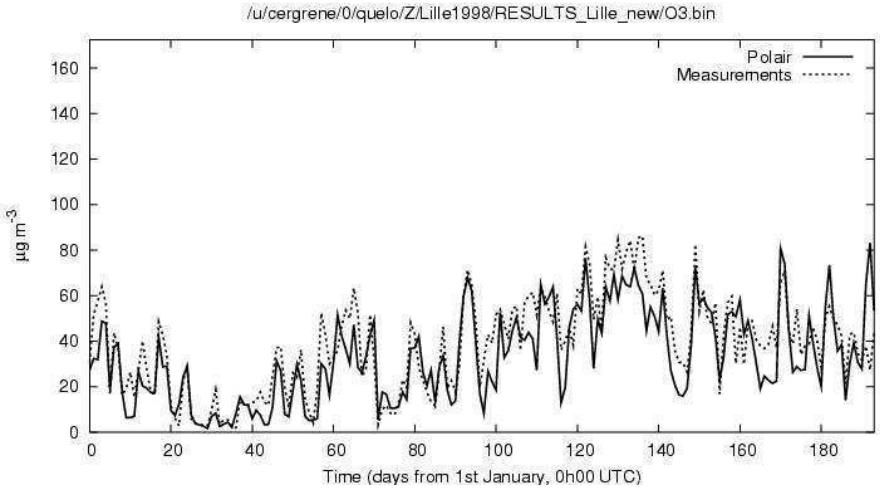

FIG. 1 – Concentrations d'ozone en moyenne journalière à la station de Roubaix. Mesures et simulations sur la région de Lille avec POLAIR3D pour l'année 1998.

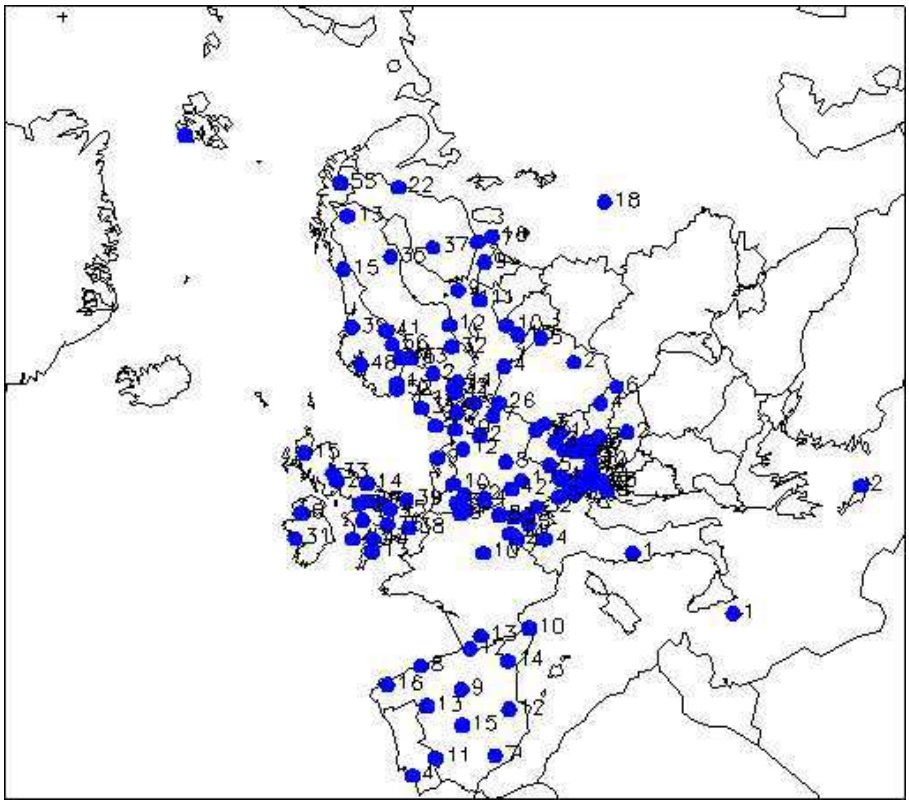

FIG. 2 – Carte des stations d'ozone du réseau EMEP.

– ou avoir une meilleure connaissance des bases de données d'entrée (notamment d'émissions pour améliorer les inventaires d'émissions).

Assimilation de données et chimie atmosphérique

L'assimilation de données est un domaine relativement récent pour la chimie atmosphérique ([Austin, 1992, Fisher et Lary, 1995, Riishojgaard, 1996] par exemple). Les mesures, de nature diverse (mesures terrestres, mesures aéroportées, mesures satellitaires) peuvent être assimilées soit dans des modèles de circulation globale (GCM) soit dans des modèles de chimie-transport à l'échelle globale ou régionale (par exemple [Elbern *et al.*, 1997, Elbern et Schmidt, 2001, Elbern *et al.*, 2000, Elbern et Schmidt, 1999]). Les mesures satellitaires sont bien sûr *a priori* particulièrement appropriées du fait de leur bonne couverture spatiale et temporelle et du grand nombre de données fournies.

Les questions méthodologiques posées par l'assimilation de données pour la chimie atmosphérique sont pour une large part "classiques". Citons, à titre d'exemples, les quelques points suivants :

- les méthodes utilisées peuvent être séquentielles ou variationnelles, voire beaucoup plus simples (nudging) ;

- l'information contenue dans une mesure influence d'autres variables du modèle via les contraintes fournies par les équations du modèle : l'assimilation d'ozone peut ainsi contraindre des champs de vent ([Riishojgaard, 1996]) via les équations de transport. D'un point de vue plus centré sur la cinétique chimique, l'information contenue dans un traceur peut contraindre d'autres espèces réactives ([Austin, 1992, Fisher et Lary, 1995]) ;

- un enjeu important reste l'assimilation directe de radiance au sein d'un modèle réactif afin de ne pas rajouter une étape supplémentaire fondée sur une climatologie d'espèces lors de la première étape d'inversion de la donnée radiative (pour des données satellitaires) ;

- un des points limitants jusque là de l'assimilation de données satellitaires a été le manque d'information sur la distribution verticale. Une solution largement utilisée est l'assimilation de données 2D, des colonnes (pour l'ozone), dans des modèles de traceurs passifs bidimensionnels (intégrés verticalement) ou tridimensionnels.
  La disponibilité de capteurs avec une résolution verticale a ouvert la possibilité d'assimiler des profils avec des modèles réactifs.

Les exemples qui suivent se distinguent selon les choix effectués pour chacun de ces

items. Ils donnent des points d'entrée typiques dans la littérature et n'ont aucune prétention d'exhaustivité.

Assimilation de colonnes d'ozone dans un modèle de transport tridimensionnel par interpolation optimale

On peut se référer par exemple à [Jeuken *et al.*, 1999] avec l'utilisation de mesures TIROS/TOVS ([Planet et al, 1984]) via une approche de type interpolation optimale dans le modèle TM3. Un autre exemple du même type est l'assimilation de données GOME/ERS2 ([Eskes et Jeuken, 1998]).

Assimilation de colonnes d'ozone dans un modèle de transport par approche variationnelle, 4D-var

On peut se référer par exemple à [Heskes *et al.*, 1999] avec l'assimilation de colonnes GOME/ERS2 dans le modèle 2D d'advection de traceur AMK, par une approche de type 4D-var (le nom étant un peu impropre dans ce cas). Ce travail fait suite à une étude dans le même contexte avec une approche de type nudging ([Lebelt *et al.*, 1996]).

Pour le cas tridimensionnel, des exemples synthétiques (observations numériques) sont donnés dans [Riishojgaard, 1996] avec un modèle simplifié.

Assimilation de profil (ozone, CO) par interpolation optimale dans un CTM

Pour l'ozone, un exemple est fourni par les données MLS/UARS (Microwave Limp Sounder on Upper Atmosphere Research Satellite), dont les profils d'ozone résultants ont été assimilés par exemple dans le modèle de transport ROSE, utilisé dans sa version off-line, par interpolation optimale ([Levelt *et al.*, 1998]).

Pour le monoxyde de carbone, un exemple est celui des données MAPS (Measurement of Air Pollution from Space, MAPS on space shuttle) avec l'assimilation de profil de CO (obtenu à l'aide d'une approche classique de type "kernel") dans une version du modèle 3D MOZART ([Hauglustaine *et al.*, 1998]) par interpolation optimale (sous sa version PSAS, Physical Space Statistical Analysis System).

Assimilation de données satellitaires par approche 4D-variationnelle dans un modèle réactif lagrangien

On peut se référer par exemple à [Khattatov et al, 1999] avec des données UARS (Upper Atmospheric Research Satellite) utilisées dans un modèle réactif de boîte lagrangienne.

Plus généralement, des études synthétiques sont données dans [Wang *et al.*, 2001, Fisher et Lary, 1995].

Assimilation de données terrestres par approche séquentielle

Différentes méthodes (interpolation statistique, krigeage) ont été appliquées dans [Blond, 2002] pour obtenir des cartes tridimensionnelles de concentrations d'ozone et de dioxyde d'azote à l'aide d'observations de surface.

Un filtre de Kalman étendu réduit a été développé dans [Segers, 2002]. Son application à des mesures de surface d'ozone permet de réduire les incertitudes des émissions de NO_x.

Assimilation de données terrestres par approche 4D-variationnelle dans un CTM

Les références incontournables sont bien sûr [Elbern *et al.*, 1997, Elbern et Schmidt, 2001, Elbern *et al.*, 2000, Elbern et Schmidt, 1999] avec une approche 4D-var appliquée au modèle EURAD sur des données terrestres, dans un contexte de modélisation de la qualité de l'air (photochimie) à l'échelle régionale/continentale.

Ces quelques exemples montrent la diversité des approches suivies et des applications visées. Même si l'utilisation préopérationnelle de tels systèmes est en train de se mettre en place, de nombreux points restent largement ouverts.

Organisation de la thèse

Cette thèse s'est plus particulièrement concentrée sur la problématique de la modélisation inverse des émissions pour les modèles de Chimie-Transport, sur la base d'observations terrestres. Les travaux réalisés ont plus spécifiquement porté sur les trois thèmes suivants :

1. la construction et la validation d'une plate-forme de Chimie-Transport, le modèle POLAIR3D ;

2. la réalisation du système d'assimilation de données variationnelles associé et l'application à des expériences de modélisation inverse d'émissions ;
3. une étude théorique sur la problématique du conditionnement à l'inversion pour de tels systèmes : les questions traitées recoupent les questions de dynamique en temps (et l'implication pour l'inversion), de sensibilité au réseau de mesure (ce que l'on appelle traditionnellement le *network design*) et de sensibilité aux paramètres non estimés (la sensibilité de second-ordre).

Cette thèse est ainsi organisée de la manière suivante.

Construction et validation d'une plate-forme de Chimie-Transport, le modèle POLAIR3D

Les modèles de Chimie-Transport et les processus physiques décrits sont exposés au chapitre 1. Un modèle de Chimie-Transport, POLAIR3D, a été développé au cours de cette thèse et on présente les principaux choix de modélisation (paramétrisations, bases de données utilisées) et de simulation numérique (algorithmes) qui ont été effectués.

Ce travail a donné lieu aux publications suivantes : [Sportisse *et al.*, 2002, Boutahar *et al.*, 2004, Sartelet *et al.*, 2002].

POLAIR3D a été utilisé, sur la base notamment de ce travail, dans de nombreuses situations, que ce soit aux échelles locales ou régionales/continentales. On présente au chapitre 2 un cas de validation qui a constitué un des "terrains de jeu" privilégiés de ce travail : les simulations effectuées sur la région de Lille pour l'année 1998. Ce travail a aussi été l'occasion de tester la sensibilité à divers inventaires d'émission, dans le cadre d'un projet du PREDIT (Programme pour la Recherche, le Développement et l'Innovation dans les Transports Terrestres), consacré à l'estimation de la qualité de l'air selon divers scénarios de Plans de Déplacement Urbain à l'horizon 2015.

Ce travail, mené en collaboration avec le CETE (Centre d'Etude Technique de l'Equipement) de Lille, au sein d'un projet du PREDIT, fait l'objet d'une publication en cours de soumission ([Quélo *et al.*, 2004]).

Ces deux chapitres permettent de fixer le cadre de la modélisation directe pour les modèles de Chimie-Transport, tels que POLAIR3D, et les limitations de ces modèles.

Construction d'un système d'assimilation de données variationnelles associé et application à des expériences de modélisation inverse d'émissions

Le recours à des approches de type assimilation de données (ou plus généralement de modélisation inverse) est une voie possible d'amélioration de la qualité des résultats. Les méthodes d'assimilation de données sont ainsi présentées dans un chapitre à nature largement bibliographique, au chapitre 3. L'accent est mis sur les approches dites "variationnelles", qui sont celles que nous avons retenues.

La construction des modèles dérivés (modèles *linéaire tangent* et *adjoint*) est une tâche lourde, sur le plan informatique, pour un modèle tridimensionnel du type de POLAIR3D. Le chapitre 4 présente cette problématique et les divers choix retenus pour notre travail. Des cas préliminaires d'application à des calculs de sensibilité sont également présentés.

L'article [Sportisse et Quélo, 2003] reprend largement ce chapitre. De manière plus générale, l'ensemble de ces points peut être trouvé dans les rapports [Sportisse et Quélo, 2004, Quélo et Sportisse, 2004].

Le chapitre 5 illustre quelques expériences de modélisation inverse, sur la base de données réelles, qui ont été effectués dans le cadre de la simulation présentée au chapitre 2. Un des points clés est la sensibilité des résultats obtenus (inversion de paramètre) vis à vis des incertitudes du modèle (des paramètres non inversés).

Problématique du conditionnement à l'inversion pour de tels systèmes

A ce stade, si l'on suppose possible la mise en œuvre d'expériences de modélisation inverse, un problème méthodologique d'un niveau supérieur est celui de la sensibilité des résultats de l'inversion à l'ensemble des paramètres du système (paramètres physiques mais aussi paramètres de l'assimilation de données, erreurs d'observations, positionnement des capteurs, etc). On dresse dans le chapitre A l'état des lieux "théorique" de cette problématique avec notamment les questions relatives au conditionnement du système à inverser et la problématique du "network design" (le positionnement "optimal" des capteurs). Un cadre théorique approprié, pour les méthodes variationnelles, est fourni par les méthodes dites du "second-ordre".

La synthèse ([Sportisse et Quélo, 2002]) reprend partiellement ces thèmes.

Pour le cas de la cinétique chimique atmosphérique, un élément clé du bon conditionnement du système est déterminé par la grande dispersion des échelles de temps liées à la

cinétique chimique. On étudie de manière théorique, dans le chapitre B, le comportement pour l'assimilation de données de modèles dits "lents/rapides" tels que rencontrés en chimie atmosphérique. Les résultats sont largement attendus mais le cadre mathématique utilisé permet de donner un sens rigoureux. Des illustrations numériques, sur un cas simple est également donné.

Les articles [Quélo *et al.*, 2002] et [Quélo et Sportisse, 2004] illustrent également cette thématique.

Les méthodes de second-ordre, qui sont difficiles à appliquer pour un cas tridimensionnel non-linéaire de grande dimension, sont utilisées dans un cas de dispersion plus simple au chapitre C : celui de la dispersion de radionucléides à petite échelle (dans le contexte du suivi d'un incident radioactif autour d'une centrale nucléaire). Les modèles utilisés sont des modèles de type gaussien et un grand nombre de calculs peut être fait explicitement.

Ce travail a été mené dans le cadre d'une étude commune avec l'IRSN (Institut de Radioprotection et de Sûreté Nucléaire) et a fait l'objet du rapport [Quélo et Sportisse, 2002] (inclu in extenso) et de l'article [Quélo *et al.*, 2005].

Enfin, une partie reprenant les principales conclusions et dressant quelques perspectives de recherche clôt cette thèse.

Chapitre 1

Le modèle de Chimie–Transport POLAIR3D

Résumé

Ce chapitre décrit les modèles de chimie–transport (CTM) utilisés pour la simulation de la dispersion des espèces réactives dans la troposphère. On détaille également la plate-forme POLAIR3D qui a été développée puis utilisée au cours de ce travail.

Sommaire

1.1 Modélisation de la pollution photochimique troposphérique . . 18
1.1.1 Les principaux processus 19
1.1.2 Cadre général des CTM 26
1.2 La plate-forme POLAIR3D 28
1.2.1 Structure d'un CTM 28
1.2.2 Equation de dispersion atmosphérique 30
1.2.3 Composantes du modèle et description des processus 31
1.2.4 Schémas numériques 38
1.2.5 Utilisation antérieure de POLAIR3D et comparaison à des mesures 41

Introduction

Les modèles de Chimie-Transport sont devenus de plus en plus performants par leur niveau de description des processus physiques mais aussi par la qualité de leur retour d'expérience par comparaison à des données mesurées. De nombreux modèles sont à présent disponibles. On peut citer par exemple : EURAD [Ebel *et al.*, 1989], CMAQ [Byun et Ching, 1999], EMEP [Barrett et Berge, 1996], CHIMERE [Vautard *et al.*, 2001], LOTOS [Blom et Roemer, 1997], etc ...

L'objectif de ce chapitre est la présentation du modèle de Chimie–Transport POLAIR3D ([Boutahar *et al.*, 2001, Sportisse *et al.*, 2002]). Ce modèle a été développé afin de pouvoir disposer de schémas numériques correspondant à l'état de l'art. Sa structure lui permet de pouvoir être différencié de manière automatique (voir chapitre 4), ce qui est un point crucial pour pouvoir utiliser des approches variationnelles en assimilation de données. Il est à présent utilisé de manière opérationnelle pour les études d'impact d'EDF R&D et par le centre de crise de l'IRSN.

Un CTM est un ensemble de relations mathématiques et algorithmiques qui permettent le calcul de l'évolution des polluants. Il inclut d'une part l'étape préliminaire de construction des entrées à partir des données météorologiques, de type de sol, d'inventaires d'émissions, et d'autre part un modèle de chimie–transport à proprement parler (POLAIR3D) pour la simulation des niveaux de concentrations de polluants. La présentation qui suit s'intéresse principalement à la partie photochimique gazeuse de POLAIR3D.

Ce chapitre est organisé de la manière suivante. Les processus jouant un rôle dans un CTM sont décrits dans la section 1 d'un point de vue physique. Le cadre général de la modélisation de la chimie troposphérique est présenté dans la section 2. La structure, l'équation de dispersion atmosphérique et les différentes composantes de POLAIR3D sont décrites dans la section 3. Des détails concernant l'implémentation et les schémas numériques employés font l'objet de la section 4.

1.1 Modélisation de la pollution photochimique troposphérique et modèles de chimie–transport

Cette partie décrit rapidement les principaux processus jouant un rôle pour la pollution photochimique (figure 1.1) : les émissions de précurseurs, la photochimie troposphérique, la dispersion météorologique et les interactions avec la surface par dépôt sec.

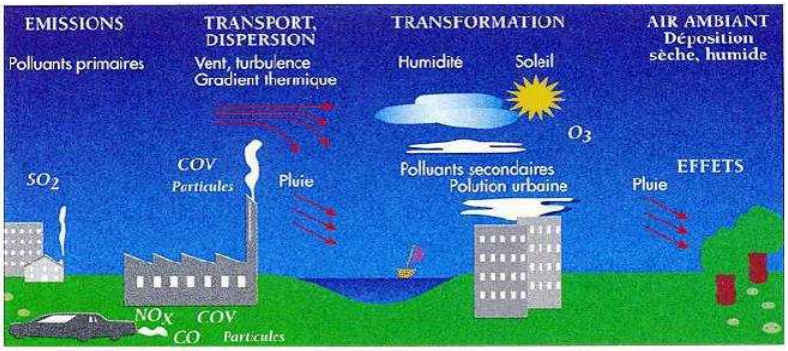

FIG. 1.1 – Les principaux processus de la pollution photochimique.

1.1.1 Les principaux processus

1.1.1.1 Les émissions de précurseurs

Les polluants primaires appelés précurseurs sont à l'origine de la pollution photochimique. Ils sont directement issus des sources de pollution. Les substances suivantes sont classiquement inventoriées :

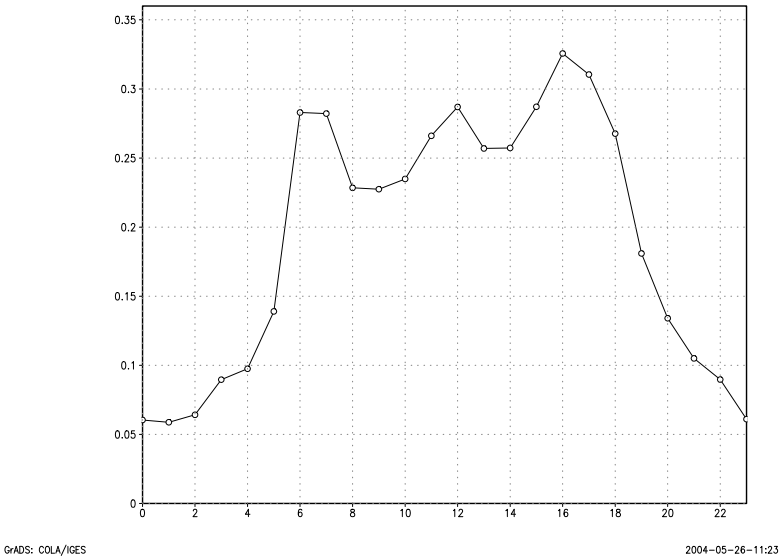

FIG. 1.2 – Chronique horaire typique des émissions d'oxydes d'azote sur une journée à Lille.

- Les *oxydes d'azote*
  Ils sont communément définis comme $NO_x = NO + NO_2$. Ils proviennent essentiellement de la combustion des combustibles fossiles et de quelques procédés industriels. Les NO_x dans l'atmosphère sont émis principalement sous forme de NO. Une fraction a été transformée en NO_2 dans le foyer de combustion (environ

10 %). En zone urbaine, l'essentiel des NO_x provient du trafic routier, avec deux pics journaliers correspondant aux déplacements pour se rendre et revenir du lieu de travail (figure 1.2). En matière d'évolution, il est à noter que le surplus d'émissions causé par l'augmentation du parc automobile est compensé depuis plusieurs années par les améliorations des processus de combustion dans les moteurs.

- Les *composés organiques volatils* (COV)
  Les sources de COV sont très nombreuses. Les émissions sont dues à certains procédés industriels et surtout aux transports. On notera également que la biomasse est fortement émettrice (forêts) pour les terpènes, sans oublier non plus les émissions liées aux produits domestiques (peinture, produits d'entretien, parfums et cosmétiques, journaux, tabac, etc.).

- le *dioxyde de soufre* (SO_2)
  Il provient de la combustion des combustibles fossiles contenant du soufre : fioul, charbon. Les émissions sont en nette diminution depuis 1956 (voir figure 1.3). C'est aussi le cas depuis 10 ans.

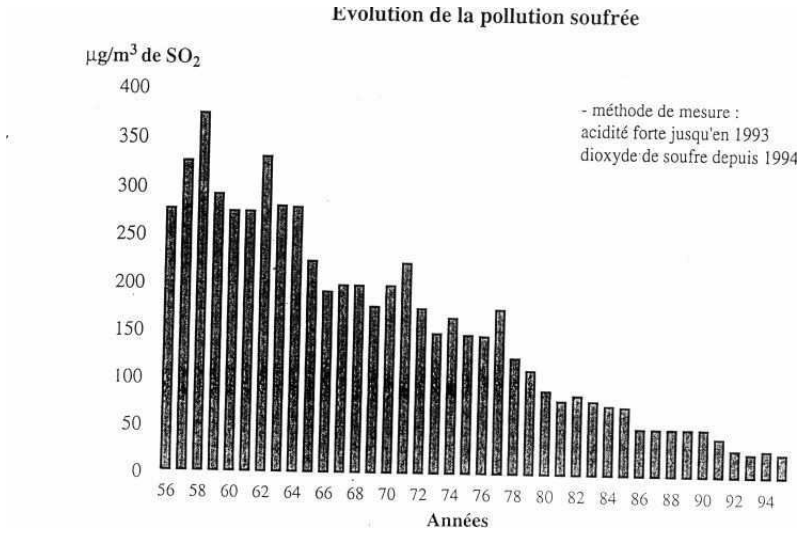

FIG. 1.3 – Evolution des émissions de SO_2 en France de 1956 à 1995.

- Le *monoxyde de carbone* (CO)
  Il provient de la combustion incomplète des combustibles et carburant.

- Les *métaux lourds* (Plomb, mercure, etc ...)
  Les sources sont d'origine anthropique et naturelle. Une part importante est due à la re-suspension (métaux lourds) et aux ré-émissions (mercure).

- Les *particules*
  Elles sont constituées d'un mélange complexe de substances organiques ou minérales

d'origine naturelle ou anthropique. Elles peuvent contenir des composés toxiques. Les émissions sont habituellement distribuées sur la partie la plus grossière du spectre en taille.

On se réfère à [Friedrich et Reis, 2004] pour une description plus complète des polluants émis. La figure 1.4 donne une ventilation typique par secteur d'activité pour l'année 1998.

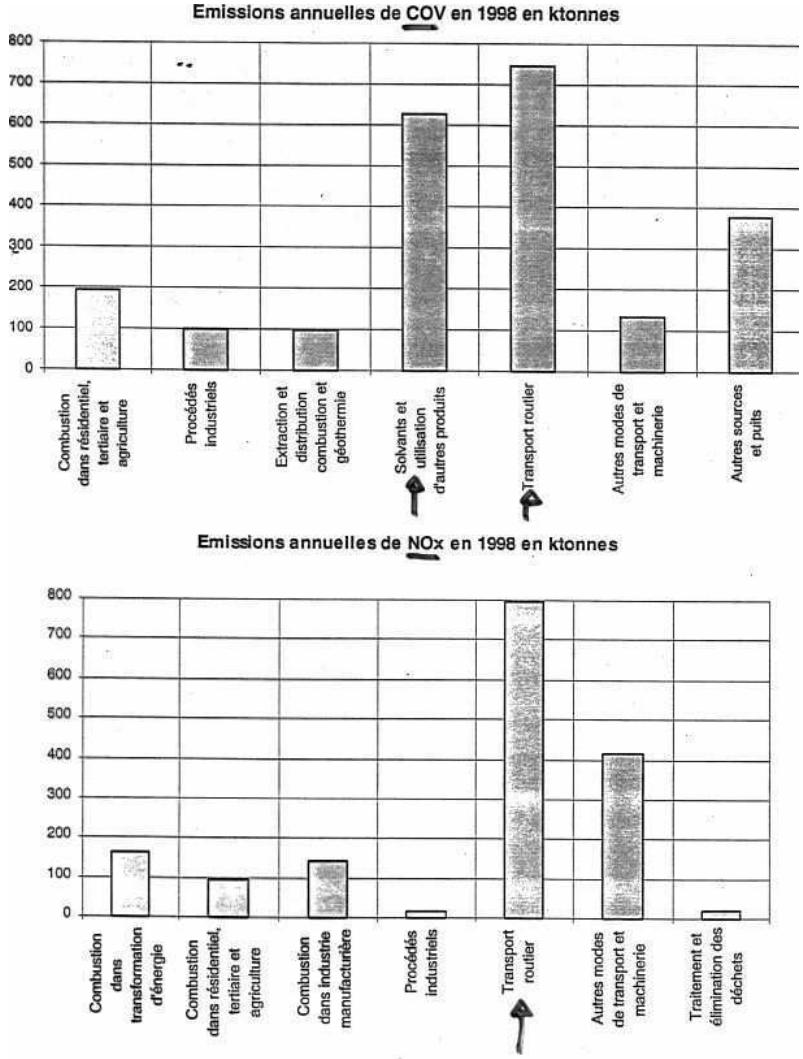

FIG. 1.4 – Emissions annuelles de NO_x et de COV en France par secteur d'activité en 1998.

1.1.1.2 La photochimie

On va maintenant s'attacher à décrire les mécanismes chimiques de formation de l'ozone et du dioxyde d'azote dans la troposphère. De manière simplifiée, la production photochimique d'ozone résulte de la photo–oxydation des COV en présence de NO_x.

1.1.1.2.1 L'équilibre photochimique La seule réaction de production de l'ozone est donnée par :

$$O_2 + O \quad \rightarrow \quad O_3 \tag{1.1}$$

Dans la troposphère, la source privilégiée d'oxygène atomique est la dissociation photolytique de NO_2 :

$$NO_2 + h\nu(300\,nm \leq \lambda \leq 400\,nm) \xrightarrow{J} NO + O \tag{1.2}$$

ce qui explique de manière immédiate les liens entre l'ozone et les oxydes d'azote dans la troposphère.

La réaction de destruction de l'ozone par oxydation de NO

$$NO + O_3 \xrightarrow{k} NO_2 + O_2 \tag{1.3}$$

permet d'équilibrer le cycle, de bilan nul, formé avec les réactions 1.2 et 1.1 : c'est le cycle de Chapman dans la troposphère. Ces réactions étant rapides, on peut en général faire l'approximation de l'équilibre quasi–stationnaire décrit par la relation :

$$J\,[NO_2] \simeq k\,[NO]\,[O_3] \tag{1.4}$$

Cet équilibre est valable en journée loin des émissions de précurseurs.

1.1.1.2.2 Oxydation des COV En s'oxydant, les COV donnent naissance à des espèces chimiques capables d'oxyder NO en NO_2 (en concurrence avec la réaction de consommation d'ozone 1.3). En bilan, sous réserve qu'il y ait une molécule de OH pour initier le processus

$$COV + 2NO + O_2 \rightarrow 2NO_2 + H_2O + RCHO \tag{1.5}$$

où RCHO désigne un aldéhyde.

Les radicaux péroxyles RO_2 et HO_2 sont des produits intermédiaires intervenant dans la réaction 1.5 qui se conservent au cours du processus (de même que OH). En réalité, selon la teneur du milieu en NO_x, ils ne sont pas toujours productifs en ozone à cause de réactions concurrentes.

1.1.1.2.3 Notion de régime chimique On peut schématiquement distinguer 2 cas limites ([Seinfeld, 1985, Honoré, 2000]) :

- Pour un rapport NO_x / COV élevé, par exemple en agglomération urbaine, les radicaux OH réagissent préférentiellement avec NO_2 pour former du HNO_3. Augmenter les NO_x a peu d'effet. On parle de régime "COV limité".

- Pour un rapport NO_x / COV faible, ce sont les réactions de recombinaisons des radicaux péroxyles qui prennent le dessus, aboutissant à la formation d'ozone. Une augmentation des COV a peu d'effet sur la production d'ozone. Dans ce cas, le régime est dit "NO_x limité".

La figure 1.5 représente le pic d'ozone (en μgm^{-3}) en fonction d'un coefficient multiplicatif appliqué aux émissions de NO_x et de COV pour la région Lilloise le 11 Mai 1998. La situation réelle (coefficients égaux à 1) correspond à un régime "NO_x limité". Selon l'endroit où l'on se place sur le graphique, le pic d'ozone décroît préférentiellement avec une réduction des émissions de COV ou des NO_x. On remarque aussi qu'une réduction des émissions de NO_x peut même conduire à une augmentation du pic d'ozone.

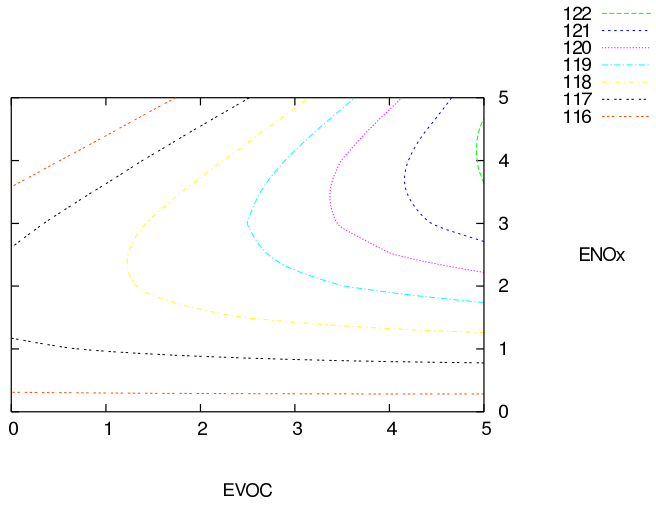

FIG. 1.5 – Pic d'ozone en fonction d'un coefficient multiplicatif appliqué aux émissions de NO_x et de COV.

Même si la photolyse n'est plus active de nuit, la chimie nocturne a une incidence non négligeable, en particulier du fait de NO_3 qui joue un rôle analogue au radical hydroxyl. Le jour, cette espèce est immédiatement photolysée. La nuit, elle peut devenir la principale espèce oxydante.

Pour synthétiser, Les éléments suivants interviennent donc dans la production d'ozone :

- Les NO_x et le radical OH sont des catalyseurs.
- Les concentrations de COV, OH et NO_x, et surtout le rapport NO_x / COV ont un effet sur la vitesse de production d'ozone.

1.1.1.3 La dispersion turbulente

Les conditions météorologiques conditionnent pour une large part le niveau de pollution dans les basses couches de la troposphère. Les polluants émis sont en effet plus ou moins rapidement dispersés selon deux phénomènes :

- le transport horizontal par le champ de vent, qui explique le déplacement des polluants les plus stables sur de longues distances.
- le brassage vertical par la turbulence atmosphérique.

L'étude de la Couche Limite Atmosphérique (CLA), la partie de la troposphère directement soumise à l'influence de la surface terrestre et surtout du cycle diurne, est donc essentielle pour la compréhension des épisodes de pollution. Elle est comprise en moyenne entre 100 et 3000 mètres.

- La partie supérieure de la CLA est appelée la couche d'Ekman. Les masses d'air y subissent à la fois le frottement de l'air sur la surface terrestre, la stratification thermique de l'air et la force de Coriolis qui entraîne une rotation du vent.

- La couche limite de surface (CLS) représente environ 10% de la CLA soit quelques dizaines de mètres. La turbulence y est homogène et la force de Coriolis est négligeable devant les forces de frottements dues au sol. La direction du vent (composante horizontale) ne varie pas avec la hauteur alors que son intensité est proportionnelle au logarithme de l'altitude (couche log).

- La partie inférieure de la couche de surface, directement en contact avec la surface terrestre est la sous-couche rugueuse. Elle se trouve au voisinage immédiat du sol. C'est une zone de mélange suite aux obstacles rencontrés par le vent. Les champs de vitesses y sont fortement hétérogènes et instationnaires. Son épaisseur varie en fonction de la nature du terrain, de quelques millimètres en mer à quelques dizaines de mètres en zone urbaine. C'est au travers de cette couche que se font les échanges d'humidité et de chaleur entre le sol et l'atmosphère. Elle est caractérisée par une longueur de rugosité dynamique et thermique.

On peut considérer que la turbulence atmosphérique a deux origines :

- La turbulence d'origine *dynamique* traduit l'agitation de l'air due au frottement avec le sol (effet de cisaillement).

– La turbulence d'origine *thermique* liée au gradient vertical de température provient des capacités d'absorption du rayonnement solaire par la surface. La variation diurne du rayonnement solaire y est directement perceptible.

Ce phénomène de turbulence fait intervenir la notion de *stabilité* de l'atmosphère. On distingue classiquement trois états de la CLA définis en fonction du gradient vertical de température potentielle :
– la CLA *neutre*,
– la CLA *instable* (air chaud en dessous de l'air froid), pour laquelle les polluants sont fortement brassés par la turbulence,
– la CLA *stable* (souvent nocturne) où les polluants restent accumulés au niveau du sol.

Le cycle diurne de la couche limite atmosphérique joue un rôle important dans les niveaux de concentrations de polluants observés. Un des paramètres déterminant est la hauteur de la CLA (ou *hauteur de mélange*), qui donne une indication du volume disponible pour la dispersion des polluants émis au niveau du sol. D'un jour sur l'autre, elle varie en fonction de la couverture nuageuse, de la pression atmosphérique et de la température de surface.

Pendant la nuit, certains polluants comme l'ozone sont stockés dans la couche réservoir qui se forme au-dessus de la couche de surface. Durant le jour, ils se dispersent à nouveau dans toute la couche limite atmosphérique et contribuent à l'augmentation des concentrations au sol.

Le phénomène d'inversion de température peut être à l'origine d'une augmentation des concentrations en polluants. Normalement la température de l'air diminue avec l'altitude (dans les basses couches de l'atmosphère), l'air chaud chargé de polluants se disperse à la verticale. Lorsque le sol s'est fortement refroidi pendant la nuit (par temps clair en hiver), et que la température à quelques centaines de mètres d'altitude est plus élevée que celle du sol, alors il y a phénomène d'inversion de la température au voisinage du sol. Les polluants se trouvent alors bloqués par cette masse d'air chaude en altitude plus communément appelée couche d'inversion.

1.1.1.4 Le dépôt

La rugosité et le revêtement du sol influencent l'écoulement des flux gazeux et détermine la quantité de polluants éliminés par dépôt sec. Les espèces chimiques se décomposent sur la surface terrestre au contact de matériaux et de plantes. Ainsi, les feuilles des plantes peuvent décomposer 10 à 20 % de l'ozone présent en une journée. L'évolution temporelle du dépôt sec lié à la végétation est déterminée par les saisons

(présence de feuilles) et l'éclairement (la nuit, les stomates des plantes se referment).

Enfin, les précipitations peuvent entraîner (lessiver) avec elles une partie des polluants. Le dépôt humide appelé aussi lessivage sous les nuages traduit ainsi l'absorption des polluants par les gouttes de pluie. C'est un processus considéré comme peu important en ce qui concerne les espèces photochimiques qui s'applique essentiellement aux particules de grandes tailles ou aux espèces les plus solubles.

1.1.2 Cadre général des CTM

Le paragraphe précédent a permis de situer la pollution atmosphérique dans un cadre global, de préciser les polluants émis, de mettre l'accent sur le rôle des conditions météorologiques pour la pollution troposphérique et de donner quelques éléments sur les principaux phénomènes physico-chimiques dans l'atmosphère. On va à présent s'intéresser à la modélisation de ces phénomènes par les modèles de chimie–transport.

Les CTM reconstituent l'information de la composition chimique de l'air en tout point d'une grille maillant le domaine d'étude grâce à la résolution d'équations physiques et chimiques qui décrivent les phénomènes présentés dans la section précédente.

Les modèles photochimiques peuvent être formulés dans les deux référentiels Lagrangien et Eulérien. Certaines applications de transport sur de longues distances ont été menées historiquement en utilisant des trajectoires Lagrangiennes, en partie pour les faibles ressources mémoires qui sont nécessaires pour suivre un volume d'air. A présent, la plupart des modèles adoptent une démarche eulérienne.

Une autre classe de modèles simplifiés est constituée par les modèles de boîtes qui simulent numériquement un réacteur chimique avec éventuellement des termes sources et des processus de surface comme les émissions surfaciques et le dépôt sec (modèle 0D). Leur intérêt réside dans leur simplicité de mise en œuvre et dans leur rapidité d'exécution. Plusieurs boîtes peuvent être connectées ensemble afin d'obtenir une description plus réaliste. Par exemple, les boîtes peuvent être empilées pour respecter la structure verticale en couche de l'atmosphère (modèle 1D).

Un point clé est l'utilisation du couplage "off-line" qui consiste à résoudre l'évolution des concentrations de polluants en utilisant des champs de forçage provenant d'un modèle météorologique. Pour pouvoir tenir compte de l'impact des polluants sur les champs météorologiques, certains modèles évoluent vers l'inclusion de la résolution des équations

décrivant le transfert radiatif et dynamique de l'atmosphère (Par exemple MESO-NH-C, [Tulet, 2000]).

1.1.2.1 Les échelles

Les phénomènes relatifs à la pollution atmosphérique se déclinent selon différentes échelles d'espace et de temps. Selon l'échelle spatiale visée, les phénomènes représentés ne sont pas les mêmes. On distingue :

- le niveau *local* ou pollution de proximité dont l'échelle de temps est de l'ordre de l'heure. On s'intéresse dans ce cas aux grandes villes et zones fortement peuplées parce que c'est là que se situe la plupart des sources de pollution due aux activités humaines et que la majeure partie de la population réside en respirant l'air pollué. La pollution provient des effets directs du chauffage, des fumées des usines et des transports.

- le niveau *régional* ou pollution à longue distance dont l'échelle de temps, de l'ordre de quelques jours, permet de prendre en compte la pollution qui se forme à des centaines de kilomètres de leur sources émettrices comme l'acidification et la pollution photochimique.

- le niveau *global* ou pollution planétaire dont l'échelle de temps est de l'ordre de quelques mois à quelques années. Les polluants les plus stables chimiquement sont étudiés. Il s'agit en particulier de la destruction de l'ozone stratosphérique et de l'effet de serre.

1.1.2.2 Les applications

Plusieurs raisons peuvent amener à l'utilisation d'un CTM. Les applications courantes sont typiquement les suivantes :

- La compréhension des phénomènes passe par l'analyse d'épisodes de pollution afin d'en évaluer les causes (Quelle est l'influence de tel processus au regard de celle de tel autre pour cette situation de pic de pollution ?). Ce sont des états réels que l'on s'attache à reproduire *a posteriori* en analysant les processus prépondérants dans une situation particulière.

- Les études d'impacts consistent à évaluer l'effet de mesures appliquées à une situation, comme par exemple, la réduction des émissions. Il peut s'agir aussi d'études prospectives à plus ou moins long terme pour évaluer l'impact en terme

de qualité de l'air de la mise en place d'un plan de déplacement urbain. Ici, c'est la relation entrée/sortie qui doit être décrite précisément. Un exemple typique est donné dans le chapitre 2.

- La représentation de la qualité de l'air sur un site donné (cartographie) peut donner une évaluation de la qualité de l'air sur des zones peu mesurées.

- La prévision opérationnelle de la qualité de l'air en terme de pic ou de valeurs moyennes de pollution commence également à être effectuée (voir par exemple les prévisions quotidiennes de PREV'AIR, `http://prevair.ineris.fr`).

1.2 La plate-forme POLAIR3D

On va à présent s'attacher à donner une description du système POLAIR3D qui simule l'évolution d'un polluant dans l'atmosphère.

1.2.1 Structure d'un CTM

Une simulation de l'évolution des polluants dans l'atmosphère peut être décomposée en trois étapes :

1. La construction des entrées du modèle : émissions, météorologie, dépôt, conditions aux limites, atténuation photolytique. Cela nécessite en particulier de disposer de champs météorologiques tels que le vent, la température, la pression, l'humidité, la radiation solaire, le contenu en eau des nuages et la couverture nuageuse. Cette étape est réalisée en amont du modèle POLAIR3D en utilisant la librairie AtmoData [Mallet, 2004], qui permet le traitement des données atmosphériques et le calcul d'entrées pour un code de chimie atmosphérique.

2. La résolution du système d'équations de conservation pour chaque polluant (POLAIR3D).

3. L'évaluation du modèle par comparaison aux observations.

Ces trois parties ont été bien séparées dans la conception de POLAIR3D et les échanges entre les trois modules se fait par l'intermédiaire de fichiers binaires.

La figure 1.6 schématise les différentes composantes d'un CTM et leurs interactions.

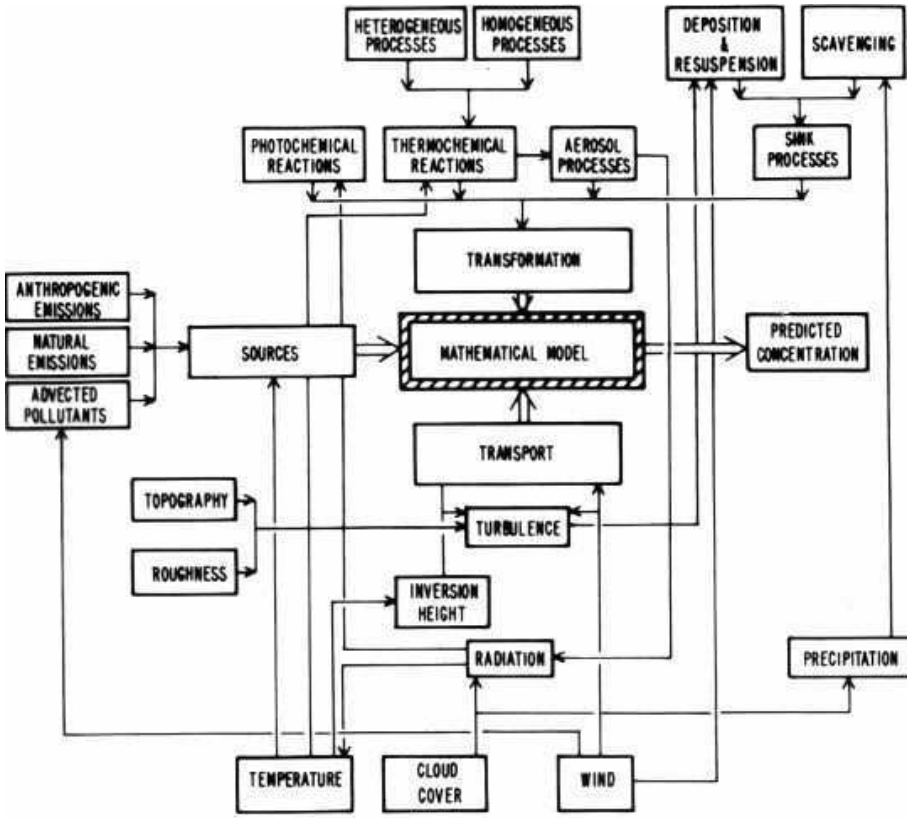

FIG. 1.6 – Composantes d'un modèle de qualité de l'air [Seinfeld, 1985]

Les données d'entrées indispensables pour l'application de POLAIR3D peuvent être ainsi classées en plusieurs catégories :

- les champs météorologiques : vent, flux solaire, température, pression, humidité, couverture nuageuse, précipitations, contenu en eau liquide des nuages.
- le type de sol : rugosité, couvert végétal.
- les inventaires d'émissions : principaux émetteurs, quantités émises par type d'activité et par secteur géographique.
- les espèces : solubilité, réactivité, diffusivité, masses molaires, conditions aux limites.

1.2.2 Equation de dispersion atmosphérique

L'équation de dispersion atmosphérique est obtenue à partir de l'équation de conservation après décomposition de Reynolds et application de la théorie–K pour modéliser les fluctuations très petites de la turbulence. Pour chaque espèce chimique indicée par i, on obtient :

$$\frac{\partial C_i}{\partial t} = -\mathrm{div}\,(UC_i) + \mathrm{div}\,(K\nabla C_i) + \chi_i\,(C) - \Lambda_i C_i + S_i \tag{1.6}$$

où C désigne le vecteur des concentrations, U est le vecteur de vent moyen, K est la diffusion turbulente, χ_i est la production–destruction chimique, Λ_i est le coefficient de lessivage sous les nuages et S_i est une émission ponctuelle. Les termes qui figurent dans cette équation sont explicités dans la suite.

Les flux au sol (dépôt et émission) interviennent comme condition au limite de la diffusion verticale, qui s'écrit pour l'espèce i :

$$\phi^i_{sol} = E_i - v^i_{dépôt}\, C_i \tag{1.7}$$

avec E_i les émissions et $v^i_{dépôt}$ la vitesse de dépôt de l'espèce i.

Le modèle adopte une approche tri-dimensionnelle eulérienne, qui consiste à diviser en mailles horizontales et verticales la région considérée et à y résoudre les différents processus en tenant compte de leurs interactions. Le domaine est dit à *aire limitée*, c'est-à-dire qu'il est centré sur la région d'intérêt. Les coordonnées sont des latitudes–longitudes permettant de prendre en compte la courbure de la Terre. Les altitudes sont définies à partir du sol et incluent donc le relief. Le niveau inférieur du modèle se situe au-dessus d'une hauteur répercutant la rugosité de la canopée. L'accent est porté sur les trois premiers kilomètres en bas de la troposphère dans lesquels se développe la couche limite atmosphérique.

En coordonnées sphériques l'équation de conservation 1.6 s'écrit :

$$\begin{aligned} \frac{\partial C_i}{\partial t} &+ \frac{1}{z\cos\phi}\left[\frac{\partial\, uC_i}{\partial\lambda} + \frac{\partial\, vC_i\cos\phi}{\partial\phi}\right] + \frac{1}{z^2}\frac{\partial\, wC_i z^2}{\partial z} \\ &= \frac{1}{z\cos\phi}\left[\frac{\partial}{\partial\lambda}\left(\frac{K_\lambda}{z\cos\phi}\frac{\partial C_i}{\partial\lambda}\right) + \frac{\partial}{\partial\phi}\left(K_\phi\cos\phi\frac{\partial c}{\partial\phi}\right)\right] + \frac{1}{z^2}\frac{\partial}{\partial z}\left(z^2\,K_z\frac{\partial c}{\partial z}\right) \\ &+ \chi_i\left(C,\lambda,\phi,z,t\right) + S_i\left(\lambda,\phi,z,t\right) \end{aligned} \tag{1.8}$$

où u, v, w sont les composantes de la vitesse du vent dans le système de coordonnées (λ, ϕ, z), λ et ϕ désignent respectivement la longitude et la latitude. z est la coordonnée verticale mesurée à partir du centre de la Terre.

Il est pratique de faire l'approximation $z = r$ où $r \simeq 6378$ kilomètres est le rayon terrestre supposé constant et d'introduire le changement de variable :

$$\begin{cases} x{=}r.\lambda \\ y{=}r.\sin(\phi) \\ z{=}z - z_{sol} \end{cases}$$

L'équation s'écrit alors :

$$\begin{aligned} \frac{\partial C_i}{\partial t} &+ \frac{\partial}{\partial x}\left(U^* C_i\right) + \frac{\partial}{\partial y}\left(V^* C_i\right) + \frac{\partial}{\partial z}\left(wC_i\right) \\ &= \frac{\partial}{\partial x}\left(K_x\frac{\partial C_i}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y\frac{\partial C_i}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z\frac{\partial C_i}{\partial z}\right) \\ &+ \chi_i\left(C,x,y,z,t\right) + S_i\left(x,y,t\right) \end{aligned} \tag{1.9}$$

où

$$U^* = u/\cos\phi\ ,\ \ V^* = v\,\cos\phi \tag{1.10}$$

$$K_x = K_\phi/\cos^2\phi\ ,\ \ K_y = K_\lambda\,\cos^2\phi \tag{1.11}$$

L'équation (1.9) est l'équation générale qui gouverne le modèle POLAIR3D. Une modification des variables d'entrées : vitesse du vent et constante de diffusion horizontale permet donc de résoudre une équation comparable à celle que nous aurions dans un repère cartésien.

1.2.3 Composantes du modèle et description des processus

1.2.3.1 Advection et diffusion

Deux termes de transport apparaissent dans l'équation (1.6). Au premier ordre, l'advection est essentiellement horizontale et la diffusion verticale. Il est important de

bien décrire ces deux termes de transport notamment lorsqu'ils sont tous les deux faibles parce que cette situation favorise l'accumulation de polluants et conduit à l'apparition de pics de pollution.

Lorsqu'elles ne sont pas directement adaptées à la résolution du maillage, les composantes horizontales du vent sont interpolées linéairement et la composante verticale est recalculée pour respecter l'équation de continuité selon :

$$\frac{U_{i+1,j,k} - U_{i,j,k}}{\Delta x_{i,j,k}} + \frac{V_{i,j+1,k} - V_{i,j,k}}{\Delta y_{i,j,k}} + \frac{W_{i,j,k+1} - W_{i,j,k}}{\Delta z_{i,j,k}} = 0 \tag{1.12}$$

A moins de résoudre toutes les échelles, il est nécessaire d'introduire la décomposition de toute grandeur physique ϕ en un champ moyen $\bar{\phi}$ et une fluctuation ϕ'. Ainsi, $\phi = \bar{\phi} + \phi'$ avec $\bar{\phi'} = 0$. Dans l'équation 1.6, le terme de corrélation $w'c'$ est classiquement modélisé en utilisant la théorie K par $-K\nabla\bar{c}$, traduisant que le flux vertical turbulent est proportionnel au gradient de concentration vertical. On notera qu'en toute rigueur, cette approche devrait aussi être appliquée pour le terme de chimie[†1].

Le coefficient K_z est déterminé à partir la paramétrisation de Louis [Louis *et al.*, 1982]. Le calcul met en jeu une fonction qui diffère selon le signe du nombre de Richardson (indicateur de la stabilité de l'atmosphère). On considère que la troposphère n'est jamais parfaitement stable ce qui impose une valeur critique maximale pour le nombre de Richardson Ri_c. Une valeur théorique de 0.21 est préconisée mais les modèles utilisent en pratique une valeur comprise entre 0.15 et 1.

La grande sensibilité des paramètres est mise en évidence sur la figure 1.7. Elle représente des valeurs horaires d'ozone en μg.m^{-3} pendant dix jours sur Europe. L'influence est très nette la nuit où un brassage vertical plus important conduit à des valeurs d'ozone plus fortes. Comparativement les pics de l'après-midi sont peu influencés par ce paramètre.

Les flux de concentrations de polluants apportés de l'extérieur du domaine sont pris en compte par l'intermédiaire de concentrations advectées au niveau des faces latérales. Elles proviennent par exemple d'une simulation à plus grande échelle. A l'échelle continentale, MOZART2 ([Hauglustaine *et al.*, 1998]) fournit des valeurs journalières typiques pour certaines espèces chimiques. Dans le cadre d'une application régionale, nous la couplons à une simulation à l'échelle continentale (one-way nesting).

[†1]Comme décrit dans [Sportisse, 1999], l'erreur introduite dépend de la diffusion et de la non-linéarité de la chimie.

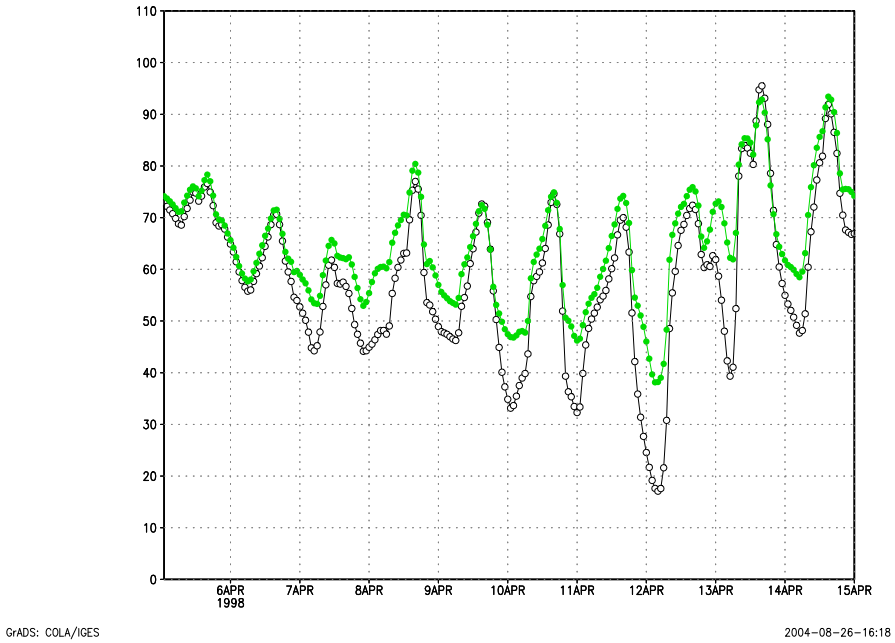

FIG. 1.7 – Concentrations horaires d'ozone à Lille (Simulation européenne) pour deux valeurs du nombre de Richardson critique. Les ronds pleins (respectivement vides) correspondent à $Ri_c = 0.25$ (respectivement $Ri_c = 1$).

1.2.3.2 Dépôt

Le flux de dépôt sec est paramétrisé classiquement par une vitesse de dépôt que multiplie une concentration. La modélisation utilisée met en jeu trois résistances en série et représente relativement bien le processus [Wesely, 1989]. Chacune des résistances traduit un phénomène devenant prépondérant au cours du transfert de l'atmosphère vers le sol.

$$\phi_{dépôt} = v_{dep}\, C = \frac{C}{R_a + R_b + R_c} \tag{1.13}$$

où :

- R_a désigne la résistance aérodynamique et est pilotée principalement par la turbulence atmosphérique. Elle est indépendante du polluant considéré.

- R_b correspond à la résistance attribuée à la sous-couche très proche de la canopée. Elle traduit l'importance de la diffusion moléculaire dans la fine couche quasi-laminaire d'air en contact avec la surface du sol.

- R_c est la résistance de la canopée. Elle est fonction du polluant, de l'occupation du sol, de l'humidité et de la saison. Elle correspond aux interactions physico-chimiques entre le polluant et la surface d'absorption. Elle inclut notamment l'action du couvert végétal. Les jeux de constantes sont attribuées en fonction de la saison et

les phases de transition avec et sans feuilles ne sont pas précisément décrites ce qui conduit à de fortes variations lors des changements de saisons.

La figure 1.8 représente des valeurs typiques de dépôt sec d'ozone pour l'année 1998.

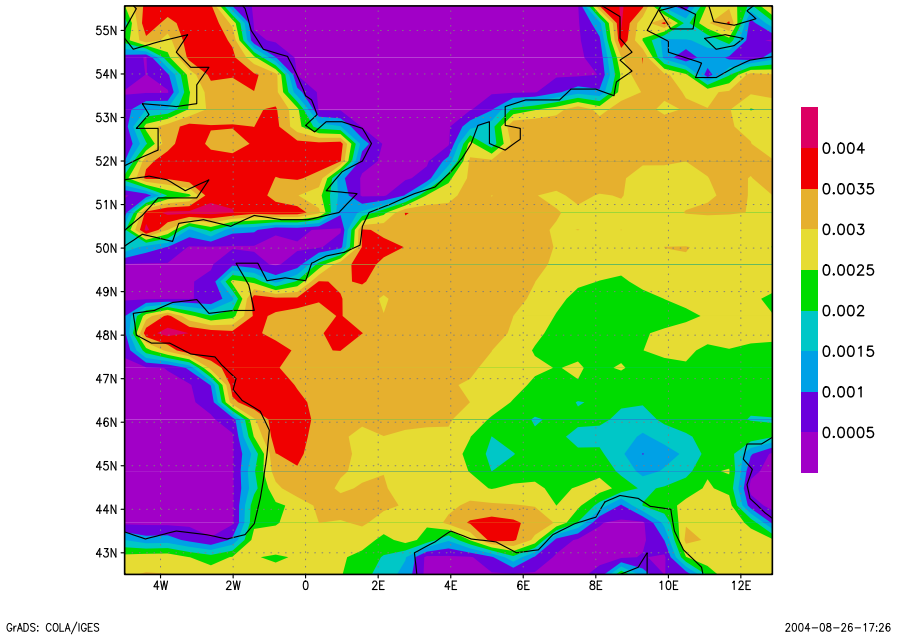

FIG. 1.8 – Moyenne annuelle de vitesse de dépôt d'ozone en $m.s^{-1}$.

1.2.3.3 Les émissions

Les cadastres d'émissions recensent l'ensemble des émissions gazeuses d'une zone géographique avec leur distribution spatiale et temporelle. Ces bases de données sont établies à partir de modèles d'émissions tenant compte de données diverses (population, indice de production, comptages routiers, consommation d'énergie, température, flux solaires, occupation des sols, ...).

Deux types d'approches sont utilisées simultanément pour constituer cette base. La première de type "top-down" consiste à collecter des données générales qui concernent la zone d'étude et à les redistribuer aux échelles spatiales et temporelles souhaitées. La seconde approche de type "bottom-up" a pour principe de collecter les données les plus fines possibles et de les rassembler pour obtenir la résolution voulue.

S'il est relativement simple d'établir un inventaire pour les sources fixes, le travail est beaucoup plus complexe pour les sources mobiles. En pratique, il est possible de réaliser

des reconstructions de cadastre à partir de mesures du trafic.

Pour utiliser ces données dans les modèles de qualité de l'air, il est nécessaire de connaître la composition chimique de chaque source d'émissions et de quantifier son flux avec une résolution spatiale et temporelle en relation avec le domaine d'étude et la période considérée, ainsi qu'avec le modèle chimique utilisé. Typiquement, un inventaire comprend les émissions totales annuelles d'espèces inventaires comme CO, NO_x, SO_2 et COV qui sont référencées par secteur d'activité selon la nomenclature SNAP (Selected Nomenclature for Air Pollution). Eventuellement, elles peuvent déjà avoir été réparties par mailles en utilisant des critères tels que les tailles de population par exemple.

Les émissions sont décomposées le plus couramment en des flux de surface qui servent de condition au limite au sol (c'est le terme E_i dans l'équation 1.7) et des flux volumiques en altitude (S_i, voir l'équation 1.6). Nous avons modifiée la distribution verticale utilisée dans le modèle EMEP (tableau 1.1) pour qu'elle soit adaptée à la résolution verticale du modèle. Elle considère que les émissions provenant du trafic routier, de l'utilisation de solvants, de l'agriculture sont exclusivement émises au sol tandis que les autres secteurs d'activités émettent à la fois au sol et en altitude. Pour l'industrie (SNAP 1), les niveaux élevés d'émissions correspondent à la prise en compte d'une sur-hauteur d'émission provenant du fait que les gaz d'éjection en sortie de cheminée montent rapidement dans l'atmosphère en raison de leur température élevée.

hauteur	sol	92 m	184 m	324 m	522 m	781 m	1106 m
SNAP 1	0	0	0	8	46	29	17
SNAP 2	16.7	33.3	50	0	0	0	0
SNAP 3	0	0	4	19	41	30	6
SNAP 4	30	60	10	0	0	0	0
SNAP 5	30	60	10	0	0	0	0
SNAP 6	100	0	0	0	0	0	0
SNAP 7	100	0	0	0	0	0	0
SNAP 8	100	0	0	0	0	0	0
SNAP 9	3.3	6.7	15	40	35	0	0
SNAP 10	100	0	0	0	0	0	0

TAB. 1.1 – Répartition verticale des émissions par SNAP utilisée dans le modèle EMEP.

Lorsqu'une hauteur d'émission est disponible, comme c'est le cas par exemple pour des cheminées industrielles, les émissions sont dispersées instantanément dans tout le volume de la maille correspondante (c'est le terme S_i dans l'équation 1.6). En pratique, cela peut conduire à des maxima de concentrations trop forts. Pour l'éviter, on peut faire appel à un sous-modèle (plume-in-grid) qui considère individuellement des bouffées lagrangiennes soumises à la dispersion et à l'évolution chimique jusqu'à ce que leur développement soit

suffisant pour être insérées dans le modèle.

La répartition horizontale appliquée par le modèle de génération d'émissions est faite selon un critère qui accorde un poids plus important aux types de sol urbain, en comparaison à la forêt ou aux autres types de sol. Pour une maille inventaire d'émission totale E comportant N_{urb} sous-mailles urbaines, N_{for} sous-mailles de forêt et N_{autres} sous-mailles des autres types de sol, l'émission $e_{<type>}$ d'une sous-maille dont le type de sol est $< type >$ (à choisir entre urbain, forêt et autres) est donnée par :

$$e_{<type>} = \frac{E}{\gamma_{urb}N_{urb} + \gamma_{for}N_{for} + \gamma_{autres}N_{autres}} \times \gamma_{<type>}$$

où $\gamma_{<type>}$ est le poids accordé à chaque type de sol. Par exemple, nos simulations sont effectuées avec $\gamma_{urb} = 12$, $\gamma_{for} = 1.6$ et $\gamma_{autres} = 1$.

La répartition temporelle est réalisée sur la base des coefficients mensuels, journaliers et horaires donnés par pays et par classes SNAP (GENEMIS). A chaque espèce inventaire correspond une spéciation en espèces réelles, puis cette description détaillée est agrégée en espèces modèles (figure 1.9). Par exemple, les COVNM sont ventilés en classes de composés de réactivités similaires puis agrégés en utilisant une procédure spécifique proposée par Middleton [Middleton *et al.*, 1990] qui fait varier les coefficients d'agrégation en fonction de l'intégrale temporelle de OH.

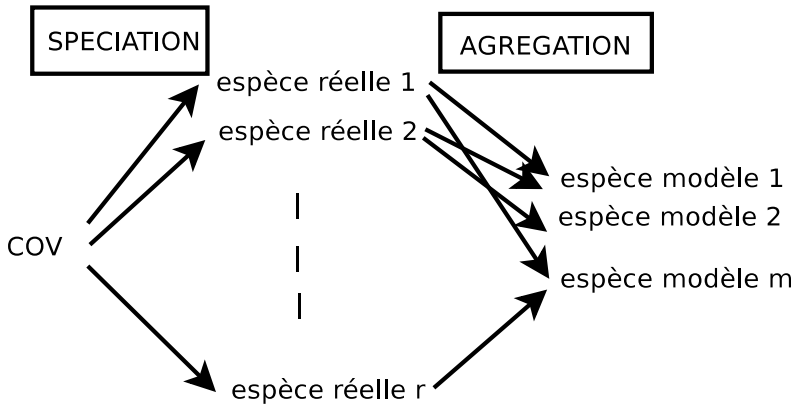

FIG. 1.9 – Spéciation-agrégation des composés organo-volatils.

Les émissions biogéniques (terpènes et isoprènes) sont calculées séparément en fonction du type de sol, de l'éclairement et de la température en suivant [Simpson *et al.*, 1999].

La figure 1.10 représente les valeurs d'émissions surfaciques de NO_x pour l'année 1998.

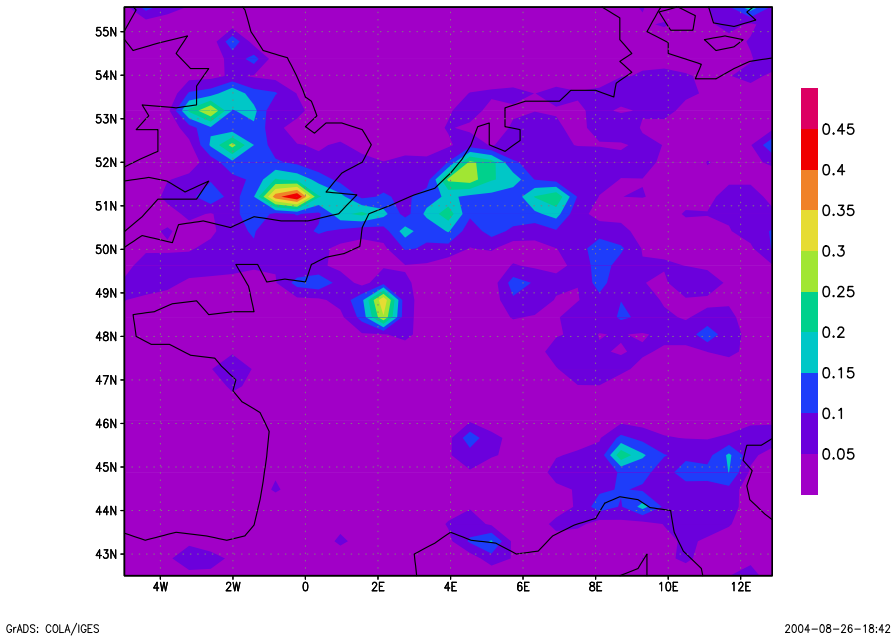

FIG. 1.10 – Emissions surfaciques moyennes de NO_x ($\mu g.m^{-2}s^{-1}$).

1.2.3.4 Chimie troposphérique

Les mécanismes gazeux étant à l'origine de la pollution photochimique sont à présent relativement bien compris et ont été implémentés dans un nombre important de schémas chimiques.

Les composés chimiques dans l'atmosphère sont de l'ordre de plusieurs milliers et les réactions chimiques se produisant entre eux sont encore plus nombreuses. Actuellement, tous ces composés ne sont pas bien connus et il est impossible de simuler toutes ces réactions chimiques dans les modèles 3D pour des raisons évidentes de coût de calcul. Les mécanismes chimiques utilisés sont donc basés sur une représentation simplifiée des espèces chimiques en les regroupant par classe de composés ayant des propriétés moléculaires ou réactionnelles semblables. Un large éventail de schémas chimiques existent comme par exemple RACM [Stockwell *et al.*, 1997b], RADM2, SAPRC, CBM-IV, MELCHIOR, etc ... Un outil de génération automatique de réactions chimiques comme SPACK (Simplified Preprocessor for Atmospheric Chemical Kinetics [Djouad *et al.*, 2002a]) s'avère utile pour générer facilement les mécanismes.

Les constantes photolytiques sont déterminantes pour la simulation des espèces photochimiques. L'approche retenue consiste à les calculer par ciel clair, ce qui permet de les tabuler uniquement en fonction de la latitude, de l'altitude et du temps. Un facteur correctif (atténuation) fonction de la couverture nuageuse leur est appliqué localement.

1.2.4 Schémas numériques

La résolution numérique soulève de nombreuses difficultés, notamment du fait de la taille du système à considérer (plusieurs dizaines d'espèces chimiques) mais également à cause de la dispersion des échelles de temps caractéristiques des phénomènes chimiques (de quelques nanosecondes pour les espèces les plus réactives à des mois ou des années pour les espèces les plus stables). Ceci induit la *raideur* de ce système, qui constitue un des principaux obstacles pour la simulation (impact en terme de coût calcul, utilisation de solveurs spécifiques *implicites*, ...).

POLAIR3D utilise une méthode de décomposition d'opérateurs (time splitting) qui consiste à résoudre les différents processus de manière découplée ce qui permet l'application d'un schéma numérique spécifique à chaque processus. Le gain en temps de calcul est significatif. L'ordre de résolution des processus est important ([Sportisse, 1999]). Considérons l'équation de dispersion atmosphérique sous la forme fonctionnelle suivante :

$$\frac{\partial c_i}{\partial t} = L_{adv}(c_i) + L_{diff}(c_i) + L_{chem}(c)_i \tag{1.14}$$

Les ordres de résolution possibles sont les suivants :

- Splitting de premier ordre :
  1. Intégration de L_{adv} sur $[t, t + \Delta t]$,
  2. Intégration de L_{diff} sur $[t, t + \Delta t]$,
  3. Intégration de L_{chem} sur $[t, t + \Delta t]$,

- Strang Splitting (ordre 2) :
  1. Intégration de L_{adv} sur $[t, t + \frac{\Delta t}{2}]$,
  2. Intégration de L_{chem} sur $[t, t + \frac{\Delta t}{2}]$,
  3. Intégration de L_{diff} sur $[t, t + \Delta t]$,
  4. Intégration de L_{chem} sur $[t + \frac{\Delta t}{2}, t + \Delta t]$,
  5. Intégration de L_{adv} sur $[t + \frac{\Delta t}{2}, t + \Delta t]$,

- AMF (Approximate Matrix Factorization, [Botchev, 2003]) :
  Cette approche utilise l'ordre de résolution précédent en résolvant de manière couplée la chimie et la diffusion verticale. Elle est basée sur un splitting au niveau algébrique des matrices intervenant dans le solveur Rosenbrock (paragraphe 1.2.4.2) :

$$(I - \gamma \Delta t A_{chem+diff_z}) \sim (I - \gamma \Delta t A_{chem})(I - \gamma \Delta t A_{diff_z}) \tag{1.15}$$

où A désigne le jacobien du système. Cette approximation permet d'éviter l'inversion du système couplé dont la dimension est le produit du nombre de points de grille par le nombre d'espèces chimiques.

Des tests préalables de comparaison ont été effectués et sont présentés dans le tableau 1.2. Ils correspondent à une simulation européenne sur un an. Les sorties horaires de O_3, NO et NO_2 présentent plus de 99% de corrélation. Par la suite, les calculs de ce manuscrit ont été réalisés avec un splitting d'ordre 1.

		AMF	Strang	Ordre 1
NO_2	max	134.2543	134.0437	132.466
	mean	21.81169	21.80232	21.7875
NO	max	78.9695	79.0501	79.244
	mean	3.727317	3.725399	3.70977
O_3	max	336.1618	336.1939	335.5024
	mean	58.9784	58.9891	58.9971

TAB. 1.2 – Comparaison de différents types de splitting : AMF, Strang et ordre 1 sur des sorties horaires pour l'année 1998 à l'échelle européenne en valeur maximale et moyenne spatio-temporelles.

Le pas de temps utilisé pour les simulations est de l'ordre de 10 minutes. Il permet à la fois de respecter la condition de CFL mais aussi de limiter les erreurs introduites par la décomposition d'opérateur. Typiquement, le temps CPU pour un calcul européen est réparti comme suit :

- 50 % : chimie,
- 25 % : advection,
- 25 % : diffusion.

1.2.4.1 Intégration de l'advection

Le schéma numérique d'advection est le schéma 'Direct Space Time' (évoqué dans [Spee, 1998]) qui fait partie de la famille des schémas de type volumes finis avec limiteur de flux. Le limiteur de flux est particulièrement adapté pour résoudre le transport de polluants inégalement répartis spatialement, comme c'est le cas au sol à cause de l'hétérogénéité des émissions. Son principe est de calculer pour chacune des faces de la maille le flux de polluants F en faisant le meilleur compromis entre précision (discrétisation ordre 3) et stabilité (discrétisation ordre 1). Il s'écrit de la manière suivante :

$$c_i^{n+1} = c_i^n + (F_{i-\frac{1}{2}} - F_{i+\frac{1}{2}}) \tag{1.16}$$

où :

– Si $u_{i+\frac{1}{2}} \geq 0$:

$$F_{i+\frac{1}{2}} = \nu_{i+\frac{1}{2}} \left(c_i + \psi \left(\nu_{i+\frac{1}{2}}, \theta_i \right) (c_{i+1} - c_i) \right) \tag{1.17}$$

– Si $u_{i+\frac{1}{2}} < 0$:

$$F_{i+\frac{1}{2}} = -\nu_{i+\frac{1}{2}} \left(c_{i+1} + \psi \left(\nu_{i+\frac{1}{2}}, \frac{1}{\theta_{i+1}} \right) (c_i - c_{i+1}) \right) \tag{1.18}$$

ψ est le limiteur de flux de Sweby [Spee, 1998] et est donné par :

$$\psi(\nu, \theta) = \max(0, \min(1, d_0(\nu) + d_1(\nu)\theta, \mu\theta) \tag{1.19}$$

où

$$d_0(\nu) = \frac{1}{6}(2 - \nu)(1 - \nu) \ , \quad d_1(\nu) = \frac{1}{6}(1 - \nu^2) \ , \tag{1.20}$$

$$\nu_{i+\frac{1}{2}} = \frac{\Delta t}{\Delta x} \left| u_{i+\frac{1}{2}} \right| \ , \quad \theta_i = \frac{c_i - c_{i-1}}{c_{i+1} - c_i} \tag{1.21}$$

et le paramètre μ dans (1.19) et pris comme dans [Verwer *et al.*, 1998] selon :

$$\mu = \frac{1 - \nu}{\nu} \tag{1.22}$$

Ce schéma possède aussi les propriétés essentielles de conservation de masse, de monotonie et de positivité. Ces propriétés ne sont respectées qu'avec un champ de vent respectant l'équation de continuité. Aussi, il est important que le champ de vent utilisé par le modèle (souvent obtenu par interpolation du champ météorologique à plus grande échelle) soit à divergence nulle. Ceci est réalisé dans POLAIR3D par un recalcul de la vitesse verticale respectant cette condition.

La stabilité est assurée par le respect de condition de Courant-Friedrichs-Lewy (CFL) qui relie le pas de temps Δt, la discrétisation spatiale Δx et la vitesse du vent u par la relation :

$$\Delta t < \frac{\Delta x}{u} \tag{1.23}$$

1.2.4.2 Solveur de Rosenbrock

L'intégration numérique temporelle de la diffusion et de la chimie se fait par la méthode de Rosenbrock ([Verwer *et al.*, 1999]), qui est une méthode semi-implicite de la famille des méthodes de Runge-Kutta particulièrement adaptée à la résolution des systèmes raides. On se réfère par exemple à [Djouad *et al.*, 2002a] pour une description plus détaillée. Cette raideur est très marquée pour la chimie à cause de la disparité des vitesses de réactions. Cette méthode est performante mais reste coûteuse car elle résout le système par une double inversion d'un système linéaire. Considérons l'équation suivante :

$$\frac{\partial c}{\partial t} = f(t, c) \tag{1.24}$$

On note $\Delta t = t_{n+1} - t_n$ et A une approximation de $\frac{\partial f}{\partial c}(t_n, c_n)$.

On considère habituellement le schéma de Rosenbrock non-autonome :

$$c_{n+1} = c_n + \Delta t b_1 k_1 + \Delta t b_2 k_2 \tag{1.25}$$

avec

$$\begin{cases} k_1 = f(t_n, c_n) + \Delta t \gamma A k_1 \\ k_2 = f(t_n + \alpha_{12}\Delta t, c_n + \alpha_{21}\Delta t k_1) + \Delta t \gamma_{21} A k_1 + \Delta t \gamma A k_2 \end{cases} \tag{1.26}$$

La méthode est d'ordre deux et est consistante pour tout A, si et seulement si :

$$b_1 = 1 - b_2\,; \qquad \gamma_{21} = -\frac{\gamma}{b_2}\,; \qquad \alpha_{21} = \frac{1}{2\,b_2}. \tag{1.27}$$

De plus, la méthode est A-stable si et seulement si $\gamma \geq \frac{1}{4}$. Elle est L-stable pour $\gamma = 1 \pm \frac{\sqrt{2}}{2}$. La méthode programmée dans POLAIR3D utilise $b_2 = \frac{1}{2}$ et $\gamma = 1 + \frac{\sqrt{2}}{2}$. On remplace k_2 par $k_2 - 2k_1$, et, dans le cas autonome, on obtient le schéma suivant :

$$c_{n+1} = c_n + \frac{3}{2}\Delta t k_1 + \frac{1}{2}\Delta t k_2 \tag{1.28}$$

avec

$$\begin{cases} (I - \gamma \Delta t A) \cdot k_1 = f(t_n, c_n) \\ (I - \gamma \Delta t A) \cdot k_2 = f(t_{n+1}, c_n + \Delta t k_1) - 2k_1 \end{cases} \tag{1.29}$$

Le caractère creux de la matrice $(I - \gamma \Delta t A)$ est mis à profit en générant automatiquement le code qui n'effectue que les opérations nécessaires de la décomposition LU, ce qui apporte un gain de temps significatif (jusqu'à 3 pour le schéma RACM où la matrice comporte environ 80 % de "zéros")).

1.2.5 Utilisation antérieure de POLAIR3D et comparaison à des mesures

Pour évaluer la qualité d'une simulation, les concentrations simulées sont comparées statistiquement aux observations disponibles. Elles sont interpolées linéairement aux endroits où se trouvent les capteurs. On suppose que les observations sont une représentation satisfaisante de la réalité, notamment qu'elles ne contiennent pas une erreur systématique instrumentale ou de représentativité.

L'évolution des modèles se fait en général sur les critères suivants :

- Le *biais* s'obtient en sommant pour toutes les mesures la différence entre les sorties de modèle et les observations. Un biais nul indique qu'en moyenne le modèle ne commet pas d'erreur.
- La *RMS* (Root Mean Square) se calcule en sommant cette même différence préalablement élevée au carré. Elle permet d'estimer la dispersion des erreurs.
- Le coefficient de *corrélation* est un nombre compris entre -1 et 1. S'il est proche de 1, la variabilité du phénomène est bien représentée.

A l'échelle régionale, POLAIR3D a été comparé aux mesures de la campagne régionale ESQUIF sur l'Ile-de-France. La simulation a montré de bonnes capacités à reproduire les observations d'ozone sur la période d'observation intensive du 7 au 9 Août 1998, pendant laquelle de fortes valeurs de pollution ont été observées [Sartelet *et al.*, 2002]. POLAIR3D participe actuellement à l'exercice d'intercomparaison de modèles sur la région Fos-Berre-Marseille durant l'été 2001 (ESCOMPTE) et a également servi à mener des études d'impacts de la centrale de production thermique de Martigues ([Taghavi *et al.*, 2004]).

A l'échelle européenne, une simulation de quatre mois (de Mai à Août 2001) a été comparée aux observations de plusieurs pays [Mallet et Sportisse, 2004]. Les scores de prévision sur les pics d'ozone à 15 heures étaient un biais moyen de $-5.7\,\mu\text{g.m}^{-3}$, une corrélation de 69 % et une RMS de $25\,\mu\text{g.m}^{-3}$ à mettre en regard des niveaux moyens de concentrations de $80\,\mu\text{g.m}^{-3}$.

POLAIR3D a également servi à faire de la modélisation d'impact ([Boutahar, 2004]) et à simuler le mercure et les métaux lourds ([Roustan, 2005]). Enfin, deux schémas d'aérosols ont été implémentés dans POLAIR3D : le modèle MAM (Modal Aerosol Model, [Sartelet, 2004]) et le modèle SIREAM (Size Resolved Aerosol Model, [Debry, 2004]) et sont en cours d'évaluation.

Chapitre 2

Modélisation de la qualité de l'air sur la région de Lille

Résumé

Le modèle de Chimie-Transport POLAIR3D a été validé dans de nombreuses applications qui vont à l'échelle continentale, et ce pour plusieurs types de traceurs considérés (métaux lourds, mercure, traceur passif, photochimie).

On présente ici le cas de la modélisation de la qualité de l'air pour la région de Lille. Cette modélisation s'inscrit dans le cadre d'un projet du PREDIT, en collaboration notamment avec le CETE de Nord-Picardie, dont l'objectif était la quantification de l'impact en terme de qualité de l'air de plusieurs scénarii de transport pour le Plan de Déplacement Urbain de Lille.

Cette étude a donc permis de bien quantifier la sensibilité aux divers scénarii d'émission d'un modèle de Chimie Transport sur de longues périodes en temps (de l'ordre de l'année).

Sommaire

2.1 Les observations 45
2.1.1 Les stations de mesures 45
2.1.2 Evolution moyenne des mesures 46
2.2 Conditions de la simulation 49
2.2.1 L'échelle continentale 49
2.2.2 L'échelle régionale 51
2.2.3 Temps de calcul 55

2.2.4 La simulation 2015 56
2.3 Résultats pour l'année de référence 1998 et pour l'exercice 2015 57
2.3.1 Année 1998 57
2.3.2 Résultats pour 2015 63
2.4 Conclusion 65

Présentation

Cette étude de modélisation de la qualité de l'air a été réalisée pour évaluer l'impact du plan de déplacement urbain de Lille en terme de concentrations de polluants à l'échelle régionale. Ces concentrations représentent des teneurs de fond de polluants parmi lesquels les oxydes d'azote ont été sélectionnés comme étant des indicateurs de la pollution atmosphérique. Elles serviront par la suite à un modèle de santé dont le rôle est de permettre une évaluation des effets à long terme de l'exposition de la population pour deux scénarios à horizon 2015 : avec et sans plan de déplacement urbain.

Le CEREA[†1] a développé des outils permettant de modéliser ces processus et d'obtenir une description spatiale et temporelle fine des niveaux de polluants notamment sur un domaine régional telle que l'agglomération Lilloise. Cette étude décrit l'obtention de la variation dans l'espace et dans le temps des concentrations de polluants pour l'année de référence 1998 et l'année 2015 sur l'agglomération lilloise. Elle s'est appuyée sur la mise en œuvre de données d'émissions correspondant au trafic routier :

- Pour l'année de référence 1998.
- A horizon 2015 sans plan de déplacement urbain.
- A horizon 2015 avec plan de déplacement urbain.

Les échelles considérées sont d'une part continentale pour décrire la pollution de fond résultant du transport à longue distance des polluants contribuant de manière sensible aux teneurs mesurées sur Lille et d'autre part régionale pour prendre en compte les sources d'émissions situées dans l'environnement immédiat.

Ce chapitre est organisé de la manière suivante. Le réseau d'observations est décrit dans le paragraphe 1. La simulation sur l'agglomération lilloise est décrite dans une partie 2. Une troisième partie présente la comparaison de l'année de référence 1998 aux mesures et les résultats des simulations pour 2015.

2.1 Les observations

2.1.1 Les stations de mesures

Les champs de concentrations obtenus en sortie de simulation sont des moyennes à un instant donné sur l'ensemble de la maille. Généralement, la comparaison à des mesures s'effectue à partir de moyennes horaires ou journalières interpolées horizontalement au point de mesure. La valeur simulée dans la première maille verticale est utilisée. Elle se compare à des mesures représentatives de la pollution au-dessus de la canopée

[†1] laboratoire commun EDF R&D / ENPC

comme cela est défini dans les stratégies de surveillance de la qualité de l'air. Elle ne représente en aucun cas une concentration à hauteur de voies respiratoires (exposition réelle).

La qualité de l'air dans l'agglomération lilloise est surveillée par le réseau de surveillance de la qualité de l'air AREMA. Ce réseau de mesures inclut trois sortes de stations définies par le type de pollution que l'on veut mesurer : 6 stations de proximité automobile ou industrielle, 6 stations périurbaines et 7 stations urbaines denses, selon la typologie de l'ADEME. Aucune station rurale n'est recensée. Leur définition est la suivante :

- *Station urbaine* :
  Cette station se trouve en milieu urbain suffisamment éloignée de sources spécifiques de pollution pour mesurer une pollution moyenne. Elle est représentative de la qualité de l'air ambiant sur l'agglomération dans un rayon de 100 mètres à 2 kilomètres.

- *Station périurbaine* :
  Cette station se trouve en dehors de la zone agglomérée dans les espaces suffisamment habités. Elle permet de mesurer les polluants secondaires qui se forment par réaction chimique en périphérie des agglomérations. Le rayon de son aire de représentativité est compris entre 1 et 5 kilomètres.

- *Station de proximité* :
  Cette station urbaine mesure la pollution aux abords d'une source ponctuelle et/ou spécifique (industrie, voies de circulation importante ...). Sa représentativité peut être très locale (10 à 50 mètres).

Les stations de fond (urbaines et périurbaines) sont placées dans des zones non influencées par des sources d'émissions locales, de façon à caractériser la pollution moyenne de tout un secteur géographique.

2.1.2 Evolution moyenne des mesures

Il est impossible de comprendre le comportement des oxydes d'azote sans évoquer la pollution photochimique. L'analyse des observations du réseau AREMA apporte des informations sur le rôle des phénomènes physiques mis en jeu sur l'agglomération lilloise. Les résultats présentés par la suite ont été vérifiés qualitativement sur le modèle.

Les figures 2.2 et 2.3 présentent des moyennes journalières de concentrations horaires d'ozone, de monoxyde d'azote et de dioxyde d'azote mesurées par le réseau de mesure pour les mois de Janvier-Février et Mai-Juin. Les évolutions temporelles présentent des similitudes mais les niveaux moyens et les amplitudes des variations sont très différents

pour ces deux périodes. Les profils d'une journée moyenne sont typiques et correspondent aux flux automobiles. On aperçoit clairement un pic matinal pour le NO et un pic en soirée moins marqué, correspondant aux départs du matin et aux retours du soir (voir la figure 2.1 pour le profil moyen des émissions de monoxyde d'azote correspondant au trafic routier). Les heures indiquées sont exprimées en temps universel. Il convient de rajouter 1 heure en hiver et 2 heures en été pour obtenir une heure locale.

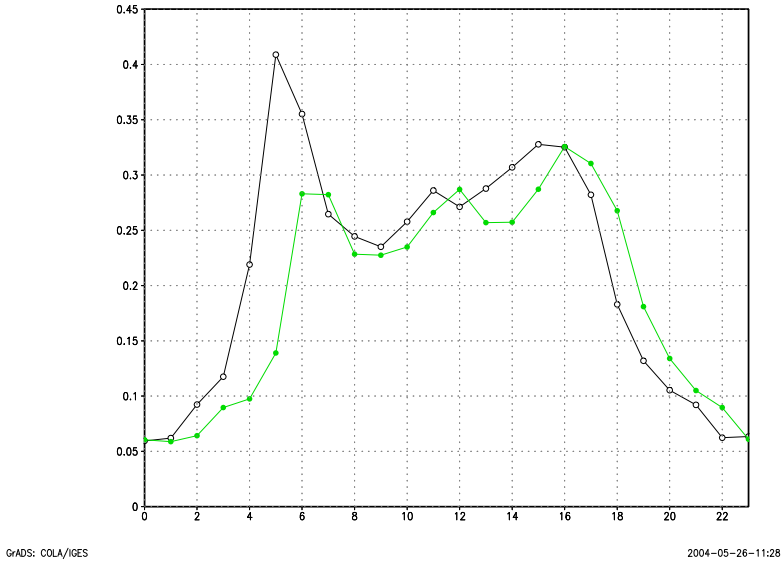

FIG. 2.1 – Evolution moyenne des émissions de NO provenant du trafic routier pour la période Janvier–Février (en vert, ronds pleins) et Mai–Juin (en noir, ronds vides).

2.1.2.1 Janvier-Février

Les oxydes d'azote ont un profil caractéristique bi-modal. Le trafic routier s'amplifie entre 4h et 6h et contribue à une amplification par un facteur 3 du niveau d'émissions d'oxydes d'azote. Les polluants émis s'accumulent dans la couche de surface stable qui s'est mise en place au cours de la nuit. Ceci entraîne une augmentation forte de NO (90 % des émissions de NO_x) à partir de 5h. Ce phénomène s'accompagne d'une destruction chimique de l'ozone par la réaction $O_3 + NO \rightarrow NO_2 + O_2$. Il en résulte une augmentation du NO_2 due principalement à la formation chimique mais aussi à l'émission directe de NO_x.

Le rayonnement solaire chauffe le sol à partir de 8h (heure de lever du soleil). Progressivement, la température du sol devient supérieure à la couche d'air au-dessus. La turbulence d'origine thermique se développe et mixe la couche de surface et la couche résiduelle pour former la couche de mélange. Le mélange vertical maximal est atteint vers 15h. Pendant cette période, les émissions de NO_x sont diluées dans un volume croissant. Aussi, les concentrations de NO et NO_2 atteignent un minimum

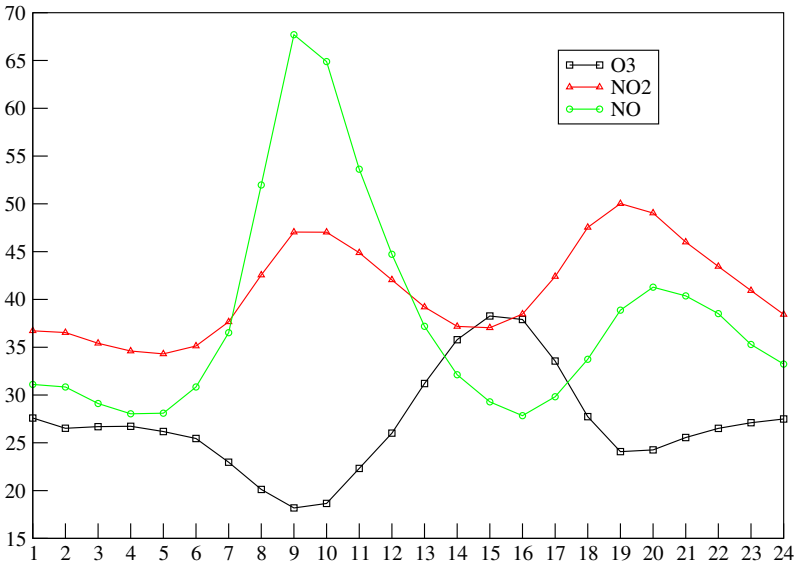

FIG. 2.2 – Evolution moyenne des polluants O_3, NO et NO_2 sur un jour durant la période Janvier–Février.

vers 15h. A l'opposé, les concentrations d'ozone au sol décrivent une évolution symétrique due à la présence abondante d'ozone dans la couche résiduelle. Parallèlement, la photolyse établit l'équilibre photo-stationnaire qui contribue à accentuer les variations. Après 15h, le brassage diminue ce qui permet aux émissions de recommencer à s'accumuler.

Après 18h, plusieurs facteurs agissent simultanément. D'une part, les émissions du trafic routier retrouvent progressivement leur niveau de nuit et d'autre part, le sol n'est plus chauffé par le rayonnement solaire (le coucher du soleil a eu lieu vers 17h) ce qui se traduit par le développement de la couche nocturne stable qui emprisonne les émissions. Progressivement, un équilibre de nuit s'établit avec l'arrivée sur le domaine de masses d'air "propres". Un pic de NO_2 est présent en soirée. A ce moment de la journée, la destruction photolytique de NO_2 s'arrête faute de rayons solaires suffisamment énergétiques et le surplus d'émissions de NO dû au retour du soir est transformé en NO_2 par l'ozone.

2.1.2.2 Mai-Juin

L'évolution horaire des trois polluants en été ressemblent aux courbes obtenues en hiver. Néanmoins, on note plusieurs différences :

- Le pic de NO_x du matin est moins marqué et celui du soir presque absent. La principale différence tient au fait que la turbulence thermique se met en place avant le pic d'émission du matin. Le lever du soleil a lieu à 4h (coucher à 20h). L'ozone atteint un maximum tard dans l'après-midi et limite le pic de NO_x en soirée.

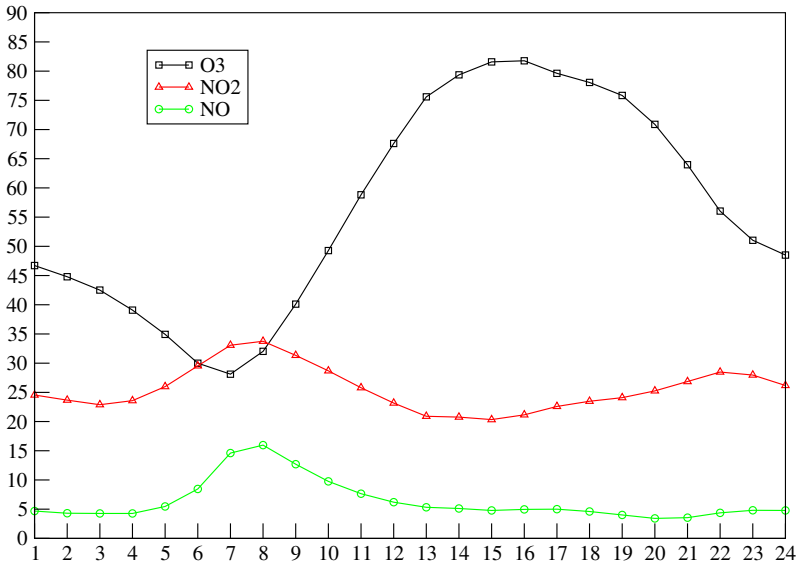

FIG. 2.3 – Mai–Juin

- Les niveaux moyens de polluants sont très différents. La photolyse bien plus importante qu'en hiver favorise la création d'ozone. Le brassage turbulent plus fort a tendance à diluer plus rapidement les émissions, ce qui conduit à des niveaux de NO plus faible.

2.2 Conditions de la simulation

2.2.1 L'échelle continentale

L'agglomération lilloise fait partie d'une zone urbaine dense avec la présence de nombreuses villes (Lille, Roubaix, Tourcoing et Villeneuve d'Ascq). Même si les pics de pollution sont observés à proximité des fortes sources d'émissions, une part importante de polluants est apportée de régions éloignées. Le site pouvant se trouver dans le panache de gros centres d'émissions (Ruhr, Londres, Paris, etc). Ceci justifie d'avoir de bonnes conditions aux limites pour cette étude régionale. Ainsi, une première étape a consisté à faire une simulation sur un domaine large afin de fournir les concentrations aux limites pour la simulation sur le domaine lillois.

Pour une utilisation simple de ces conditions aux limites, le domaine européen a été choisi de sorte qu'une maille coïncide exactement avec le domaine lillois. La figure 2.4 représente les émissions annuelles de dioxyde d'azote pour l'année 1998 sur le domaine européen.

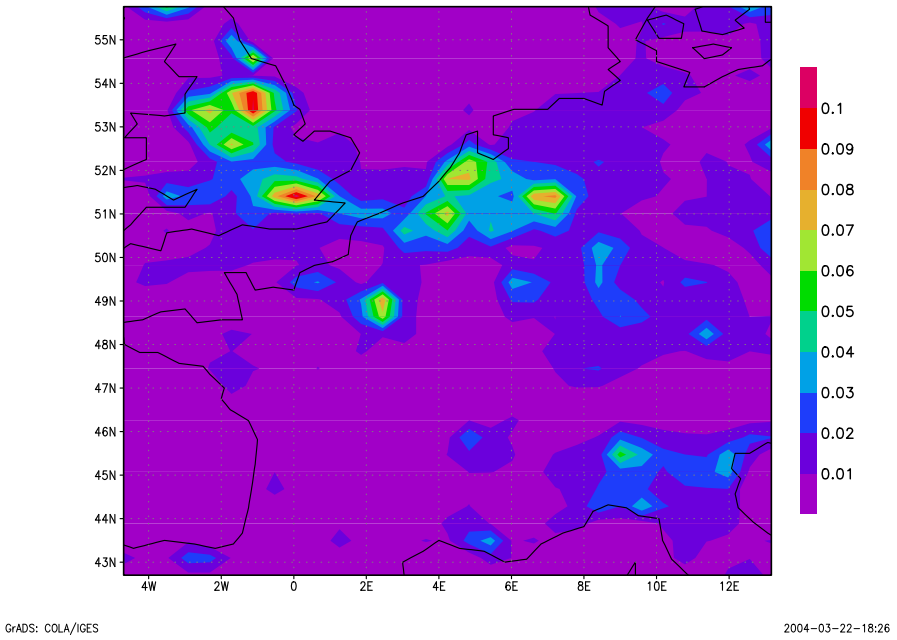

FIG. 2.4 – Moyenne annuelle des flux d'émissions de dioxyde d'azote. (microgrammes par mètre carré par seconde).

Neuf niveaux verticaux ont été définis entre le sol et trois mille mètres, ce qui est suffisant pour décrire les processus qui ont lieu dans la couche limite atmosphérique. Le premier niveau est défini au-dessus de la sous-couche rugueuse à une hauteur correspondant à la somme des hauteurs de déplacement (hauteur caractéristique de bâtiments par exemple et de rugosité (typiquement quelques mètres). La première maille a une hauteur de trente mètres, qui correspond à une hauteur typique de couche de surface. Le pas de temps temporel est de 10 minutes. Les champs de forçage en entrée de la chaîne de modélisation sont interpolés linéairement pour être utilisables à cette résolution. A chaque pas de temps, les concentrations de toutes les espèces sont sauvegardées pour servir de conditions aux limites pour la simulation à l'échelle régionale (one-way nesting).

La simulation d'une durée d'un an inclut une période supplémentaire de six jours, permettant aux modèle de se mettre à l'équilibre et de ne plus dépendre des conditions initiales (notion classique de spin-up).

Le mécanisme chimique utilisé est RACM. Il décrit les réactions chimiques de 72 espèces gazeuses par l'intermédiaire de 237 réactions chimiques. On ne tient pas compte des réactions chimiques en phase aqueuse.

Les conditions aux limites de la simulation européenne sont fournies par le modèle

à l'échelle globale MOZART2 ([Hauglustaine *et al.*, 1998]). Elles correspondent à des valeurs journalières typiques pour certaines espèces chimiques.

Les champs météorologiques proviennent des réanalyses du CEPMMT (Centre Européen de Prévision à Moyen Terme) qui réalise des prévisions météorologiques à l'échelle globale. Ces données ont une résolution temporelle de trois heures. Le modèle du CEPMMT possède une résolution spectrale T511 en ré-analyse, ce qui correspond à une résolution horizontale d'environ 1.125 degrés (de l'ordre de 150 kilomètres) et 60 niveaux verticaux dont environ 16 pour la couche limite atmosphérique (en-dessous de 3000 mètres).

EMEP dispose d'émissions annuelles à l'échelle européenne à une résolution de 50 kilomètres. Ces émissions ont été spatialisées et temporalisées par des profils horaires typiques provenant du projet Européen GENEMIS comme décrit dans le chapitre 1.

Les vitesses de dépôt sec sont calculées pour 14 espèces à partir des données d'occupation de sol de USGS (`http://lpdaac.usgs.gov/glcc/euras_int.asp`) ayant une résolution au kilomètre carré Lambert.

2.2.2 L'échelle régionale

2.2.2.1 Domaine d'étude

Le domaine de simulation locale englobe l'agglomération de Lille dans son entier et s'inscrit dans un rectangle de 42 par 44 km, avec une résolution de 1 km. La résolution verticale est la même que la simulation européenne. La taille du maillage est la plus fine possible compte tenu de la disponibilité des données (occupation des sols, etc). Elle permet de bien rendre compte de la finesse de la modélisation trafic/émission et d'estimer correctement les quantités de polluants. Le domaine ne permet pas d'observer complètement la pollution secondaire (ozone) induite par Lille qui a souvent lieu à quelques dizaines de kilomètres des agglomérations. Par contre il est bien adapté pour la pollution primaire qui a lieu à proximité des sources d'émissions.

2.2.2.2 Données météorologiques

Nous avons fait l'hypothèse que les données à grande échelle de météorologie étaient suffisamment représentatives. Aussi, les champs météorologiques ont été interpolés sur la grille fine. L'application du modèle CHIMERE à l'Ile-de-France (ESQUIF) avec des champs météorologiques de grande échelle a montré que cette hypothèse était

pertinente ([Vautard *et al.*, 2001]). Cette simplification par rapport à l'utilisation d'un modèle météorologique à méso-échelle n'est valide que dans le cas où la dynamique grande échelle n'est pas trop perturbée par des effets locaux (topographie, proximité côtière, etc ...) ce que l'on peut supposer pour Lille qui est située dans une région au relief plat. Néanmoins, cette hypothèse présente certaines limites pour le milieu urbain :

- La rugosité dynamique y est plus importante en raison de la présence de bâtiments ce qui entraîne un ralentissement du vent dans la couche de surface.
- La diffusion turbulente y est quasiment toujours instable (surtout en hiver et la nuit) du fait du phénomène d'îlot de chaleur urbain : la ville a des capacités plus grandes pour emmagasiner la chaleur, ce à quoi peut s'ajouter l'effet du chauffage.

La résolution tri-horaire des données météorologiques ne permet pas de décrire fidèlement l'évolution journalière de la couche de limite atmosphérique. On peut s'interroger sur la validité d'une description des processus météorologiques toutes les trois heures avec une résolution spatiale d'environ un degré comme c'est le cas pour les données du CEPMMT disponibles. Il serait intéressant d'avoir une comparaison avec une météo à l'échelle régionale pour quantifier les erreurs introduites par cette approximation, mais ce point dépasse le cadre de notre travail.

La carte 2.5 représente le vent moyen au sol sur l'Europe pour l'année 1998. On observe que Lille est balayée en général par des vents canalisés par la Manche.

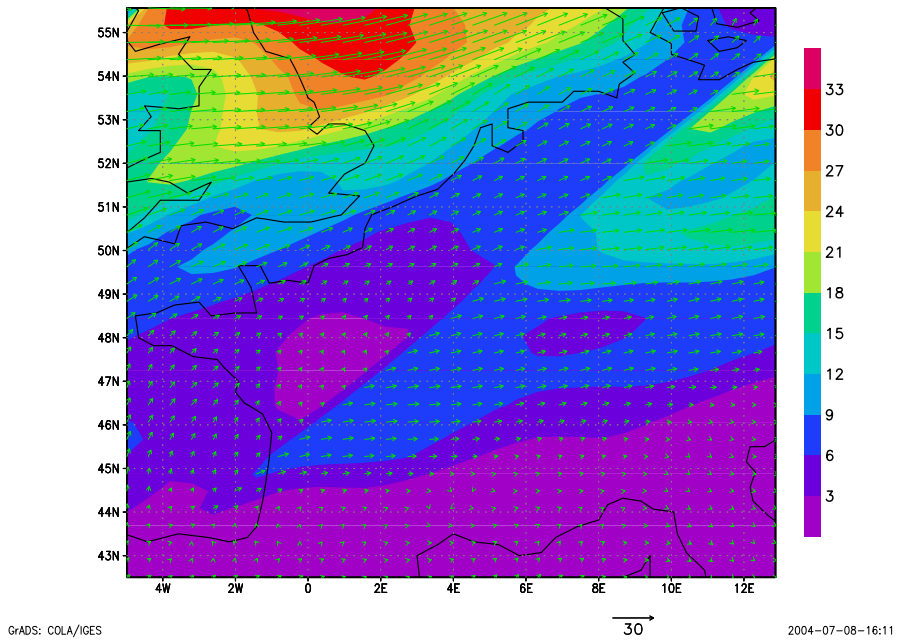

FIG. 2.5 – Direction et module moyens du vent au sol pour l'année 1998.

2.2.2.3 Données d'émissions

Les données d'émissions ont plusieurs origines :

- Les données d'émissions EMEP sont le résultat d'une approche "top-down" qui redistribue spatialement et temporellement des quantités globales comme par exemple la consommation d'essence ou de produits d'entretien. Elles sont fournies pour l'Europe en totaux annuels par classes SNAP (Selected Nomenclature for Air Pollution) sur des mailles de 50 km^2.

- Les émissions provenant du trafic routier spatialisées sur les mailles contenues dans l'agglomération administrative lilloise sont fournies par heure pour une journée type. Elles sont accompagnées de coefficients correctifs horaires issus du comptage routier pour une répartition sur l'année.

- Les émissions ponctuelles des principaux gros émetteurs appartenant aux secteurs d'activité du chauffage urbain, de l'industrie et de l'incinération de déchets sont inventoriées et leurs hauteurs de rejet sont disponibles. Ces données proviennent de la DRIRE Nord - Pas-de-Calais (Direction Régionale de l'Industrie, de la Recherche et de l'Environnement).

Le domaine lillois recoupe le maillage EMEP en quatre mailles. Les données d'émissions les plus fines sont utilisées lorsqu'elles sont disponibles et elles sont complétées le cas échéant par les données EMEP.

– Pour générer les données d'émissions surfaciques, une première étape consiste à soustraire des données EMEP la contribution des grandes sources ponctuelles. Une mise à zéro est effectuée dans les cas où cette différence s'avère négative. Cet inventaire est ensuite distribué sur les mailles du domaine lillois en utilisant les caractéristiques d'occupation des sols. Par exemple, un poids 12 fois plus important est attribué aux zones urbaines. La distribution verticale est indiquée dans le tableau 2.1. La répartition horaire se fait en appliquant un coefficient mensuel, hebdomadaire et horaire. Ces facteurs sont fournis par le projet Européen GENEMIS.

– Lorsque les données fines du trafic routier sont disponibles, elles remplacent les émissions précédemment obtenues.

La carte 2.6 représente une moyenne annuelle des émissions de dioxyde d'azote. Les principaux émetteurs sont localisés au niveau des axes routiers et du centre urbain.

Le tableau 2.2 regroupe les émissions annuelles d'oxydes d'azote rapportées au domaine lillois. Les émissions issues des données EMEP uniquement s'accordent bien

Secteur d'activité (SNAP)	sol	15 m	90 m	250 m	490 m	802.5 m	1167.5 m
1. Energie	0	0	0	14	52	27	7
2. Combustion hors industrie	16.7	10.8	54.	18.5	0	0	0
3. Combustion industrielle	0	0	2.5	25.9	48.1	21	2.5
4. Procédé de production	30	19.6	46.7	3.7	0	0	0
5. Extraction de combustibles fossiles	30	19.6	46.7	3.7	0	0	0
6. Solvants	100	0	0	0	0	0	0
7. Transport routier	100	0	0	0	0	0	0
8. Autres sources mobiles	100	0	0	0	0	0	0
9. Traitement des déchets	3.3	2.2	14	50.1	30.4	0	0
10. Agriculture	100	0	0	0	0	0	0

TAB. 2.1 – Distribution verticale des émissions (en %).

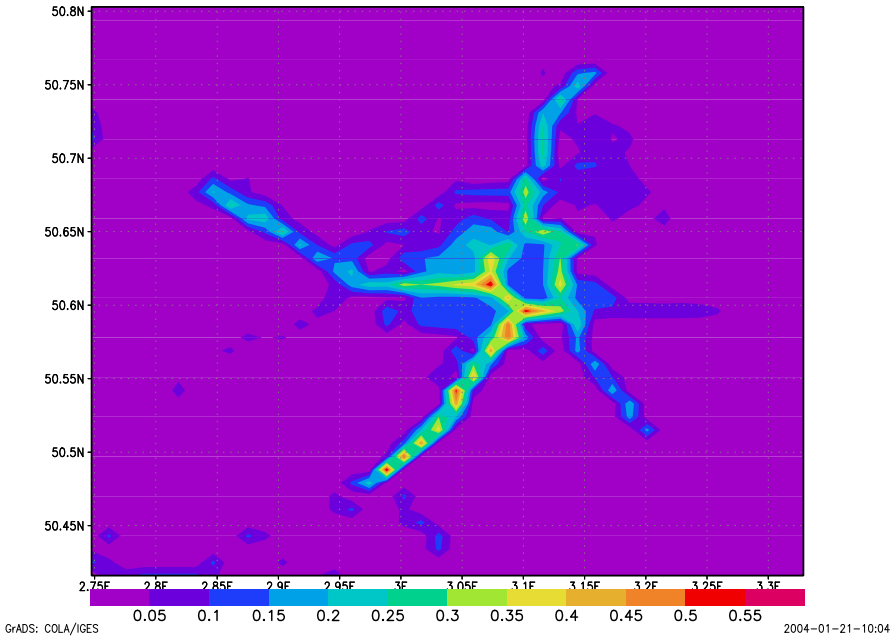

FIG. 2.6 – Moyenne annuelle des émissions de dioxyde d'azote pour l'année 1998.

(moins de 6 % d'écart) avec l'inventaire qui a été construit selon la méthodologie décrite précédemment. Le secteur du trafic routier contribue très largement aux émissions totales d'oxydes d'azote (60 %) et constitue donc l'information la plus importante.

EMEP spatialisé	inventaire complet	trafic routier	GSP
26130	27625	16472 (14288)	995

TAB. 2.2 – Emissions annuelles d'oxydes d'azote rapportées au domaine lillois (en tonnes d'équivalent NO_2 par an).

La figure 2.7 représente le profil horaire moyens d'émissions surfaciques de monoxyde d'azote pour la période hivernale de Janvier-Février. Les émissions de NO contribuent le jour à 36 $\mu gm^{-3}h^{-1}$ contre 12 la nuit.

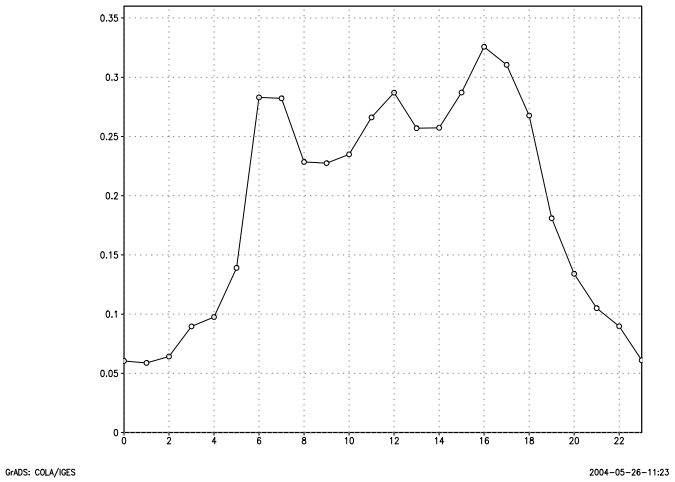

FIG. 2.7 – Profils horaires moyens des émissions de monoxyde d'azote. (microgrammes par mètre carré par seconde).

2.2.3 Temps de calcul

Un jour de simulation européenne demande 4 minutes de temps CPU sur un ordinateur PC Pentium 4 à 2.8 GHz, soit environ 30 heures pour un an simulé.

La simulation locale comporte un sous-cyclage de l'advection afin de respecter la condition sur le nombre de CFL, schématiquement, $u.\delta t/\delta x < 1$ où δt est le pas de temps d'intégration, u une vitesse de vent et δx une taille caractéristique de maille. Pour la simulation locale la taille des mailles de 1 km^2 impose de choisir un pas de temps d'intégration de l'ordre de 15 secondes, valide pour toutes les situations de

vent de l'année 1998. Le temps CPU pour effectuer un jour de calcul est alors de 40 minutes, ce qui correspond à une durée totale supérieure à 10 jours pour une simulation d'un an. Pour réduire ce temps de calcul, chaque mois de simulation est résolu séparément sur une machine. Une période de six jours est ajoutée en début du calcul de chaque mois permettant aux conditions initiales d'être totalement évacuées du domaine.

2.2.4 La simulation 2015

Il est prévu que les émissions continuent de baisser fortement d'ici l'année 2015, ce qui va s'accompagner d'une modification de la composition des masses d'air en Europe et en particulier celles qui vont être advectées jusqu'au domaine lillois. Une première étape a consisté à simuler la qualité de l'air à l'échelle européenne par le biais d'une projection d'un inventaire d'émissions à horizon 2015. L'exercice de modélisation européenne CITY-DELTA (une inter-comparaison des réponses à long terme de la part des modèles à des scénarios de réduction des émissions urbaines) a construit des scénarios d'émissions pour l'année 2010 à l'échelle européenne. En particulier, nous avons utilisé l'inventaire basé sur la législation actuelle de réduction des émissions. Pour établir l'inventaire de 2015 à partir de celui de 2010, un coefficient de réduction a été appliqué pour le trafic routier à partir du rapport [Ntziachristos *et al.*, 2002], qui estime des valeurs d'émissions du trafic routier par pays pour 2010 et 2020 en se basant sur des données de l'année 2000.

- 0.8 pour les NOX.
- 0.85 pour les composés organo-volatiles.

Ces coefficients sont le résultat d'une interpolation des réductions prévues entre 2010 et 2020.

Pour la simulation locale de 2015, Les émissions sont modifiées à partir des projections d'inventaires du trafic routier avec et sans plan de déplacement urbain. Les grandes sources ponctuelles sont conservées à l'identique. La même méthode que celle décrite pour l'année 1998 est appliquée pour obtenir des émissions correspondant à tous les secteurs d'activité.

2.3 Résultats pour l'année de référence 1998 et pour l'exercice 2015

2.3.1 Année 1998

2.3.1.1 Qualité de la simulation européenne

La composition des masses d'air qui traversent le domaine lillois doit être évaluée de façon précise pour bien décrire la pollution photo-chimique locale, essentiellement pour l'ozone qui est un polluant à échelle continentale. La figure 2.8 représente un diagramme de comparaison entre les niveaux de concentrations au sol à l'intérieur du domaine et sous le vent. La plupart du temps, le NO_2 est créé sur Lille et l'ozone est détruit en raison des fortes émissions de NO sur l'agglomération (phénomène de titration).

Les concentrations simulées ont été comparées statistiquement aux observations disponibles pour évaluer la qualité de la simulation européenne, tout au moins la validité des niveaux de polluants de la première maille. Les concentrations simulées par la simulation européenne ont été interpolées linéairement aux endroits où se trouvent les stations. On suppose que les observations sont une représentation satisfaisante de la réalité, notamment qu'elles ne contiennent pas une erreur systématique instrumentale ou de représentativité. Les mesures statistiques présentées dans le tableau 2.3 sont classiquement utilisées pour évaluer la qualité de la simulation :

- Le *biais* s'obtient en sommant pour toutes les mesures la différence entre les sorties de modèle et les observations. Un biais nul indique qu'en moyenne le modèle ne commet pas d'erreur.
- La *RMS* (Root Mean Square) se calcule en sommant cette même différence préalablement élevée au carré. Elle permet d'estimer la dispersion des erreurs.
- Le coefficient de *corrélation* est un nombre compris entre -1 et 1. S'il est proche de 1, la variabilité du phénomène est bien représentée.

Espèces chimiques	nombre de mesures	moyenne des mesures	concentrations simulées	biais	RMS	corrélation
O_3	105548	38.6	46.9	7.3	22.5	75%
NO_2	152368	29.8	24.9	-4.9	17.3	54%
NO	153962	14.5	5.4	-9.1	28.6	59%

TAB. 2.3 – Statistiques des comparaisons entre les concentrations simulées à l'échelle européenne et les mesures. Elles sont calculées sur des valeurs horaires. L'unité est le μgm^{-3}.

La comparaison s'effectue uniquement sur les stations de fond en comparaison horaire. On rappelle que les points de mesures sont localisés en majorité à proximité du centre urbain ce qui n'est pas très pertinent vu l'hétérogénéité des émissions dans ces zones.

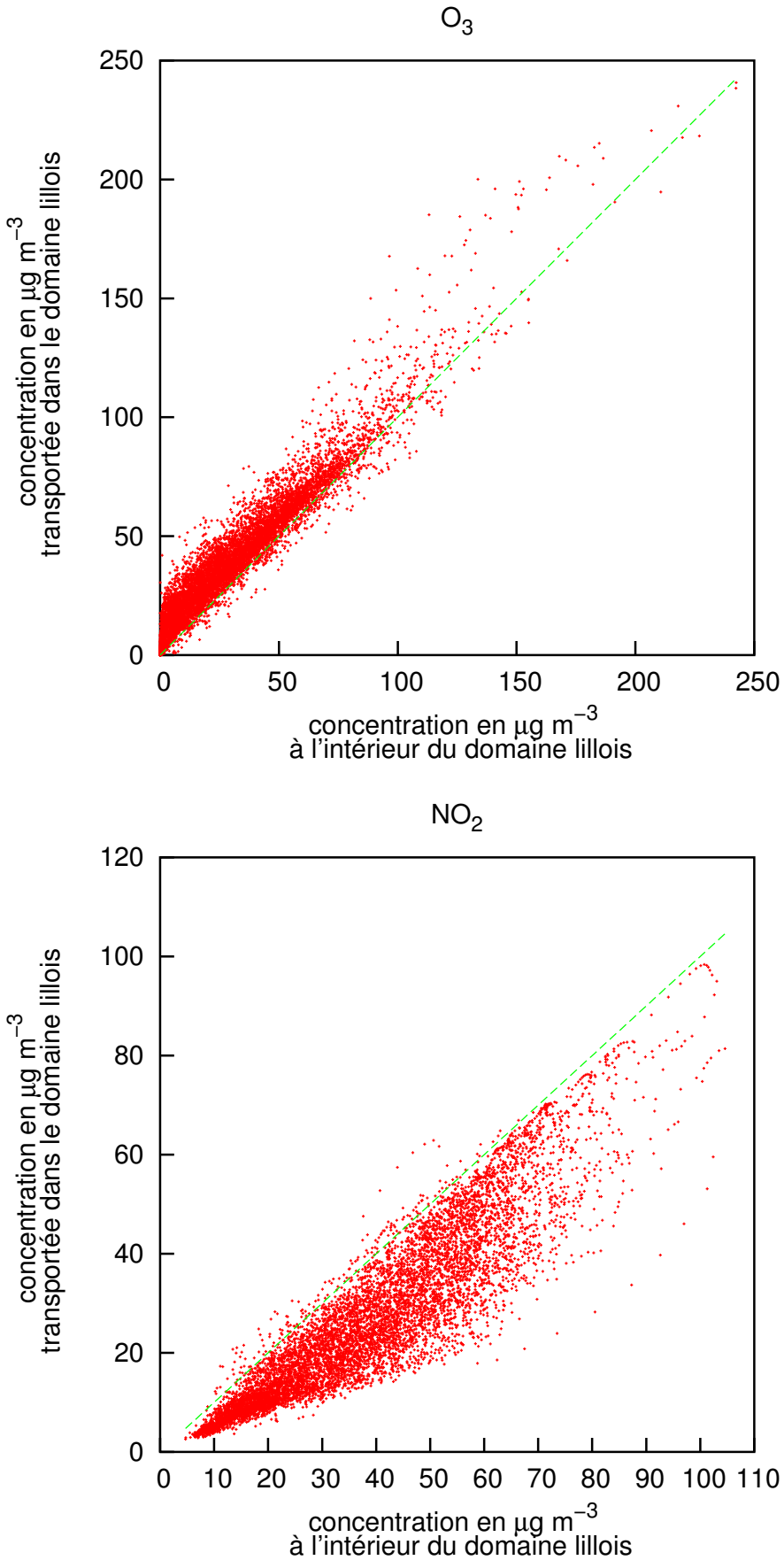

FIG. 2.8 – Diagramme de comparaison des niveaux de polluants à l'intérieur du domaine et ceux apportés de l'extérieur. (μgm^{-3}).

Néanmoins, l'objectif est de vérifier qu'il n'y a pas d'anomalies.

Globalement, le modèle sous-estime les oxydes d'azote et il sur-estime l'ozone. Les corrélations élevées montrent que les variations sont bien représentées. Les écarts de niveaux observés entre la simulation européenne et les mesures ne sont pas surprenants et ont plusieurs origines, notamment liées à la résolution de l'inventaire d'émissions. Lille se distingue par une variabilité importante des caractéristiques d'émissions à l'échelle de quelques kilomètres (cas typiquement de certaines stations urbaines et péri-urbaines) qui n'est pas prise en compte par la résolution du modèle (environ 40 kilomètre). Dans ce type de contexte, les sources d'émissions ont tendance à être diluées dans les grandes mailles du domaine et les concentrations modélisées de polluants primaires sont généralement sous-estimées par rapport aux valeurs des stations de fond implantées préférentiellement dans les zones denses en émissions même si elles ne sont pas influencées directement par des sources locales d'émissions. Par effet de titration, les concentrations simulées en ozone dans ces mailles sont un peu sur-estimées par rapport aux valeurs mesurées.

On peut donc considérer que les valeurs de fond d'ozone et d'oxydes d'azote sont satisfaisantes pour forcer la simulation locale.

2.3.1.2 Evaluation de la simulation lilloise

Les concentrations simulées de dioxyde d'azote ont été comparées aux mesures du réseau AREMA.

Pour donner une idée des performances du modèle, les évolutions annuelles des moyennes journalières des concentrations de dioxyde d'azote ont été tracées sur la figure 2.9 pour les trois types de stations.

Ces séries temporelles suggèrent que POLAIR3D est à même de reproduire la variabilité journalière du dioxyde d'azote. Cependant, le modèle montre des difficultés à reproduire les journées les plus polluées, notamment en hiver. Ces périodes sont marquées par des températures basses (inférieures à 0° C) qui peuvent être la cause d'une mauvaise description d'un processus physique (dépôt sec, mélange vertical, ...) ou d'une augmentation des émissions dues au chauffage.

Une carte des concentrations annuelles de dioxyde d'azote est représentée sur la figure 2.10. Les concentrations simulées sont homogènes avec des maxima à proximité des routes et du centre urbain alors que les mesures présentent une répartition très

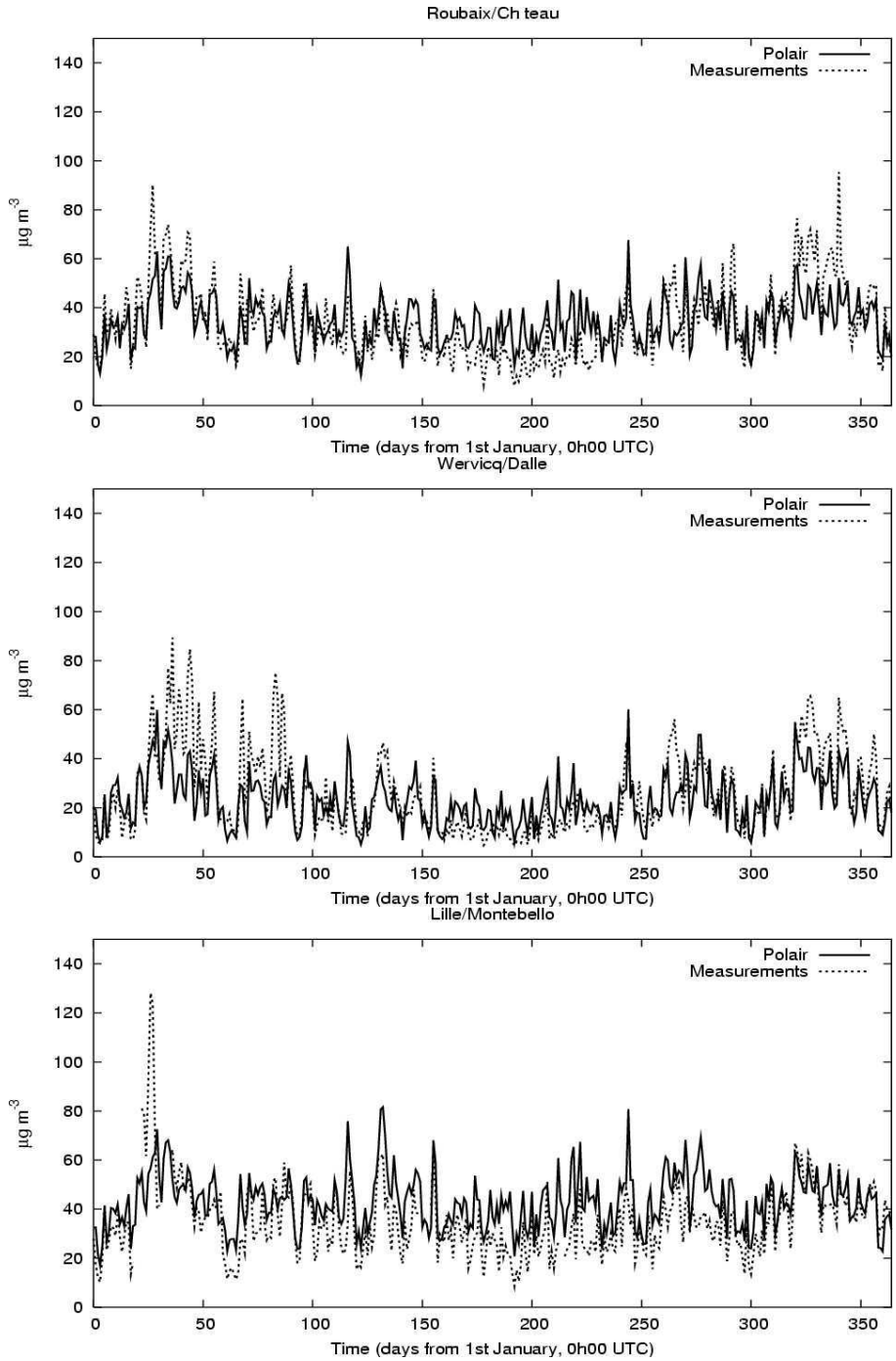

FIG. 2.9 – Série des moyennes journalières des concentrations de dioxyde d'azote observées (courbe pointillée) et simulées par POLAIR3D (courbe pleine) pour l'année 1998 sur trois stations de mesure : Roubaix/Château (urbaine), Werwicq (périurbaine) et Lille/Montebello (proximité automobile).

irrégulière notamment dans le centre urbain. On atteint ici vraisemblablement les limites de description de la variabilité spatiale d'un modèle chimie-transport sans description explicite de l'îlot de chaleur urbain.

Des valeurs statistiques (biais, RMS, corrélation) sont reportées dans le tableau 2.3.1.2, station par station.

station	nombre de mesures	moyenne des mesures	concentrations simulées	biais	RMS	corrélation
urbaine dense						
Armentières	8731	27.9	22.9	-5.0	13.5	67%
Haubourdin	8664	27.2	30.3	3.1	14.9	61%
Marcq-en-Baroeul	6870	35.9	31.1	-4.8	17.8	58%
Lille/Fives	8665	37.9	36.2	-1.7	16.2	65%
Roubaix/Château	8710	34.2	34.7	0.5	15.4	57%
Tourcoing	8597	34.2	32.0	-2.2	19.6	54%
périurbaine						
Halluin	8696	23.2	26.3	3.1	15.6	57%
Salomé	8562	23.1	21.5	-1.6	11.9	65%
Wervicq	8689	25.9	22.8	-3.1	15.3	65%
proximité						
La Bassée	4547	25.1	18.0	-7.1	18.6	48%
Lille/Faidherbe	8646	35.6	36.3	0.7	18.1	59%
Lille/Liberté	8230	41.7	36.8	-4.9	20.1	41%
Lille/Pasteur	8687	45.3	46.5	1.2	20.6	47%
Villeneuve d'Ascq	8216	33.6	45.6	12.0	21.9	52%
Lille/Montebello	8209	34.8	45.1	10.3	20.0	57%
moyenne						
	245438	32.5	32.8	0.3	17.4	59%

TAB. 2.4 – Statistiques des comparaisons entre les concentrations de NO_2 simulées à l'échelle locale et les mesures. Elles sont calculées sur des valeurs horaires, station par station. L'unité est le μgm^{-3}.

On observe qu'il existe des différences significatives entre les niveaux de dioxyde d'azote mesurés par les stations, qui ne sont pas dues uniquement à la typologie des stations. Les corrélations se situent au-dessus de 50 % ce qui montre que le modèle reproduit assez bien la variabilité temporelle. Seule la station Lille/Liberté a une corrélation plus faible.

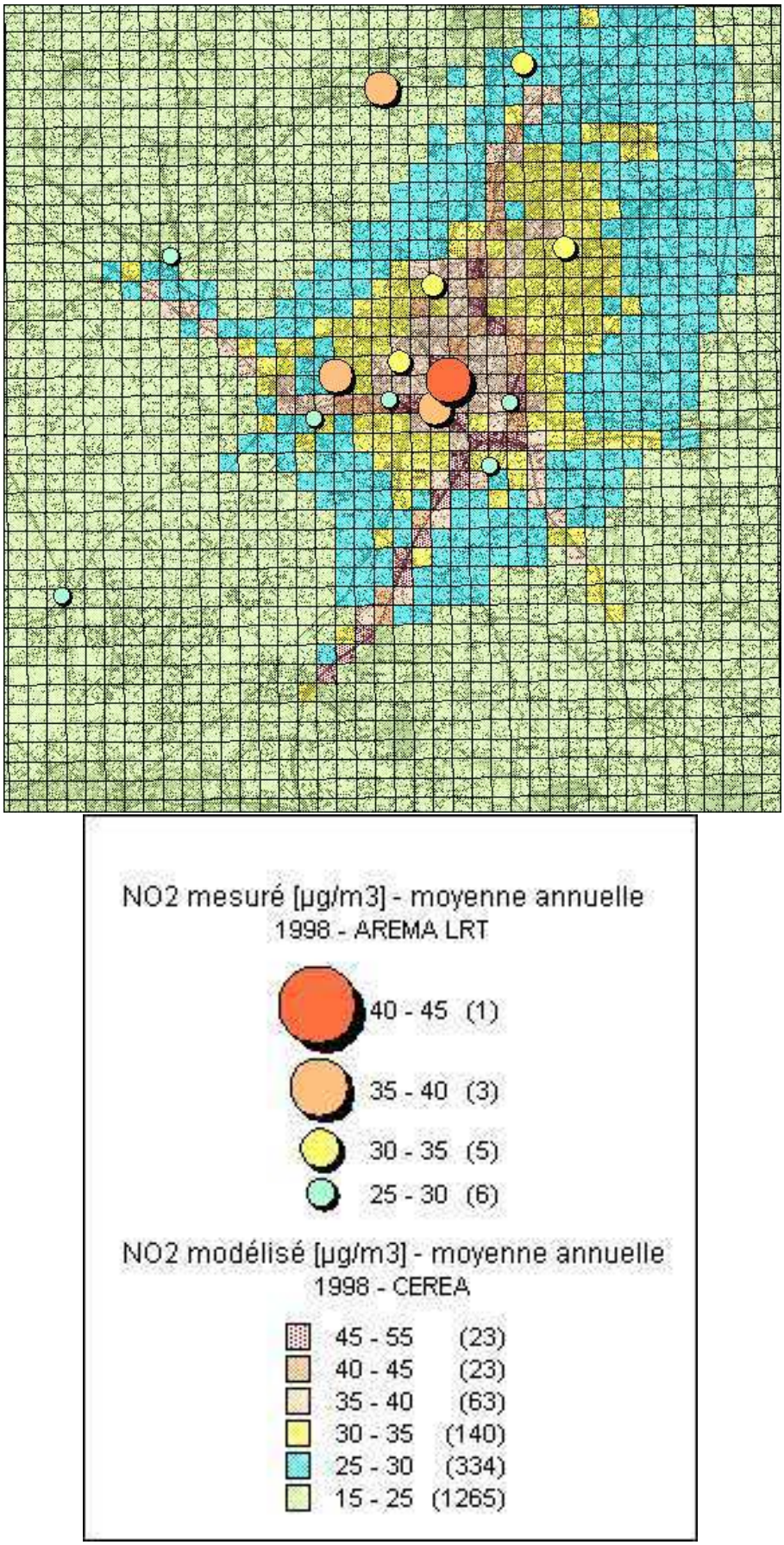

FIG. 2.10 – Moyenne annuelle du dioxyde d'azote.

2.3.2 Résultats pour 2015

La comparaison des résultats obtenus pour 1998, 2015 sans plan de déplacement urbain et 2015 avec plan de déplacement urbain permet d'évaluer l'impact :

- de l'évolution des émissions du trafic due à la croissance du parc automobile compensée par les réglementations de réduction d'émissions.
- des modifications de trafic engendrées par le plan de déplacement urbain sur la qualité de l'air à l'échelle régionale.

2.3.2.1 Comparaison des inventaires d'émissions

On observe des différences mineures entre les deux inventaires pour les NO_x. La comparaison se base sur des cartes de différence absolue (2.11) ou relative (2.12) de totaux annuels et sur une évolution journalière moyenne (2.13). Globalement, l'inventaire avec PDU apporte une réduction d'émissions de NO_x.

Les différences principales sont localisées au niveau de l'axe Nord-Sud avec des écarts maximum de 20 %. Le PDU permet de réduire faiblement les émissions de NO_x en journée notamment en fin d'après midi.

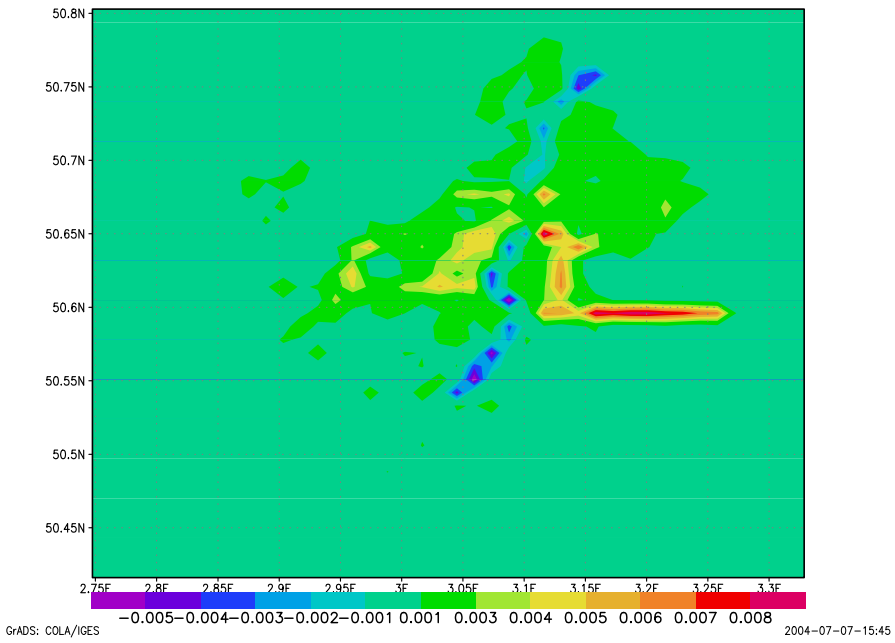

FIG. 2.11 – Différence des émissions surfaciques de dioxyde d'azote entre les inventaires 2015 avec et sans PDU (microgrammes par mètre carré par seconde).

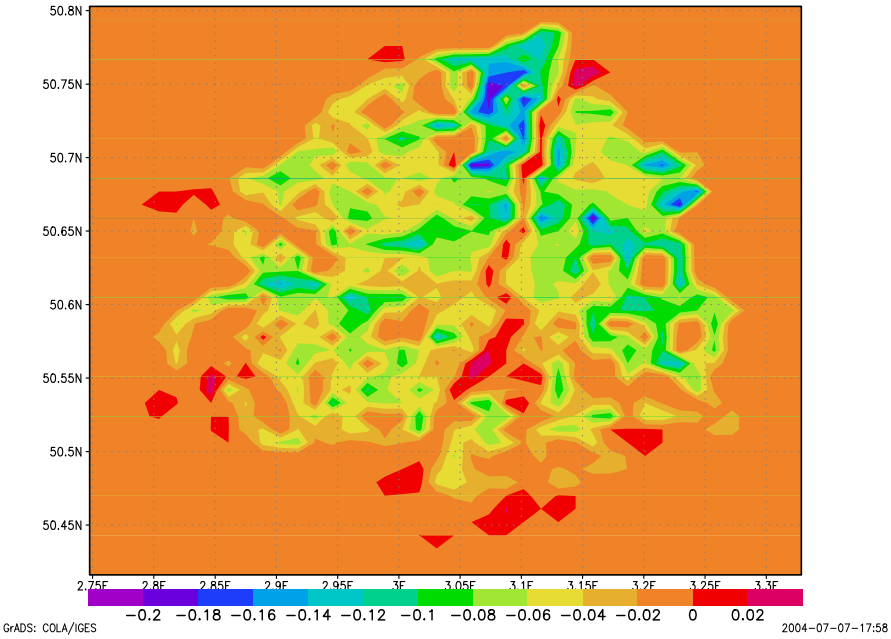

FIG. 2.12 – Différence relative des émissions surfaciques de dioxyde d'azote entre les inventaires 2015 avec et sans PDU.

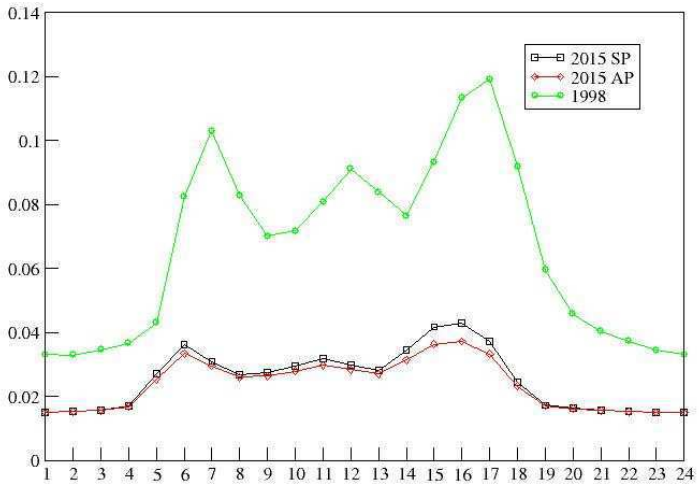

FIG. 2.13 – Evolution horaire moyenne des émissions surfaciques de dioxyde d'azote pour les inventaires 1998, 2015 avec PDU, 2015 et sans PDU (microgrammes par mètre carré par seconde).

2.3.2.2 Cartes de concentrations

De manière cohérente avec la faible différence des émissions entre les deux scénarii, le même type de comparaison sur les concentrations de NO_2 indique des écarts faibles : 5 % au maximum sur les totaux annuels par maille (figures 2.14 et 2.15).

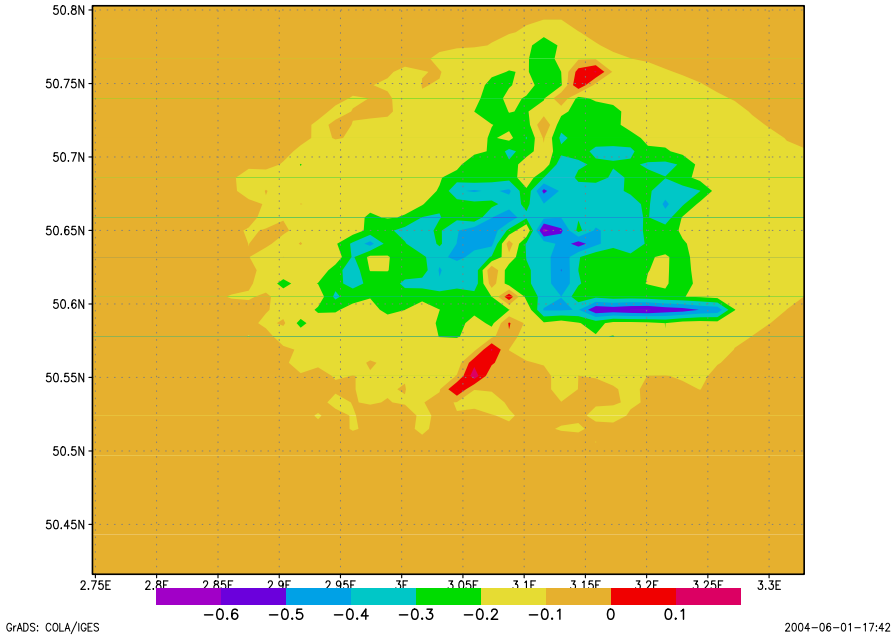

FIG. 2.14 – Différence des concentrations de dioxyde d'azote simulées entre les inventaires 2015 avec et sans PDU (microgrammes par mètre cube par seconde).

2.4 Conclusion

Ce travail a permis de modéliser la qualité de l'air sur la région de Lille sur de longues périodes, comme requis par des études d'impact (échelle annuelle).

La simulation de référence sur l'année 1998 a permis de montrer, par comparaison aux mesures disponibles, la capacité de la chaîne de modélisation (POLAIR3D) à décrire la pollution photochimique. La simulation à l'échelle continentale donne déjà de relatifs bons résultats mais l'utilisation d'un inventaire d'émissions spatialisé plus fin (notamment pour le trafic routier) au sein d'une simulation à l'échelle régionale reste nécessaire pour décrire localement la pollution. La difficulté à améliorer de façon significative les résultats par rapport à la la simulation continentale est probablement liée au défaut de description de la ville.

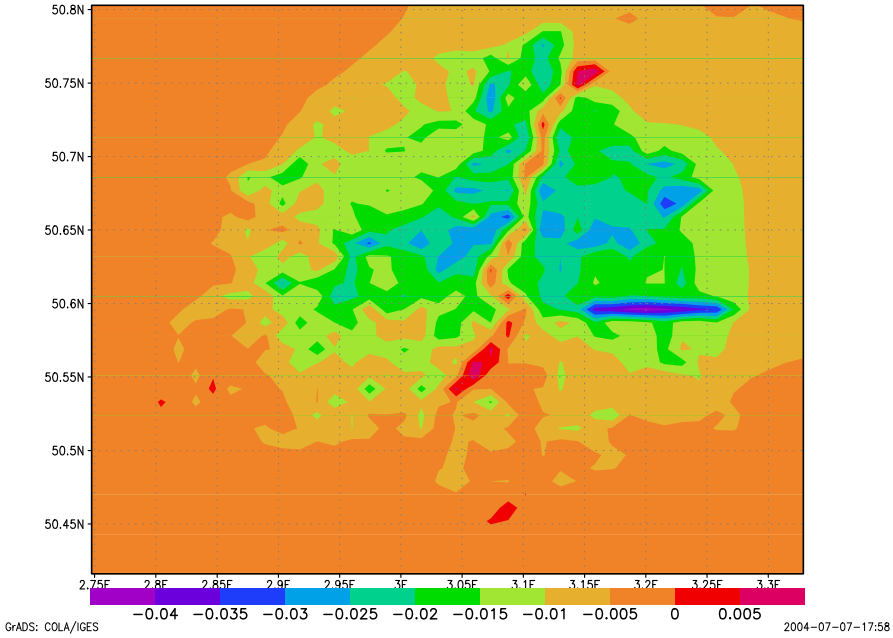

FIG. 2.15 – Différence relative des concentrations de dioxyde d'azote simulées entre les inventaires 2015 avec et sans PDU.

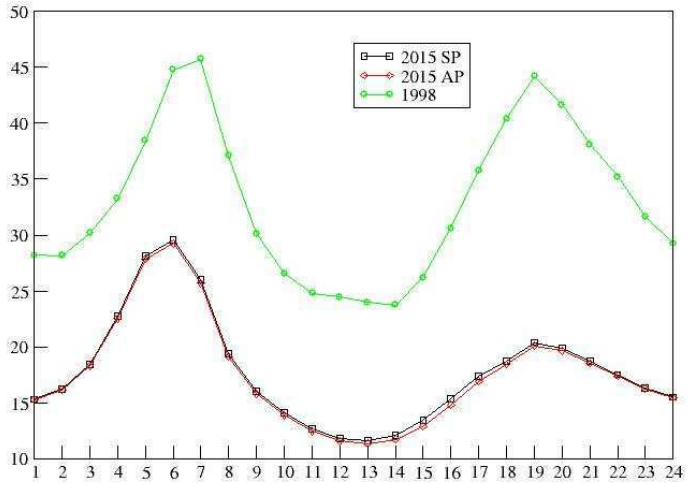

FIG. 2.16 – Evolution horaire moyenne des concentrations de dioxyde d'azote pour les inventaires 1998, 2015 avec PDU, 2015 et sans PDU (microgrammes par mètre cube).

L'ensemble de la chaîne a été par ailleurs utilisé pour quantifier l'impact de deux scénarii d'émissions liés à la mise en place ou non d'un PDU à l'échelle 2015. Les différences entre les deux inventaires étant relativement modérées, les résultats sur les concentrations sont logiquement proches.

Chapitre 3

Assimilation de données

Résumé

L'objectif de ce chapitre, à caractère largement méthodologique, est de faire un point sur la problématique générale de l'assimilation de données et sur ses applications récentes à la modélisation de la chimie atmosphérique. On détaille notamment les principaux angles d'approche (théorie de l'estimation et contrôle optimal) et les méthodes classiquement utilisées (approches séquentielles et variationnelles). Un accent particulier est mis sur le conditionnement des systèmes d'assimilation de données, que ce soit en fonction de la dynamique des systèmes physiques étudiés, des caractéristiques des réseaux de mesure ou des sensibilités dites du second-ordre.

Sommaire

3.1 Quelques notations et définitions 73
3.1.1 Modèle direct 73
3.1.2 Modèle linéaire tangent 74
3.1.3 Observations 74
3.1.4 Fonctionnelle de coût 76
3.1.5 Cadre stochastique 77
3.1.6 Un rapide historique 79
3.2 Modèles stationnaires 81
3.2.1 Approche bayésienne 81
3.2.2 Interpolation optimale 82
3.2.3 Moindres carrés et régularisation de Tikhonov 83
3.2.4 Approche variationnelle : 3D-var 84
3.3 Modèles d'évolution 84
3.3.1 Méthodes séquentielles 84
3.3.2 Méthode 4D-variationnelle 88

3.4 Conditionnement et qualité du résultat 95
3.4.1 Conditionnement et dynamique du modèle : systèmes stables versus systèmes chaotiques 97
3.4.2 Network design 100
3.4.3 Sensibilité de second-ordre 100

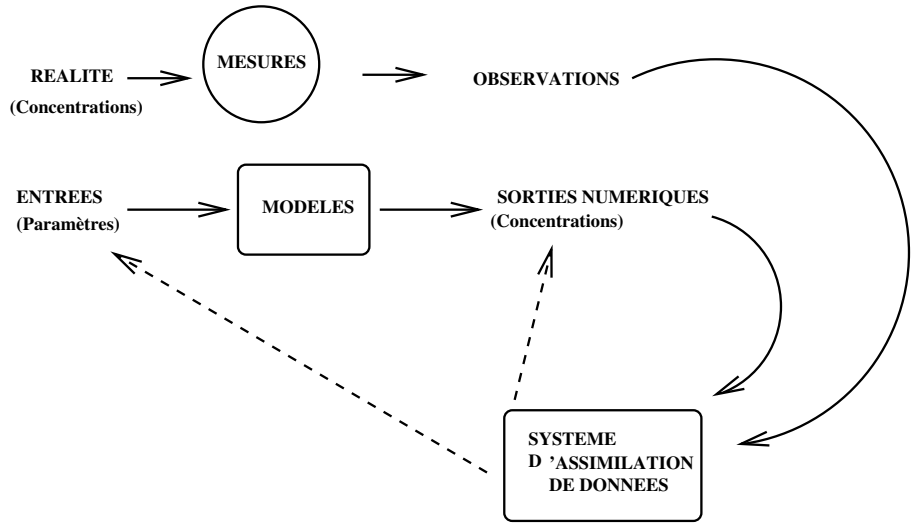

FIG. 3.1 – Système d'assimilation de données.

Introduction

Au niveau de l'état de l'art scientifique, on peut considérer à présent que nous disposons de modèles physiques détaillés décrivant les processus pilotant la pollution atmosphérique et de leur implémentation numérique sous forme de codes appropriés (les modèles de Chimie-Transport). A-t-on pour autant un outil utilisable pour des applications "réelles" ?

Tout dépend bien sûr des applications visées :
- si l'objectif est de mener des études de compréhension des processus (quelle est l'influence de tel processus au regard de celle de tel autre ?), la réponse est oui.
- si l'objectif est de mener à bien des études d'impact, reposant sur un grand nombre de simulations numériques, la réponse reste positive.
- si par contre, l'objectif est de faire de la *prévision* environnementale, le recours à des méthodes d'assimilation de données sera indispensable en prenant en compte, couplés aux sorties de modèles numériques, des résultats de mesures, provenant d'un réseau d'observations.

La raison est bien entendu la grande incertitude qui peut caractériser les données d'entrée du modèle : conditions initiales, conditions aux limites, paramètres de forçage. Il convient néanmoins de distinguer les situations selon les cas d'applications : pour simplifier, on dira que la situation pour la prévision météorologique relève d'un problème aux conditions initiales (très forte sensibilité aux conditions initiales) alors que la prévision de la pollution atmosphérique relève plutôt d'un problème aux conditions aux limites (très forte sensibilité aux flux au sol, dépôt et émissions).

De manière schématique, un problème météorologique pourra être vu comme un système dynamique sous la forme

$$\frac{dc}{dt} = \lambda\, c \ , \quad c(0) = c_0 \ , \quad \lambda > 0 \tag{3.1}$$

dont la solution $c(t) = c_0\, e^{\lambda t}$ est très sensible aux conditions initiales.

Dans le même style, un modèle de pollution atmosphérique sera de la forme :

$$\frac{dc}{dt} = -\lambda\, c + E \ , \quad c(0) = c_0 \ , \quad \lambda > 0 \tag{3.2}$$

de solution $c(t) = \frac{E}{\lambda} + (c_0 - \frac{E}{\lambda})\, e^{-\lambda t}$ et, cette fois-ci, la condition initiale est rapidement "oubliée" alors que le terme de flux E joue un rôle prépondérant.

L'objectif de l'assimilation de données va, dans ce contexte, consister à utiliser des mesures de l'état que l'on cherche à décrire afin de réduire les incertitudes sur les conditions initiales ou/et les conditions aux limites. Il est à noter que dans certaines applications (notamment pour la pollution atmosphérique), l'objectif de la modélisation peut être non pas la prévision d'un état mais directement les flux d'émissions (par exemple dans un contexte de surveillance d'une réglementation de plafonnement d'émission). On est alors typiquement dans une problématique de *modélisation inverse* et l'utilisation du modèle que l'on a construit, en tant que tel, n'est pas suffisante.

Précisons à présent l'organisation de ce chapitre. On va introduire quelques notations et définitions puis dresser un rapide historique dans la section 3.1. Le cas des systèmes sans dynamique en temps sera traité en section 3.2 : ceci permettra de se familiariser avec les méthodes d'interpolation optimale et 3D-var. Le cas des systèmes avec une évolution en temps sera traité en section 3.3 avec les méthodes séquentielles (filtres de Kalman) et 4D-var. La problématique de construction de l'adjoint d'un modèle (nécessaire pour les approches variationnelles) sera également abordée. Enfin, on terminera ce chapitre avec un aperçu sur la problématique du conditionnement des systèmes d'assimilation de données : ceci recoupe à la fois les problèmes de comportement dynamique des systèmes, d'influence du réseau de capteurs ("network design", optimisation d'un réseau d'observation), ou de sensibilité du second-ordre (la sensibilité des paramètres inversés à des paramètres incertains ou à des composantes du système d'assimilation de données).

Les références [Tarantola, 1987], [Kalnay, 2003], [Talagrand, 1997], [Bouttier et Courtier, 1999] ou [Bennett, 1992] sont, à des titres divers, des références incontournables sur ce sujet.

3.1 Quelques notations et définitions

3.1.1 Modèle direct

On suppose que l'on dispose d'un modèle décrivant un état $c \in R^n$ (des champs dynamiques, des champs de concentration d'espèces chimiques [†1]). Ce modèle dépend de paramètres d'entrée $u \in R^m$ et s'écrit sous la forme d'une relation fonctionnelle :

$$c = \Phi(u) \tag{3.3}$$

qui est à voir comme la solution d'un code informatique en pratique. Pour un CTM, u représente les conditions initiales, les champs de forçage (notamment météorologiques) et des paramètres physiques ; le modèle est alors donné par une équation de dispersion réactive (ou sa discrétisation numérique) typiquement de la forme :

$$\frac{\partial c_i}{\partial t} + \mathrm{div}(V(x,t)\, c_i) = \mathrm{div}(K(x,t)\nabla c_i) + \chi_i(c, T(x,t)) \tag{3.4}$$

$$-\Lambda_i(x,t)c_i + S_i(x,t)\ , \quad c(0) = c_0 \tag{3.5}$$

où c_i représente la concentration chimique de l'espèce i et χ_i les taux de production/consommation de cette espèce donnés par le mécanisme chimique choisi. Les grandeurs de nature météorologique sont V, le champ de vitesse du vent, K une paramétrisation de la diffusion turbulente, et T le champ de température ; elles sont données par un modèle météo dans le cadre d'un couplage off-line, ce qui est l'approche usuelle aux échelles régionales/continentales.

Λ_i désigne une paramétrisation des processus de lessivage et S_i un terme source ponctuel (typiquement un rejet de cheminée d'usine).

A cette équation de champs sont classiquement ajoutées des conditions aux limites, notamment au sol pour décrire les processus de dépôt sec et d'émissions surfaciques :

$$-K\,\nabla c_i \cdot \vec{n} = E_i(x,t) - v_i^{dep}(x,t)c_i \tag{3.6}$$

avec $\vec{n}$ le vecteur unitaire vertical ascendant, E_i un terme d'émission surfacique et v_i^{dep} une paramétrisation de la vitesse de dépôt sec.

Φ est ce que l'on appellera par la suite le *modèle direct* : sur la base des entrées (des fichiers) on calcule directement les sorties (figure 3.2).

[†1] On a fait le choix de prendre cette notation pour essayer d'être le plus compatible possible avec les modèles de Chimie-Transport, dans lesquels c désigne classiquement une concentration chimique. Ceci au risque de ne pas suivre les notations recommandées en assimilation de données par [Ide *et al.*, 1997].

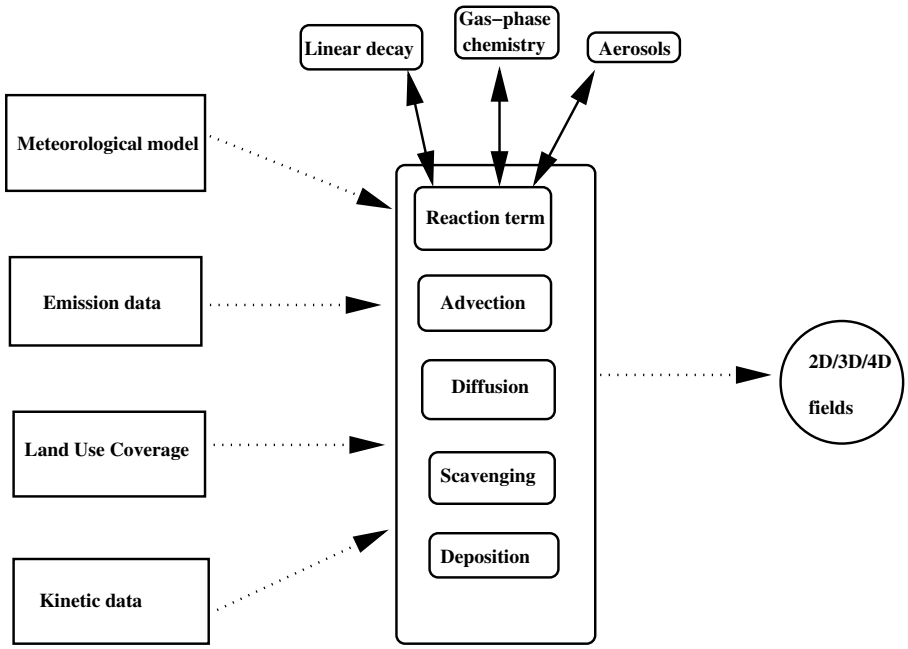

FIG. 3.2 – Structure entrée/sortie d'un CTM.

3.1.2 Modèle linéaire tangent

On peut être intéressé par la sensibilité des sorties aux entrées du modèle, ce que donnera le calcul de la matrice jacobienne :

$$s(u) = \frac{\partial \Phi}{\partial u} \tag{3.7}$$

la fonction étant *a priori* non linéaire. Au premier ordre, ceci permet de calculer la propagation de petites perturbations selon :

$$\delta c = s(u)\delta u \tag{3.8}$$

On parle habituellement de *modèle linéaire tangent*.

3.1.3 Observations

Supposons à présent qu'un réseau d'observation délivre une mesure $obs \in \mathbf{R}^p$ qui est reliée à l'état du système observé c via une relation $obs = H(c)$. H est classiquement appelé *opérateur d'observation*.

- **Exemple 1 :** dans le cas d'un CTM, on peut disposer de capteurs terrestres. Auquel cas, c désigne la concaténation des champs 4D (temps/espace) pour les espèces

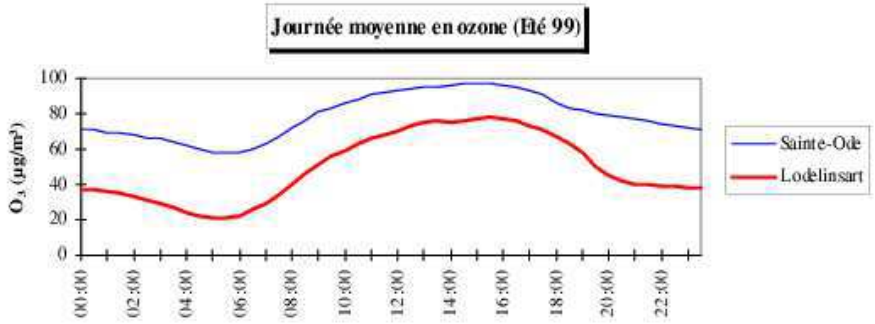

FIG. 3.3 – Ozone - journée moyenne - Comparaison milieu urbain/rural

chimiques. *obs* est formé de quelques espèces mesurables en quelques points (les positions des capteurs) et à quelques instants (les temps de mesures). H est alors une matrice de projection.
On distingue classiquement plusieurs types de station selon leur proximité aux sources d'émissions. De manière schématique, on peut lister :

- les *stations de proximité* placées aux abords d'une source spécifique (usine, voie de circulation) et dont la représentativité est de l'ordre de quelques dizaines de mètres.
- les *stations urbaines*, situées en milieu urbain, mais dont l'éloignement à des sources spécifiques permet de les considérer comme représentatives d'un état moyen de la qualité de l'air dans un rayon de quelques centaines de mètres.
- les *stations rurales* (ou *de fond* selon le contexte) assez éloignées des sources d'émission pour pouvoir mesurer les polluants secondaires formés par les réactions chimiques. Leur représentativité est de l'ordre de quelques kilomètres.

On se réfère par exemple à [Hoelzemann *et al.*, 2001, Hoelzemann *et al.*, 2004, Tilmes et Zimmermann, 1998] pour une classification des stations au regard de l'assimilation de données.

La figure 3.3 montre la différence des niveaux de concentrations d'ozone pour une journée moyenne de l'été 1999 pour les stations de mesures belges de Sainte-Ode et Lodelinsart.

- **Exemple 2 :** les observations peuvent aussi être de nature satellitaire. Dans ce cas, pour un capteur radiatif, on va typiquement mesurer une radiance à une certaine longueur d'onde λ, I_λ. H va alors être défini par les processus d'interaction entre la matière (dont les espèces chimiques) et le rayonnement. En négligeant la diffusion pour un rayonnement dans l'infra-rouge ainsi que les émissions thermiques, on a alors :

$$obs = I_\lambda = I_0 \, e^{-\int_0^h \sum_i \sigma_\lambda^{(i)} c_i(z) dz} \tag{3.9}$$

avec h l'altitude du capteur, i un index pour les espèces chimiques, $c_i(z)$ un profil de concentration pour les espèces chimiques et σ_λ une section efficace d'absorption

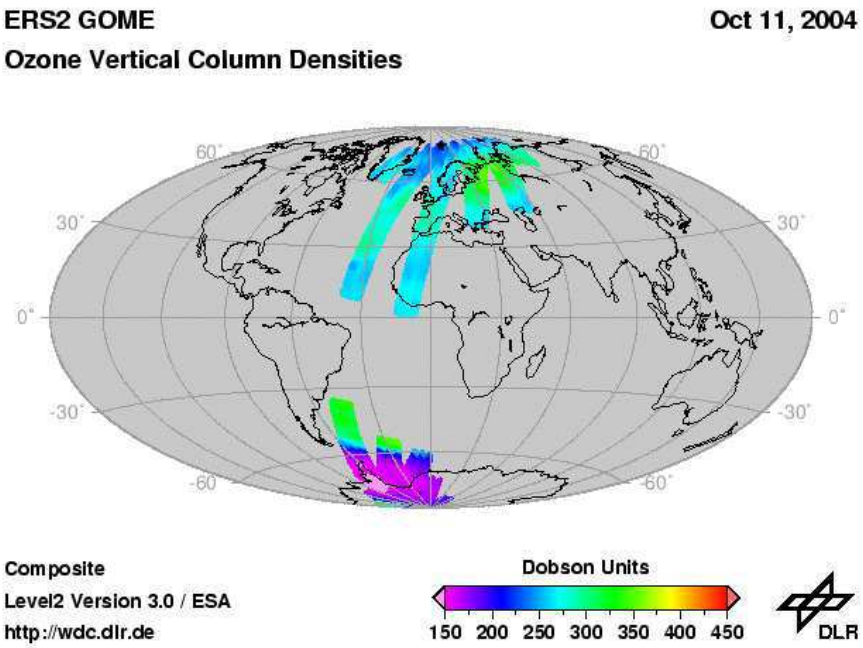

FIG. 3.4 – Courbe typique GOME.

pour la longueur d'onde λ. Ceci définit H comme une fonction des profils de concentrations $c(z)$.

La figure 3.4 est une représentation des mesures effectuées par le satellite GOME (Global Ozone Monitoring Experiment) pour la journée du 11 octobre 2004.

3.1.4 Fonctionnelle de coût

On désire à présent améliorer la connaissance de l'état c en prenant en compte de manière conjointe les deux sources d'information disponibles, le modèle direct et les observations.

Il est à noter, de manière quelque peu spécieuse (les démarches conduisant mathématiquement à la même méthodologie), qu'il y a deux démarches possibles :

- du point de vue du réseau de mesures, on peut vouloir "interpoler" entre les données d'observation en général lacunaires (en espèces, en temps, en espace) sous la contrainte de la physique (le modèle direct) ;
- du point de vue du modélisateur, on peut vouloir réduire les incertitudes sur les entrées u du modèle, sous contrainte des mesures, pour affiner les sorties c ou parce que l'on est de toute manière intéressé par les entrées u (modélisation

inverse/identification de paramètres).

Modulo la définition d'une mesure de l'écart entre *obs* (la mesure réelle) et $H(\Phi(u))$ (l'équivalent numérique de la mesure), une approche logique est de chercher à minimiser par rapport à u la fonction :

$$J(u) = ||obs - H(\Phi(u))||_o^2 \tag{3.10}$$

avec $||.||_o$ une norme qui reste à choisir. J s'appelle habituellement la *fonction coût*.

Dans un cas vectoriel, il est logique de pondérer chacune des composantes par la confiance accordée à la mesure.

On peut par ailleurs disposer d'une estimation *a priori* de la valeur recherchée, que l'on notera u_b. On parle classiquement d'*ébauche* ("background" en anglais). Pour assurer un compromis entre l'information donnée par les observations et l'information donnée par cette valeur d'ébauche, la fonctionnelle J devient alors :

$$J(u) = ||obs - H(\Phi(u))||_o^2 + ||u - u_b||_b^2 \tag{3.11}$$

avec $||.||_b$ une norme qui reste à choisir.

On définit, pour la suite, de la même manière une fonctionnelle sur l'état :

$$\tilde{J}(c) = ||obs - H(c)||_o^2 + ||c - c_b||_b^2 \tag{3.12}$$

où de manière impropre on a gardé la même notation pour la métrique d'ébauche.

3.1.5 Cadre stochastique

3.1.5.1 Propagation d'incertitudes

Il est évidemment beaucoup plus riche de considérer l'état c comme une variable stochastique C, donnée par une densité de probabilité, que l'on notera $P_C(.)$. $P_C(c)$ est la probabilité que l'état C ait pour réalisation c. De même, dans ce cadre, l'entrée est supposée suivre une loi de probabilité $P_U(.)$.

Un calcul de propagation des incertitudes revient à calculer la loi P_C en fonction de la loi P_U. De manière usuelle, cela va revenir à considérer des modèles de type Fokker-Planck ([Risken, 1984]). Par exemple, si on considère un modèle dynamique donné par :

$$\frac{dc}{dt} = f(c) \ , \quad c(0) = c_0 \tag{3.13}$$

La fonction Φ (on écrira Φ_t pour marquer sa dépendance au temps) est alors donnée, si l'on considère comme seule variable d'entrée les conditions initiales, par :

$$c(t) = \Phi_t(c_0) \stackrel{\Delta}{=} c_0 + \int_0^t f(c(s))ds \tag{3.14}$$

La propagation des incertitudes sur les conditions initiales est donnée par une équation de Fokker-Planck qui décrit l'évolution au cours du temps de la densité de probabilité $P_C(c,.)$:

$$\frac{\partial P_C}{\partial t} + \mathrm{div}_c(P_C\, f) = 0\ , \quad P_C(c,0) = P_C^{(0)}(c) \tag{3.15}$$

Nous allons à présent rajouter une erreur modèle q (la dérive de la dynamique du modèle) avec le nouveau système :

$$\frac{dc}{dt} = f(c) + q\ , \quad c(0) = c_0 \tag{3.16}$$

Dans ce cas, les paramètres d'entrée sont définis par q et c_0. Si on suppose que q (l'erreur du modèle) suit une loi brownienne, alors $P_C(c,.)$ est une fonction du temps donnée par :

$$\frac{\partial P_C}{\partial t} + \mathrm{div}_c(P_C\, f) = \frac{1}{2}\mathrm{div}_c(Q\nabla_c\, P_C)\ , \quad P_C(c,0) = P_C^{(0)}(c) \tag{3.17}$$

avec $Q = V(qq^T)$ qui joue le rôle d'une diffusion. Avec un résultat classique des équations d'advection-diffusion, il est clair dans ce cas là que lorsque l'erreur modèle augmente (Q "élevé"), l'incertitude augmente (P_C devient uniformément faible au cours du temps).

La résolution de l'équation de Fokker-Planck est bien entendu extrêmement difficile en pratique puisqu'elle est posée sur un espace de dimension le nombre de degrés de liberté de l'état c (pour un CTM, le produit du nombre de mailles par le nombre d'espèces, de l'ordre de 10^6 typiquement). Une des solutions praticables est alors le recours à des méthodes de type Monte-Carlo (on le verra par la suite dans la partie consacrée à la prévision d'ensemble).

3.1.5.2 Hypothèses de normalité

On suppose que les valeurs réelles de l'état et des entrées sont données respectivement par c_t et u_t [†2].

L'ébauche u_b est supposée être une variable stochastique à laquelle on associe l'*erreur d'ébauche*, $\epsilon_b = u_b - u_t$ de moyenne supposée nulle et dont la matrice de covariance est supposée donnée et sera notée B :

$$B = cov(\epsilon_b) =< \epsilon_b \epsilon_b^T > \tag{3.18}$$

[†2]t pour "true". Impérialisme anglo-saxon des notations standards et confusion éventuelle avec le temps.

avec $< ., . >$ l'opération de moyenne au sens des variables stochastiques.

On définit de même l'*erreur d'observation* $\epsilon_o = obs - H(c_t)$ de moyenne supposée nulle et dont la matrice de covariance est supposée donnée et sera notée R :

$$R = cov(\epsilon_o) =< \epsilon_o \epsilon_o^T > \tag{3.19}$$

Les erreurs d'ébauche et d'observation sont supposées être décorrélées, ce qui signifie que $< \epsilon_b \epsilon_o^T >= 0$.

Notons que l'on ne définit jamais l'espace de probabilité sur lequel on travaille effectivement ...

3.1.5.3 Quelques remarques sur les matrices de covariance

La définition des matrices de covariance d'erreur est en pratique un point particulièrement délicat, notamment en météorologie.

Les matrices de covariance d'erreur d'ébauche sont souvent modélisées à l'aide de fonctions d'influence, sous la forme classique de Balgovind selon :

$$B(r) = \left(1 + \frac{r}{L}\right) \exp\left(-\frac{r}{L}\right) \tag{3.20}$$

On fait ici une hypothèse d'isotropie (r désignant la distance entre deux positions spatiales). L est un rayon d'influence. Cette forme est par exemple prise dans [Elbern *et al.*, 1997] et les travaux qui suivent pour la chimie atmosphérique.

Les matrices de covariance d'erreur d'observation mettent en jeu deux notions différentes ([Cohn, 1997] par exemple) :

- les erreurs de mesures à proprement parler (de l'ordre du ppb pour un capteur terrestre d'ozone par exemple) ;
- les erreurs de représentativité par ailleurs.

Le second point est d'autant plus crucial pour la chimie atmosphérique que de très fortes hétérogénéités existent à proximité des sources d'émissions.

3.1.6 Un rapide historique

On fait par exemple référence à [Bouttier et Courtier, 1999].

Les méthodes de "nudging" (plus généralement de corrections successives) reviennent à contraindre un modèle d'évolution par une relaxation vers une donnée connue par ailleurs (observation, climatologie, ébauche). Par exemple, notre modèle d'évolution (3.13) est alors corrigé selon :

$$\frac{dc}{dt} = f(c) - \alpha(H(c) - obs) \tag{3.21}$$

en supposant connu (par souci de clarté) un continuum en temps d'observation $obs(t)$. α est le coefficient de nudging à estimer ([Zou *et al.*, 1992, Stauffer et Bao, 1993] pour une approche théorique).

Une méthode d'*interpolation optimale* va revenir à calculer directement un état solution (que l'on désigne classiquement comme l'*analyse*) de la minimisation de la fonctionnelle :

$$\tilde{J}(c) = ||obs - Hc||_o^2 + ||c - c_b||_b^2 \tag{3.22}$$

sous la forme $c_a = c_b + K(Hc - obs)$ où K est la *matrice de gain* (on a supposé que l'opérateur d'observation et le modèle étaient linéaires). L'extension de cette approche à un modèle d'évolution est alors fournie par les méthodes de type filtres de Kalman. On parle dans ce cadre de *méthodes séquentielles* (voir les sections suivantes).

Une alternative, éventuellement moins coûteuse en temps calcul (notamment pour des systèmes de grande taille non linéaires) est le recours aux *méthodes variationnelles*, qui reviennent à utiliser un algorithme itératif pour calculer la solution du problème de minimisation. Lorsqu'il n'y a pas d'évolution en temps, on parle de méthode 3D-variationnelle (ou *3D-var*) ; lorsqu'il y a une dynamique en temps, on parle de méthode 4D-variationnelle (ou *4D-var* [Le Dimet et Talagrand, 1986]). Un des ingrédients essentiels est alors le calcul du gradient ∇J de la fonction coût si on utilise un algorithme de type gradient. Les sections suivantes vont détailler ces approches.

Les méthodes précédentes (séquentielles et variationnelles) permettent de retrouver une solution c_a (ou u_a pour de la modélisation inverse) et également (on ne détaille pas ici) une estimation de la confiance accordée à cette solution, en pratique une matrice de covariance. Définir la variable en toute rigueur stochastique par un état (en pratique un état moyen) et une matrice de covariance revient à considérer que cette variable suit une loi gaussienne. Comme les incertitudes sont supposées rester gaussiennes, cela revient implicitement à ne traiter que des modèles linéaires (ou linéarisés).

Une autre classe de méthodes désormais utilisées de manière opérationnelle repose sur une extension au contexte de l'assimilation de données de la problématique de propagation des incertitudes. Ce sont les méthodes de *prévision d'ensemble* ([Hamill, 2002] par exemple, dont on parlera par la suite, qui ne reposent plus sur ce strict cadre

gaussien/linéaire).

Enfin, une problématique émergente est celle du "network design". Comment optimiser un réseau de mesures en fonction d'une cible de prévision (d'assimilation de données) ? Comment déployer en temps réel des moyens supplémentaires d'observation ? On fait référence au Chapitre A pour une revue synthétique de cette problématique.

3.2 Modèles stationnaires

On va oublier ici l'existence d'un modèle d'évolution et chercher à retrouver l'état c à partir des seules observations $obs = H(c)$ et d'une ébauche c_b. Notons que H et c_b peuvent être (sont en pratique) donnés par des modèles.

Nous allons d'abord introduire le cadre bayésien classiquement utilisé avant de présenter les méthodes de l'interpolation optimale et du 3D-Var. On finira cette section par quelques remarques sur l'interpolation au sens des moindres carrés et la régularisation de Tikhonov.

Pour compléter cette présentation basique pour l'assimilation de données, on fait par exemple référence à [Tarantola, 1987].

3.2.1 Approche bayésienne

Dans un cadre bayésien, on va chercher à trouver l'état c le plus probable connaissant l'observation obs et l'ébauche c_b. La formule de Bayes sur les probabilités conjointes de l'état et de l'observation s'écrit :

$$P(c, obs) = P(c|obs)P(obs) = P(obs|c)P(c) \tag{3.23}$$

$P(c)$ correspond ici à la connaissance *a priori* sur c et est définie par l'ébauche ; $P(obs|c)$ est déterminée à partir du modèle et de l'opérateur d'observation via $obs = H(c)$; enfin, l'analyse *a posteriori* est associée à $P(c|obs)$.

$P(obs)$ étant non contrôlable (elle ne dépend pas de l'état), on peut se ramener pour simplifier à :

$$P(c|obs) \sim P(obs|c)P(c) \tag{3.24}$$

La meilleure estimation de c va alors dépendre du critère d'optimalité choisi :

– le *maximum a posteriori* est défini par

$$c_a = \text{argmax}_c \, P(c|obs) \tag{3.25}$$

– le *maximum de vraisemblance* est défini à partir l'observation estimée

$$c_a = \text{argmax}_c P(obs|c) \tag{3.26}$$

– l'*estimation de variance minimale* par :

$$c_a = \int_c cP(c|obs)dc \tag{3.27}$$

Si on fait des hypothèses gaussiennes sur les distributions de probabilité, ces estimateurs se confondent. On peut écrire par exemple (en omettant les facteurs préexponentiels de normalisation) :

$$\begin{aligned} P(c) &\sim \exp\left(-\frac{1}{2}(c-c_b)^T B^{-1}(c-c_b)\right) \\ P(obs|c) &\sim \exp\left(-\frac{1}{2}(obs-H(c))^T R^{-1}(obs-H(c))\right) \end{aligned} \tag{3.28}$$

Pour le maximum a posteriori, on va donc chercher à maximiser $P(c|obs)$ en fonction de c, ou de manière équivalente sachant que $P \in [0,1]$:

$$-\log P(c|obs) = \frac{1}{2}(obs-H(c))^T R^{-1}(obs-H(c)) + \frac{1}{2}(c-c_b)^T B^{-1}(c-c_b) \tag{3.29}$$

$$= \tilde{J}(c) \tag{3.30}$$

Notons que nous avons, de cette manière, naturellement donné un sens aux normes $||.||_o$ et $||.||_b$.

Une solution peut alors être obtenue directement (méthode de l'interpolation optimale) ou de manière itérative (approche variationnelle, 3D-Var).

3.2.2 Interpolation optimale

Un minimum local de $\tilde{J}$ est donné par $\nabla_c \tilde{J} = 0$, ce qui définit ce que l'on appelle classiquement l'*analyse*, c_a :

$$c_a = c_b + K(obs - H(c_b)) \tag{3.31}$$

avec K la *matrice de gain* :

$$BH^T(HBH^T + R)^{-1} = (B^{-1} + H^T R^{-1} H)^{-1} R^{-1} H \tag{3.32}$$

Le vecteur $obs - H(c_b)$ est parfois appelé *vecteur innovation* (ce qu'amène l'information conjointe observation/ébauche). La matrice K donne une pondération relative en fonction de la confiance accordée respectivement aux observations et à l'ébauche. Notons, en faisant

un calcul de coin de table, que si $R \gg B$ (peu de confiance à l'observation), $c_a \simeq c_b$. A l'inverse, si $R \ll B$ (peu de confiance à l'ébauche), on retrouve $H(c_a) \simeq obs$ via

$$HK = I - R(HBH^T + R)^{-1} \tag{3.33}$$

Après calcul, on a également une estimation de la variance de l'erreur d'analyse, $A = cov(c_a)$, avec :

$$A = (I - KH)B \tag{3.34}$$

En pratique, l'interpolation optimale est aisée à mettre en œuvre mais un certain nombre de problèmes restent posés : H doit être linéaire et le coût en algèbre linéaire peut être élevé avec des matrices de grande taille à inverser.

3.2.3 Moindres carrés et régularisation de Tikhonov

Une autre interprétation est d'utiliser une approche de type *moindres carrés*. On veut pouvoir inverser le problème :

$$Hc = obs \tag{3.35}$$

qui est souvent en pratique un problème sous-contraint : on dispose de beaucoup moins d'observations que de degrés de liberté recherchés ($n \gg p$). Une manière de résoudre cet écueil est de régulariser (pénaliser) le problème en rajoutant une information supplémentaire via l'ébauche c_b. Dans ce contexte, on cherche alors à minimiser une fonction du type :

$$\tilde{J}(c) = ||H(c) - obs||^2 + \epsilon||c - c_b||^2 \tag{3.36}$$

avec ϵ un paramètre de régularisation (indiquant la pondération entre les deux termes).

Un théorème classique (admis ici, [Kern, 2003]) donne l'estimation entre la solution réelle c_t et la solution c_n^ϵ correspondant au problème régularisé avec ϵ et des observations bruitées selon $obs = Hc_t + \delta_n$:

$$||c_n^\epsilon - c_t|| \leq O\left(\frac{\delta_n}{\epsilon^2}\right) + O(\epsilon) \tag{3.37}$$

Le premier terme de ce contrôle d'erreur montre que la régularisation doit être effectuée en fonction de l'erreur d'observation : si l'erreur d'observation est forte, ϵ devra être choisi grand (mais au risque de détériorer le second terme) ; à l'inverse si l'erreur d'observation est faible, on ne peut choisir un ϵ trop petit (pour diminuer le second terme) car ceci aurait un impact sur le premier terme.

Cette formule illustre en quelque sorte le compromis entre complexité du modèle et précision. Un modèle complexe sera plus précis mais les observations seront probablement entachées d'erreur ; un modèle simple sera plus imprécis mais les observations associées seront plus robustes.

3.2.4 Approche variationnelle : 3D-var

L'idée est d'éviter le calcul explicite et l'inversion de la matrice de gain par l'utilisation d'un algorithme de minimisation, par exemple du type gradient (ou ses algorithmes dérivés). On calcule alors une suite de valeurs pour l'état estimé selon :

$$c_a^{k+1} = c_a^k - \alpha_k \nabla_c \tilde{J}(c_a^k) \tag{3.38}$$

avec α_k un paramètre numérique qui dépend de la méthode choisie. A convergence, la valeur donne l'analyse c_a. Notons que ceci nécessite le calcul du gradient de la fonction coût (voir plus loin).

Un des avantages de cette approche est l'extension au cas des opérateurs d'observation non linéaires et l'extension aux problèmes de type modélisation inverse.

3.3 Modèles d'évolution

On considère à présent un modèle dynamique que l'on va écrire sous la forme itérative discrétisée suivante pour passer du temps t_i au temps t_{i+1} :

$$c(t_{i+1}) = M(t_i, t_{i+1})c(t_i) + q_i \tag{3.39}$$

avec q_i une erreur modèle supposée suivre une loi normale de moyenne nulle (pas de biais) et de matrice de covariance Q_i. Ce faisant, on s'est restreint au cas d'un modèle linéaire (ou linéarisé) de t_i à t_{i+1}.

On suppose disposer également d'observations distribuées en temps, $(obs(t_i))_i$, associées à un jeu d'opérateurs d'observation non constant a priori, $(H(t_i))_i$, et à des matrices de covariances d'erreur d'observation $(R(t_i))_i$. Pour simplifier les notations, on s'autorisera également les notations c_i, obs_i, H_i et R_i.

3.3.1 Méthodes séquentielles

3.3.1.1 Filtre de Kalman

L'extension naturelle de la méthode d'interpolation optimale est le filtre de Kalman ([Jazwinski, 1970] par exemple). L'idée est de réaliser des interpolations optimales successives (de manière *séquentielle*) en propageant entre deux échéances d'observation les informations (état et covariance d'erreur) selon le modèle.

Dans ce contexte, l'ébauche à t_{i+1} est remplacée par une valeur fournie par le modèle en *prévision* de t_i à t_{i+1}, à partir de la meilleure estimation de l'état en t_i, à savoir l'analyse $c_a(t_i)$. On remplacera dans les notations c_b par c_f ("f" pour forecast) et B par P_f. Par souci de cohérence, on remplace également A par P_a.

L'extension directe des formules de l'interpolation optimale entre t_i et t_{i+1} conduit au schéma suivant :

1. Etape de prévision de l'état :

$$c_f(t_{i+1}) = M(t_i, t_{i+1}) c_a(t_i) \tag{3.40}$$

2. Propagation de l'erreur :

$$P_f(t_{i+1}) = M(t_i, t_{i+1}) P_a(t_i) M(t_i, t_{i+1})^T + Q(t_i) \tag{3.41}$$

3. Analyse en t_{i+1} :

$$c_a(t_{i+1}) = c_f(t_{i+1}) + K(t_{i+1}) \left(obs(t_{i+1}) - H(t_{i+1}) c_f(t_{i+1}) \right) \tag{3.42}$$

avec la matrice de gain :

$$K(t_{i+1}) = P_f(t_{i+1}) H(t_{i+1})^T \left(H(t_{i+1}) P_f(t_{i+1}) H(t_{i+1})^T + R(t_{i+1}) \right)^{-1} \tag{3.43}$$

4. Calcul de la covariance d'analyse en t_{i+1} :

$$P_a(t_{i+1}) = \left(I - K(t_{i+1}) H(t_{i+1}) \right) P_f(t_{i+1}) \tag{3.44}$$

Quelques remarques :

– Une des hypothèses clés pour établir ces relations est l'indépendance des erreurs d'observation, de prévision et de modèle.

– Une extension au cas d'un système non-linéaire de dynamique :

$$c(t_{i+1}) = F_i(c(t_i)) \tag{3.45}$$

est de considérer le modèle linéarisé avec $M(t_i, t_{i+1}) = \frac{\partial F_i}{\partial c}$. On parle alors de filtre de Kalman *étendu* (Extended Kalman Filter : EKF).

3.3.1.2 Filtres réduits

Une des difficultés d'implémentation de cette méthode est le coût prohibitif de l'étape de propagation de la matrice de covariance d'erreur (de l'ordre de $2n$ fois le coût du modèle direct avec n le nombre de degrés de liberté de l'état !). En pratique, cette étape est souvent remplacée par une modélisation de l'erreur de covariance.

Une autre approche est donnée par ce que l'on appelle communément les *filtres réduits* qui revient à projeter la matrice de covariance sur quelques *modes* (les méthodes vont différer selon le choix de ces modes).

Notons d'abord que les matrices de covariance peuvent s'écrire (en omettant les indices temporels) selon :

$$P_f = S_f S_f^T \ , \quad P_a = S_a S_a^T \ , \quad Q = S_q S_q^T \ , \quad R = S_r S_r^T \tag{3.46}$$

A partir de la définition des matrices de covariance, il est direct de vérifier que les racines carrées S peuvent se propager selon :

$$S_f = (MS_a \ , \quad S_q) \ , \quad K = S_f \Psi (\Psi^T \Psi + R)^{-1} \tag{3.47}$$

$$S_a = S_f (I - \Psi (R + \Psi^T \Psi)^{-1} \Psi^T)^{1/2} \tag{3.48}$$

où $\Psi^T = HS_f$.

Un des avantages de propager les racines carrées S plutôt que les matrices de covariances P est la garantie de conserver le caractère symétrique défini positif. De plus, en notant $S = (s_1 \dots s_{n_s})$ (n_s dépend de la matrice, s_i est une colonne), on a :

$$SS^T = \sum_{i=1}^{n_s} s_i {s_i}^T \tag{3.49}$$

et modulo un rangement des colonnes par ordre décroissant de $s_i^T s_i$, on définit aisément des troncatures de S.

Pour le filtre RRSQRT ([Verlaan, 1998]), les modes sont directement définis à partir de la matrice $S_a^T S_a$ (qui a les mêmes valeurs propres que P_a). Si $S_a^T S_a = V diag(\lambda_i) V^T$ avec V la matrice de changement de base orthonormale des vecteurs propres v_i (associés à λ_i), on a aisément :

$$SS^T = S(VV^T)S^T = SV(SV)^T = \sum (Sv_i)(Sv_i)^T \tag{3.50}$$

Or avec $||Sv_i|| = \sqrt{\lambda_i}$, on peut réécrire :

$$SS^T = \sum \lambda_i w_i w_i^T \ , \quad w_i = \frac{Sv_i}{\sqrt{\lambda_i}} \ , \quad ||w_i|| = 1 \tag{3.51}$$

Si on ordonne les valeurs propres par ordre décroissant, le filtre réduit naturel est alors de tronquer la somme, ou de manière équivalente de ne garder que les colonnes correspondantes dans V, ce qui définit $\tilde{V}$. On définit alors :

$$\tilde{S} = S\tilde{V} \tag{3.52}$$

ce qui constitue la réduction du filtre initial à ses modes dominants.

Le filtre SEEK (Singular Evolutive Extended Kalman, [Pham *et al.*, 1998]) ou sa variante SEIK (Singular Evolutive Interpolated Kalman, [Pham, 1996]) est fondé sur une réduction liée au comportement "statistique" du modèle : les modes dominants sont alors calculés à partir des EOFs qui contiennent l'information du système le long de trajectoires calculées. On ne détaille pas plus dans ce cadre.

3.3.1.3 Filtres d'ensemble

Une alternative est le recours à des méthodes de prévision d'ensemble en cherchant à estimer la covariance d'erreur de prévision à l'aide d'une méthode de type Monte–Carlo :

$$P_f =< (c - c_t)(c - c_t)^T >\simeq \frac{1}{N}\sum_{j=1}^{j=N}(c^{(j)} - \bar{c})(c^{(j)} - \bar{c})^T \tag{3.53}$$

avec un *ensemble* de N états possibles $\left(c^{(j)}\right)$ de moyenne $\bar{c}$.

Notons que vis-à-vis de l'équation 3.17 de propagation des densités de probabilité, cela revient à résoudre le terme d'advection par le modèle via une étape de type prévision et à approcher le terme de diffusion par N filtres de Kalman.

Une étape clé est bien entendu la génération des perturbations. On se réfère par exemple à [Durbiano, 2001] pour une synthèse des approches actuellement utilisées.

3.3.1.4 Lisseur de Kalman

Notons que l'approche variationnelle est générale : la variable de contrôle peut être l'état (ce qui a été présenté pour plus de clarté) mais toute autre variable également (par exemple des termes de flux/émissions). Ceci n'est pas possible avec le filtre de Kalman tel qu'il a été présenté, mais une variante existe avec la notion de *Kalman Smoother* qui revient à rajouter dans l'état la valeur du paramètre recherché telle qu'elle est estimée après les analyses déjà effectuées (et qui est affinée au fur et à mesure des analyses successives).

De manière générale, un des inconvénients majeurs de ces approches est son caractère séquentiel et les discontinuités successives que ceci introduit sur l'estimation de l'état du système (figure 3.5). Une manière de *lisser* est d'utiliser une méthode variationnelle en prenant en compte les observations de manière conjointe sur une *fenêtre d'assimilation*.

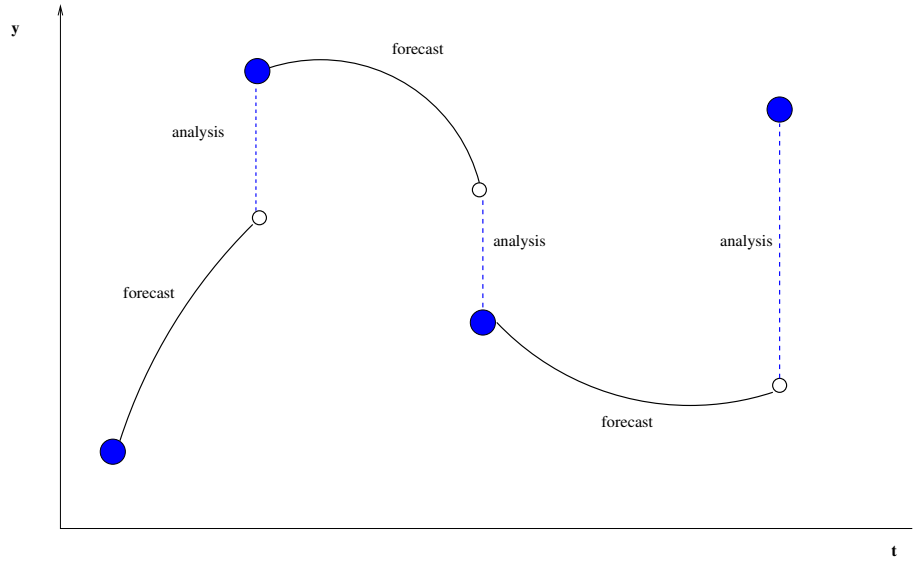

FIG. 3.5 – Méthode séquentielle : succession des étapes de prévision/analyse.

3.3.2 Méthode 4D-variationnelle

On reprend l'approche par minimisation d'une fonctionnelle en définissant pour un jeu d'observations distribuées en temps aux instants successifs $t_0 \dots t_{n_o}$:

$$\begin{aligned} \tilde{J}(c(t_0), \dots, c(t_{n_o})) &= \sum_{i=0}^{i=n_o} (H(t_i)c(t_i) - obs(t_i))^T R(t_i)^{-1} (H(t_i)c(t_i) - obs(t_i)) \\ &\triangleq \sum_{i=0}^{i=n_o} \tilde{J}_i(c(t_i)) \end{aligned} \tag{3.54}$$

Si l'on suppose que le modèle est *parfait* (on parle aussi de *contrainte forte*), c'est à dire si l'on suppose que les états successifs sont reliés par le modèle dynamique :

$$c(t_{i+1}) = F_i(c(t_i)) \tag{3.55}$$

la fonction n'est plus qu'une fonction de la condition initiale (ou par extension de tout autre paramètre que l'on chercherait à estimer : par exemple des flux d'émission). Si u est l'ensemble des variables à estimer, on note la dépendance $(c(t_i))\,(u) = c_i(u)$ et on définit :

$$J(u) = \tilde{J}(c_0(u), \dots, c_{n_o}(u)) \tag{3.56}$$

Une méthode variationnelle va alors consister à utiliser, de même que pour l'approche 3D-var, un algorithme itératif de minimisation. En pratique, le gradient de la fonction coût $\nabla_u J$ devra être calculé.

3.3.2.1 Modèle adjoint

Une première approche revient à calculer ce gradient par une approche de type "différences finies" :

$$(\nabla_u J)_k \simeq \frac{J(u+\epsilon e_k) - J(u)}{\epsilon} \tag{3.57}$$

avec $\epsilon \in \mathbf{R}$ une perturbation infinitésimale et e_k le k-ième vecteur unitaire de la base canonique.

Le calcul est irréalisable en pratique puisqu'il nécessite un nombre d'évaluations de J (donc du modèle direct) égal à la dimension de u ! Un autre écueil pratique est le choix approprié de la perturbation ϵ (assez petite pour l'hypothèse de linéarisation, assez grande pour ne pas avoir des problèmes d'erreur d'arrondi numérique).

Une seconde approche est reliée à ce que l'on appelle communément *le modèle adjoint*. En notant la relation $J(u) = \tilde{J}(c(u))$, on a pour une perturbation δu :

$$\begin{aligned} \delta J &= (\nabla_c \tilde{J}, \delta c) && (3.58)\\ &= \left(\nabla_c \tilde{J}, \frac{\partial c}{\partial u}\delta u\right) && (3.59)\\ &= \left((\frac{\partial c}{\partial u})^T \nabla_c \tilde{J}, \delta u\right) \end{aligned}$$

ce qui définit :

$$\nabla_u J = (\frac{\partial c}{\partial u})^T \nabla_c \tilde{J} \tag{3.60}$$

où la transposition est intrinsèquement liée au choix des produits scalaires. On note $V_c^{'}$ et $V_u^{'}$ le espaces linéaires tangents associés respectivement à l'espace des états et des paramètres, munis d'un produit scalaire, et $M : V_u^{'} \rightarrow V_c^{'}$. La transposition est alors définie par :

$$(M\delta u, \delta\tilde{c})_{V_c^{'}} \stackrel{\Delta}{=} (\delta u, M^T \delta\tilde{c})_{V_u^{'}} \tag{3.61}$$

L'extension au cas d'un modèle itératif est directe avec la loi de dérivation de fonctions composées. Par exemple, avec un modèle à deux itérations $u \rightarrow c_1 \rightarrow c_2$, on a directement :

$$\nabla_u J = (\frac{\partial c_1}{\partial u})^T (\frac{\partial c_2}{\partial c_1})^T \nabla_{c_2} \tilde{J} \tag{3.62}$$

L'algorithme de calcul est clair en notant que la séquence $u \rightarrow c_1 \rightarrow c_2 \rightarrow J$ a été inversée :

1. Développer le *modèle adjoint* :

$$z \rightarrow \left(\frac{\partial c_i}{\partial c_{i-1}}\right)^T z \tag{3.63}$$

2. Utiliser une séquence inverse.

La généralisation est immédiate au cas d'une fonction coût générale de la forme (3.54). On se réfère au chapitre 5 pour l'application au cas d'un modèle de Chimie-Transport.

Quelques remarques :

1. Cette approche ne nécessite pas de propager des matrices de covariance et peut s'appliquer à des modèles non linéaires.
2. Sous certaines conditions (notamment de linéarité du modèle et des observations), il y a équivalence entre le filtre de Kalman et l'approche 4D-var pour l'analyse finale $c_a(t_{n_o})$ correspondant à la condition initiale restaurée ([Bouttier et Courtier, 1999, Li et Navon, 2001]).

3.3.2.2 Cas d'une EDP

On va illustrer l'extension de la notion d'adjoint au cas d'une EDP avec l'exemple simplissime de l'équation de la chaleur stationnaire posée sur un ouvert borné Ω :

$$-\Delta c_u = S + u \ , \quad c_u \in \ H_0^1(\Omega) \tag{3.64}$$

La fonction c est la sortie du modèle, que l'on note c_u en référence au paramètre d'entrée u. S désigne le terme source connu, u le terme source inconnu. On note $obs(x) = H(x)c(x)$ le champs d'observation qui permet de définir la fonction coût :

$$J(u) = \frac{1}{2}\int_\Omega (Hc_u(x) - obs(x))^2\, dx = \tilde{J}(c_u) \tag{3.65}$$

Pour trouver $\nabla_u J$, on cherche à obtenir un développement :

$$J(u+\delta u) = J(u) + < \nabla_u J, \delta u > + O(||\delta u||^2) \tag{3.66}$$

où $< \phi, \psi >= \int_\Omega \phi(x)\psi(x)dx$ désigne le produit scalaire de $L^2(\Omega)$.

Par linéarité (principe de superposition), $c_{u+\delta u} = c_u + \delta c$ où δc est solution du problème :

$$-\Delta \delta c = \delta u \ , \quad \delta c \in \ H_0^1(\Omega) \tag{3.67}$$

Ceci donne directement :

$$\begin{aligned} J(u+\delta u) &= \frac{1}{2}\int_\Omega (Hc_u(x) - obs(x))^2\, dx \\ &+ \int_\Omega (Hc_u(x) - obs(x)) H\delta c\, dx + \frac{1}{2}\int_\Omega (H\delta c)(x))^2\, dx \end{aligned} \tag{3.68}$$

soit

$$J(u+\delta u) = J(u) + < H^T(Hc_u - obs), \delta c > + O(||\delta u||^2) \tag{3.69}$$

avec un résultat classique d'analyse fonctionnelle (Inégalité de Poincaré).

Remarquons que ceci est inutilisable puisque la seule information disponible est l'action du gradient de J dans toutes les directions δu :

$$< \nabla_u J, \delta u >=< H^T(Hc_u - obs), \delta c > \tag{3.70}$$

ceci via la connaissance préalable de δc (ie via la résolution d'une EDP, dans notre cas trivial, donnée par (3.67)). Autrement dit, la connaissance du gradient passe par la résolution d'une *infinité* de problèmes directs !

La notion de *modèle adjoint* permet de simplifier considérablement le calcul du gradient. En notant $\delta c = M\delta u$ (c'est le modèle linéaire tangent), on a par définition :

$$< M\delta u, H^T(Hc_u - obs) >=< \delta u, M^T H^T(Hc_u - obs) > \tag{3.71}$$

ce qui définit $\nabla_u J = M^T H^T(Hc_u - obs)$.

Dans ce cas précis, $M = M^T$ (autoadjoint) puisque :

$$< M\phi, \psi >=< \Delta(M\phi), M\psi >=< \phi, M\psi > \tag{3.72}$$

en intégrant deux fois par partie, en utilisant les conditions aux limites nulles et $-\Delta(M\psi) = \psi$.

Autrement dit, $\nabla_u J$ peut être calculé directement à l'aide d'un seul modèle direct via la résolution du problème :

$$-\Delta(\nabla_u J) = H(Hc_u - obs) \ , \quad \nabla_u J \in H_0^1(\Omega) \tag{3.73}$$

qui apparaît comme la traduction de la relation (3.60).

$$\nabla_u J = (\frac{\partial c}{\partial u})^T \nabla_c \tilde{J}$$

3.3.2.3 Cas d'une EDO

Pour finir d'illustrer la notion de modèle adjoint, on va étudier le modèle dynamique de la forme :

$$\frac{dc_u}{dt} = f(t, c_u) \ , \quad c_u(0) = u \tag{3.74}$$

avec pour paramètre de contrôle la condition initiale. On cherche alors à minimiser une fonction coût de la forme :

$$J(u) = 1/2 \int_0^T (obs(t) - Hc_u(t))^2 \, dt = \tilde{J}(c_u) \tag{3.75}$$

où T désigne l'intervalle d'intégration en temps.

On note $\delta c_u(t) \triangleq c_{u+\delta u}(t) - c_u(t)$ solution de :

$$\frac{d\delta c_u}{dt} = \frac{\partial f}{\partial c}\delta c_u \ , \quad \delta c_u(0) = \delta u \tag{3.76}$$

On a alors directement :

$$J(u+\delta u) = J(u) + \int_0^T (obs - Hc_u, H\delta c_u)\ dt + O(||\delta u||^2) \tag{3.77}$$

Ce qui donne :

$$(\nabla_u J, \delta u) = \int_0^T (obs(t) - Hc_u(t), H\delta c_u(t))\ dt \tag{3.78}$$

Sous cette forme, le calcul du gradient n'est pas utilisable puisqu'il nécessite de résoudre pour chaque perturbation le système linéaire tangent.

On va procéder de la même manière que précédemment. Dans la suite, $\langle ., . \rangle$ désigne le produit scalaire de $L^2(0,T)$ et $(.,.)$ le produit scalaire euclidien. On note de plus $\delta c_u = M\delta_u$ avec $M\delta u$ solution de :

$$\frac{d(M\delta u)}{dt} = \frac{\partial f}{\partial c}(M\delta u) \ , \quad (M\delta u)(0) = \delta u \tag{3.79}$$

ce qui donne :

$$(\nabla_u J, \delta u) = \left\langle H^T(obs - Hc_u), M\delta u \right\rangle \ , \quad \nabla_u J = M^T \left(H^T(obs - Hc_u) \right) \tag{3.80}$$

L'adjoint de M est défini par :

$$\forall \phi \ , \ \psi(.), \langle M\phi, \psi \rangle = \left(\phi, M^T\psi\right) \tag{3.81}$$

On introduit pour effectuer une intégration par partie $c^\star(t)$ telle que $dc^\star/dt = \psi(t) + k(t)$, où pour le moment $k(t)$ n'est pas spécifié, et on a aisément :

$$\begin{aligned} \langle M\phi, \psi \rangle &= \int_0^T ((M\phi)(t), \psi(t))\ dt \qquad (3.82) \\ &= \int_0^T \left(M\phi, \frac{dc^\star}{dt} - k(t) \right) dt \\ &= (c^\star(T), (M\phi)(T)) - (c^\star(0), \phi) - \int_0^T \left(M\phi, \frac{\partial f}{\partial c}^T c^\star + k(t) \right) dt \end{aligned}$$

En prenant $k(t) = -\frac{\partial f}{\partial c}^T c^\star$, c'est à dire :

$$\frac{dc^\star}{dt} = -\frac{\partial f}{\partial c}c^\star + \psi(t) \ , \quad c^\star(T) = 0 \tag{3.83}$$

on obtient :

$$\langle M\phi, \psi\rangle = (\phi, -c^\star(0)) \tag{3.84}$$

ce qui définit le modèle adjoint :

$$M^T\psi = -c^\star(0) \tag{3.85}$$

Finalement, on a $\nabla_u J = -c^\star(0)$ avec :

$$\frac{dc^\star}{dt} = -\frac{\partial f}{\partial c}c^\star + H^T(obs - Hc_u)\ , \quad c^\star(T) = 0 \tag{3.86}$$

3.3.2.4 Modèle adjoint et lagrangien

Une manière élégante d'introduire l'état adjoint est de définir le lagrangien :

$$L(u, c, \hat{c}) \triangleq \tilde{J}(c) + \int_0^T \left(\hat{c}, \frac{dc}{dt} - f(c)\right) dt + (c(0) - u, \hat{c}(0)) \tag{3.87}$$

avec $\hat{c}$ le multiplicateur de Lagrange. Ceci revient à pénaliser la fonction coût par les contraintes que sont le modèle dynamique et la condition initiale.

Remarquons ensuite que si c satisfait le modèle, $L(u, c, \hat{c}) = J(u)$, et que l'on a alors le long de cette contrainte :

$$\nabla_u J = \frac{\partial L}{\partial u} + \frac{\partial L}{\partial c}\frac{\partial c}{\partial u} \tag{3.88}$$

Le premier terme est aisé à calculer : $\frac{\partial L}{\partial u} = -\hat{c}(0)$.

Imposons à présent pour $\hat{c}$:

$$\frac{\partial L}{\partial c} = 0 \tag{3.89}$$

La clé est que l'on a aisément par intégration par partie :

$$L(u, c, \hat{c}) = \tilde{J}(c) - \int_0^T (c, \frac{d\hat{c}}{dt})dt - \int_0^T (\hat{c}, f(c))dt + (\hat{c}(T), c(T)) - (u, \hat{c}(0)) \tag{3.90}$$

et la contrainte devient si on impose de plus $\hat{c}(T) = 0$:

$$\frac{\partial L}{\partial c} = 0 = \nabla\tilde{J} - \frac{d\hat{c}}{dt} - (\frac{\partial f}{\partial t})^T\hat{c} \tag{3.91}$$

Dans ces conditions, $\nabla_u J = \frac{\partial L}{\partial u} = -\hat{c}(0)$.

Notons que la variable adjointe $c^\star$ s'avère être exactement le multiplicateur de Lagrange $\hat{c}$.

3.3.2.5 Autre forme de pénalisation

On n'a pas parlé de la pénalisation de la fonction coût qui peut être effectuée avec une ébauche ou, de manière un peu moins classique, par une régularisation sur des dérivées de l'état recherché ([Lin *et al.*, 2002]).

3.3.2.6 Contrainte faible et erreur modèle

L'extension au cas d'une erreur modèle revient à prendre comme variable de contrôle les états successifs c_i ou de manière analogue les erreurs modèles q_i définies par $c_{i+1} = F_i(c_i) + q_i$. La fonction à contrôler est alors de la forme :

$$J(c_1, \ldots, c_{n_o}) \tag{3.92}$$

Ceci revient à augmenter considérablement la taille de l'espace de contrôle et la méthode est en général non utilisable en l'état. Les deux alternatives habituellement utilisées sont :

- une modélisation de l'erreur modèle sur la base par exemple d'une étude théorique des processus non résolus (paramétrisation sous-maille) ;
- une réduction de l'espace de contrôle des états.
  Une approche couramment utilisée, notamment en océanographie, est fondée sur l'utilisation des bases d'EOFs, définies comme les modes les plus importants pour l'information statistique du modèle. On projette alors les états (ou l'erreur modèle dans une base réduite), par exemple selon :

$$c_i = \sum_{j=1}^{j=n_{eof}} \lambda_i^j \pi_j^{eof} \tag{3.93}$$

  avec $(\pi_1^{eof} \ldots \pi_{n_{eof}}^{eof})$ une base prédéterminée. Bien entendu, $n_{eof} \ll n$ pour qu'il y ait un intérêt. La fonction de contrôle se restreint alors (par abus de notation) à :

$$J(\lambda_1, \ldots, \lambda_{n_{eof}}) \tag{3.94}$$

  On se réfère par exemple à [Durbiano, 2001] pour une étude méthodologique et l'application à un modèle quasi-géostrophique en océanographie.

3.3.2.7 Codage pratique de l'adjoint

En pratique, le modèle adjoint doit être construit à partir du code décrivant le modèle direct. L'approche usuelle est de construire l'adjoint du code associé au modèle direct (et non pas le code associé au modèle adjoint) pour garantir une cohérence entre le calcul du modèle direct et de la fonction coût et le calcul du gradient (voir chapitre 4.

Le code adjoint peut être soit écrit "à la main" soit généré de manière automatique [†3] par, ce que l'on appelle, des outils de différenciation automatique (comme par exemple le logiciel Odyssée de l'INRIA).

Prenons par exemple le modèle simplissime (une instruction informatique) :
J=X*X+X*Y

Le modèle linéaire tangent associé est de la forme :
DJTL=2*X*DXTL+X*DYTL+Y*DXTL
et le gradient ∇J peut être calculé composante par composante en prenant successivement comme jeu de conditions initiales pour les perturbations (DXTL,DYTL) (1,0) et (0,1), et en récupérant la valeur dans DJTL. Ceci nécessite donc deux utilisations du code linéaire tangent.

Le modèle adjoint est construit en considérant la relation :

$$\begin{bmatrix} x \\ y \\ J \end{bmatrix} \rightarrow \begin{bmatrix} x \\ y \\ x*x+x*y \end{bmatrix} \tag{3.95}$$

et en appliquant la transposée de la matrice jacobienne de cette transformation au vecteur de variables *adjointes* (DXCL,DYCL,DZCL), ce qui donne comme code adjoint :

DXCL=DXCL+(2X+Y)*DZCL
DYCL=DYCL+X*DZCL

Pour une initialisation DZCL=1, on récupère donc en une seule utilisation du code adjoint le gradient ∇J.

Notons pour terminer cette partie que l'implémentation d'un adjoint doit aussi tenir compte du caractère spécifique des parties linéaires du modèle (qui ne sont pas à dériver mais simplement à transposer).

3.4 Conditionnement et qualité du résultat

Un point clé est bien entendu de pouvoir estimer la confiance accordée au résultat (l'équivalent de la matrice de covariance d'erreur d'analyse pour les approches bayésiennes).

[†3] Toute chose étant égale par ailleurs, le code direct devant être écrit de manière propre pour ce qui concerne la structure entrée/sortie.

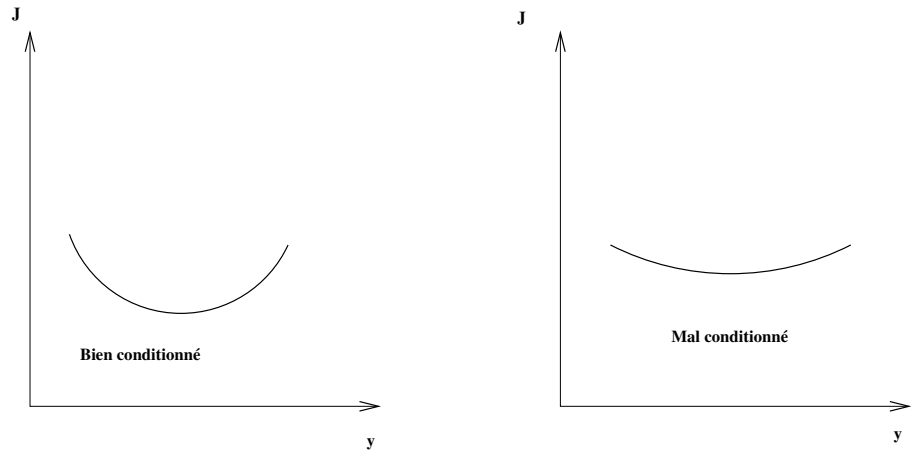

FIG. 3.6 – Conditionnement à l'inversion.

Prenons un exemple simple de fonction coût en revenant à une approche de type 3D-var avec $\tilde{J}(c) = \frac{1}{2}\, c^T Ac - b^T c$, A une matrice symétrique définie positive nxn et $b \in \mathbf{R}^n$. La condition d'optimalité s'écrit bien évidemment :

$$Ac_a = b \tag{3.96}$$

et la "qualité" du résultat va dépendre de la sensibilité de c_a à des erreurs sur b ou/et sur A. Un résultat classique d'algèbre linéaire indique que ceci est lié au conditionnement de la matrice A pour l'inversion, défini par exemple pour la norme euclidienne comme :

$$\kappa_2(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)} \tag{3.97}$$

avec $\lambda_{max}(A)$ et $\lambda_{min}(A)$ les valeurs propres respectivement maximale et minimale de A. Bien sûr, si on avait diagonalisé, on aurait obtenu composante par composante un système du type :

$$\lambda_i (c_a)_i = b_i \tag{3.98}$$

avec λ_i une valeur propre de A et il est clair que l'inversion est mal conditionnée pour les petites valeurs propres (figure 3.6).

L'extension au cas général se fait (après avoir noté que A est dans l'exemple précédent la matrice hessienne de J) avec l'analyse spectrale de la matrice hessienne de J.

De manière générale, le conditionnement à l'inversion de J est déterminé par :

- le modèle et sa dynamique en espace et en temps ;
- le positionnement du réseau de capteurs d'observation.

3.4.1 Conditionnement et dynamique du modèle : systèmes stables versus systèmes chaotiques

Il convient de bien distinguer les situations au regard de l'assimilation de donnée suivant le comportement dynamique des systèmes considérés. On va pour schématiser distinguer deux situations :

- le cas de la météorologie pour lequel le problème est essentiellement un problème aux conditions initiales.
- le cas de la chimie atmosphérique qui est essentiellement un problème de conditions aux limites, les conditions initiales étant très rapidement "oubliées".

3.4.1.1 Systèmes aux conditions initiales

L'exemple type ([Daley, 1991]) est fourni par le modèle $c_{i+1} = a\, c_i - c_i^2$ pour lequel l'horizon de prévisibilité est limité, la solution présentant une très forte dépendance aux valeurs des paramètres physiques (a) et à la condition initiale c_0 (figure 3.7).

Cet exemple est emblématique de la situation rencontrée en météorologie. L'assimilation de données sur les conditions initiales est bien entendu indispensable pour éviter que les erreurs sur les modes les plus instables ne perturbent trop la qualité de la solution.

3.4.1.2 Systèmes "stables"

Un autre exemple simple est donné par l'équation de la chaleur.

On se donne un corps Ω, supposé infini (par exemple une barre) et on mesure sa température $c(x,t)$ pour un certain temps $t > 0$ en toute position x. L'objectif est de retrouver la température à l'instant initial $c(x,0)$. Bien entendu :

$$\frac{\partial c}{\partial t} = k\Delta c \tag{3.99}$$

avec k la conductivité thermique de la barre. Si $c(x,0) = \sum a_n\, sin(nx)$, on a aisément :

$$c(x,t) = \sum a_n\, sin(nx) exp(-n^2 t) \tag{3.100}$$

Les coefficients a_n s'obtiennent par une analyse de Fourier :

$$a_n = \frac{2}{\Pi} \int_0^{\Pi} c(x,0) sin(nx) dx \tag{3.101}$$

Les coefficients $\tilde{a}_n$ de la solution à t s'obtiennent à partir de a_n par :

$$\tilde{a}_n = exp(-n^2 t) a_n \tag{3.102}$$

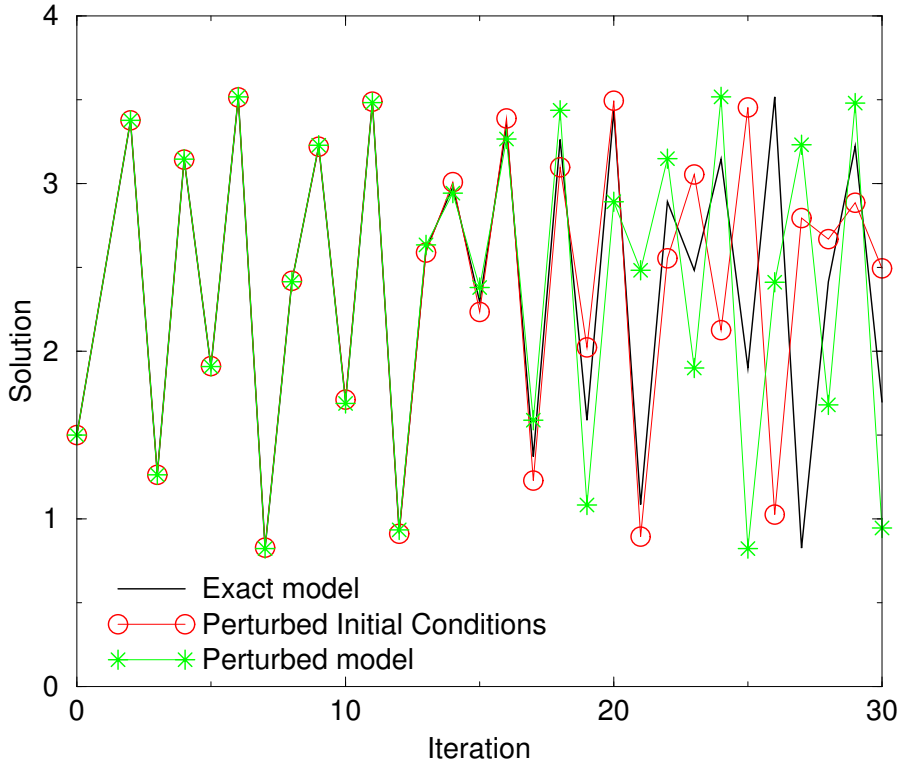

FIG. 3.7 – Instabilité du système $c_{i+1} = a\, c_i - c_i^2$ pour $c_0 = 1.5$ et $a = 3.75$. Les perturbations sont de 0.001.

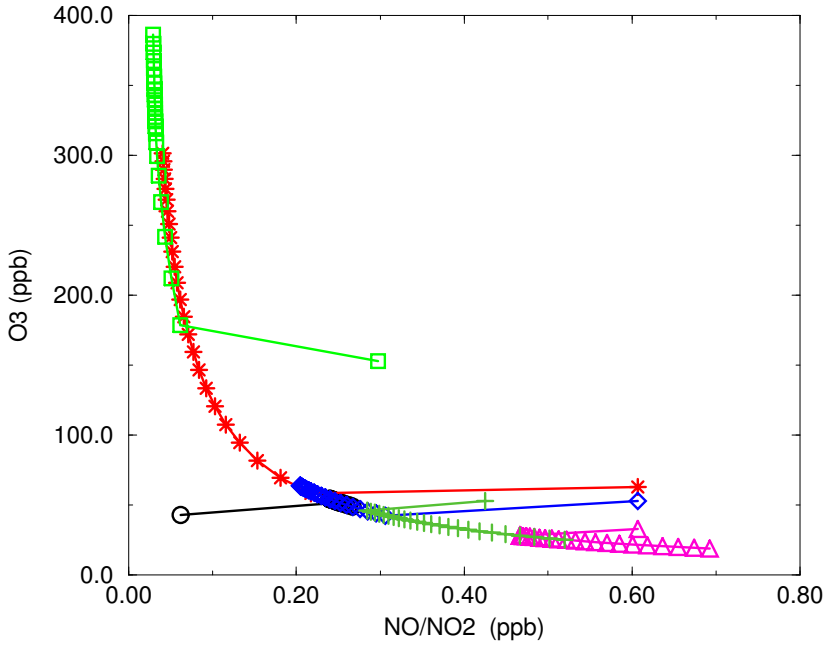

FIG. 3.8 – Systèmes lent/rapide pour le cycle de Chapman troposphérique. Convergence de plusieurs trajectoires vers l'équilibre photostationnaire.

Si l'on étudie l'évolution de la température, ceci signifie ($t > 0$) qu'une erreur sur le mode n va s'atténuer en temps et ce d'autant plus pour les grands modes. A l'inverse, si l'on observe la température à t, et que l'on cherche à calculer la température initiale, les erreurs vont être amplifiées et ce d'autant plus pour les grands modes puisque :

$$\delta a_n = exp(n^2 t)\delta\tilde{a}_n \tag{3.103}$$

Ceci est bien entendu lié au caractère *dissipatif* de l'équation de la chaleur.

Le cas de la cinétique chimique est similaire. Le meilleur exemple est fourni par les modèles de type lent/rapide caractéristiques du comportement dynamique de la chimie atmosphérique (figure 3.8). Il est clair que la dépendance aux conditions initiales des espèces rapides est nulle, autrement dit que des erreurs sur ces conditions initiales n'auront pas d'incidence ultérieure (la contrepartie est aussi que l'on ne sera pas capable de remonter à ces concentrations). On se réfère plus généralement à [Sportisse, 1999] pour l'étude du comportement dynamique de la cinétique chimique, notamment atmosphérique.

On renvoie également au chapitre B pour une analyse du comportement des systèmes dissipatifs de type lent/rapide comme ceux que l'on rencontre en chimie atmosphérique.

3.4.2 Network design

L'influence du réseau d'observation dans la qualité de l'assimilation de données est claire et est directement reliée au conditionnement à l'inversion.

De manière générale, la matrice hessienne dépend de l'opérateur d'observation H. Les méthodes de network design, sous leurs diverses déclinaisons, reviennent à optimiser un critère de bon conditionnement à l'inversion du hessien. le cadre général est fourni par la théorie de l'optimal design, classiquement utilisé en régression ([Fedorov et Hackl, 1997]) et les applications classiques à l'assimilation de donnée vont des "adaptive observations" (utilisées en météorologie pour le déploiement en temps réel de moyens de mesures supplémentaires à un réseau fixe de mesures) aux techniques de sensibilité.

On renvoie au chapitre A pour une synthèse méthodologique des techniques habituellement utilisées.

3.4.3 Sensibilité de second-ordre

L'assimilation de données conduit à la minimisation d'un système sous la forme suivante :

$$\min_u J(u, p, obs, h) \tag{3.104}$$

où l'on a, par abus de notation, désigné la dépendance de la fonction coût aux paramètres à estimer (le cadre précédent) mais également à des paramètres non estimés mais jugés incertains (p), aux valeurs des observations (obs) et à des paramètres du réseau de mesure (h).

Dans le cas de la chimie atmosphérique, les paramètres estimés sont typiquement les concentrations ou des flux de forçage (émissions), les paramètres incertains sont notamment les constantes de paramétrisation utilisées pour les phénomènes sous-mailles (la résolution horizontale d'une maille d'un modèle de Chimie-Transport est typiquement, à l'échelle continentale, de l'ordre de quelques dizaines de kilomètres), h est une densité spatio-temporelle du réseau de mesures.

La solution au problème de minimisation (3.104) dépend, par le théorème des fonctions implicites, des autres paramètres (p, obs, h) :

$$u_a(p, obs, h) = \text{argmin}_u J(u, p, obs, h) \tag{3.105}$$

Un point essentiel est bien entendu, pour estimer la qualité des résultats, d'estimer la dépendance de u_a en ces paramètres non estimés. Cette sensibilité est donnée, par exemple

pour les paramètres incertains p, par ce que l'on appelle la *sensibilité du second ordre* que l'on a directement par :

$$\frac{\partial u_a}{\partial p} = -(\mathrm{Hess}_u J)^{-1} \frac{\partial^2 J}{\partial p \partial u} \tag{3.106}$$

Ceci montre le caractère clé joué par la matrice hessienne de J. On renvoie par exemple à [Le Dimet *et al.*, 2002, Wang *et al.*, 1992, Ngodock, 1996] pour plus de détails, notamment numériques.

L'implémentation numérique des sensibilités de second-ordre, dans un cadre variationnel, conduit à "dériver une seconde fois" les modèles différentiés. L'approche est donc particulièrement lourde pour des codes tridimensionnels complexes même si elle reste faisable pour des applications académiques (le code POLAIR3D a un hessien différentié par exemple). On se réfère au chapitre C pour une application à un exemple simple dans le cadre de la modélisation inverse de traceurs radioactifs à petite échelle (ie donnés par des modèles simples de type gaussiens).

Chapitre 4

Modèles linéaire tangent et adjoint de POLAIR3D

Résumé

Les modèles adjoint et linéaire tangent associés à un modèle physique direct sont les composantes indispensables d'une approche variationnelle. L'objectif de ce chapitre est de faire le point sur les difficultés rencontrées pour les modèles de Chimie-Transport et les choix retenus pour le modèle POLAIR3D dont les modèles dérivés sont obtenus par différenciation automatique.

On détaille notamment les approches choisies numériquement et les différences entre les approches discrète et continue. On présente enfin quelques applications à des études de sensibilité.

Sommaire

4.1 Equations dérivées continues pour les modèles de dispersion réactive 106
4.2 Approche continue versus approche discrète 108
4.2.1 Cas d'une équation différentielle : application à la cinétique chimique 109
4.2.2 Cas d'un modèle d'advection-diffusion 111
4.3 Algorithmes mis en œuvre dans POLAIR3D 113
4.3.1 Approche retenue 113
4.3.2 Différentiation automatique et contraintes de développement . . 113
4.3.3 Le préprocesseur SPACK 116
4.3.4 Boucle en temps et implémentation pratique dans POLAIR3D . 117
4.4 Application à des tests de sensibilité 118
4.4.1 Chimie diphasique 119

4.4.2 Calcul 3D ESQUIF 121

Introduction

Les calculs de sensibilité aux paramètres d'entrée du modèle (conditions initiales, paramétrisations physiques, flux -émissions, ...-) peuvent passer par le calcul des dérivées des sorties du modèle par rapport à ses entrées. Considérons le modèle, une fois discrétisé en espace (c'est le cas en pratique pour les modèles de Chimie-Transport comme POLAIR3D qui adoptent une "méthode des lignes" pour la discrétisation en espace puis la résolution numérique d'une équation différentielle écrit sous la forme :

$$\frac{dc}{dt} = f(t, c, k) \ , \quad c(0) = c_0 \tag{4.1}$$

avec $c \in \mathrm{R^n}$ le vecteur des variables d'état (en pratique les champs 3D d'espèces chimiques), $k \in \mathrm{R^m}$ le vecteur des paramètres et c_0 la condition initiale.

Un calcul de sensibilité locale aux paramètres k revient par exemple à estimer la matrice $n \times m$:

$$S(t) = \frac{\partial c(t)}{\partial k} \tag{4.2}$$

Evidemment, une "grande valeur" de la colonne S_i correspond à une composante des paramètres dont l'influence est *localement* (autour du point de fonctionnement courant dans l'espace des paramètres) importante pour les sorties du modèle.

Pour la modélisation inverse par des approches variationnelles, la sensibilité doit être estimée pour des fonctions scalaires appliquées à la sortie du modèle (une fonction coût, notée $J(c(T))$ en cohérence avec les chapitres précédents) :

$$\nabla_k J = S(T)^T \nabla_c J(T) \tag{4.3}$$

Les deux approches utilisées couramment pour calculer ces sensibilités sont :
- par le modèle linéaire tangent (on parle aussi souvent de mode "forward"),
- ou par le modèle adjoint (on parle aussi de mode "reverse").

Un premier choix est à faire entre l'utilisation de la discrétisation numérique des modèles "dérivés" associés aux équations continues sous-jacentes au modèle physique (on parlera par la suite d'*approche continue*) et celle des modèles "dérivés" associés aux modèles discrétisés (on parlera d'*approche discrète*). La seconde approche est en général préférée car elle permet notamment, dans le contexte de la modélisation inverse, d'avoir une compatibilité entre les fonctions coût calculées et les gradients associés. C'est cette approche que nous avons choisie pour POLAIR3D.

La mise en œuvre pratique de cette approche peut nécessiter des développements lourds pour de "gros" modèles tridimensionnels. Le codage en dur ("à la main") de ces

modèles dérivés à partir du modèle physique initial est une approche encore largement utilisée. Une alternative est l'utilisation d'outils de différenciation, comme Odyssée ([Faure et Papegay, 1997]), Tapenade, TAMC ([Giering et Kaminski, 1998]), ADIFOR ([Bischof *et al.*, 1992, He et al, 2000]), ..., permettant de générer des codes de manière automatique à partir d'un code source. C'est cette approche que nous avons choisie afin de pouvoir garantir la pérennité de la plate-forme de calcul POLAIR3D et l'obtention aisée de ses codes périphériques (dont les modèles linéaires tangents et adjoints). Ceci ne va pas néanmoins sans un certain nombre de contraintes sur l'écriture du modèle direct, qui doit essentiellement être très précis sur les relations fonctionnelles entrées/sorties.

Ce chapitre, à portée largement "technologique" (mais clé pour l'ensemble de ce travail), est organisé de la manière suivante. On rappelle rapidement dans la section 1, sur des exemples simples, la construction des modèles linéaire tangent et adjoint pour les modèles de Chimie-Transport. La problématique des approches continue et discrète est esquissée en section 2. Les algorithmes effectivement implémentés dans POLAIR3D sont présentés en section 3. Le couplage au préprocesseur chimique SPACK (Simplified Preprocessor for Atmospheric Chemical Kinetics, [Djouad *et al.*, 2002a]) permet une génération automatique rapide des modèles dérivés associés à une cinétique chimique quelconque (gazeuse et diphasique eau/gaz). Enfin, des tests d'illustration sont présentés en section 4.

4.1 Equations dérivées continues pour les modèles de dispersion réactive

On part de l'équation de dispersion réactive des espèces chimiques, en supposant que la densité de l'air est constante :

$$\frac{\partial c_i}{\partial t} + \mathrm{div}(V\, c_i) = \mathrm{div}(K \nabla c_i) + \chi_i(c) - D_i c_i + E_i \ , \quad c(0) = c_0 \tag{4.4}$$

où pour plus de simplicité, on a incorporé les conditions aux limites de dépôt (D_i est une vitesse de dépôt) et d'émissions (E_i) dans les équations de champs.

Par exemple, une perturbation δc_0 sur les conditions initiales et δE sur les émissions va se propager selon l'équation linéarisée :

$$\frac{\partial \delta c}{\partial t} + \mathrm{div}(V \delta c) = \mathrm{div}(K \nabla \delta c) + F(c)\delta c - D\delta c + \delta E \ , \quad \delta c(0) = \delta c_0 \tag{4.5}$$

avec D une matrice diagonale formée à partir des vitesses de dépôt et F la matrice jacobienne $\partial\chi/\partial c$ associée aux termes de cinétique chimique [†1]. Ce modèle est la version continue du modèle linéaire tangent associé à l'équation de dispersion réactive.

[†1] Notée habituellement J, ce qui est ambigu dans le contexte de l'assimilation de données.

Considérons à présent une fonction coût J définie à partir de la sortie du modèle c. En toute généralité on prend J définie sur un intervalle d'intégration en temps $[0, T]$:

$$J(c) = \int_0^T j(c(t))dt \tag{4.6}$$

La perturbation induite sur J est donnée au premier ordre par :

$$\delta J = J(c + \delta c) - J(c) = \int_0^T (\nabla_c j, \delta c(t))dt \tag{4.7}$$

Cette équation définit $\nabla_{E(t)} J$ et $\nabla_{c_0} J$ selon :

$$\int_0^T (\nabla_c j, \delta c(t))dt = (\nabla_{c_0} J, \delta c_0) + \int_0^T (\delta E(t), \nabla_{E(t)} J)dt \tag{4.8}$$

En faisant le produit scalaire L^2 en espace de l'équation d'évolution de la perturbation $\delta c(t)$ par la variable adjointe $c^\star(t)$ (définie ci-après) et en intégrant en temps, on obtient aisément après intégration par partie) :

$$\int_0^T \left(-\frac{\partial c^\star}{\partial t} - V\mathrm{div}(c^\star) - \mathrm{div}(K\nabla c^\star) - F^T(c)c^\star + Dc^\star, \delta c(t) \right) dt \tag{4.9}$$

$$+ (c^\star(T), \delta c(T)) = (c^\star(0), \delta c_0) + \int_0^T (\delta E(t), c^\star(t))dt \tag{4.10}$$

En prenant comme équation adjointe :

$$-\frac{\partial c^\star}{\partial t} - V\mathrm{div}(c^\star) = \mathrm{div}(K\nabla c^\star) + F^T(c)c^\star - Dc^\star + \nabla_c j \ , \quad c^\star(T) = 0 \tag{4.11}$$

on a directement par identification :

$$\nabla_{c_0} J = c^\star(0) \ , \quad \nabla_{E(t)} J = c^\star(t) \tag{4.12}$$

Notons que pour le cas d'une densité de l'air constante (ρ), le champ de vent vérifie $\mathrm{div}V = 0$, et on se ramène aisément à :

$$-\frac{\partial c^\star}{\partial t} - \mathrm{div}(Vc^\star) = \mathrm{div}(K\nabla c^\star) + F^T(c)c^\star - Dc^\star + \nabla_c j \ , \quad c^\star(T) = 0 \tag{4.13}$$

Il est à noter que, dans le cas d'un modèle de traceurs passifs ou linéaires, F est alors une matrice diagonale, éventuellement nulle, et $F^T = F$. Pour le modèle continu, les variables adjointes sont aisément obtenues à l'aide du modèle initial en inversant le signe du temps (en pratique les chronologies de champs de forçage) et le signe des champs de vent et en prescrivant comme terme d'émission $\nabla_c j$.

Lorsque la densité de l'air (ρ) n'est pas constante, l'équation de dispersion de départ doit être réécrite, le terme de diffusion devenant :

$$\mathrm{div}(\rho\, K\nabla \frac{c}{\rho}) \tag{4.14}$$

Le produit scalaire définissant le cadre fonctionnel sous-jacent doit être redéfini afin de pouvoir avoir une propriété du même type. Si on néglige les termes de dérivées en temps dans l'équation de continuité ($\mathrm{div}(\rho V) = 0$), il est relativement direct de vérifier que le "bon" produit scalaire est alors pour les concentrations :

$$(\phi, \psi) = \int \frac{\phi(x)\psi(x)}{\rho} dx \tag{4.15}$$

permettant d'obtenir de nouveau pour l'équation adjointe (4.13). De manière similaire, on aurait pu travailler en fraction massique $x_c = c/\rho$ (le terme de diffusion devenant alors $\mathrm{div}(\rho K \nabla x_c)/\rho$) avec le produit scalaire :

$$(\phi, \psi) = \int \rho(x)\phi(x)\psi(x) dx \tag{4.16}$$

On se réfère à [Issartel et Baverel, 2002, Roustan et Bocquet, 2006] pour plus de détails et l'application aux méthodes de rétropanaches.

4.2 Approche continue versus approche discrète

Dans le cas d'une cinétique chimique non-linéaire, sur la base des équations précédentes, une approche serait de discrétiser les équations continues précédentes pour obtenir les modèles linéaire tangent et adjoint (on parlera d'*approche continue*, le modèle dérivé étant obtenu pour un modèle continu). Cette approche est rarement utilisée en pratique car elle ne permet pas de garantir une compatibilité entre la fonction coût calculée numériquement (via les schémas numériques associés au calcul du modèle direct) et les valeurs des gradients.

Une approche alternative est d'utiliser les modèles dérivés associés aux modèles directs discrétisés (on parlera par la suite d'*approche discrète*).

Dans [Sirkes et Tziperman, 1997], les différences entre ces deux approches sont étudiées, notamment sur la nécessité d'avoir recours à des schémas numériques appropriés pour l'approche continue. Par exemple, dans le cas de la cinétique chimique, la "chimie adjointe" garde fondamentalement des propriétés similaires au modèle direct : le modèle est raide (avec une grande dispersion des échelles de temps, les valeurs propres de F et de F^T), ce qui conduira à privilégier des algorithmes d'intégration en temps implicites.

Un point clé est la non-symétrie en temps (et en espace pour des modèles de type advection-diffusion) des algorithmes numériques utilisés qui conduit à des différences entre les approches discrète et continue.

4.2.1 Cas d'une équation différentielle : application à la cinétique chimique

Pour illustrer les différences entre les deux approches, on a choisi d'étudier le cas d'une cinétique chimique donnée par un modèle écrit sous la forme :

$$\frac{dc}{dt} = f(c,k) \ , \quad c(0) = c_0 \tag{4.17}$$

pour l'intervalle de temps $[0,T]$.

Le paramètre k contient les paramètres du modèle (éventuellement c'est d'ailleurs une fonction du temps $k(t)$). L'ensemble des variables par rapport auxquelles on souhaite dériver est alors donné par $u = (k, c_0)$ (ce qui serait la variable de contrôle dans un contexte de modélisation inverse/assimilation de données).

Afin de bien noter la dépendance aux paramètres, on note $c(t,u)$ la concentration c au temps t. Le calcul va être identique au calcul précédent avec une dépendance en les paramètres k quelconque.

On se donne dans un premier temps une fonction $J(c(T,u))$. Une perturbation $\delta u = (\delta k, \delta c_0)$ conduit à une perturbation :

$$\delta J = J(c(T,u+\delta u)) - J(c(T,u)) \simeq (\nabla_c J, \delta c(T,u)) \tag{4.18}$$

Or les perturbations de c vérifient le modèle d'évolution linéarisé :

$$\frac{d\delta c}{dt} = F\delta c + \frac{\partial f}{\partial k}\delta k \ , \quad \delta c(0) = \delta c_0 \tag{4.19}$$

avec $F = \partial f/\partial c$ le jacobien.

De la même manière que précédemment, en intégrant sur $[0,T]$ après multiplication par la variable adjointe $c^\star(t)$ (sans plus de précision à ce niveau), on obtient alors :

$$(c^\star(T), \delta c(T,u)) - (c^\star(0), \delta c_0) = \int_0^T \left[(\frac{dc^\star}{dt} + F^T\, c^\star, \delta c) + (\frac{\partial f}{\partial k}^T\, c^\star, \delta k) \right] dt \tag{4.20}$$

Si le modèle adjoint est défini par :

$$\frac{dc^\star}{dt} = -F^T\, c^\star \ , \quad c^\star(T) = \nabla_c J \tag{4.21}$$

on a directement :

$$\delta J = (\nabla_c J, \delta c(T,u)) = (c^\star(0), \delta c_0) + \int_0^T (\frac{\partial f}{\partial k}^T\, c^\star, \delta k) dt \tag{4.22}$$

ce qui montre que :

$$\nabla_{c_0} J = c^\star(0) \ , \quad \nabla_k J = \int_0^T \frac{\partial f}{\partial k}^T c^\star dt \tag{4.23}$$

De manière directe, si k dépend du temps et si l'on cherche à calculer la sensibilité par rapport à un état du modèle à t (et non plus à la condition initiale), on obtient :

$$\nabla_{c(t,u)} J = c^\star(t) \ , \quad \nabla_{k(t)} J = \frac{\partial f}{\partial k}^T c^\star(t) \tag{4.24}$$

Un point clé est donc, dans tous les cas, l'intégration du système (4.21).

Utilisons à présent une méthode d'intégration implicite, en l'occurrence le schéma d'Euler implicite linéarisé appliqué à une EDO non autonome. L'extension à d'autres schémas (d'ordre plus élevé) est directe. L'intégration du modèle direct (4.17) s'écrit sous la forme :

$$\frac{c_{n+1} - c_n}{\Delta t} = f(c_n, k_n) + J_n(c_n)(c_{n+1} - c_n) \tag{4.25}$$

avec c_n et k_n les approximations au temps t_n de c et k. On a noté $J_n(c_n)$ une approximation numérique du jacobien pour marquer sa dépendance en l'état pour les cas non linéaires. Δt est le pas de temps supposé constant.

Dans le cas linéaire $f(c_n, k_n) = F_n\, c_n$ et le passage de c_n à c_{n+1} peut donc s'écrire sous la forme $c_{n+1} = M_n c_n$. Il est alors trivial de vérifier que le schéma itératif adjoint associé est :

$$\tilde{c}_n^\star = {M_n}^T \tilde{c}_{n+1}^\star \tag{4.26}$$

On a directement après quelques manipulations algébriques :

$$\tilde{c}_n^\star = (I + (I - J_n \Delta t)^{-1} F_n \Delta t)^T \tilde{c}_{n+1}^\star \tag{4.27}$$

Ce schéma correspond à la version adjointe du modèle discrétisé (approche discrète).

Considérons à présent le modèle adjoint continu. Il s'écrit :

$$\frac{dc^\star}{dt} = -F^T c^\star \tag{4.28}$$

Une discrétisation par le schéma d'Euler implicite linéarisé (si les équations sont raides, le modèle adjoint garde cette propriété) conduit à :

$$\frac{c_n^\star - c_{n+1}^\star}{\Delta t} = F_{n+1}^T c_{n+1}^\star + J_{n+1}^T \left(c_n^\star - c_{n+1}^\star\right) \tag{4.29}$$

On obtient alors aisément le schéma itératif suivant :

$$c_n^\star = (I + (I - J_{n+1} \Delta t)^{-1} F_{n+1} \Delta t)^T {c^\star}_{n+1} \tag{4.30}$$

Ceci constitue la version discrétisée du modèle adjoint continu. On remarque bien entendu que le schéma est similaire à celui obtenu par l'approche discrète si ce n'est que la non autonomie est prise en compte dans un cas de manière explicite pour le schéma rétrograde (F_{n+1}, approche continue), de manière implicite dans l'autre cas (F_n, approche discrète).

Dans le cas d'un système non-autonome (même linéaire), les deux approches ne sont donc pas équivalentes, du fait de la non-symétrie en temps induit par la discrétisation numérique. Pour le schéma numérique utilisé ici pour cette illustration (Euler implicite linéarisé), il est probable que l'approche discrète conviendra mieux à des systèmes avec de forts gradients en temps des constantes de forçage (implicite). Si le schéma numérique initial avait été le schéma d'Euler implicite sans linéarisation, les résultats auraient été strictement inversés et le comportement en terme de stabilité des deux approches est donc à étudier au coup par coup.

Une autre implication est bien entendu que même si le modèle continu est auto-adjoint ($F = F^T$), cette propriété n'est plus conservée numériquement (du fait du défaut de non-autonomie). On peut par exemple se référer à [Roustan et Bocquet, 2006] pour une analyse plus détaillée dans le cas du modèle POLAIR3D.

4.2.2 Cas d'un modèle d'advection-diffusion

On se réfère à [Sei et Symes, 1995, Sirkes et Tziperman, 1997] pour l'étude approfondie d'une équation d'advection-diffusion. Un point clé est l'étude de stabilité des schémas numériques d'advection, pour les équations adjointes discrètes ou continues.

On se propose ici d'illustrer le comportement en terme de *consistance* des deux approches numériques. On considère une équation d'évolution linéaire continue du type :

$$\frac{\partial c}{\partial t} + Mc = S(k) \tag{4.31}$$

avec $c \rightarrow Mc$ un opérateur de type advection-diffusion et $S(k)$ un terme source fonction des paramètres k. Un point clé est l'étude numérique des modèles adjoints associés à M.

L'équivalent, pour ce cas, du comportement non-symétrique en temps présenté sur l'exemple précédent (schémas implicite/explicite), est par exemple la non-symétrie en espace d'une approche de type upwind.

Avec des notations standards en analyse numérique, on note V l'espace fonctionnel sur lequel c est défini (typiquement L^2) et V_h un sous-espace de discrétisation (h est typiquement un paramètre de discrétisation qui va tendre vers 0). Soit Π_h l'opérateur de

discrétisation de V dans V_h et M_h l'opérateur discret associé à M défini sur V_h (le schéma numérique appliqué au problème discrétisé).

Une propriété cruciale du schéma numérique choisi est la propriété de consistance qui assure que l'erreur locale tend vers 0 lorsque le pas de discrétisation h tend vers 0. Si on utilise le diagramme suivant, dans lequel une flèche horizontale correspond à l'application de l'opérateur et une flèche verticale à une discrétisation :

$$\begin{array}{lcl} c \in V & \longrightarrow & Mc \in V \\ \downarrow & & \downarrow \\ & & \Pi_h M c \in V_h \\ c_h = \Pi_h c \in V_h & \longrightarrow & M_h \Pi_h c \in V_h \end{array} \tag{4.32}$$

le schéma est dit consistant si on a asymptotiquement la propriété de commutation :

$$M_h \Pi_h - \Pi_h M \stackrel{h \to 0}{\longrightarrow} 0 \tag{4.33}$$

L'équivalent pour le cas du modèle adjoint M^T (défini en fonction du produit scalaire sur V) n'a pas de raison d'être vérifié, autrement dit on n'a pas nécessairement la propriété :

$$M_h^T \Pi_h - \Pi_h M^T \stackrel{h \to 0}{\longrightarrow} 0 \tag{4.34}$$

Un bon exemple est fourni par l'équation d'advection discrétisée par un schéma de type "upwind". On considère le cas scalaire ($x \in \mathbf{R}$ est une coordonnée d'espace) :

$$\frac{\partial c}{\partial t} + V(x) \frac{\partial c}{\partial x} = S(k) \tag{4.35}$$

avec $V(x)$ un champ de vent. On ne traite pas ici les conditions aux limites (le développement étant largement formel). On a pris $(Mc)(x) = V(x) \frac{\partial c}{\partial x}$. Le schéma upwind est alors défini par :

$$\frac{c_i^{n+1} - c_i^n}{\Delta t} + V_i^+ \frac{c_i^n - c_{i-1}^n}{\Delta x} + V_i^- \frac{c_{i+1}^n - c_i^n}{\Delta x} = S(k) \tag{4.36}$$

avec les notations classiques $V_i^+ = \max(0, V(x_i))$ et $V_i^- = \min(0, V(x_i))$.

L'équation directe continue est directement donnée par :

$$-\frac{\partial c^\star}{\partial t} - \frac{\partial (V(x) c^\star)}{\partial x} = 0 \tag{4.37}$$

Un calcul fastidieux d'intégration par partie discrète permet de se convaincre que l'adjoint discret obtenu à partir de (4.36) est :

$$-\frac{(c^\star)_i^{n+1} - (c^\star)_i^n}{\Delta t} - \frac{V_{i+1}^+ (c^\star)_{i+1}^{n+1} - V_i^+ (c^\star)_i^{n+1}}{\Delta - x} - \frac{V_i^- (c^\star)_i^{n+1} - V_{i-1}^- (c^\star)_{i-1}^{n+1}}{\Delta - x} = 0 \tag{4.38}$$

Si le champs de vitesse $V(x)$ s'annule dans la cellule i (ce qui implique $V_i^+ = V_i^- = 0$), on note alors directement que :

$$\frac{V_{i+1}^+ (c^\star)_{i+1}^{n+1} - V_i^+ (c^\star)_i^{n+1}}{\Delta x} + \frac{V_i^- (c^\star)_i^{n+1} - V_{i-1}^- (c^\star)_{i-1}^{n+1}}{\Delta x} \simeq 2 \frac{\partial (V(x) c^\star)}{\partial x} \tag{4.39}$$

ce qui prouve la non-consistance avec l'équation adjointe continue (4.37).

4.3 Algorithmes mis en œuvre dans POLAIR3D

4.3.1 Approche retenue

En pratique, on a choisi dans POLAIR3D d'utiliser une approche discrète (ce sont les schémas numériques et leur implémentation sous forme de code qui sont différenciés). Il y a essentiellement deux raisons qui ont motivé ce choix :

- selon les schémas numériques utilisés (qui peuvent évoluer sur une grosse plate-forme de calcul), aucune approche n'est définitivement supérieure en terme de stabilité et un choix doit donc être fait si l'on souhaite avoir une approche générique ;
- l'avantage d'une approche discrète est de pouvoir être effectuée par une méthode de différenciation automatique, à même d'alléger les contraintes de développement informatique.
- Ceci permet d'avoir une compatibilité entre valeur de la fonction coût et valeur du gradient.

4.3.2 Différentiation automatique et contraintes de développement

4.3.2.1 Contraintes

De nombreux logiciels de différentiation automatique (Odyssée [Faure et Papegay, 1997], TAMC [Giering et Kaminski, 1998], ADIFOR [Bischof *et al.*, 1992], TAPENADE, ...) ont été développés ces dernières années. Le principe est de prendre en entrée un code (l'implémentation informatique de la discrétisation numérique du modèle physique direct : on parlera de *code direct*) et de générer les codes dérivés associés (code *linéaire tangent* et code *adjoint* ou *cotangent*).

Un inconvénient majeur est la gestion de la trajectoire calculée par le modèle direct. En effet, les dérivées de fonctions non-linéaires nécessitent la connaissance des états le long desquels elles doivent être calculées (fournie par le calcul direct). Le recalcul de la trajectoire dans chaque routine du code nécessite *a priori* l'emploi d'un grand nombre de variables locales afin de stocker l'historique du modèle direct. Une telle gestion induit alors des codes dérivés extrêmement coûteux en temps calcul et en place mémoire.

Un second élément important est lié aux contraintes de développement qui peuvent être associées à l'utilisation d'un logiciel de différentiation automatique. Dans le cas d'Odyssée ([Charpentier, 2000] par exemple), les principales restrictions sont les suivantes :

– les instructions de saut (**GOTO** en Fortran) ne peuvent souvent pas être traitées ;
– les tableaux de taille variable ne sont pas reconnus ;
– les boucles d'arrêt conditionnel (**DO WHILE** en Fortran) ne sont pas traitées, ...

On notera un principe (bien compréhensible par ailleurs) : l'exécution du modèle direct doit être la "plus figée possible" (de fait le cheminement de la trajectoire parmi l'arborescence des routines).

4.3.2.2 Schémas de check-points

Rappelons pour mémoire que des méthodes de gestion de la trajectoire ont été développées (notamment à la suite des travaux de Griewank, [Griewank et Walther, 1999]). L'idée est de sauver l'état du modèle direct en jalonnant l'intervalle de "check-points" (de points où l'on va recalculer des morceaux de la trajectoire directe). Les divers algorithmes de check-points vont alors se distinguer les uns des autres selon le placement de ces instants (méthodes FISC, Periodis, bisection, Treeverse, TwiCE, etc).

Dans notre cas, la gestion de la trajectoire de POLAIR3D ne nécessite pas le recours à de telles approches.

4.3.2.3 Validation des modèles dérivés

Deux tests sont couramment utilisés afin de valider les codes dérivés obtenus. On considère une fonctionnelle J (de R^{m} dans R, m étant le nombre de paramètres par rapport auxquels la différentiation a lieu).

4.3.2.3.1 Test du gradient Le premier test, communément appelé *test du gradient*, consiste à vérifier pour toute direction de perturbation δu que :

$$\lim_{\epsilon \to 0} \frac{J(u + \epsilon \delta u) - J(u)}{\epsilon (\nabla J(u), \delta u)} = 1 \tag{4.40}$$

en se fondant sur le développement de Taylor

$$J(u + \epsilon \delta u) = J(u) + \epsilon (\nabla J(u), \delta u) + O(\epsilon^2) \tag{4.41}$$

Dans les équations précédentes, ∇J est la valeur fournie par les modèles dérivés. Le comportement attendu est celui d'une "courbe en cloche" : pour des petites valeurs de la perturbation ϵ, les erreurs d'arrondis dominent informatiquement et le ratio calculé est loin de 1 ; pour des grandes valeurs de ϵ (pour des fonctions non linéaires), l'hypothèse de linéarisation n'est plus valide et le ratio est à nouveau loin de 1.

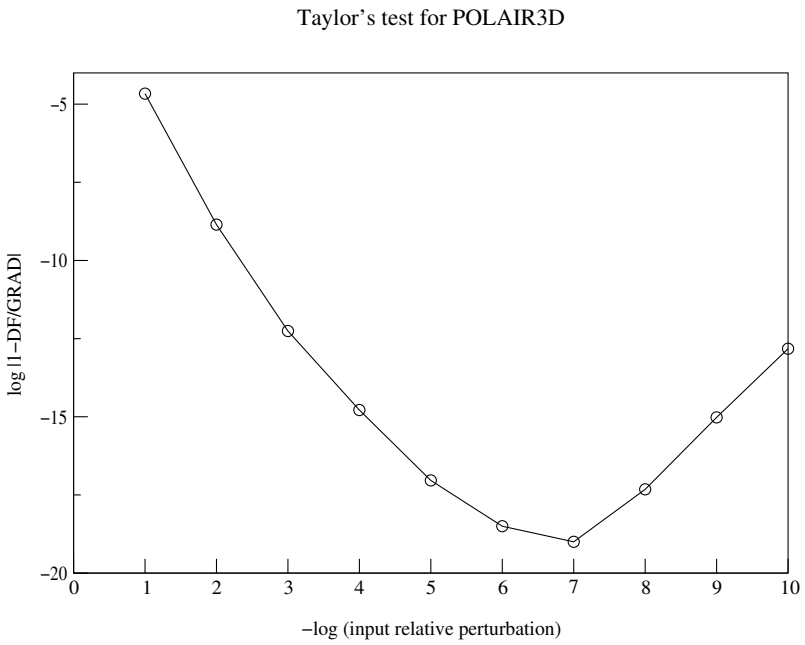

FIG. 4.1 – Un test de Taylor typique pour un paramètre du modèle de Chimie-Transport POLAIR3D.

La courbe 4.1 donne un exemple typique de courbe de Taylor obtenue pour le modèle POLAIR3D.

4.3.2.3.2 Test du produit scalaire Un second test utilise conjointement les codes linéaire tangent et adjoint. Considérons à présent (en toute généralité pour ne pas se restreindre au cas d'une fonctionnelle scalaire) une "partie de code" associant un vecteur de paramètres $u \in \mathrm{R^m}$ à un vecteur de variables $y \in \mathrm{R^n}$ par la relation $y = \Phi(u)$.

Le code linéaire tangent va alors associer à la perturbation $\delta u \in \mathrm{R^m}$, la perturbation $\delta y \in \mathrm{R}^n$ donnée par $\delta y = \Phi^{'}(u)\delta u$. On notera : $\delta y = LT(u, \delta u)$.

Le code adjoint permet de calculer le produit scalaire de la perturbation δy avec tout vecteur $\delta\tilde{y} \in \mathrm{R^n}$ via la relation :

$$(\delta y, \delta\tilde{y}) = (AD(u, \delta\tilde{y}), \delta u) \tag{4.42}$$

De fait, $AD(u, \delta\tilde{y}) = (\Phi^{'}(u))^T\delta\tilde{y}$.

Si l'on prend $\delta\tilde{y} = \delta y$, on doit pouvoir vérifier pour toute perturbation des paramètres δu :

$$(LT(u, \delta u), LT(u, \delta u)) = (AD(u, LT(u, \delta u)), \delta u) \tag{4.43}$$

ce qui constitue le test du produit scalaire.

4.3.2.4 Traitement des parties linéaires

L'utilisation d'un code de différentiation automatique n'exclut bien entendu pas le recours à un traitement approprié de certaines parties du modèle, notamment les parties linéaires. Considérons par exemple la routine qui inverse le système $A.y = b$ (y étant l'inconnue, A une matrice carrée inversible et b un second membre). Dans ce cas, en cohérence avec les notations précédentes, les paramètres sont définis par $u = (b, A)$ et $F(u) = A^{-1}b$. De fait, la routine est de la forme $ROUTINE(A, b, y)$ avec comme arguments d'entrée (A, b) et y comme argument de sortie.

On note δy, δA et δb les dérivées directionnelles et on a évidemment pour le code linéaire tangent :

$$A.\delta y = \delta b - \delta A.y \tag{4.44}$$

La routine initiale peut être utilisée sous la forme $ROUTINE(A, \delta b - \delta A.y, \delta y)$ qui renvoie $\delta y(\delta A, \delta b)$.

Pour le code adjoint, on utilise la relation de définition :

$$(\delta y, \delta\tilde{y}) = (A^\star(\delta\tilde{y}), \delta A) + (b^\star(\delta\tilde{y}), \delta b) \tag{4.45}$$

Or :

$$(\delta y, \delta\tilde{y}) = (A^{-1}(\delta b - \delta A.y), \delta\tilde{y}) \tag{4.46}$$

et l'on a directement :

$$b^\star(\delta\tilde{y}) = (A^{-1})^T\delta\tilde{y}\ , \quad A^\star(\delta\tilde{y}) = -y(b^\star(\delta\tilde{y}))^T \tag{4.47}$$

Autrement dit, on utilise cette fois-ci la routine sous la forme $ROUTINE(A^T, \delta\tilde{y}, b^\star)$ puis on en déduit $A^\star$.

4.3.3 Le préprocesseur SPACK

La génération des mécanismes chimiques sur lesquels reposent les modèles de Chimie-Transport peut être réalisée de manière automatique et couplée avec des procédures de différenciation automatique afin de pouvoir disposer de modèles adjoints et linéaires tangents associés.

C'est l'approche que nous avons choisie, qui est fondée sur le couplage entre un préprocesseur chimique (SPACK pour Simplified Preprocessor for Atmospheric Chemical

Kinetics) et un outil de différentiation automatique (Odyssée, [Faure et Papegay, 1997]).

SPACK permet notamment de réaliser les opérations suivantes :
- le passage d'une représentation symbolique (décrivant les réactions chimiques entre les espèces chimiques considérées) à une représentation mathématique (des équations différentielles ordinaires décrivant la cinétique chimique) ;
- la génération automatique des lumpings d'espèces lorsque ceux-ci sont utilisés dans les mécanismes (c'est particulièrement le cas pour les modèles diphasiques que l'on rencontre dans la chimie des gouttes de nuage) ;
- la génération en Fortran 77 des routines associées pour la simulation numérique (a priori insérables dans n'importe quel code hôte) ;
- la génération en Fortran 77 des routines dérivées pour les modèles adjoint et linéaire tangent associés.

C'est cette approche extrêmement modulaire que nous avons adoptée et qui permet d'obtenir, modulo une description standardisée des mécanismes chimiques, les routines décrivant les modèles directs, linéaire tangent et adjoint de n'importe quel mécanisme.

En pratique, les schémas dont nous disposons sont par exemple CBM IV ([Gery *et al.*, 1989]), RADM2 ([Stockwell *et al.*, 1990]), RACM ([Stockwell *et al.*, 1997a]), etc...

4.3.4 Boucle en temps et implémentation pratique dans POLAIR3D

Un modèle de Chimie-Transport va être décrit au niveau algorithmique par la séquence suivante :

1. initialisation des données indépendantes du temps ;
2. initialisation de la fonction coût : $J(c_0) = 0$ (ou un terme d'ébauche éventuel) ;
3. boucle en temps (indexée par $1 \leq i \leq n_o$) :
   - lecture des données de forçage ϕ_{i-1} (nécessaires au modèle pour le calcul de l'état au temps t_i) ;
   - calcul du nouvel état $c_i = \Phi(c_{i-1}, \phi_{i-1})$;
   - calcul de la fonction coût au temps t_i : $J_i(c_i)$;
   - mise à jour de la fonction coût : $J = J + J_i$.

Les données de forçage ϕ_i comprennent l'ensemble des données météorologiques (en réalité les sorties des paramétrisations calculées à partir des champs météos et leur interpolation sur la grille du CTM, ainsi que les paramétrisations physico-chimiques).

Un calcul direct de ∇J pourrait être fait mais il nécessiterait de stocker toutes les données de forçage $(\phi_0, \ldots, \phi_{n_o-1})$ et la trajectoire de la solution $(c_0, \ldots, c_{n_o})$, ce que l'on

ne peut pas faire en pratique. Une approche plus pratique est de noter que $\nabla_u J = \sum_i \nabla_u J_i$, c'est à dire :

$$\nabla_u J = \sum_{1 \leq j < i \leq n_o} (\frac{\partial c_j}{\partial u})^T_{|t_{j-1}} \dots (\frac{\partial c_i}{\partial c_{i-1}})^T_{|t_{i-1}} (\frac{\partial J_i}{\partial c_i})^T_{|t_i} + \sum_{1 \leq i \leq n} (\frac{\partial c_i}{\partial u})^T_{|t_{i-1}} (\frac{\partial J_i}{\partial c_i})^T_{|t_i} \tag{4.48}$$

Le modèle adjoint permet de calculer pour tout vecteur z :

$$(\frac{\partial c_i}{\partial u})^T_{|t_{i-1}} z \ , \quad (\frac{\partial c_i}{\partial c_{i-1}})^T_{|t_{i-1}} z \tag{4.49}$$

Ceci nécessite de stocker (par exemple dans des fichiers) la trajectoire $(c_0, c_1, \dots, c_n)$. Pour le cas où l'on travaille sur les conditions initiales ($u = c_0$), la formule précédente devient :

$$\nabla_{c_0} J = \sum_{1 \leq i \leq n_o} (\frac{\partial c_1}{\partial c_0})^T_{|t_0} \dots (\frac{\partial c_i}{\partial c_{i-1}})^T_{|t_{i-1}} (\frac{\partial J_i}{\partial c_i})^T_{|t_i} \tag{4.50}$$

$$= \dots \times \left[(\frac{\partial J_{n_o-2}}{\partial c_{n_o-2}})^T + (\frac{\partial c_{n_o-1}}{\partial c_{n_o-2}})^T [(\frac{\partial J_{n_o-1}}{\partial c_{n_o-1}})^T + (\frac{\partial c_{n_o}}{\partial c_{n_o-1}})^T (\frac{\partial J_{n_o}}{\partial c_{n_o}})^T] \right] \tag{4.51}$$

et l'algorithme s'écrit à présent :

1. initialisation : $c^\star_{n_o} \stackrel{\Delta}{=} (\frac{\partial J_{n_o}}{\partial c_{n_o}})^T$,
2. boucle en temps rétrograde (indexée par $n_o \geq i \geq 1$) :
   - lecture des données de forçage ϕ_{i-1},
   - lecture de la trajectoire sauvegardée c_{i-1},
   - calcul de :
   $$c^\star_{i-1} = (\frac{\partial c_i}{\partial c_{i-1}})^T_{|c_{i-1}, \phi_{i-1}} c^\star_i \tag{4.52}$$
   - mise à jour de $c^\star_{i-1} = c^\star_{i-1} + (\frac{\partial J_{i-1}}{\partial c_{i-1}})^T$.
3. Sortie de : $\nabla_{c_0} J = c^\star_0$.

Notons que le coût du calcul est de l'ordre du coût de calcul d'un modèle direct (mais certainement pas multiplié par le nombre de degrés de liberté recherché).

4.4 Application à des tests de sensibilité

Cette section présente des cas d'application préliminaires au calcul de la sensibilité, notamment de l'ozone. Le premier cas concerne un modèle diphasique présenté dans [Djouad *et al.*, 2002a], qui décrit le transfert de masse entre l'air et les gouttes d'eau de nuage d'une part, la chimie en phases gazeuse et aqueuse d'autre part. Le second cas présente un calcul 3D effectué sur la base des simulations ESQUIF effectuées avec POLAIR3D au-dessus de la région parisienne ([Sartelet *et al.*, 2002]).

4.4.1 Chimie diphasique

Pour des raisons de cohérence avec la littérature du domaine, on a gardé les notations standards (au risque éventuel d'avoir des ambiguïtés avec les notations précédentes).

L'évolution des espèces chimiques dans la phase gazeuse et dans les gouttes de nuage est donnée par le système d'EDOs couplées suivant ([Lelieveld et Crutzen, 1991]) :

$$\frac{dG}{dt} = \chi_G(G) - L\,k_{mt}\,G + \frac{k_{mt}\,A}{H\,R\,T}\ ,\ \frac{dA}{dt} = \chi_A(A) + L\,k_{mt}\,G - \frac{k_{mt}\,A}{H\,R\,T} \tag{4.53}$$

avec G et A respectivement les concentrations dans la phase gaz et dans la phase eau, exprimées en $molecule/cm^3$ d'air. χ_G et χ_A sont les taux de production/consommation liés à la cinétique chimique dans les deux phases. L est le contenu en eau liquide (liquid water content), défini comme le ratio du volume d'eau sur le volume d'air ; H est la constant de Henry, R la constante des gaz parfaits et T la température. k_{mt} est le coefficient de transfert de masse paramétrisé par exemple selon [Schwartz, 1986] ou [Djouad *et al.*, 2002a] :

$$k_{mt} = \left(\frac{a^2}{3D_g} + \frac{4a}{3\alpha\,v}\right)^{-1} \tag{4.54}$$

avec a le rayon de la goutte, D_g la diffusivité moléculaire dans le gaz, α le coefficient d'accommodation. v est la vitesse d'agitation thermique donnée par :

$$v = \sqrt{\frac{8\,R\,T}{\pi\,M}} \tag{4.55}$$

avec M la masse molaire. On se réfère par exemple à [Schwartz, 1986, Gregoire et Chaumerliac, 1994, Audiffren *et al.*, 1998] pour plus de détails.

De nombreux paramètres physiques interviennent donc et sont associés à des incertitudes :

- les constantes cinétiques définissant χ_G et χ_A ;
- les constantes permettant le calcul du coefficient de transfert de masse : $H(T)$, α, D_g et D_a ;
- les caractéristiques de la distribution des gouttes : dans notre cas simplifié (population monodispersée) le rayon de la goutte a ;
- les variables météorologiques comme la température T ou le flux actinique (pour les réactions de photolyse).

L'utilisation de SPACK a permis de générer les fonctions à intégrer dans les EDOs définissant la cinétique chimique.

Le tableau 4.1 donne quelques coefficients de sensibilité L^1 et L^2 (au sens de l'intégration en temps) pour des espèces gazeuses importantes vis à vis des paramètres L, pH, α, a et T. On se réfère à [Djouad *et al.*, 2002b] pour la définition des conditions

numériques et une interprétation physique, cet exemple n'ayant qu'une valeur illustrative. On voit néanmoins les différences d'ordre de grandeurs des impacts de chaque paramètre d'entrée. Par exemple, l'ozone gazeux est relativement peu impacté et la température et le diamètre de la goutte sont des paramètres clés.

Species	L	pH	α	a	T
O_3(g)	$5.5\,10^{-4}(-5.3\,10^{-4})$	$1.7\,10^{-4}(1.6\,10^{-4})$	$3.1\,10^{-7}(-2.9\,10^{-7})$	$7.9\,10^{-4}(7.5\,10^{-4})$	$10^{-1}(9.9\,10^{-2})$
NO	$4\,10^{-2}(-0.031)$	$7.9\,10^{-3}(-5.5\,10^{-3})$	$2.1\,10^{-19}(-1.4\,10^{-19})$	$8.9\,10^{-2}(6.4\,10^{-2})$	$1.7\,10^{1}(7.6)$
NO_2	$2.6\,10^{-2}(-1.9\,10^{-2})$	$5.7\,10^{-3}(-3.8\,10^{-3})$	$3.3\,10^{-14}(-1.7\,10^{-14})$	$6\,10^{-2}(4\,10^{-2})$	$1.7\,10^{1}(10)$
HO_2	$5.1\,10^{-2}(-5\,10^{-2})$	$5.\,10^{-3}(4.9\,10^{-3})$	$1.1\,10^{-3}(-1.1\,10^{-3})$	$10^{-1}(9.9\,10^{-2})$	5.3(4)
OH	$5.8\,10^{-3}(5.7\,10^{-3})$	$3.4\,10^{-3}(3.3\,10^{-3})$	$7.1\,10^{-06}(-6.9\,10^{-06})$	$10^{-2}(-9.9\,10^{-3})$	1.4(1.3)

TAB. 4.1 – Normes L_2 et (L_1) pour les coefficients de sensibilité adimentionnés de quelques espèces gazeuses en fonction de paramètres microphysiques dans le cas de référence.

Un point clé de l'étude de l'impact de la chimie en phase aqueuse dans les nuages est donné par le ratio HO_2/OH, dont les sensibilités au coefficient de transfert de masse et aux taux de photolyse de NO_2 et O_3 peuvent être calculées et sont données dans les figures 4.2 et 4.3, respectivement.

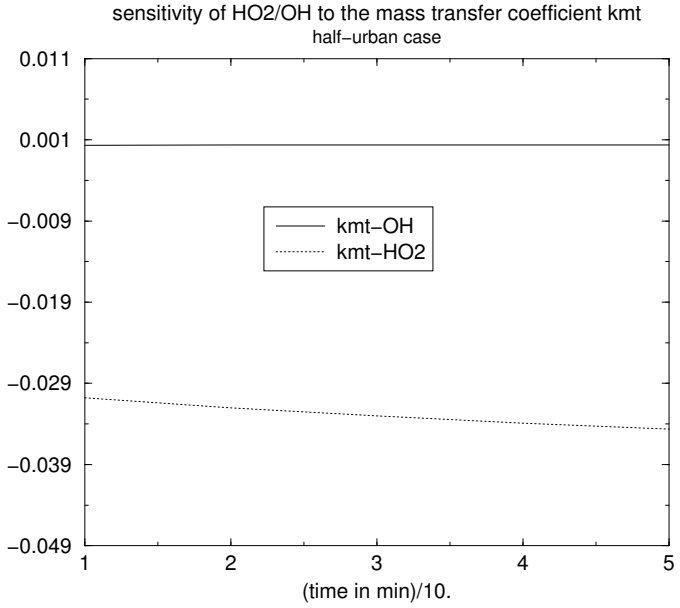

FIG. 4.2 – Sensibilité du ratio $\frac{HO_2}{OH}$ au coefficient de transfert de masse k_{mt}.

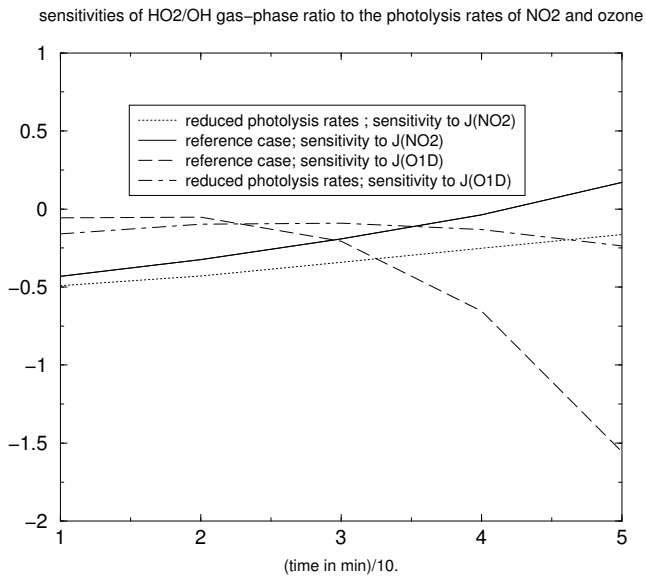

FIG. 4.3 – Sensibilité du ratio $\frac{HO_2}{OH}$ aux taux de photolyse de NO_2 et O_3.

4.4.2 Calcul 3D ESQUIF

L'objectif de la campagne de mesures ESQUIF ([Menut *et al.*, 2000]) était d'étudier la pollution photochimique sur l'Ile-de-France. La période d'observations intensives n°2 (POI2) a duré du 7 au 9 Août 1998 et correspond à un épisode important de pollution à l'ozone. Elle est caractérisée par des vents faibles, un ciel clair et des températures élevées.

POLAIR3D a montré de bonnes aptitudes à reproduire les concentrations d'ozone sur cette période. On se réfère à [Sartelet *et al.*, 2002] pour plus de détails. On rappelle que le domaine de simulation de 150 km × 150 km est divisé en mailles carrées de 6 km de longueur, soit 25 × 25 mailles. La résolution verticale comprend 9 niveaux.

Le mécanisme chimique RADM2 a été utilisé dans cette étude. L'inventaire d'AIRPARIF, fourni par classe SNAP, a été utilisé pour obtenir des émissions temporelles, spatialisées et par espèce modèle selon la procédure décrite au chapitre 1.

On souhaite identifier les espèces dont les émissions jouent le rôle le plus important lors de cette période. En pratique, on a calculé les sensibilités des concentrations d'O_3, NO et NO_2 dans la première maille verticale (simulées par POLAIR3D) par rapport aux émissions de tous les polluants. La sensibilité est calculée par un modèle linéaire tangent[†2].

[†2]Choix motivé par le plus grand nombre de sorties comparativement au nombre d'entrées.

Pour pouvoir comparer les sensibilités entre elles, nous avons choisi le critère suivant :

$$\frac{1}{N_i\, N_j} \sum_{i,j} e_i \frac{\partial c_j}{\partial e_i} \tag{4.56}$$

qui correspond à une sensibilité relative et donne la variation absolue de la concentration c en fonction de la perturbation relative de l'émission e. Les sensibilités sont moyennées en espace et en temps par l'intermédiaire des indices i et j qui font référence au moment d'émissions (N_i pas de temps) et aux temps de sorties de modèle (N_j pas de temps).

Le tableau 4.2 rapporte les résultats. On remarque l'importance prépondérante des émissions de NO sur les concentrations d'O_3, NO et NO_2.

Espèce émise	Sensibilité de O_3	Espèce émise	Sensibilité de NO	Espèce émise	Sensibilité de NO_2
NO	2.1e-03	NO	5.4e-04	NO	1.3e-03
XYL	3.2e-04	NO_2	3.2e-05	NO_2	1.9e-04
HCHO	2.3e-04	HCHO	2.8e-05	XYL	5.5e-05
OL_2	1.9e-04	XYL	2.6e-05	HCHO	4.6e-05
TOL	1.5e-04	MGLY	1.4e-05	TOL	2.6e-05
OLT	1.4e-04	OLI	1.2e-05	MGLY	2.3e-05
MGLY	1.1e-04	TOL	1.1e-05	OLI	2.3e-05
OLI	1.1e-04	OL_2	1.1e-05	OL_2	1.8e-05
NO_2	9.8e-05	OLT	1.0e-05	OLT	1.7e-05
CO	9.2e-05	GLY	6.8e-06	GLY	1.0e-05
HC_5	6.3e-05	CO	4.5e-06	CSL	9.8e-06
GLY	5.5e-05	HC_5	2.5e-06	CO	6.8e-06
HC_3	4.3e-05	CSL	2.4e-06	HC_5	3.8e-06
HC_8	3.6e-05	HC_3	1.9e-06	ALD	3.3e-06
ISO	2.8e-05	ALD	1.9e-06	HC_3	2.7e-06
ALD	2.6e-05	HC_8	1.2e-06	HC_8	2.1e-06
CSL	1.7e-05	ISO	1.0e-06	ISO	2.0e-06
SO_2	1.4e-05	KET	6.3e-07	KET	1.4e-06
KET	1.0e-05	SO_2	3.3e-07	SO_2	8.7e-07
ETH	8.8e-07	ETH	4.2e-08	ETH	5.9e-08
CH_4	6.4e-07	CH_4	3.0e-08	CH_4	4.6e-08

TAB. 4.2 – Sensibilités relatives moyennes de O_3, NO et NO_2 aux émissions des espèces primaires.

Chapitre 5

Modélisation inverse des émissions pour un modèle de Chimie-Transport à l'échelle locale : application au cas de Lille.

Résumé

L'objectif de ce chapitre est d'améliorer les performances du modèle sur le cas de la modélisation de Lille. Des études de sensibilité aux divers paramètres d'entrée ont été effectuées et le choix a été fait de réaliser une modélisation inverse des émissions jugées incertaines, en l'occurrence les émissions issues de l'étude PREDIT présentée au chapitre 2. On étudie notamment dans ce chapitre la robustesse de l'inversion des émissions vis-à-vis des incertitudes affectant les autres paramètres.

Sommaire

5.1 Cadre de l'étude 126
5.1.1 Domaine 126
5.1.2 Réseau d'observation 126
5.2 Etude de sensibilité 127
5.2.1 Sensibilité de la fonction coût 127
5.2.2 Impact temporel et spatial des émissions de NO_x 132
5.3 Modélisation inverse des émissions : expériences synthétiques 135
5.3.1 Cadre de la modélisation inverse 135
5.3.2 Résultats avec des observations synthétiques non perturbées . . . 137
5.3.3 Résultats avec des expériences synthétiques perturbées 137

5.4 Résultats sur le cas réel 142
5.4.1 Inversion sur la semaine du 11 au 15 Mai 142
5.4.2 Vérification 145
5.4.3 Optimisation sur d'autres périodes 145
5.4.4 Sensibilité du second-ordre 145

Introduction

Les flux d'émissions de polluants primaires font partie des paramètres d'entrée des modèles de chimie–transport qui sont connus avec le moins de certitudes, que ce soit en terme de spatialisation (il peut y avoir de très fortes hétérogénéités spatiales des sources d'émissions) en terme de temporalisation (liée à une variabilité de l'activité "naturelle" pour les émissions biogéniques et une variabilité des conditions de trafic par exemple pour les émissions anthropogéniques) ou en terme de spéciation chimique (la correspondance entre espèces inventoriées, espèces chimiques réelles et espèces modèles étant loin d'être directe).

Par ailleurs, indépendamment de l'impact des émissions sur les concentrations modélisées, on peut être intéressé à connaître avec précision les émissions en tant que telles, par exemple dans un contexte de surveillance des émissions autorisées (un exemple emblématique étant bien sûr le protocole de Kyoto pour les gaz à effets de serre ou la convention sur la pollution atmosphérique transfrontière à longue distance en Europe).

Pour toutes ces raisons, la modélisation inverse des émissions est une approche qui est de plus en plus utilisée. Au niveau global, pour des polluants essentiellement passifs (CO, CH_4), on peut par exemple se référer aux travaux de [Kaminski, 1998, Bergamashi *et al.*, 2000, Bousquet *et al.*, 1999]. De même pour des traceurs linéaires (radioactifs), de nombreuses études ont déjà été réalisées ([Hourdin et Issartel, 2000]). La donne est largement différente pour les modèles de chimie–transport, du fait notamment de la complexité des mécanismes chimiques, de la grande dimension des espaces de contrôle et des coûts de calculs associés. On peut se référer par exemple à [Elbern *et al.*, 2000] pour des tests de faisabilité en observations synthétiques et des travaux à l'échelle continentale [Elbern et Schmidt, 2001, Mendoza-Dominguez et Russel, 2001, van Loon *et al.*, 2000]. Des calculs académiques sont menés dans [Massart, 2003] à l'aide d'un modèle 2D simplifié (une quinzaine d'espèces chimiques avec des paramétrisations constantes).

L'objet de ce chapitre est de réaliser des expériences de modélisation inverse sur des données réelles dans le cas de la simulation lilloise présentée au chapitre 2. Notre démarche est la suivante : du fait du faible nombre de capteurs disponibles, nous avons restreint les paramètres de contrôle à une chronique temporelle (de résolution horaire) appliquée aux émissions de NO_x. Cette chronique a été optimisée sur une semaine d'observations puis validée par comparaison aux mesures sur les semaines suivantes.

Ce chapitre est organisé de la manière suivante. La première section présente le cadre de l'étude. Dans la deuxième section, des études de sensibilité sur la fonction coût mesurant l'écart aux observations sont menées. Cela permet de mieux cerner les paramètres

contrôlant de façon prépondérante les sorties choisies du modèle. On précise le cadre de la modélisation inverse envisagé et des expériences de modélisation inverse en expériences synthétiques sont effectuées et rapportées dans la section 3. On montre en particulier l'influence des paramètres pas ou mal connus sur les paramètres inversés. Le cas réel fait l'objet de la section 4 avec l'inversion d'une chronique temporelle appliquée aux émissions de NO_x. Des tests de validation (vérification a posteriori) et de robustesse à des paramètres mal connus (turbulence) sont également présentés afin de tester la qualité de la modélisation inverse.

5.1 Cadre de l'étude

5.1.1 Domaine

Le domaine est centré sur l'agglomération Lilloise. Pour des raisons de coût de calcul, il a été réduit par rapport au chapitre 2. Il est composé de 21 × 24 × 9 mailles.

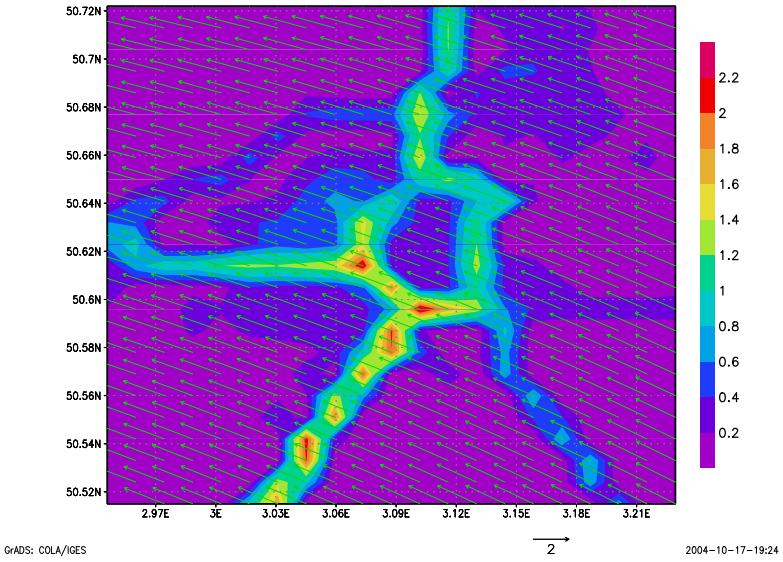

FIG. 5.1 – Emissions surfaciques moyennes de NO en $\mu gm^{-2}s^{-1}$ et vent moyen au sol en $m.s^{-1}$ pour le 11 Mai.

Le mois de Mai a été plus particulièrement étudié. Le vent moyen pour le 11 Mai est représenté sur la figure 5.1. Son module est de l'ordre de 2 $m.s^{-1}$ ce qui correspond à un temps de résidence des polluants dans le domaine de l'ordre de 2-3 heures.

5.1.2 Réseau d'observation

Les stations de mesures du réseau AREMA sont représentées sur la figure 5.2 pour l'ozone et la figure 5.3 pour les oxydes d'azote (une mesure de NO plus une mesure de

NO_2 par station).

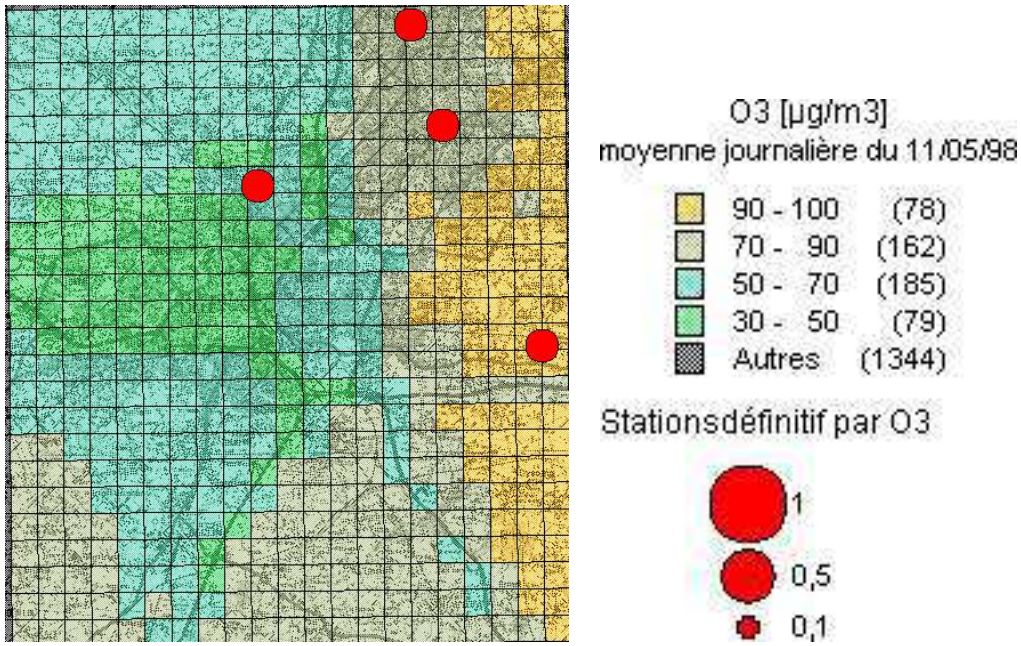

FIG. 5.2 – Réseau de mesures et concentrations simulées d'ozone à Lille pour le 11 Mai 1998 (μgm^{-3}).

5.2 Etude de sensibilité

Dans un premier temps, on étudie le caractère multifactoriel des modèles de chimie–transport en estimant la sensibilité des sorties du modèle (en l'occurrence la fonction coût) à l'ensemble des paramètres d'entrée.

On se réfère à [Menut, 2003] pour une étude de sensibilité par la méthode de l'adjoint sur Paris et à [Segers, 2002] pour une étude similaire sur l'Angleterre.

5.2.1 Sensibilité de la fonction coût

Ce paragraphe présente une étude de sensibilité par rapport aux paramètres d'entrée du modèle (notés k dans la suite) : conditions aux limites (figure 5.4), diffusion verticale (figure 5.5), constante cinétique de la réaction O_3 + NO -> NO_2 (figure 5.6), émissions (figure 5.7), dépôt (figure 5.8) et photolyse (figure 5.9). Pour chacun de ces paramètres, on peut s'attendre en effet à des incertitudes liées :

- aux résultats du calcul de forçage continental,

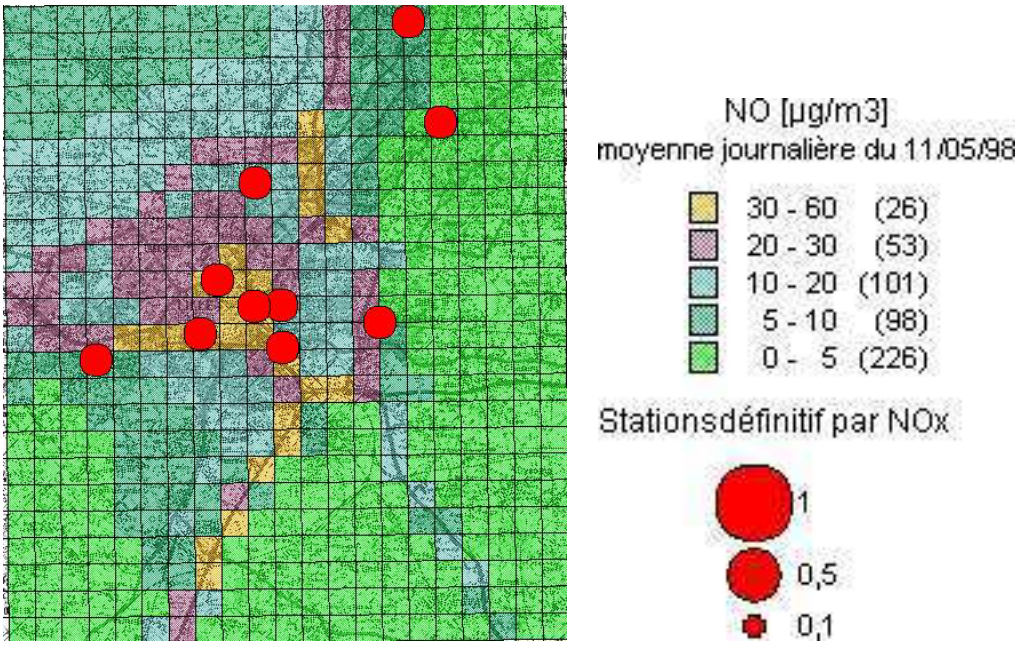

FIG. 5.3 – Réseau de mesures de NO_x et concentrations simulées de NO à Lille pour le 11 Mai 1998 (μgm^{-3}).

- à la représentation du mélange turbulent,
- aux effets de ségrégation pour la cinétique chimique,
- à la qualité des inventaires d'émissions,
- à la paramétrisation du dépôt,
- à la paramétrisation de l'atténuation pour la photolyse.

L'expérience de sensibilité est présentée pour la journée du 11 Mai 1998. Les résultats sont bien sûr dépendants du jour choisi mais ils sont relativement bien représentatifs de la période printemps–été sur le plan qualitatif (tests non reportés ici).

On définit artificiellement une fonction coût de la façon suivante :

$$J(\alpha) = \sum_i (f_i(\alpha k_i) - obs_i)^2 \tag{5.1}$$

où l'indice i fait référence au temps, f_i est le modèle POLAIR3D et obs_i désigne les observations disponibles de O_3, NO et NO_2 provenant du réseau de mesures AREMA. α est un coefficient multiplicatif appliqué à l'entrée k_i qui varie de 0.5 à 1.5 ce qui se traduit par des perturbations de plus ou moins 50 % sur k_i. La fonction coût définie ci-dessus est une moyenne temporelle (sur une journée) et spatiale (sur le réseau d'observations). Ainsi, l'impact d'un paramètre ayant un effet localisé spatialement et temporellement sera lissé.

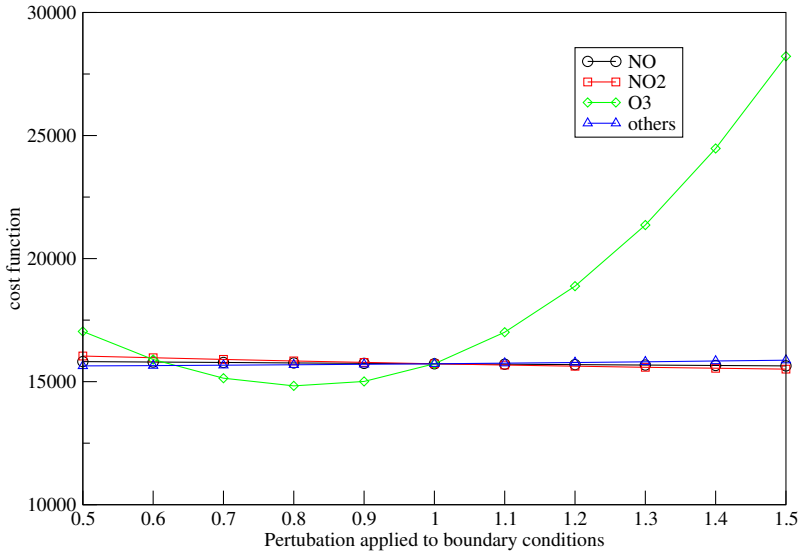

FIG. 5.4 – Fonction coût en fonction d'un coefficient multiplicatif sur les conditions aux limites

L'ozone est un polluant dont l'échelle spatiale caractéristique est supérieure à la taille de notre domaine. On n'est donc pas surpris de l'importance des conditions aux limites d'ozone pour notre simulation régionale. Comparativement, les conditions aux limites de NO_x ont peu d'influence sur la fonction coût, l'apport en NO_x étant principalement dû aux émissic

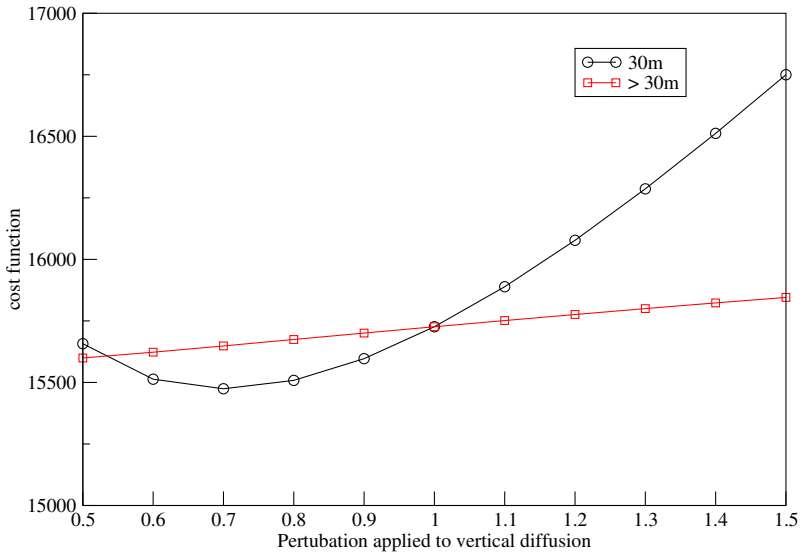

FIG. 5.5 – Fonction coût en fonction d'un coefficient multiplicatif sur la diffusion verticale

Le mélange vertical turbulent est un processus important, que ce soit au premier niveau (30 m) ou au-dessus (voir figure 5.5).

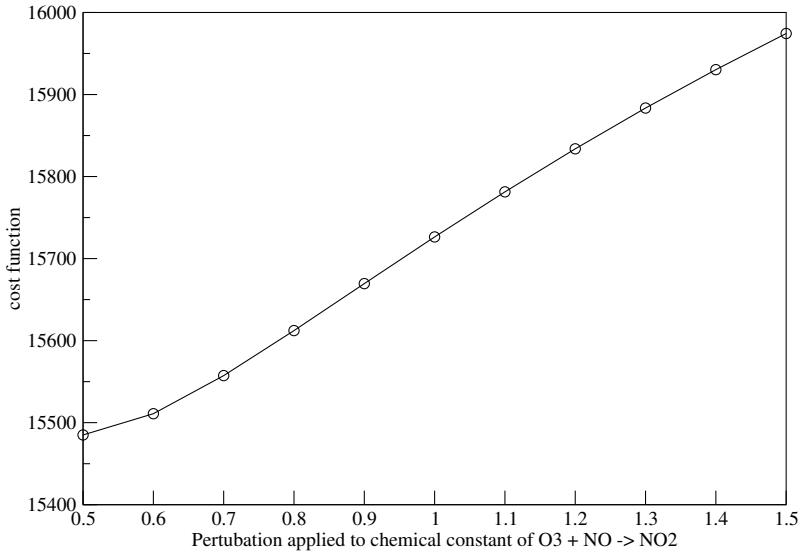

FIG. 5.6 – Fonction coût en fonction d'un coefficient multiplicatif sur la constante cinétique O_3 + NO - > NO_2

L'influence de la constante cinétique de la réaction de formation de NO_2 à partir de O_3 et NO est mise en évidence sur la figure 5.6. Sa valeur est très incertaine à cause des effets de sé

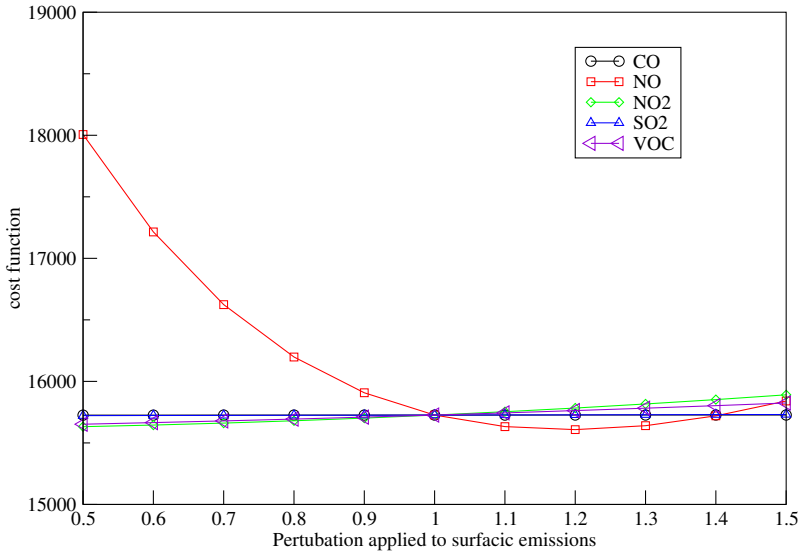

FIG. 5.7 – Fonction coût en fonction d'un coefficient multiplicatif sur les émissions

Parmi les émissions, on note le rôle primordial joué par les émissions de NO (voir figure 5.7). Elles influencent directement les concentrations de NO mais aussi celle de NO_2 et O_3 par la cinétique chimique rapide. Ceci est à mettre en regard des résultats déjà obtenus au chapitre 4 avec les calculs de sensibilité aux espèces émises réalisés pour le cas d'ESQUIF.

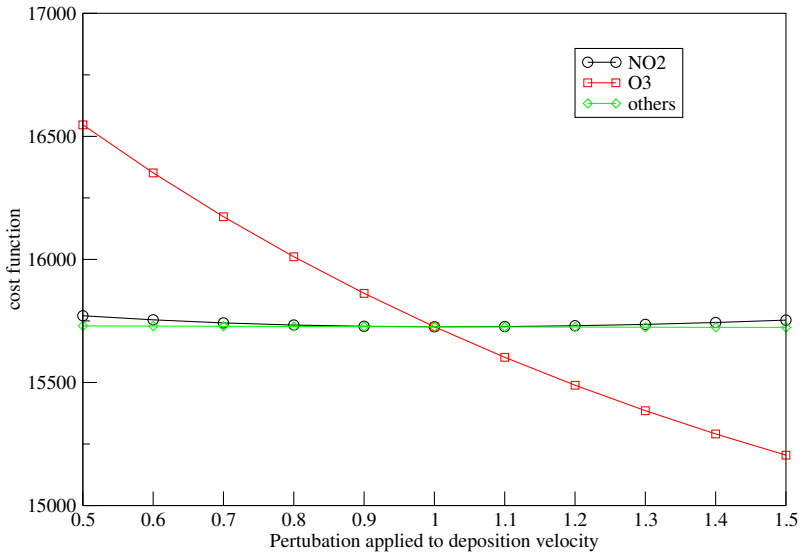

FIG. 5.8 – Γ dépôt sec

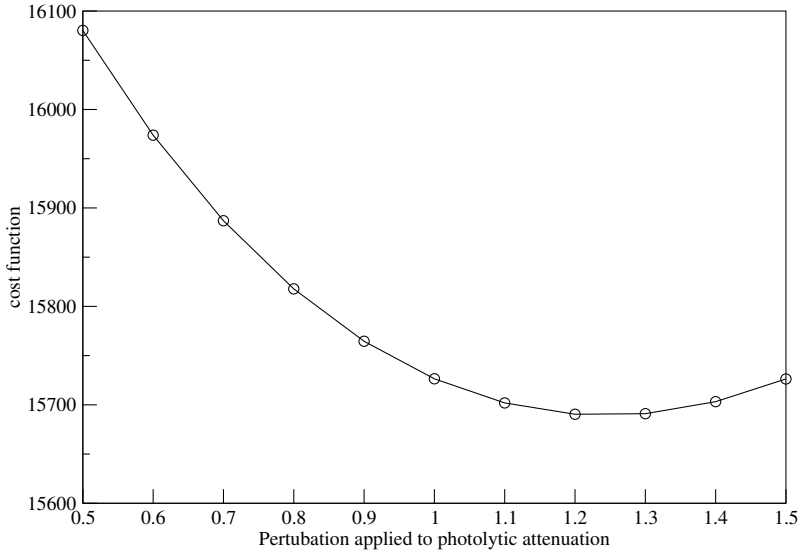

FIG. 5.9 – Fonction coût en fonction d'un coefficient multiplicatif sur la photolyse

Le dépôt sec d'ozone est un paramètre important (figure 5.8) et l'atténuation photolytique dont la paramétrisation est relativement mal connue est une source d'incertitude importante. En effet, sa paramétrisation fait intervenir un diagnostic de présence de nuages difficile à mettre en œuvre. La variation de la fonction coût en fonction d'un coefficient appliqué à l'atténuation photolytique est représentée sur la figure 5.9.

En résumé, les résultats montrent l'importance prépondérante des conditions aux limites d'ozone, du mélange vertical, de la constante cinétique de la réaction O_3 + NO ->

NO_2, des émissions de NO, du dépôt d'O_3 et de la photolyse.

Les courbes obtenues correspondent à des coupes de la fonction coût dans les directions associées aux paramètres perturbés. On remarque que dans l'intervalle de variation considéré [-50 %, +50%], il peut ne pas y avoir de minima locaux ce qui met en évidence le caractère linéaire local du modèle.

5.2.2 Impact temporel et spatial des émissions de NO_x

Afin de préparer la modélisation inverse des émissions, on a cherché à réduire la dimension de l'espace de contrôle (a priori des champs 2D instationnaires par espèces émises). Les calculs précédents ont montré que dans le régime chimique étudié (prépondérant sur Lille), la sensibilité aux NO_x était la plus forte. On va étudier dans la suite, la "zone d'influence" spatiale et temporelle d'une émission localisée.

On procède de la manière suivante :

- Pour l'impact spatial, on perturbe l'émission de NO_x dans la maille $(15, 10)$ de +30 % à tous les pas de temps et on trace les cartes des différences des moyennes journalières de O_3 (figure 5.12), NO (figure 5.10) et NO_2 (figure 5.11).
- Pour l'impact temporel, on perturbe à 3 heures les émissions de NO_x de +30 % dans toutes les mailles du domaine. On trace l'évolution horaire de la différence des moyennes spatiales de concentrations.

L'impact d'une perturbation des émissions de NO_x est très localisé en espace. La différence la plus importante pour les concentrations des trois espèces est observée directement dans la maille d'émission. En comparaison, les mailles directement autour subissent une perturbation 20 fois inférieure pour NO et 6 fois inférieure pour NO_2 ou O_3. Au delà de quelques kilomètres, les émissions de NO_x n'ont relativement plus d'influence.

La figure 5.13 montre le temps d'influence d'une perturbation des émissions de NO_x sur les concentrations des polluants mesurés. Au bout de 3 heures, les différences de concentrations sont négligeables. Ceci est du même ordre que le temps nécessaire aux masses d'air pour traverser le domaine.

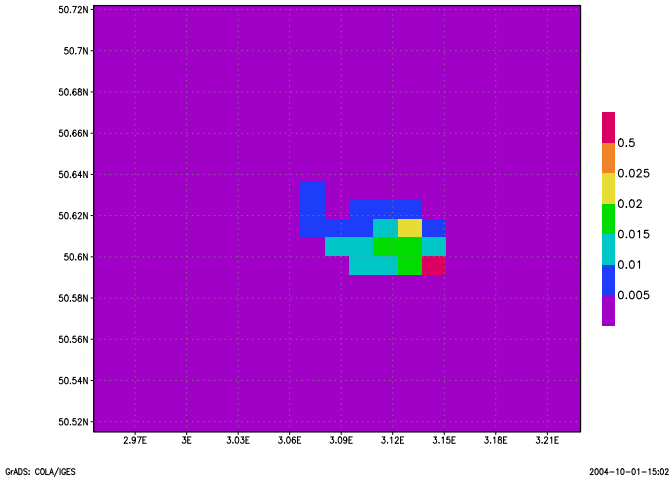

FIG. 5.10 – Carte de différence de concentration moyenne de NO engendrée par une perturbation de l'émission de NO_x en une maille fixée.

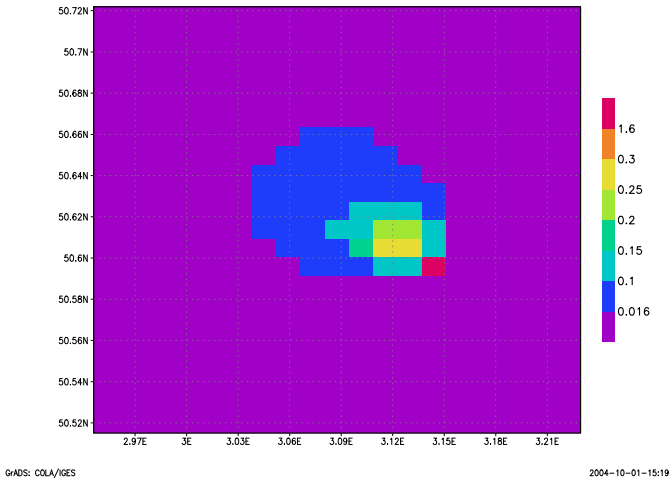

FIG. 5.11 – Carte de différence de concentration moyenne de NO_2 engendrée par une perturbation de l'émission de NO_x en une maille fixée.

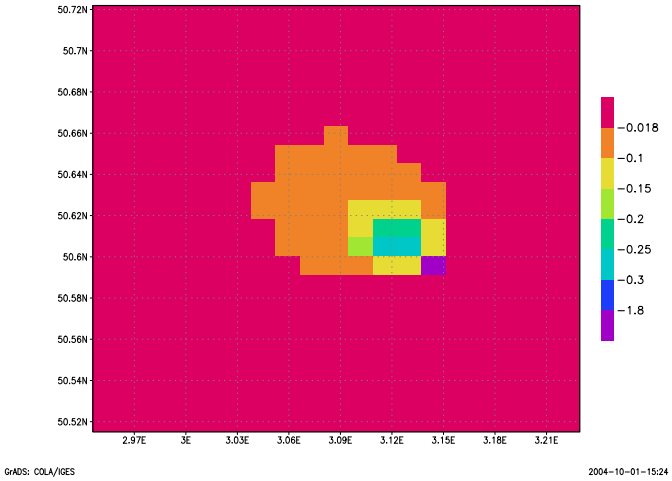

FIG. 5.12 – Carte de différence de concentration moyenne de O_3 engendrée par une perturbation de l'émission de NO_x en une maille fixée.

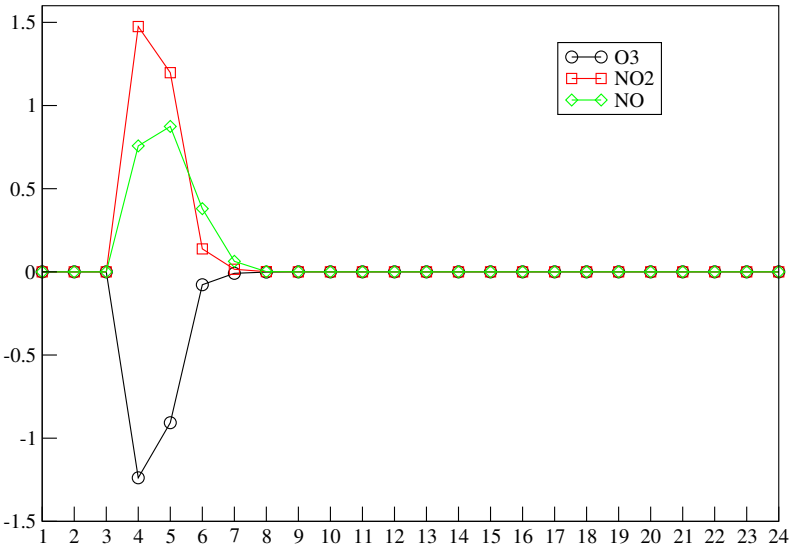

FIG. 5.13 – différence de concentration moyenne de O_3, NO et NO_2 engendrée par une perturbation de l'émission de NO_x à 3h.

5.3 Modélisation inverse des émissions : expériences synthétiques

5.3.1 Cadre de la modélisation inverse

5.3.1.1 Variables de contrôle

Les émissions sont des entrées de modèle jugées parmi les plus incertaines. Outre les incertitudes sur les totaux annuels, la répartition temporelle des émissions est loin d'être précisément connue : les valeurs horaires sont en général obtenues à partir de coefficients mensuels, journaliers et horaires qui reflètent un comportement moyen et ne s'appliquent pas nécessairement pour la région de Lille. Dans cette étude, les émissions du trafic routier sont calculées à partir de coefficients horaires issues du comptage routier sur les axes principaux.

On va représenter les émissions de la manière suivante :

$$E_{\mathrm{NO_x}}(x,t) = \alpha(t)\tilde{E}_{\mathrm{NO_x}}(x,t) \tag{5.2}$$

où $\tilde{E}_{\mathrm{NO_x}}(x,t)$ sont les émissions de l'inventaire et $\alpha(t)$ sont des coefficients multiplicatifs temporels appliqués aux émissions sur tout le domaine spatial donnés en pratique par pas de temps horaires, soit 24 coefficients pour une journée.

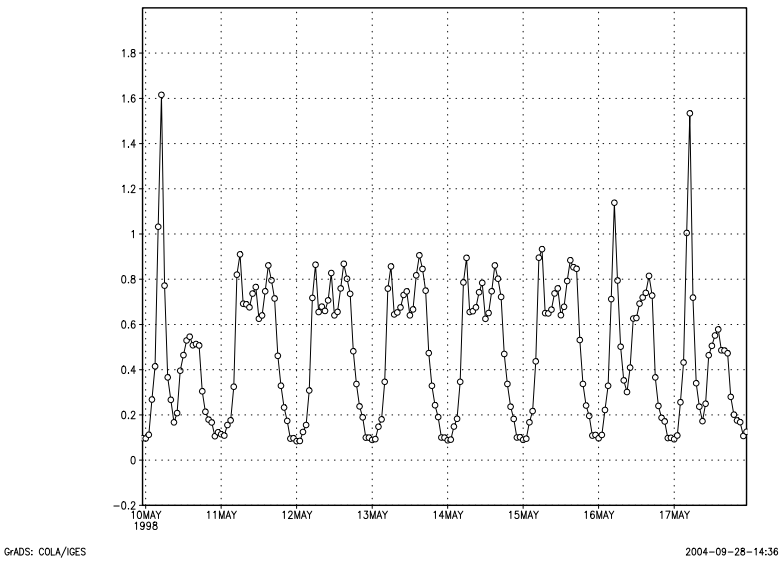

FIG. 5.14 – Chronique horaire moyenne d'émissions de $\mathrm{NO_x}$ sur Lille du Dimanche 10 Mai au Dimanche 17 Mai.

Le choix des paramètres d'inversion repose sur deux critères :

- Leur impact sur les concentrations simulées d'ozone et d'oxydes d'azote est fort (voir section 5.2).
- On suppose qu'ils peuvent être réappliqués pour un autre jour du même type[†1] (caractère reproductible). Par exemple, les chroniques horaires d'un jour de semaine sont très semblables (figure 5.14).

L'objectif est donc obtenir une chronique journalière moyenne sur Lille valable pour un jour de semaine.

5.3.1.2 Fonction coût

L'assimilation de données classique combine trois sources d'information provenant : des paramètres, du modèle et des observations. Les erreurs respectives associées doivent être spécifiées. Cette description des erreurs est difficile et joue cependant un rôle clé dans le résultat de l'assimilation de données. On a choisit ici de se placer dans le cadre du contrôle optimal et d'inverser une fonction coût de la forme :

$$J(\alpha) = \sum_i \left(f_i(\alpha) - obs_i\right)^2 \tag{5.3}$$

Le modèle est pris comme une contrainte forte (pas d'erreur modèle), un poids égal est donné à toutes les observations (en l'absence d'informations précises) et on n'utilise pas d'ébauche (le problème étant sur-contraint).

5.3.1.3 Temps de calculs

Le problème de la minimisation de la fonction coût est résolu de manière itérative par l'algorithme BFGS ([H. Byrd et Zhu, 1995]). Cela nécessite de disposer du gradient de la fonction coût par rapport aux émissions. Celui-ci est obtenu par la méthode adjointe décrite au chapitre 4.

Un calcul adjoint pour une journée de simulation nécessite environ 21 minutes de temps CPU sur un processeur à 3 GHz, ce qui correspond approximativement à un facteur 7 par rapport au temps CPU d'un calcul direct (environ 3 minutes). Chaque itération de la procédure d'optimisation requiert une (ou plusieurs) évaluation de la fonction coût et de son gradient. Par exemple, 50 itérations de BFGS nécessitent un minimum de 20 heures de calculs.

Un test d'arrêt est nécessaire à la procédure itérative de minimisation. Dans cette étude, nous ne nous sommes pas placés dans un contexte opérationnel et nous avons

[†1] On ne s'intéresse qu'aux jours de semaines, les valeurs de week-end étant différentes.

volontairement fixé une condition d'arrêt proche de la précision machine. Il est à noter que l'essentiel de la convergence s'effectue dans les premières itérations de minimisation et qu'il est possible de limiter ce nombre d'itérations sans dégrader le résultat de convergence.

5.3.2 Résultats avec des observations synthétiques non perturbées

L'objectif de ce paragraphe est d'abord de valider la méthode sur une expérience synthétique (observations numériques). La procédure consiste à remplacer les mesures fournies par le réseau d'observations par des sorties numériques du modèle. On se donne des valeurs *a priori* pour les variables de contrôle différentes de celles ayant servies à générer les observations numériques. Ici, on a comme valeur de référence $\alpha_i^t = 1$ et on a choisi de pa

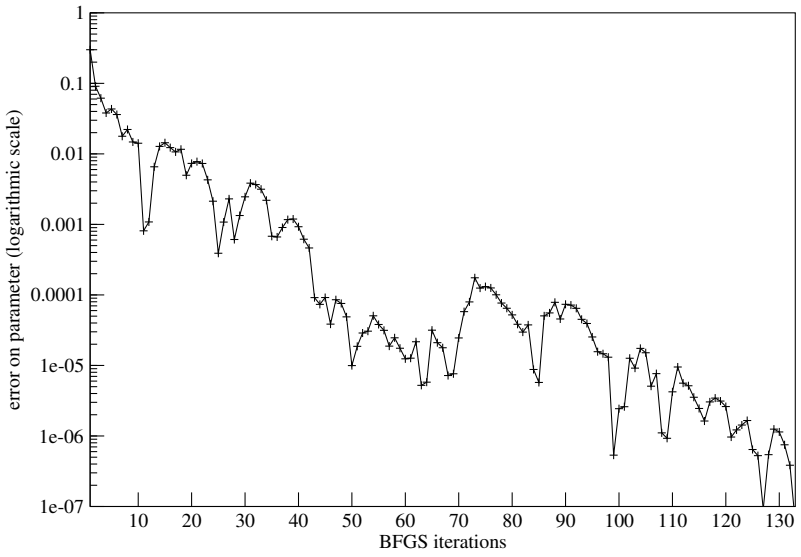

FIG. 5.15 – Erreur sur un paramètre α_8 en fonction des itérations de BFGS.

La figure 5.16 montre la décroissance de la fonction coût en fonction des itérations de l'algorithme de minimisation BFGS indiquant que la trajectoire du modèle après optimisation coïncide avec les observations. Les valeurs de α_i sont parfaitement retrouvées (à la précision numérique près). C'est ce qu'illustre par exemple la figure 5.15 pour le paramètre α_8.

5.3.3 Résultats avec des expériences synthétiques perturbées

Ce paragraphe constitue une étude préliminaire indispensable avant de se confronter à des observations "réelles". Ceci permet notamment d'avoir une estimation de la qualité

[†2]Indices t pour true et b pour background

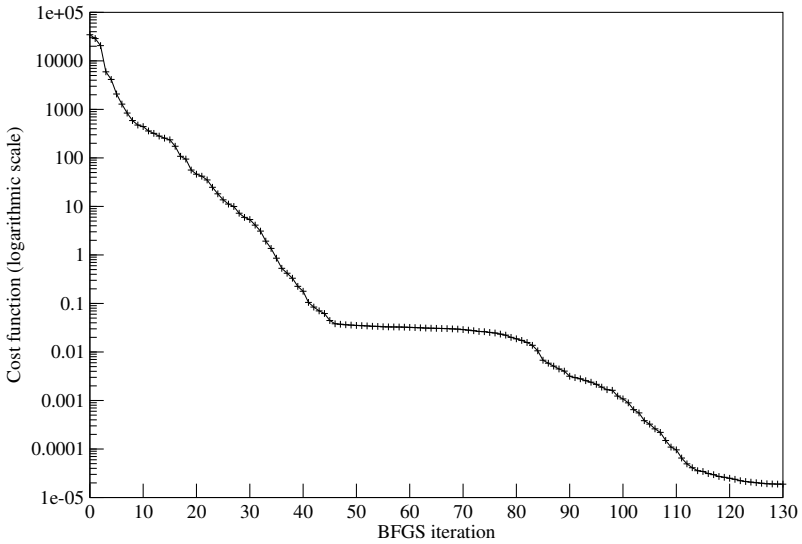

FIG. 5.16 – Fonction coût en fonction des itérations de BFGS.

que l'on peut accorder aux résultats obtenus avec des données réelles.

Nous avons donc envisagé deux expériences :
- la première consiste à perturber directement les observations pour rendre compte d'une erreur d'observation.
- la seconde consiste à lever l'hypothèse de modèle parfait en perturbant les données d'entrées (champs de forçage) utilisées par le modèle. Ceci est une façon simple de générer une erreur modèle.

5.3.3.1 Observations perturbées

Il est possible de perturber les observations en ajoutant une erreur d'observation gaussienne aléatoire aux sorties du modèle. On rappelle que cette erreur contient à la fois une erreur de mesure et une erreur de représentativité qui traduit par exemple le fait que la mesure ponctuelle est comparée à la valeur moyenne d'une maille. La seconde partie est vraisemblablement la plus importante du fait de la chimie rapide autour des sources d'émissions et de l'inhomogénéité des champs à petites échelles. On a appliqué une perturbation gaussienne aux sorties du modèle de la façon suivante :
- décorrélées en espèce et espace.
- corrélées en temps : ceci permet d'éviter que les fluctuations de concentrations soient irréalistes. On imagine par exemple difficilement que les erreurs d'observations varient de +10% à -10% en quelques heures.

Un exemple d'erreur d'observations générées aléatoirement est représentée sur la figure 5.17. On a considéré que l'erreur d'observation de NO était deux fois plus importante que

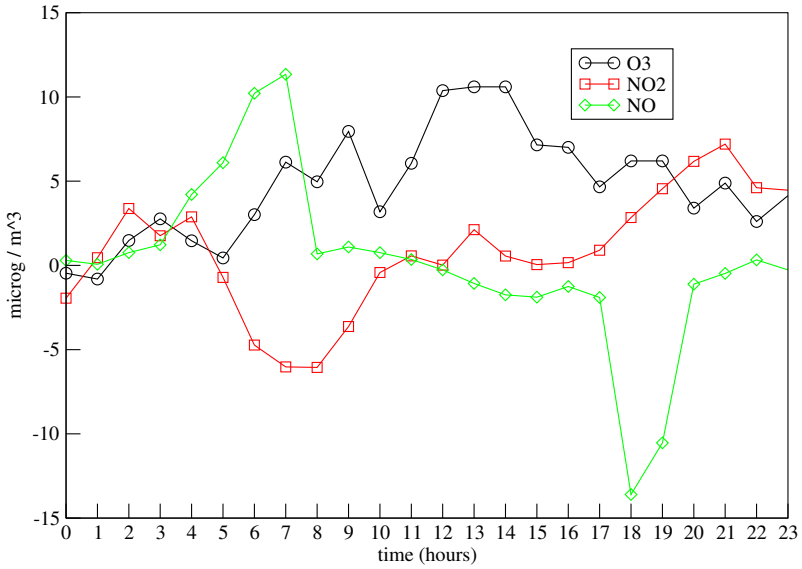

FIG. 5.17 – Erreurs d'observations générées aléatoirement pour une station de mesure de Lille le 11 Mai.

pour NO_2 et O_3 à cause de sa représentativité spatiale plus faible.

Après assimilation, la fonction coût est réduite dans ce cas de 90 % (de 83629.9 à 9155.73). On peut en conclure que la procédure d'assimilation de données a fonctionné correctement et que la trajectoire du modèle est à présent proche des observations. Ceci ne garantit pas la qualité de l'inversion des coefficients α et on peut juste avoir réalisé un bon "fitting" aux observations. La figure 5.18 montre néanmoins que les émissions retrouvées sont assez proches de la "réalité".

5.3.3.2 Erreur modèle

On peut distinguer trois types d'erreur modèle, selon leur origine :
- les champs de forçage : météo, dépôt, conditions aux limites, qui sont incertains,
- les limites de la modélisation : ségrégation sous-maille, hypothèse d'émissions parfaitement mélangées instantanément, etc ..
- Erreurs numériques : le schéma de transport introduit par exemple de la diffusion numérique.

D'une manière ou d'une autre, ces incertitudes contribuent aux différences entre les sorties du modèle et les observations. Par conséquent, la réduction de la fonction coût avec la seule prise en compte des paramètres d'émissions peut conduire à des valeurs erronées. Par exemple, le résultat de l'assimilation peut être identique pour une réduction des émissions de NO_x ou pour une augmentation des vitesses de dépôt de NO et NO_2.

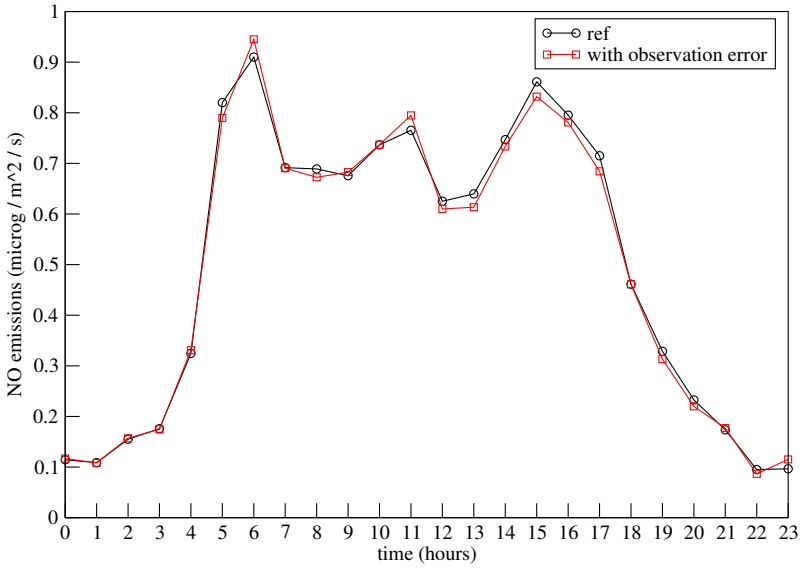

FIG. 5.18 – Emissions de NO : Référence et valeurs obtenues après optimisation dans le cas d'observations synthétiques perturbées aléatoirement.

L'erreur modèle est générée ici en appliquant une incertitude aux paramètres d'entrée listés dans la section 5.2 exceptés les émissions de NO_x (figures 5.19, 5.20, 5.21). On a supposé que les incertitudes suivaient une loi gaussienne d'écart-type 50 %, ce qui est proche des valeurs mentionnées dans [Schmidt et Martin, 2003] ou [Hanna *et al.*, 2001].

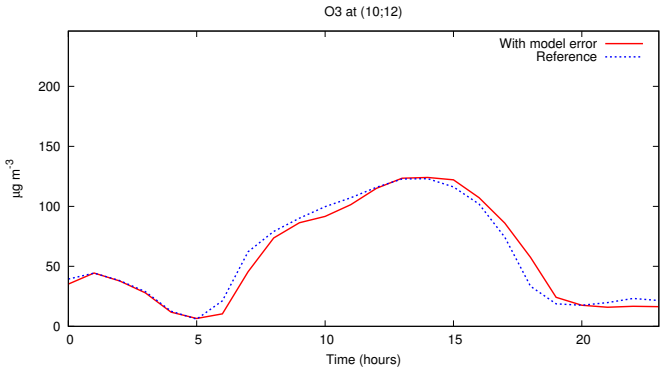

FIG. 5.19 – Concentrations d'O_3 à une station de Lille le 11 Mai. Valeurs de référence et valeurs avec une erreur modèle.

La minimisation apporte une réduction de 85% de la fonction coût et converge vers le profil d'émission représentée sur la figure 5.22. La période 6-7 heures est très sensible et la modélisation inverse peut conduire à des erreurs allant jusqu'à 50 % par rapport aux valeurs réelles.

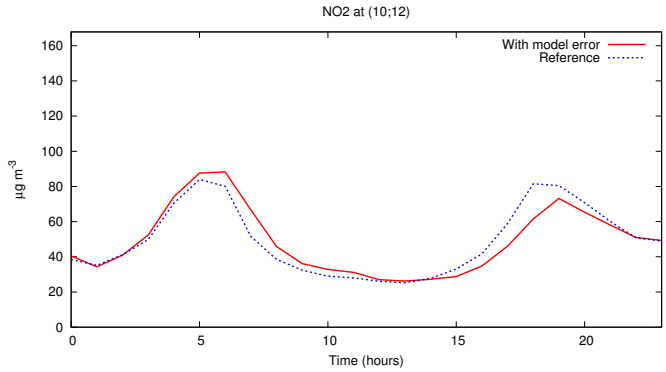

FIG. 5.20 – Concentrations de NO_2 à une station de Lille le 11 Mai. Valeurs de référence et valeurs avec une erreur modèle.

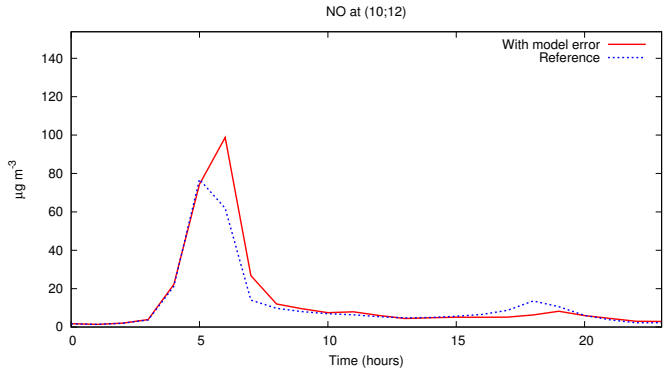

FIG. 5.21 – Concentrations de NO à une station de Lille le 11 Mai. Valeurs de référence et valeurs avec une erreur modèle.

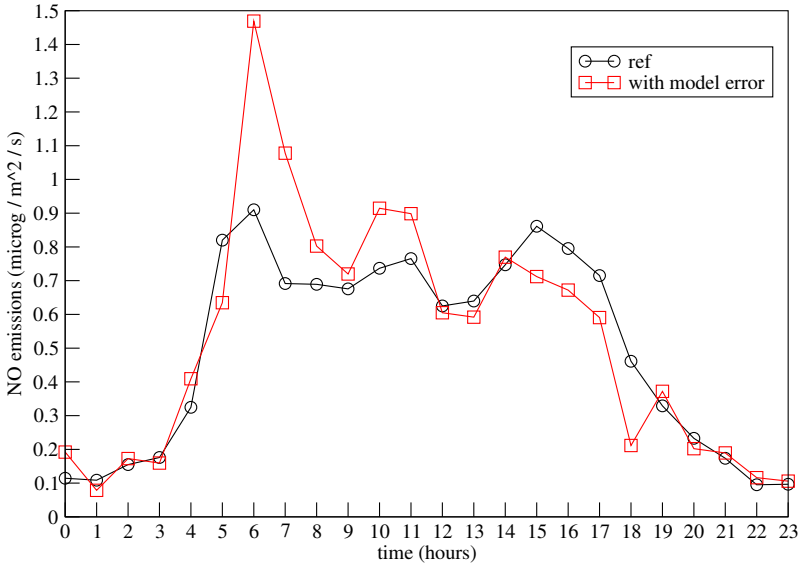

FIG. 5.22 – Emissions de NO : Référence et valeurs obtenues après optimisation dans le cas d'ajout d'une erreur modèle.

5.4 Résultats sur le cas réel

L'étude préliminaire avec des observations synthétiques a montré que :
- La chaîne de calcul était validée (cas des observations non perturbées).
- La modélisation inverse était possible avec des erreurs d'observations ou une erreur modèle.

Dans ce paragraphe, les observations sont issues des stations de mesure du réseau AREMA. Le modèle présente des difficultés sur cette période à reproduire les observations de NO et nous avons décidé dans un premier temps de mener les expériences de modélisation inverse sans ces observations (voir la section 5.4.4.2 pour la prise en compte de ces observations). Par pas de temps horaire, nous disposons donc de 4 mesures d'ozone et 10 mesures de NO2.

5.4.1 Inversion sur la semaine du 11 au 15 Mai

Notre démarche a été la suivante :
- Une chronique temporelle est optimisée sur une période de référence (une semaine servant de base d'apprentissage).
- Elle est ensuite appliquée pour d'autres semaines où l'on vérifie qu'elle a améliorée ou non les résultats (base de vérification).

Les tests ont donc d'abord été menés sur la période du 11 au 15 Mai. Deux types d'expériences sont présentés, afin de tester la variabilité quotidienne des chroniques opti-

misées :

- Chaque jour de la semaine est inversé séparément (5 simulations d'un jour).
- Les coefficients temporels sont optimisés sur toute la période (1 simulation de 5 jours)

Er pour
minin

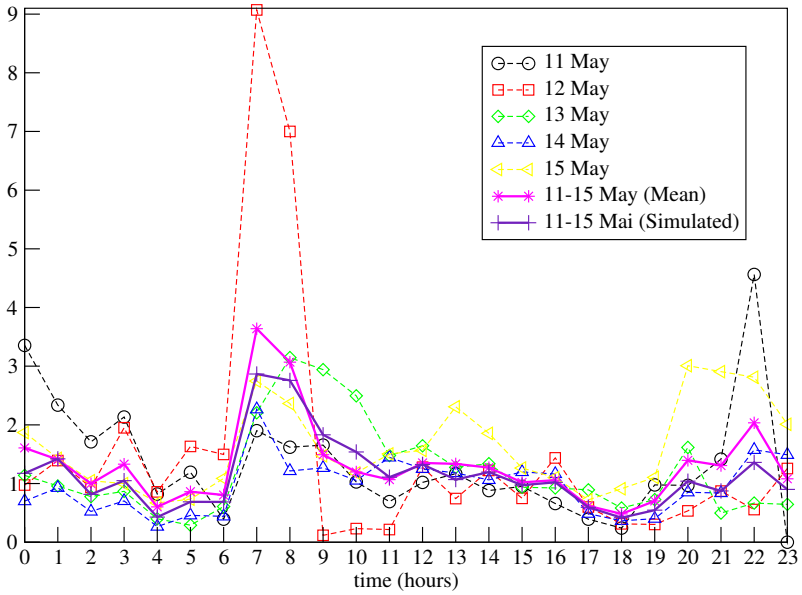

FIG. 5.23 – Chronique journalière optimisée des coefficients multiplicatifs α sur la période du 11 au 15 Mai 1998.

La figure 5.23 représente les coefficients α obtenus par modélisation inverse pour chacune des expériences mentionnées ci-dessus. Pour les simulations d'une journée, la convergence est obtenue après un nombre d'itérations de BFGS compris entre 59 et 128 pour une réduction de la fonction coût allant de 25 % à 66 % (tableau 5.1). L'inversion du 11 au 15 Mai a nécessité 42 itérations et a permis une réduction de la fonction coût de 20 %.

	11 Mai	12 Mai	13 Mai	14 Mai	15 Mai	11-15 Mai
Fonction coût initiale	5.38e+5	9.50e+5	7.65e+5	3.85e+5	2.96e+5	1.6e+6
Fonction coût finale	2.55e+5	3.03e+5	4.40e+5	2.91e+5	2.06e+5	1.28e+6
Itérations de BFGS	101	128	59	64	55	42

TAB. 5.1 – Valeurs initiales et finales de la fonction coût et nombre d'itérations de BFGS pour les jours compris entre le 11 et le 15 Mai 1998.

Les 5 simulations d'un jour font apparaître une variabilité relativement forte sur les coefficients horaires α. Néanmoins, elles présentent des similitudes :

- Les coefficients correspondant aux heures 7 et 8 sont tous au-dessus de 1.
- Les heures 17 à 19 présentent une sur-estimation.

Ces deux périodes sont critiques en terme de modélisation car elles correspondent au pic d'émission du matin et du soir (départ et retour du travail) et en même temps à l'activation-désactivation de la photolyse et de la turbulence verticale. On a également déjà vu dans le paragraphe précédent que la période du matin (7-8 h) est très sensible en terme d'erreur modèle.

Le 12 Mai est un jour mal représenté par le modèle (fonction coût initiale plus élevée) où l'erreur n'est certainement pas due en majorité aux émissions mais à un autre paramètre d'entrée. On observe une sous-estimation des coefficients de l'inventaire dans la nuit du Dimanche au Lundi et dans la nuit du Vendredi au Samedi qui pourrait correspondre aux départs et retours de week-ends.

La courbe étiquetée par "11-15 May (Mean)" correspond à la moyenne effectuée sur les 5 simulations journalières. Elle se compare favorablement à la courbe étiquetée "11-15 May (Simulated)" qui correspond à la résolution couplée sur 5 jours. On peut concevoir que dans un cadre opérationnel, la résolution séparée en parallèle des journées soit envisageable.

La chronique temporelle des émissions du 11 au 15 Mai après optimisation est représentée sur la figure 5.24. La non symétrie entre le matin et le soir (départ et retour du travail) pourrait indiquer que la modélisation inverse a reporté sur les émissions d'autres erreurs du modèle comme par exemple un déclenchement de la turbulence. On peut aussi penser qu[illegible]soir du fait de la

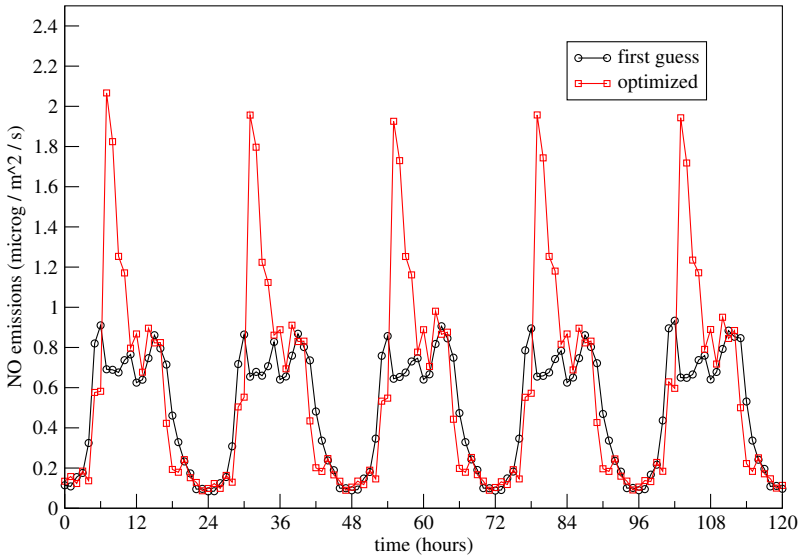

FIG. 5.24 – Chronique journalière des émissions de NO pour la période du 11 au 15 Mai 1998. Référence et optimisation.

5.4.2 Vérification

La question qui est posée dans ce paragraphe est de savoir si la chronique optimisée améliore les prévisions sur les autres semaines. Les RMS et les corrélations pour la période optimisée et deux semaines au-delà sont regroupées dans le tableau 5.2. Tous les scores sont meilleurs avec les émissions optimisées, excepté la RMS du NO pour la période du 25-30 Mai.

	11-15 Mai (ref)	(optim)	18-13 Mai (ref)	(optim)	25-30 Mai (ref)	(optim)
O_3	32.7 (0.83)	29.6 (0.87)	19.9 (0.82)	18.9 (0.85)	20.8 (0.61)	18.1 (0.69)
NO_2	29.4 (0.36)	25.5 (0.52)	19.9 (0.49)	18.4 (0.59)	19.4 (0.24)	17.4 (0.29)
NO	27.4 (0.61)	26.6 (0.66)	19.7 (0.49)	17.7 (0.66)	20.3 (0.38)	21.5 (0.40)

TAB. 5.2 – RMS et corrélations (entre parenthèses) avant (ref) et après optimisation (optim) des émissions sur la période du 11 au 15 Mai.

5.4.3 Optimisation sur d'autres périodes

L'objectif est d'observer la variabilité des paramètres optimisés sur d'autres semaines. On suppose que les émissions des semaines suivantes sont a peu près identiques mais que la situation météorologique a changé. Cela revient donc à tester la robustesse de la méthode par rapport à la situation météo.

La figure 5.25 présente les valeurs de α optimisées pour 3 semaines : la semaine du 11 au 15 Mai et les deux suivantes. On retrouve des caractéristiques communes : valeurs fortes entre 7 et 10 heures et valeurs faibles entre entre 17 et 19 heures. Ceci corrobore les résultats positifs du test de vérification précédent.

5.4.4 Sensibilité du second-ordre

5.4.4.1 Rôle du first guess

L'expérience suivante teste la sensibilité des coefficients temporels optimisés à la valeur initiale ("first guess") dont ce sert l'algorithme itératif BFGS. On a choisi de réaliser la même expérience de modélisation inverse en partant d'une part de la connaissance a priori qui nous est donnée par l'inventaire (coefficients horaires moyen pondérés par des données de comptage routier) et d'autre part du cas extrême où tous les coefficients sont pris égal à zéro (absence totale d'information sur les émissions supposées initialement nulles).

La figure 5.26 montre que l'algorithme converge vers la même chronique temporelle (à quelques différences minimes près). Evidemment le processus de convergence demande un

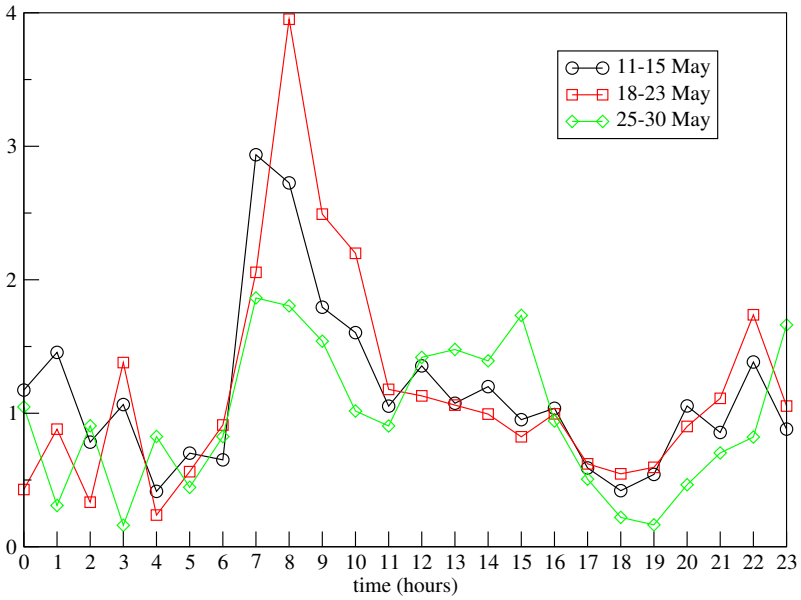

FIG. 5.25 – Chroniques journalières optimisées des coefficients multiplicatifs α pour 3 semaines distinctes.

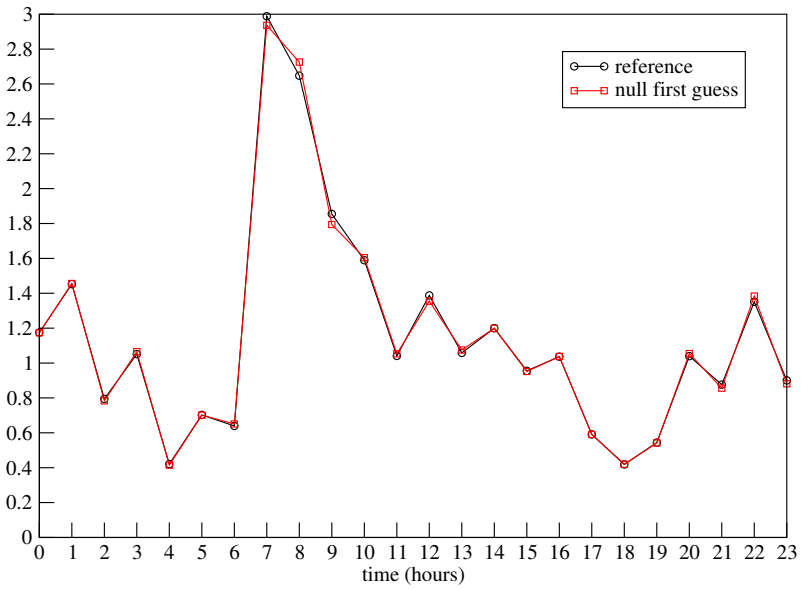

FIG. 5.26 – Comparaison des chroniques journalières optimisées des coefficients multiplicatifs α en partant d'une connaissance a priori et en partant de valeurs nulles.

nombre d'itérations plus important en ayant un point de départ plus éloigné des valeurs optimisées : 27 itérations de BFGS si l'on part de valeurs nulles contre 22 dans le cas où l'on se sert de la connaissance a priori.

L'enseignement principal est qu'une information a priori n'est pas nécessaire pour retrouver la chronique temporelle optimisée. La modélisation inverse peut donc être appliquée même si on ne dispose pas d'inventaire d'émissions. Ceci est un résultat extrêmement intéressant puisqu'il montre d'une part la capacité de la méthode à trouver une information délivrée par un inventaire fin sur la seule base du modèle et des observations et d'autre part que l'algorithme est "robuste".

5.4.4.2 Sensibilité aux espèces observées

La modélisation inverse des coefficients horaires d'émissions a été menées en observant également l'espèce NO (figure 5.27). Un poids égal a été gardé pour toutes les mesures bien que les observations de NO aient une représentativité beaucoup moins grande. La périod ée est peu al

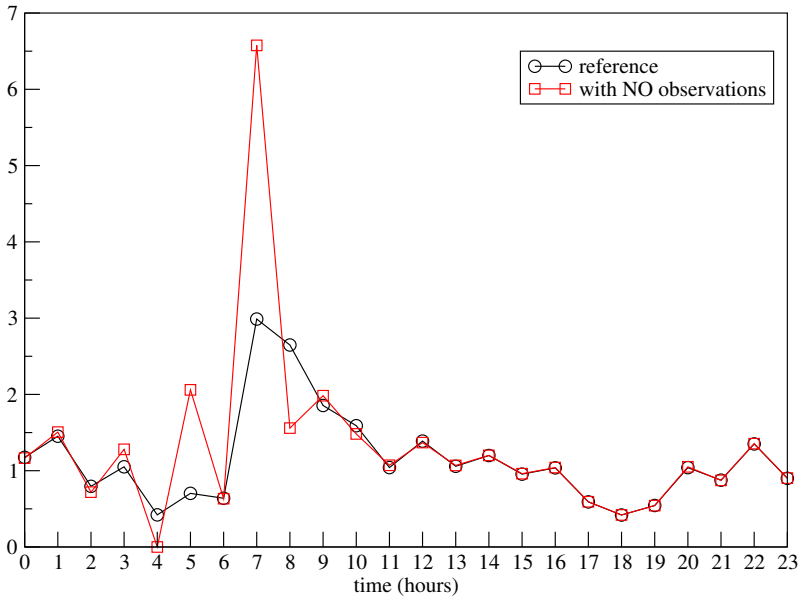

FIG. 5.27 – Chronique journalière optimisée des coefficients multiplicatifs α avec et sans prises en compte des mesures de NO.

5.4.4.3 Robustesse vis à vis d'un décalage temporel de la turbulence

Les expériences précédentes ont montré que le moment de passage de la couche nocturne stable à la couche mélangée est déterminante pour la modélisation inverse des

coefficients temporels d'émissions. La simulation repose sur une paramétrisation de la turbulence calculée avec des champs météorologiques tri-horaires ce qui est selon toute vraisemblance insuffisant pour bien détecter le moment de brassage.

Pour tester la sensibilité du résultat de l'optimisation à la chronique temporelle de la diffusion verticale, nous avons artificiellement décalé la chronique temporelle de diffusion verticale d'une heure. Ceci permet que le brassage de la couche limite nocturne se déclenche une heure plus tard et facilite l'accumulation des polluants pendant une heure supplémentaire.

La figure 5.28 présente les coefficients temporels optimisés sur la journée du 11 Mai avec et sans décalage de la diffusion verticale. Comme prévu, les différences s'observent dans les deux phases critiques de l'évolution de la couche limite : le brassage vertical de la couche nocturne le matin vers 8 heures et la mise en place de la couche nocturne le soir ver ce qui crédibil

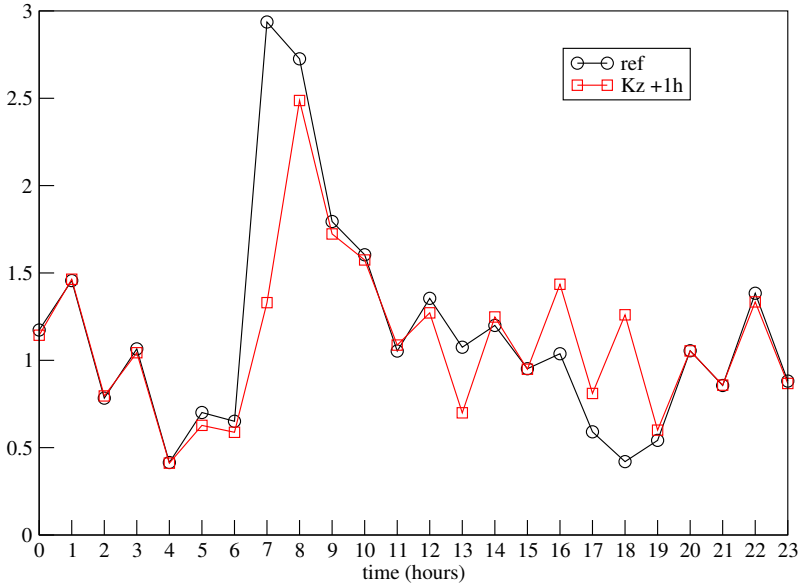

FIG. 5.28 – Chronique journalière optimisée pour le 11 Mai des coefficients multiplicatifs α avec et sans décalage temporel de la diffusion verticale.

Conclusion

Nous avons effectué une modélisation inverse des émissions dans le cas de la simulation régionale de la qualité de l'air sur la région de Lille. Les paramètres de contrôle choisis définissent une chronique temporelle (de pas horaire) appliquée aux émissions de NO_x qui s'avèrent être les paramètres d'émissions les plus sensibles vis-à-vis des espèces chimiques

photochimiques O_3, NO et NO_2.

Des expériences synthétiques (sans et avec perturbations des observations, avec et sans erreur modèle) ont montré la faisabilité de la méthode. L'application à des données réelles pour le mois de Mai 1998 a montré l'apport de l'approche, la chronique optimisée sur une semaine permettant une amélioration notable des résultats sur les deux semaines suivantes, utilisées comme ensemble de validation. Enfin, des tests de robustesse (aux conditions météorologiques et aux incertitudes sur la paramétrisation de la turbulence) ont permis de vérifier que la chronique inversée était "robuste".

Conclusion

Synthèse des travaux réalisés

Ce travail de thèse s'est articulé autour de trois parties :

- *la construction et la validation d'un modèle de Chimie-Transport tridimensionnel, POLAIR3D*, à même de simuler la qualité de l'air aux échelles régionales et continentales (en parallèle avec [Boutahar, 2004, Mallet, 2005]).

  POLAIR3D a été validé et utilisé au cours de ce travail pour la modélisation de la qualité de l'air sur la région de Lille et des études d'impact à divers scénarios de transport. Par ailleurs, POLAIR3D est maintenant largement utilisé pour de nombreuses autres applications.

- *la construction d'un système d'assimilation de données variationnel (4D-var)* qui s'appuie sur la différenciation automatique et la réalisation de tests préliminaires de modélisation inverse d'émission à l'échelle régionale.

  La plate-forme POLAIR3D a été spécifiquement conçue dans une telle perspective ce qui ne va pas sans un certain nombre de contraintes en terme de conception mais permet de disposer d'un environnement d'assimilation de données "immédiat".
  Les tests préliminaires de modélisation inverse sur Lille ont montré la faisabilité, que ce soit sur des données synthétiques que sur des données réelles. Néanmoins, nos tests ont également illustré la grande sensibilité des résultats aux incertitudes sur les autres paramètres, et ce point reste encore largement ouvert.

- *une réflexion sur la "sensibilité de second-ordre"*, qui dépasse pour une part le strict cadre de la dispersion atmosphérique.

  Nous avons proposé un cadre théorique (les systèmes lent-rapide et la théorie de la perturbation singulière) qui permet d'interpréter le comportement des systèmes d'assimilation de données pour des modèles avec une grande dispersion des échelles

de temps (ce qui est caractéristique de la chimie atmosphérique).
Un travail à nature synthétique a permis de restituer la sensibilité au choix du réseau d'observation (plus généralement la thématique du "network design") dans le cadre de théorie de l'optimal design. Cela a permis notamment de faire le lien entre les approches liées au conditionnement du hessien de la fonction coût et celles liées à la sensibilité aux observations.
Enfin, nous avons utilisé et illustré les possibilités des méthodes dites du second-ordre (et fondées sur le calcul du hessien de la fonction coût) dans un cas simplifié, celui de la dispersion de radioéléments à l'échelle locale (avec un modèle gaussien).

Perspectives

De nombreux points restent bien entendu largement ouverts. On citera notamment les points suivants :

- *l'utilisation de formulations faibles pour la chimie atmosphérique et l'introduction d'une erreur modèle :*

  Les modèles de Chimie-Transport sont, on l'a vu, faiblement dépendant aux conditions initiales et les deux incertitudes majeures concernent d'une part les bases de données (notamment pour les flux d'émission) et d'autre part les paramétrisations sous-mailles. Pour le second point, le recours à des formulations faibles peut être une voie prometteuse.
  Ceci revient à introduire une erreur modèle dans la formulation variationnelle ([Durbiano, 2001]) alors que le modèle a été considéré jusque là comme parfait.
  Un point clé est bien entendu de pouvoir réduire l'espace de contrôle et la voie classique, fondée sur l'utilisation des EOFs, peut être prometteuse en chimie atmosphérique. On se réfère par exemple, dans un contexte largement différent, à [Fiore *et al.*, 2003]) et à [Boutahar, 2004] pour l'apport de telles méthodes pour la problématique des modèles de chimie–transport.
  Une voie alternative serait de proposer une paramétrisation de l'erreur modèle (qui est essentiellement liée aux processus sous-mailles non résolus). A ce titre, une convergence avec les travaux consacrés aux effets de ségrégation dans les modèles de chimie-transport pourrait être intéressante.

- *une comparaison avec les méthodes séquentielles :*

  Dans ce travail de thèse, nous n'avons utilisé que des méthodes variationnelles. L'approche séquentielle, essentiellement via des filtres réduits ou des filtres d'ensemble ([Segers, 2002]), a déjà été utilisée et a montré ses apports.

Le montage d'une expérience, sur la base d'un modèle direct identique (POLAIR3D), confrontant les deux approches pourrait être d'une grande utilité, la comparaison de travaux menés sur des modèles, des situations et des bases de données différentes étant d'une faible utilité.

- *l'utilisation de simulations Monte-Carlo pour la modélisation inverse :*

  Nos tests préliminaires sur Lille ont montré la grande sensibilité des résultats de la modélisation inverse aux autres paramètres (notamment la paramétrisation de la turbulence).
  Nos tests ont diagnostiqué et illustré cet état de fait mais n'ont pas apporté de solutions pratiques satisfaisantes. Une voie bien entendu intéressante serait le recours à des simulations Monte-Carlo ([Mallet, 2005]).

- *l'extension à la chimie multiphasique (aérosols) :*

  Nos travaux ont exclusivement porté sur des traceurs gazeux réactifs voire passifs (le cas académique avec les radioéléments). Une thématique d'intérêt croissant est bien entendu liée à la modélisation des aérosols.
  De nombreuses questions seraient posées par l'extension de ce travail à la modélisation des aérosols, au premier rang desquelles la dimension de l'espace de contrôle (les distributions d'aérosols décrivent des espèces chimiques réparties sur un spectre de taille) et les discontinuités des modèles directs (effets de seuil et changement de phase des modèles d'équilibre thermodynamique utilisés pour la paramétrisation des flux de transfert de masse en condensation/évaporation).

Annexe A

Observational network design, adaptive observations and targeting strategies in atmospheric data assimilation : a brief methodological review

Résumé

L'objet de ce chapitre est de faire un point sur la thématique de l'optimisation des réseaux d'observation pour l'assimilation de données. On rappelle notamment dans une première partie la théorie de l'"optimal design" et on présente les principales méthodes appliquées en géophysique.

Ce chapitre, à caractère synthétique, est issu des actes de [Sportisse et Quélo, 2002].

Sommaire

A.1 Optimal design theory 158
A.1.1 Notation 158
A.1.2 The design criterion 160
A.1.3 General Equivalence Theorem of Kiefer-Wolfowitz 161
A.1.4 Theory of optimal location of sensors for dynamical systems . . . 162
A.2 A direct application to data assimilation : the targeting problem 163
A.2.1 Background 163
A.2.2 The targeting problem 164

A.2.3 The design criterion 165
A.2.4 An application of A-design 165
A.2.5 The need for sub-optimal adaptive observations 167
A.3 Singular Vector techniques 167
A.3.1 Introduction 167
A.3.2 Some definitions 168
A.3.3 Choice of different metrics 169
A.3.4 Quasi-inverse and pseudo-inverse 169
A.3.5 Singular Vectors and A-design 171
A.4 Sensitivity techniques 171
A.4.1 Introduction 171
A.4.2 Analysis sensitivity field. First-order sensitivity 171
A.4.3 Observation sensitivity field. Second-order sensitivity 172
A.4.4 Impact of sub-optimal data assimilation schemes 174
A.4.5 Extended observation sensitivity targeting technique 175
A.4.6 Serial strategy 176
A.5 Ensemble approaches 176
A.6 Time scheduling 178

Abstract

The questions of *network design* and of *adaptive* (or *targeting*) observations are becoming a key issue in modern data assimilation systems. In addition to *routine* fixed observation systems, the objective is to define supplementary observations in order to improve the results provided by the data assimilation systems, more generally to optimize the observational network with respect to a given criterion.

The objective of this article is to review the main techniques used in this field. Among them, one can cite adjoint sensitivity techniques with respect to analysis and observation and singular vector approaches. The theoretical background is mainly given by the classical theory of *experimental optimal design*.

Introduction

In the last two decades data assimilation procedures for environmental forecast (meteorology, oceanography and air pollution) have been increasingly improved by the use of different approaches : Kalman filters ([Cohn, 1997]), variational approaches ([Le Dimet et Talagrand, 1986]) and ensemble techniques (following [Epstein, 1969]).

The key point is to couple numerical forecasts and observations provided by a network. The observation data may come from different sources : terrestrial sensors, airborne data, ships, satellital data, aso. All these observations define what we will call the *routine* observations.

A further question may concern the use of supplementary data in order to improve the forecast-observation-assimilation system : this is the question of *adaptive* observations, which is a part of the general problem of *optimal network design*. This *mobile* component of the network may be for instance provided by flights to be specified, unmanned aerial vehicles (UAV), more generally by mobile sensors. Following [Hansen, 1998], one can cite the US Weather Research Program ([Dabberdt et Schlatter, 1996]) : "*... the overarching recommendation ... is that inputs for weather forecast models can best be obtained through the use of composite observing systems together with* **adaptive** *(or* **targeted***) observing strategies employing both in situ and remote sensing...*".

These topics have been already investigated in the past for hurricane forecast. More recently, some measurement campaigns have focussed on such questions : we can cite for instance the Fronts and Atlantic Storm Track Experiment (FASTEX, [Joly et al, 1997]), NORPEX (North Pacific Experiment, [Langland et al, 1999], WSRP (Winter Storm Reconnaissance Program, [Szunyogh et al, 2000]).

For the general question of network design, many issues have to be addressed : what are the "best" spatial and time locations of the observations ? How to measure the impact

of new observations ? What are the appropriate design criteria ? What are the interactions between the forecast system, the data assimilation scheme and the network design strategy ?

A large set of techniques are now available and can be compared even if the best operational design strategy is not known (if existing). We can cite for instance sensitivity analysis and singular vector techniques. The objective of this article is (*to try*) to review these approaches with a methodological focus (that is to say : we will not investigate the specific application cases, especially in meteorology).

This article is organized as follows. In Section 1 we give the background and we present the key features of the theory of optimal experimental design, which provides the underlying theory for such problems. A key result is the so-called General Equivalence Theorem of Kiefer-Wolfowitz, which states the equivalence between the two main approaches of network design (criterion based on the estimated parameter or criterion based on the estimated observations).

The singular vector approach, the sensitivity techniques and ensemble methods are presented in Sections 2, 3 and 4, respectively. Even if Singular Vector approaches may be included in the class of sensitivity techniques, we have decided to isolate them as they are strongly related to the chaotic behavior of atmospheric dynamics. We finish by some (brief) remarks related to time scheduling, that may be pertinent especially for atmospheric chemistry.

We strongly recommend to the reader some key articles, among many others [Berliner *et al.*, 1999, Bishop, 2001] and [Baker et Daley, 2000] since they underline key methodological aspects (the relation to statistical optimal design theory and the need to use observation sensitivity field, respectively). [Fedorov et Hackl, 1997] is one of the key references for optimal design.

A.1 Optimal design theory

A.1.1 Notation

In the following, $N(\bar{z}, Z)$ designs a random Gaussian process whose mean (resp. covariance matrix) is $\bar{z}$ (resp. Z).

The theory of experimental optimal design has been developed for many engineering applications, especially for defining the optimal parameters of a controlled system. We refer for instance to the book [Fedorov et Hackl, 1997].

Let us consider the following linear problem. A sensor labelled by λ_i (typically a spatial position) provides a set of measurements

$$y_i = H(\lambda_i)x + \epsilon^{(o)}(\lambda_i) \tag{A.1}$$

$x \in R^p$ stands for the unknown parameters or the state vector. $\epsilon^{(o)}$ is the observational error and is supposed to follow a normal law, the covariance matrix for the measurement at point λ_i being given by $R(\lambda_i)$.

Let n be the number of observations, associated to the observational network $(\lambda_1, \ldots, \lambda_n)$. In a variational framework, we want to minimize the cost function :

$$J(x) = \sum_{i=1}^{i=n} (y_i - H(\lambda_i)x)^T R(\lambda_i)^{-1} (y_i - H(\lambda_i)x) \tag{A.2}$$

We have supposed that we have enough observations and that no regularization (background term) is necessary.

The solution of the minimization problem (*the analysis*) is given by :

$$x_a = (\mathrm{Hess}_n)^{-1} \left(\sum_{i=1}^{i=n} H(\lambda_i)^T R(\lambda_i)^{-1} y_i \right) \tag{A.3}$$

with the Hessian matrix of the cost function :

$$\mathrm{Hess}_n = \sum_{i=1}^{i=n} H(\lambda_i)^T R(\lambda_i)^{-1} H(\lambda_i) \tag{A.4}$$

Under some classical assumptions (especially that the observations are not correlated), the analysis covariance matrix is given by $P_a = (\mathrm{Hess}_n)^{-1}$.

The theory of optimal design is based on the search for observational networks that "minimize" the analysis covariance matrix. There are two possible (equivalent) interpretations :

- for the *statistical* point of view, this is equivalent to a reduction in uncertainties ;
- for the *numerical* point of view, this is equivalent to lower the condition number of the Hessian matrix (to be inverted) in order to handle with less ill-posed problems.

In order to develop the optimal design theory, it is useful to rewrite the discrete set of observations in a *continuous* way. Let Λ be the space of all possible values for the sensor locations (or parameters) ; let ξ_n be the measure on Λ which puts mass $1/n$ at each of the design points λ_i : that is to say $\xi_n(\lambda) = 1/n$ if λ belongs to the network $(\lambda_1, \ldots, \lambda_n)$, $\xi_n(\lambda) = 0$ otherwise. We rewrite :

$$\mathrm{Hess}_n(\xi_n) = n \int_\Lambda H(\lambda)^T R(\lambda)^{-1} H(\lambda) \xi_n(d\lambda) \tag{A.5}$$

More generally, we define for any positive (discrete or continuous) measure $\xi(d\lambda)$ on Λ ($\xi(\lambda) \geq 0$ and $\int_\Lambda \xi(d\lambda) = 1$), the matrix $M(\xi)$ by :

$$M(\xi) = \int_\Lambda H(\lambda)^T R(\lambda)^{-1} H(\lambda) \xi(d\lambda) \tag{A.6}$$

and we have $\mathrm{Hess}_n(\xi_n) = n\, M(\xi_n)$.

A.1.2 The design criterion

All the standard optimal design criteria are of the following form : choose the network measure ξ in order to maximize $\Phi(M(\xi))$ where $\Phi(.)$ is a functional on the space of non-negative definite symmetric $p \times p$ matrices. The classical criteria are among the following ones :

- *D-optimality* : in this case, $\Phi(M) = -\log(\det(M))$.
  The objective is then to minimize the volume of a confidence ellipsoid (the eigenvalues of M are inversely proportional to the axes of the confidence ellipsoid around x_a).
- *A-optimality* : in this case, $\Phi(M) = \mathrm{Tr}\left(M^{-1}\right)$.
  This may be interpreted as the minimization of an average variance of the components of x.
- *G-optimality* : this is a very interesting case as this is heuristically the simplest criterion for building an algorithm (as shown below with the consequences of the Kiefer-Wolfowitz theorem).
  In this case, the functional $\Phi(.)$ is given by :

$$\begin{aligned} \Phi(M(\xi)) &= \max_{\lambda \in \Lambda} s(\lambda, \xi) \\ s(\lambda, \xi) &= \mathrm{Tr}\left(R(\lambda)^{-1} H(\lambda) M(\xi)^{-1} H(\lambda)^T\right) \end{aligned} \tag{A.7}$$

  $s(\lambda, \xi)$ may be viewed as the variance of the estimated observation at point λ ($H(\lambda)M(\xi)^{-1}H(\lambda)^T$), weighted by the observation confidence matrix at this point. It is sometimes referred as the *sensitivity function of estimated observations*. Intuitively, if $s(\lambda, \xi)$ is large, this means that much information (not too noisy) should be added by using the observation at point λ. $\Phi(.)$ is then defined by the points that should be added to the network. If $\Phi(\xi)$ has a small value, this means that there aren't valuable observational points to be added to the current network ξ and the network is optimal.
  If we integrate over the design space Λ, we obtain after easy algebraic manipulations :

$$\int_\Lambda s(\lambda, \xi) \xi(d\lambda) = \mathrm{Tr}\left(M^{-1}(\xi) \int_\Lambda H(\lambda)^T R(\lambda)^{-1} H(\lambda) \xi(d\lambda)\right) = p \tag{A.8}$$

  This proves that $\max_{\lambda \in \Lambda} s(\lambda, \xi) \geq p$. With the remarks raised before, it is logical to define the network G-efficiency through :

$$e(\xi) = \frac{p}{\max_{\lambda \in \Lambda} s(\lambda, \xi)} \leq 1 \tag{A.9}$$

The Kiefer-Wolfowitz theorem (next subsection) states that for the optimal design $\xi^\star$, $e(\xi^\star) = 1$.

In practice, the properties of G-optimality justify most of the sensitivity techniques that may be used (see next sections).

A.1.3 General Equivalence Theorem of Kiefer-Wolfowitz

A key result is the so-called equivalence theorem of Kiefer-Wolfowitz ([Fedorov et Hackl, 1997]), which states the equivalence between optimal design probability measures and location of maximal sensitivity of the observations. **It also states the equivalence between the G-stability (a response estimation criterion : minimize the variance of the estimated observation) and the D-optimality (a parameter estimation criterion : minimize the variance of the estimated parameter).**

We must first define the notion of derivative of the functional $\Phi(M(\xi))$ with respect to the network measure ξ. If we write δ_λ the unit point mass at location λ, we consider perturbation of the network through the new measure $\xi_\epsilon = (1-\epsilon)\xi + \epsilon\delta_\lambda$ with ϵ supposed to be a vanishing non-negative parameter. The Frechet derivative of $\Phi(M(\xi))$ in the direction δ_λ is then given by :

$$F_\Phi(\xi, \delta_\lambda) = \lim_{\epsilon \to 0} \left(\Phi(M((1-\epsilon)\xi + \epsilon\delta_\lambda)) - \Phi(M(\xi))\right) \tag{A.10}$$

The key result asserts that the three following conditions are equivalent under the condition that Φ is concave on the space of non negative definite symmetric matrices :

1. The network $\xi^\star$ minimizes $\Phi(M(\xi))$.
2. $F_\Phi(\xi^\star, \delta_\lambda) \geq 0$ for all λ.
3. $F_\Phi(\xi^\star, \delta_\lambda)$ attains its minimum at points λ belonging to the optimal network (that is to say : that have positive point measure under $\xi^\star(d\lambda)$).

We do not give the proof (we refer for instance to [Fedorov et Hackl, 1997]) but we illustrate this result with the D-design, which gives a sketch of the algorithms that may be derived from this result.

Let us start from a network located at points $(\lambda_1, \ldots, \lambda_n)$. Let us create a new network by adding a point at λ. The new measure is defined on the basis of the former one by :

$$\xi_{n+1} = \frac{n}{n+1}\xi_n + \frac{1}{n+1}\delta_\lambda \tag{A.11}$$

We have easily with obvious notations :

$$\text{Hess}_{n+1} = \text{Hess}_n + H(\lambda)^T R(\lambda)^{-1} H(\lambda) \tag{A.12}$$

As $n\,M_n = \mathrm{Hess}_n$ and $\Phi(M) = -\log(\det(M))$, this leads to :

$$M_{n+1} = \frac{n}{n+1} M_n \left(I + \frac{1}{n} M_n^{-1} H(\lambda)^T R(\lambda)^{-1} H(\lambda) \right) \tag{A.13}$$

and

$$\Phi_{n+1} - \Phi_n = p\,\log(\frac{n+1}{n}) - \log\,\det\left(I + \frac{1}{n} M_n^{-1} H(\lambda)^T R(\lambda)^{-1} H(\lambda) \right) \tag{A.14}$$

For n large enough, we may approximate :

$$\Phi_{n+1} - \Phi_n \simeq p\,\log(\frac{n+1}{n}) - \frac{1}{n}\mathrm{Tr}\left(M_n^{-1} H(\lambda)^T R(\lambda)^{-1} H(\lambda) \right) \tag{A.15}$$

as $\log\det(1+\epsilon A) \simeq \epsilon\,\mathrm{Tr}(A)$ for a small ϵ.

After obvious algebraic manipulations :

$$\Phi_{n+1} - \Phi_n \simeq p\,\log(\frac{n+1}{n}) - \frac{1}{n} s(\lambda, \xi_n) \tag{A.16}$$

The perturbation $\epsilon \to 0$ used for the Frechet derivative is equivalent to $n \to \infty$. The derivative of ξ at point λ is then given by :

$$\lim_{n\to\infty} (n(\Phi_{n+1} - \Phi_n)) = p - s(\lambda, \xi) \tag{A.17}$$

This means that the optimal network is related to maximal values of the sensitivity function for observations $s(\lambda, \xi)$. This proves as well the relation between D-optimality and G-optimality for continuous measure designs (the equivalence does not necessarily hold for discrete networks, met in practice).

A *natural algorithm* for building an optimal network is then to add to the existing network (let say ξ) the point λ that maximizes the sensitivity function $s(\lambda, \xi)$. A crucial point is to notice that this requires to have a large dimension of the current network (n large enough). In practice, this strategy generally leads to a very slow rate of convergence and these results have to be viewed as "limiting" results for large networks. This gives however an interesting justification of the naive approach.

We refer for instance to [Fadale *et al.*, 1995] for a numerical application of such approaches.

A.1.4 Theory of optimal location of sensors for dynamical systems

For some special cases of criterion Φ (especially the A-design), some theoretical results exists for linear evolution systems. We refer the reader to the work of Bensoussan ([Bensoussan, 1972]) and the book [Omatu et Seinfeld, 1989].

For the model :

$$\frac{dx}{dt} = L(t)x + \epsilon^{(m)}(t) \tag{A.18}$$

we perform the observation $y(t) = H(t,\lambda)x(t)+\epsilon^{(o)}(t,\lambda)$ with λ a space location (a decision variable $\lambda \in \Lambda$) and the observation error $\epsilon^o \sim N(0,R)$.

The evolution of the analysis covariance matrix is then given by a Ricatti equation :

$$\frac{dP_a}{dt} = P_a L^T + LP_a - P_a D(t,\lambda)P_a - D(0,\lambda)\ , \quad P_a(0) = P_0 \tag{A.19}$$

where $D(t,\lambda) = H^T(t,\lambda)R^{-1}H(t,\lambda)$. The key result of [Bensoussan, 1972] is the study of the control problem (which is related to A-design) :

$$\inf_{\lambda\in\Lambda} \mathrm{Tr}(P_a(t,\lambda)) \tag{A.20}$$

Apart the existence and uniqueness results given in [Bensoussan, 1972], some numerical illustrations can be found in [Omatu et Seinfeld, 1989, Omatu et Seinfeld, 1983] with an application to the mono-dimensional heat equation.

A more generalized problem is given by the so-called actuator-sensor problem : one searches simultaneously for sensor location and actuator (typically unknown source terms in geophysical flows) locations. We refer to [Omatu et Seinfeld, 1983] for some theoretical results (especially the opportunity of uncoupling these problems by dual approaches).

A.2 A direct application to data assimilation : the targeting problem

A.2.1 Background

x will stand for the n-dimensional state vector, describing for instance the meteorological fields or chemical species. $x(t_i)$ designs the value of x at time t_i.

We have at our disposal :

– a model giving a *forecast* $x_f(t)$. The nonlinear model reads from time t_i to time t_{i+1} :

$$x_f(t_{i+1}) = F_i(x(t_i),t) + \epsilon_i^{(m)} \tag{A.21}$$

with $\epsilon_i^{(m)}$ the model error drawn from a random Gaussian process $N(0,Q_i)$. The initial condition $x(t_i)$ may be either a forecast $x_f(t_i)$ or an analysis $x_a(t_i)$.
The tangent linear model associated to F is written as :

$$\delta x_f(t_{i+1}) = M_i \delta x(t_i) \tag{A.22}$$

FIG. A.1 – The targeting problem

– An observation network provides some (routines and/or adaptive) observations $y(t)$:

$$y(t_i) = H(t_i)x(t_i) + \epsilon_i^{(o)} \tag{A.23}$$

with $\epsilon_i^{(o)}$ the random observation error drawn from a random Gaussian process $N(0, R_i)$.

– A data assimilation scheme coupling the forecast and the observation gives (whatever the data assimilation is) an analysis x_a. For instance, in a sequential data assimilation scheme :

$$x_a = x_f + K(y - Hx_f) \tag{A.24}$$

where H is the observation operator and K is the gain matrix (related to error statistics of the forecast and of the observation) through :

$$K = P_f H^t (H P_f H^t + R)^{-1} = (P_f^{-1} + H^t R^{-1} H)^{-1} H^t R^{-1} \tag{A.25}$$

P_a (resp. P_f) will stand for the analysis (resp. forecast) covariance matrix in the context of Kalman Filter. We refer more generally to the nomenclature at the end of the article. We have tried to be as close as possible to [Ide *et al.*, 1997].

A.2.2 The targeting problem

In the same vein as [Berliner *et al.*, 1999], we now define the *targeting* problem in the following way (figure A.1) :

– at the *decision time* t_0, the decision to add new observations has to be taken on the basis of some criteria to be defined,
– at the *targeted time* t_1, adaptive observation are made,
– at the *verification time* t_2, a forecast is supposed to be improved.

Of course, we have : $t_0 \leq t_1 \leq t_2$. The decision to launch new observations at time t_1 is equivalent with specifying an observation operator H (more precisely : a modification of the *routine* observation operator). The analysis at time t_1 will then be modified and we hope that this will improve the forecast at t_2, made on the basis of the analysis at t_1.

The main objective is to define, giving all the information available at t_0, the observations needed at t_1 in order to make the forecast at t_2 as good as possible. This is what we will call a *targeting* or an *adaptive* strategy.

The extension to the case of targeting at a sequence of times $(t_{1,1}, t_{1,2} \dots t_{1,n_t})$ is technical but does not really change the key issues investigated here (see [Bishop *et al.*, 2001, Bishop, 2001] for instance or Section 4).

A.2.3 The design criterion

The choice of the *design criterion* is of course a key issue : that is to say, on which basis do we specify a value for H ?

Notice that the *true* observations made at t_1 are not available at t_0. This means that the design criterion does not have to depend on observations. This will lead us to perform the so-called *plug-in* approximation by linearizing future trajectories needed for the network design at the forecast values (and not at the analysis value, obtained after the observational work).

A logical choice is to take a scalar function of an estimation of the forecast error at t_2. The forecast error at t_2 is usually related to the forecast covariance matrix $P_f(t_2)$, which depends on the choice of H (the observation operator at the targeting time t_1). We write in order to underline this dependence :

$$(P_f(t_2))(H) \tag{A.26}$$

In the same way as for the optimal design theory, let Φ be a scalar convex decreasing function of any non singular nonnegative symmetric matrix N (this means that $\Phi(N_1) \leq \Phi(N_2)$ if $N_1 \geq N_2$) among the following choices (respectively reffered as *D-design*, *A-design and E-design*) :

$$\Phi(N) = -\log\left(\det(N)\right) \ , \quad \Phi(N) = \mathrm{Tr}(N^{-1}) \ , \quad \Phi(N) = \lambda_{max}(N^{-1}) \tag{A.27}$$

Once, the criterion Φ has been chosen, the optimal design problem is to find H^* such that :

$$H^* = \mathrm{argmax}\, \Phi\left((P_f(t_2))(H)\right) \ , \quad H \in V_H \tag{A.28}$$

where V_H is the space of possible designs (in practice not all the locations are allowed).

Let us notice that an economic cost may be added to the previous criterion.

A.2.4 An application of A-design

We now follow [Berliner *et al.*, 1999] with an application to the adaptive observation problem in a Kalman filter context.

Let $X_0 \sim N(x_a(t_0), P_a(t_0))$ the knowledge at initial time t_0. Let us recall that we want to minimize (through a given design function Φ) the forecast error covariance matrix $P_f(t_2)$ at verification time.

We have ([Cohn, 1997]) :

$$P_f(t_2) = M_1 P_a(t_1) M_1^t + Q_1 \tag{A.29}$$

with M_1 the tangent linear model between t_1 and t_2, linearized around the trajectory issued from analysis at t_1 ($x_f(t|x_a(t_1))$). Q_1 is the related error model covariance matrix. The analysis error covariance matrix at target time t_1 is a function of adaptive observation design $(H(t_1), R(t_1))$ through :

$$P_a(t_1) = \left(P_f(t_1)^{-1} + H(t_1)^t R(t_1)^{-1} H(t_1)\right)^{-1} \tag{A.30}$$

with $H(t_1)$ the (unknown) observation matrix and $R(t_1)$ the related observation error covariance matrix. The forecast (background) error covariance matrix $P_f(t_1)$ at target time t_1 is given by :

$$P_f(t_1) = M_0 P_a(t_0) M_0^t + Q_0 \tag{A.31}$$

The notations are coherent with those used in Eq. (A.29).

We now choose a design criterion Φ and we try to minimize $\Phi(P_f(t_2))$ with respect to the adaptive observation operator $H(t_1)$.

Given as such this problem cannot be solved *a priori* since it requires the knowledge of $M_1 = (\frac{\partial F_1}{\partial x})_{x_f(t|x_a(t_1))}$ where F_1 is the nonlinear model between t_1 and t_2. The analysis at target time indeed depends on observations $y(t_1)$ (not known at decision time t_0) :

$$x_a(t_1) = x_f(t_1) + K_1(y(t_1) - H(t_1)x_f(t_1)) \tag{A.32}$$

with K_1 the gain matrix at time t_1.

In order to make the strategy tractable at the decision time, one systematically replaces in the above equations the tangent linear model M_1 at the analysis trajectory by the tangent linear model $\tilde{M}_1$ at the forecast trajectory ($\tilde{M}_1 = (\frac{\partial F_1}{\partial x})_{x_f(t|x_f(t_1))}$).

For the specific case of A-design, this problem is tractable (in theory). For the single observation case, with

$$H_i(t_1) = (0\,0 \ldots 0\,1\,0 \ldots 0) \ , \quad R_i(t_1) = \epsilon_i^2 \tag{A.33}$$

the design criterion leads after some easy computations (purely algebraic manipulations) to *maximize* over i :

$$\frac{((\tilde{M}_1 P_f(t_1))^t (\tilde{M}_1 P_f(t_1)))_{ii}}{\epsilon_i^2 + (P_f(t_1))_{ii}} \tag{A.34}$$

Notice the trade-off between the dynamical effect (propagation of error through the tangent linear model) and analysis error related to the adaptive observation.

For the multi-observations case (let p_a be the number of adaptive observations), the problem has a combinatorial nature as $C_n^{p_a}$ possibilities exist. However, in the perfect observation case, the A-optimal $H_A(t_1)$ may be computed through the generalized eigenvalue problem :

$$\tilde{M}_1^t \tilde{M}_1 \mathbf{u_i} = \lambda_i \, P_f(t_1)^{-1} \mathbf{u_i} \tag{A.35}$$

and :

$$H_A(t_1) = \begin{bmatrix} \mathbf{v_1}^t \\ \dots \\ \mathbf{v_{p_a}}^t \end{bmatrix} \quad , \quad \mathbf{v_i} = (P_f(t_1))^{-1}\mathbf{u_i} \tag{A.36}$$

Notice that this is an idealized solution with strong requirements : only one target time, perfect observations and before all no constraints for the observation operator (in practice, H will be surely composed by 0 and 1, that is with H_i-like lines).

A.2.5 The need for sub-optimal adaptive observations

As a conclusion to this mathematical overview, the theoretical background for adaptive design is well founded. Up to our knowledge, there are no applications of the works of [Bensoussan, 1972] in the area of optimal design for data assimilation. The same occurs for the use of iterative numerical strategies as presented in [Fedorov et Hackl, 1997].

In an ideal world, one could compute for a given discrete set of Q possible observations $H^{(q)}$ the criteria $\Phi(H^q)$ and search for the minimal value. If the target area is not highly constrained, this is of course not realistic for atmospheric applications, as this leads to a combinatorial problem.

As a practical consequence, we have to use *sub-optimal* strategies for our adaptive observation problem.

A.3 Singular Vector techniques

A.3.1 Introduction

For chaotic systems, the definition of initial conditions is of course a crucial issue, since small initial modifications may degenerate in large errors. Adaptive techniques related to singular vectors techniques ([Buizza et Montani, 1999]) are deeply related to this framework : one searches for the initial perturbations that may have the largest growth. A study of the key role played by singular vectors for the meteorological dynamics may be found in [Buizza et Palmer, 1995].

Let us notice that an underlying assumption is that the evolution of the initial error may be estimated with the tangent linear model in short term (typically two days, as discussed in [Pu *et al.*, 1997]).

In the following we will consider a perfect model ($Q = 0$). We refer more generally to [Hansen, 1998] for the impact of model errors in this context.

A.3.2 Some definitions

Let us consider the tangent linear model :

$$\delta x_2 = M \delta x_1 \tag{A.37}$$

where M stands for the linear propagator of perturbations from time t_1 (the targeting time) to time t_2 (the verification time). For the sake of clarity, we omit the time dependency ($M = M_{1\rightarrow 2}$). If the nonlinear model is for $t_1 \leq t \leq t_2$:

$$x(t) = F_1(x(t_1), t) \tag{A.38}$$

remember that, after use of adaptive observations at t_1, which defines the analysis $x_a(t_1)$, the appropriate propagator would be :

$$M = (\frac{\partial F_1}{\partial x})_{x(t|x_a(t_1))} \tag{A.39}$$

where $x(t|x_a(t_1)) = F_1(x_a(t_1), t)$ is the trajectory issued from $x_a(t_1)$. As this depends on the observations, this is of course not usable in an *a priori* computation, and we rather use :

$$\tilde{M} = (\frac{\partial F_1}{\partial x})_{x(t|x_f(t_1))} \tag{A.40}$$

where $x(t|x_f(t_1)) = F_1(x_f(t_1), t)$ is the trajectory issued from $x_f(t_1)$. $\tilde{M}$ can be computed at the design time, since it does not depend on (future) observations. This is the *plug-in* assumption, as mentioned in [Berliner *et al.*, 1999].

Before assessing the amplitude of perturbations, one has of course to define a *metric* (that is a particular weighting of the perturbations). We suppose *a priori* that the initial metric and the final metric are not the same ones : let W_1 and W_2 be these metrics, respectively. They are symmetric positive matrices. For instance, $(u, W_1\, u)$ designs a measure of an initial perturbation u (in the sequel, (.,.) stands for the usual Euclidean scalar product). We will note $(.,.)_1$ (resp. $(.,.)_2$) the scalar products defined at target (resp. verification) time.

We have :

$$(\delta x_2, \delta x_2)_2 = (\delta x_2, W_2 \delta x_2) = (\tilde{M} \delta x_1, W_2\, \tilde{M} \delta x_1) \tag{A.41}$$

and therefore :

$$(\delta x_2, \delta x_2)_2 = (\tilde{M}^T W_2\, \tilde{M} \delta x_1, \delta x_1) = (W_1^{-1} \tilde{M}^T W_2\, \tilde{M} \delta x_1, W_1 \delta x_1) \tag{A.42}$$

that is to say :

$$(\delta x_2, \delta x_2)_2 = (W_1^{-1} \tilde{M}^T W_2\, \tilde{M} \delta x_1, \delta x_1)_1 \tag{A.43}$$

As we want to maximize the ratio $(\delta x_2, \delta x_2)_2 / (\delta x_1, \delta x_1)_1$, this gives raise to a Rayleigh-like problem, whose solution is given by the eigenvalue problem :

$$W_1^{-1} \tilde{M}^T W_2\, \tilde{M}\, \mathbf{u_i} = \sigma_i^2 \mathbf{u_i} \tag{A.44}$$

or in an equivalent formulation :

$$\tilde{M}^T W_2 \, \tilde{M} \, \mathbf{u_i} = \sigma_i^2 W_1 \mathbf{u_i} \tag{A.45}$$

$\mathbf{u_i}$ (resp. σ_i^2) is referred as a singular vector (resp. a singular value). They depend of the choice of the metrics W_1 and W_2.

Let us notice to conclude that the search for this (weighted) singular value decomposition is equivalent with the search for the Euclidean singular value decomposition of $W_2^{1/2} \tilde{M} W_1^{-1/2}$. If $\tilde{u}_i$ are the singular vectors for this SVD, one has easily $u_i = W_1^{-1/2} \tilde{u}_i$.

A.3.3 Choice of different metrics

Solving the SVD problem (A.45) is deeply related to the choice of metrics. The choice $W_2 = I$ and $W_1 = P_a^{-1}$ (where P_a is the analysis error covariance matrix at initial time) is often recommended ; let us notice that this is quite logical to weight the initial perturbation by the confidence given in the analysis at initial time (whose estimation is provided by P_a). Such a metric is usually designed as the Mahalanobis metric ([Stephenson, 1997]).

In this case, the targeting problem may be compromised since it is extremely difficult to estimate the analysis covariance matrix P_a (let us notice moreover that, usually, we only know the inverse of P_a, namely the Hessian matrix in a variational context). Many approaches have already been proposed in order to circumvent this problem :

- Simpler metrics are sometimes used. An initial metric based on total energy (resp. kinetic energy) is usually advocated ([Palmer *et al.*, 1998]), which defines the so-called TESV (resp. KESV) for Total Energy Singular Vectors (resp. Kinetic Energy Singular Vectors).
- the inverse of P_a may be replaced in a variational context by the Hessian matrix of the cost function, or even by (truncated) eigenvalue representations ([Barkmeijer *et al.*, 1998] and the notion of Hessian Singular Vectors).
- Some appropriate algorithms are recommended for solving the generalized eigenvalue problem (A.45) when the operators are not explicitly known ([Barkmeijer *et al.*, 1998]).

A.3.4 Quasi-inverse and pseudo-inverse

Singular vector techniques have been used in order to trace back forecast perturbations at t_2 to the original error at targeting time t_1. The forecast perturbation may be either a forecast error (the difference between a forecast and a verifying analysis) in an *a posteriori* use, or a difference between two forecasts in an *a priori* use.

We refer for instance to [Reynolds *et al.*, 2000] or [Gelaro *et al.*, 1998]. For the sake of clarity, let the metrics be equal : $W_1 = W_2 = W$. We note $\tilde{M}_W = W^{1/2} \tilde{M} W^{-1/2}$. It is

easy to prove that ($v_i = \tilde{M}_W\, u_i/\sigma_i$) is an orthonormal basis at time t_2 (for the Euclidean scalar product) ; we can then write :

$$\tilde{M}_W = \sum_{i=1}^{i=n} \sigma_i\, \mathbf{v_i}\mathbf{u_i}^t \tag{A.46}$$

which defines the so-called Moore-Penrose *pseudo-inverse* $\tilde{M}^\dagger$ ([Golub et Van Loan, 1986]) as :

$$\tilde{M}_W^\dagger = \sum_{i=1}^{i=n} \frac{1}{\sigma_i}\, \mathbf{u_i}\mathbf{v_i}^t \tag{A.47}$$

with the classical properties, such as $\tilde{M}\tilde{M}^\dagger\tilde{M} = \tilde{M}$. For $\tilde{M}$, we have then :

$$\tilde{M}^\dagger = W^{-1/2} \sum_{i=1}^{i=n} \frac{1}{\sigma_i}\, \mathbf{u_i}\mathbf{v_i}^t W^{1/2} \tag{A.48}$$

By keeping the first $N < n$ dominant singular values (supposed to be ordered in a decreasing way), we can approximate :

$$\tilde{M}_N^{-1} \simeq W^{-1/2} \sum_{i=1}^{i=N} \frac{1}{\sigma_i}\, \mathbf{u_i}\mathbf{v_i}^t W^{1/2} \tag{A.49}$$

This truncated development defines the *pseudo-inverse* of $\tilde{M}$.

Let us suppose now we want to track back to the initial time t_1 :

– either a forecast error at t_2 in an *a posteriori* use,

– or a forecast difference between two models, for instance in the framework of an ensemble prediction system ([Pu et Kalnay, 1999]), in an *a priori* use.

Let δx_2 be this forecast "perturbation"; we can compute :

$$\delta x_1 = \tilde{M}_N^{-1}\, \delta x_2 \tag{A.50}$$

Areas in which δx_1 is large are supposed to reveal target zones.

Another methods of the same kind are the quasi-inverse methods ([Pu et Kalnay, 1999, Pu *et al.*, 1997]) : the backward operator is then approximated by reversing time and changing the sign of dissipative terms in the forward operator. Comparison between these approximated inverse operators is investigated for instance in [Reynolds *et al.*, 2000, Pu *et al.*, 1997].

Let us underline the fact that the key underlying assumption is the dominance of "reversible" (non dissipative and adiabatic) dynamics in the validity domain of such studies ([Pu *et al.*, 1997]).

A.3.5 Singular Vectors and A-design

As noticed in [Berliner *et al.*, 1999], the singular vector technique may be related to the A-design described in Section A.2 with perfect observation. With $W_2 = I$ and $W_1 = P_f(t_1)^{-1}$, the observation operator is chosen for the first p_a singular eigenvalues as :

$$H_{SV} = \begin{bmatrix} \mathbf{u_1}^t \\ \dots \\ \mathbf{u_{p_a}}^t \end{bmatrix} \tag{A.51}$$

where $\mathbf{u_i}$ are solutions of Eq. (A.35). We have then directly through Eq. (A.36) :

$$H_A = H_{SV} P_f(t_1)^{-1} \tag{A.52}$$

with H_A the observation operator for the A-design.

A.4 Sensitivity techniques

A.4.1 Introduction

The key idea is to locate sensors in places towards which the forecast values of the verification area have the strongest sensitivity. We refer for instance to the work of [Langland et Rohaly, 1996]. The key point is then to choose a so-called *forecast aspect*, a scalar value, we will write h. For instance, h is the average vorticity in a vertical column inside a particular region related to a cyclone location in [Langland et Rohaly, 1996] or enstrophy in [Bergot *et al.*, 1999]. More generally, in another context than meteorological forecast, h stands for any "key" feature of forecast.

With our notations, h is a function of $x_f(t_2) = x(t_2|x_f(t_1))$.

A.4.2 Analysis sensitivity field. First-order sensitivity

The first approach computes the so-called *analysis sensitivity field* of h, that is to say the magnitude of the sensitivity of h to variations of initial conditions at t_1. This is estimated by computing the field :

$$\nabla_{x_a} h \tag{A.53}$$

where x_a stands for the initial condition (analysis). Let us underline that, despite this quite ambiguous notation, we implicitly use $x_a = x_f(t_1)$ in the following (plug-in assumption).

We have then quite easily for a perturbation δx_a the resulting perturbation for the forecast aspect :

$$\delta h = (\nabla_{x_f(t_2)} h, \delta x_f(t_2)) = (\nabla_{x_f(t_2)} h, \tilde{M} \delta x_a) = (\tilde{M}^t \nabla_{x_f(t_2)} h, \delta x_a) \tag{A.54}$$

which defines :

$$\nabla_{x_a} h = \tilde{M}^t \nabla_{x_f(t_2)} h \tag{A.55}$$

Adjoint models are then a key tool for computing such sensitivity fields. Let us notice once more that we have implicitly chosen Euclidean metrics, otherwise with our previous notations :

$$\nabla_{x_a} h = W_1^{-1} \tilde{M}^t W_2 \nabla_{x_f(t_2)} h \tag{A.56}$$

Notice the relation with the pseudo-inverse and quasi-inverse approaches as introduced in the previous section. In [Rabier *et al.*, 1996], this sensitivity pattern is used in order to adjust a forecast difference δx_2 with respect to the initial conditions. If h is the forecast difference (see previous section), adjustment with respect to x_a may be performed with a gradient method (h being the cost function). This leads to iterative perturbations of x_a proportional to $\partial h / \partial x_a$. Notice the deep interaction with singular vectors ([Gelaro *et al.*, 1998] or [Pu *et al.*, 1997]) ; if $h = (\delta x_2, \delta x_2)_2$, one has easily :

$$\nabla_{x_a} h = W_1^{-1} \tilde{M}^t W_2 \delta x_2 \tag{A.57}$$

With Eq. (A.46), this proves that the perturbation is mainly related to dominant singular vectors, on the contrary to the pseudo-inverse technique (Eq. (A.48)). We refer to [Pu *et al.*, 1997, Kalnay et Pu, 1996] for the comparison of such approaches.

Let us mention for the sake of completeness the notion of *influence function* introduced in the context of atmospheric chemistry by [Fisher et Lary, 1995] and extended by [Daescu et Carmichael, 2002]. With rather sloppy notations, the state vector x is now composed by chemical concentrations $c_i(\mathbf{X}, t)$ (species i, time t and position $\mathbf{X}$). The influence function is then :

$$\Gamma_{i,x,t_n}(j, \mathbf{X}', t_m) = \frac{\partial c_i(\mathbf{X}, t_n)}{\partial c_j(\mathbf{X}', t_m)} \frac{c_j(\mathbf{X}', t_m)}{c_i(\mathbf{X}, t_n)} \tag{A.58}$$

in order to assess the influence of species j at time t_m at position $\mathbf{X}'$ for species i at time t_n and position $\mathbf{X}$. Adjoint models are of course useful for computing such functions. We refer to [Daescu et Carmichael, 2002] for an application to the comparison of two existing networks through the use of influence function.

A.4.3 Observation sensitivity field. Second-order sensitivity

As investigated in [Bergot, 2001], the impact of adaptive observation on forecast error strongly depends on the situation and the data assimilation scheme. For instance, for 3D-var methods, some failures have occurred during FASTEX ([Bergot, 2001]) on the

contrary to the case, for which 4D-var approaches are used. In these cases, the use of adaptive observations has increased the forecast error.

Following the work of Baker ([Baker et Daley, 2000]), the interaction between the forecast system, the observations and the data assimilation system has actually to be taken into account. This covers for instance :

– the coupling of a *routine* existing network and adaptive observations,

– the need for taking into account observation statistics in the adaptive strategy.

For instance, one may wonder if defining target area in an area towards which there is a large analysis sensitivity but with a rather poor performance of the observation system is really appropriate.

As a consequence, the more appropriate sensitivity field is the *observation sensitivity field*, since it indicates the impact of observations to be taken.

Following [Baker et Daley, 2000] in a sequential context, one has at target time t_1 :

$$x_a = x_b + Kd \ , \quad d = y - Hx_b \tag{A.59}$$

with d the innovation vector, x_b the background state (typically a forecast), K the gain matrix given by :

$$K = BH^t(HBH^t + R)^{-1} = (B^{-1} + H^tR^{-1}H)^{-1}H^tR^{-1} \tag{A.60}$$

with B, R the error covariance matrices for background and observation, respectively. H is the observation operator to be defined for the adaptive observations.

We have then easily for a perturbation δy of the observations :

$$\delta h = (\nabla_{x_a} h, \delta x_a) = (\nabla_{x_a} h, K\delta y) = (K^t \nabla_{x_a} h, \delta y) \tag{A.61}$$

The *observation sensitivity field* is given by :

$$\nabla_y h = K^t \nabla_{x_a} h = R^{-1} H P_a \nabla_{x_a} h \tag{A.62}$$

since $P_a = (B^{-1} + H^tR^{-1}H)$.

Notice that there is a strong relation to the G-optimality defined in optimal design theory if $h(x_a) = Hx_a$ (the sensitivity function of optimal design theory is then given by $Tr(\nabla_y h)$).

This means that the observation sensitivity strongly depends on the data assimilation system (for instance through the observation statistics). A key relation, as noted in [Baker et Daley, 2000] or [Baker, 2000], is the analysis space projection of the observation sensitivity field, which gives an insight of the relative sensitivity to the observation as compared to the background sensitivity :

$$\nabla_{x_a} h = H^t \nabla_y h + \nabla_{x_b} h \tag{A.63}$$

In [Baker et Daley, 2000], some strategies are proposed in order to use the observation sensitivity field for adaptive issues :

– *single-observation* fields may be (easily) computed.
One computes the observation sensitivity field for each single observation labelled by i :

$$H_i = (0 \dots 0\,1\,0 \dots 0) \ , \quad R_i = \epsilon_i^2 \tag{A.64}$$

We obtain after the use of Eq. (A.62) and some (simple but tedious) algebraic manipulations :

$$\frac{\partial h}{\partial y_i} = \frac{1}{\epsilon_i^2 + B_{ii}} \sum_{j=1}^{i=n} B_{ij} (\frac{\partial h}{\partial x_a})_j \tag{A.65}$$

– *marginal single-observation* may be (less easily) computed.
One tries to estimate the *marginal* impact of an adaptive observation added to a routine data assimilation system, defined by an observation operator $H^{(r)}$ (with the notations of [Bishop *et al.*, 2001]). If p_r is the dimension of the routine observation, one has then to compute for each i in Eq. (A.62) the sensitivity field with the observation operator :

$$H_i^{(tot)} = \left[\begin{array}{c} H^{(r)} \\ H_i \end{array} \right] \tag{A.66}$$

Let us notice that $(p_r + 1) \times (p_r + 1)$ systems have to be solved for each i.

The way analysis sensitivity and observation sensitivity are coupled is deeply investigated in [Baker et Daley, 2000] and [Baker, 2000] with a focus on scales impact (in the same vein as [Daley, 1991], Section 4.5). The notion of *supersentivity* is defined : situations which for, observation sensitivity is much larger than analysis sensitivity. The key result is that supersensitivity is furthered by large scale target, accurate observations, and similar magnitude of background (forecast) and analysis sensitivity scales.

Let us notice to conclude that the extension of this approach to the 4D-var case is deeply related to the notion of second-order adjoint model as defined in [Wang *et al.*, 1992] or [Ngodock, 1996]. In a 4D-var context, sensitivity observations are deeply related to the inverse matrix of Hessian of J (the cost function). This can be illustrated in a 3D-var context with Eq. (A.62) by keeping in mind that $P_a = Hess(J)^{-1}$.

In this sense, there is a strong correlation between a "good" observational network (to be adjusted by adaptive techniques) and a "nice" numerical behavior of a variational data assimilation system (indicated for instance by a condition number for the inversion of magnitude 1).

A.4.4 Impact of sub-optimal data assimilation schemes

Let us investigate now the question of sub-optimal data assimilation schemes, that is to say practical cases for which the background (forecast) covariance matrix is only propaga-

ted along time with some approximation. We refer to [Bishop *et al.*, 2001] for a complete investigation. The key result is the better behavior of 4D-var approaches ([Bergot, 2001]).

Estimation theory provides an appropriate framework for investigating this point : why may observations increase error variance when associated with sub-optimal schemes ?

In a Kalman filtering context, we define the *signal* as the impact of the assimilation of adaptive observation (at t_1), through :

$$s = x(t_2|x_a(t_1)) - x(t_2|x_f(t_1)) \simeq M(x_a(t_1) - x_f(t_1)) = MKd \tag{A.67}$$

with $d = y - Hx_f(t_1)$ the innovation vector at t_1 and M the tangent linear model from t_1 to t_2.

A classical computation leads to the following estimation of the signal covariance matrix (parameterized by observation design H and R) :

$$S(R,H) =< ss^t >= \tilde{M}K(R + HP_f(t_2)H^t)^{-1}K^t\tilde{M}^t \tag{A.68}$$

Let us now compute the forecast covariance matrix at the verification time t_2) :

$$P_f(t_2) = MP_a(t_1)M^t + Q \tag{A.69}$$

with Q the model error covariance matrix ([Cohn, 1997]) and $P_a(t_1)$ the analysis error covariance matrix (at target time t_1) :

$$P_a(t_1) = P_f(t_1) - P_f(t_1)H^t(HP_f(t_1)H^t + R)^{-1}HP_f(t_1) \tag{A.70}$$

We have then easily with the appropriate insertion :

$$P_f(t_2) = MP_f(t_1)M^t + Q - S(R,H) \tag{A.71}$$

This implies that, with an optimal scheme ($P_a(t)$ and $P_f(t)$ exactly computed), the assimilation of (adaptive) observation always reduces the forecast error covariance matrix (in the sense of nonnegative symmetric matrices).

If the propagation equation for the forecast covariance matrix Eq. (A.69) is not solved and one rather uses an approximation $\tilde{P}_f \simeq P_f$ for the gain matrix K, the sign of $P_f(t_2) - (MP_f(t_1)M^t + Q)$ is not necessarily negative. This means that observations may increase the forecast error.

A.4.5 Extended observation sensitivity targeting technique

In [Bishop, 2001], the computation of the variance of the perturbation δh as a function of perturbation in the innovation vector is proposed (this is referred as the *extended observation sensitivity targeting technique*). With (obviously) sloppy notations, we write $h(d)$ the dependence of the forecast aspect with respect to adaptive observations. After linearization and some easy calculations :

$$< (\delta h)^2 >=< (\frac{\partial h}{\partial d}, d)^2 >= (\frac{\partial h}{\partial x_f(t_2)})^t S(\frac{\partial h}{\partial x_f(t_2)}) \tag{A.72}$$

Notice the relations with the points above.

A.4.6 Serial strategy

A sub-optimal strategy has been proposed in [Bishop *et al.*, 2001]. The basic idea is to choose in a sequential way single-observation one by one. The key advantage is that computations (see above) are easy to perform.

One first chooses an optimal position i_1 with a single observation strategy. We then search for a new optimal single observation under the constraint that the observation i_1 has been added to the routine observation network.

The algorithm is then the following (quite logical) one :

$\mathbf{H_0^{(tot)}} = \mathbf{H^{(r)}}$
DO $j = 1, p_a$

1. **Find the optimal location** i_j **:**

$$i_j = argmin\, \Phi(j, H_{j-1}^{(tot)}) \tag{A.73}$$

2. **Compute the new** *total* **observation operator** $H_j^{(tot)}$ **:**

$$H_j^{(tot)} = \begin{bmatrix} H_{j-1}^{(tot)} \\ H_{i_j} \end{bmatrix} \tag{A.74}$$

ENDDO

Such strategies (deeply related to optimization algorithms over *discrete* sets) may be promising tools.

A.5 Ensemble approaches

Ensemble approaches have been used in order to compute rapidly the error covariance matrices in the context of Kalman Filters. We refer for instance to [Bishop et Toth, 1999] and the works devoted to the so-called Ensemble Transform Kalman Filters ([Bishop *et al.*, 2001, Bishop, 2001, Majumdar *et al.*, 2001, Etherton, 2001]).

The key idea is to be able to compute rapidly for each possible deployment of the adaptive observations a design criterion (for instance the estimation of a reduction of the forecast error covariance matrix trough S, see Eq. (A.68)). In a brute-force approach, computing the criterion for each possible choice of adaptive observations (let say q with $1 \leq q \leq Q$), makes necessary the computation of the evolution of covariance matrices, which is of course unaffordable in practice.

Let us suppose now that there is a sequence of target times $t_{1,1}, t_{1,2}, \ldots, t_{1,n_t}$ (up to now, we have always taken $n_t = 1$). With the same kind of notations as in [Bishop *et al.*, 2001], the targeting problem for the time sequence $(t_{1,i})$ leads to choose a sequence of observation operators $(H^{(q)}(t_{1,i}))$ among a set of Q possible observation sequence.

Let us use a sequential context. In order to compute the design criterion, one has then, for instance, to compute an error covariance matrix (see above) at verification time t_2 for each possible sequence q. We have then to compute the evolution of the forecast error covariance matrix and of the analysis error covariance matrix along the sequence $(t_{1,i})$.

For the choice of the q adaptive deployment, the usual equations become for a perfect model :

$$P_a^{(q)}(t_{1,i}) = P_f^{(q)}(t_{1,i}) - P_f^{(q)}(t_{1,i}) H^{(q)}(t_{1,i}) \times$$
$$(H^{(q)}(t_{1,i}) P_f^{(q)}(t_{1,i}) H^{(q)}(t_{1,i})^t + R^{(q)}(t_{1,i}))^{-1} H^{(q)}(t_{1,i}) P_f^{(q)}(t_{1,i}) \quad \text{(A.75)}$$

$$P_f^{(q)}(t_{1,i+1}) = M(t_{1,i+1}) P_a^{(q)}(t_{1,i}) M(t_{1,i+1})^t \quad \text{(A.76)}$$

This is of course intractable in practice.

By using ensemble methods, the computational cost may be drastically reduced. The key point is to run a set of n_e forecasts, labeled by k $(x^{(k)})$ and to estimate an error covariance matrix by :

$$P(t) = \tilde{X}(t) \tilde{X}(t)^t \quad \text{(A.77)}$$

where the perturbation matrix is :

$$\tilde{X}_{ik} = \frac{(x_i^{(k)} - \bar{x}_i)(x_i^{(k)} - \bar{x}_i)^t}{\sqrt{n_e - 1}} \quad \text{(A.78)}$$

with the mean value of the ensemble $\bar{x}$:

$$\bar{x} = \frac{1}{n_e} \sum_{k=1}^{k=n_e} x^{(k)} \quad \text{(A.79)}$$

The way initial perturbations are produced is of course a key issue, since this is supposed to reflect the uncertainties in the initial conditions. Many methods are used : singular vector techniques (European Center for Medium Weather Forecast : [Molteni *et al.*, 1996]), the so-called breeding method (US National Center for Environmental Prediction : [Toth et Kalnay, 1997]) or the multi-replication method (Canadian Meteorological Center : [Houtemaker *et al.*, 1996]). This is anyhow out the scope of this paper.

A key result is that one can estimate these covariance matrices with the ensemble approach :

$$P_f^{(q)}(t_{1,i+1}) = Z_f^{(q)}(t_{1,i}) Z_f^{(q)}(t_{1,i})^t \quad \text{(A.80)}$$

$$P_a^{(q)}(t_{1,i}) = Z_f^{(q)}(t_{1,i}) T^{(q)}(t_{1,i}) T^{(q)}(t_{1,i})^t Z_f^{(q)}(t_{1,i})^t \quad \text{(A.81)}$$

where the *transformed* perturbation matrix is :

$$Z_f^{(q)}(t_{1,i}) = \tilde{X}(t_{1,i})T^{(q)}(t_{1,i-1}) \tag{A.82}$$

The sequence may therefore be computed if the $n_e \times n_e$ matrices T are easy to calculate. We refer to [Bishop *et al.*, 2001, Bishop, 2001]. By omitting (at the end of this paper) all sub/superscripts :

$$T = C(\Gamma + I)^{-1/2} \tag{A.83}$$

with C (resp. the diagonal matrix Γ) listing the eigenvectors (resp. the eigenvalues) of $Z_f^T H^t R^{-1} H Z_f$. The key point is that this matrix is only of dimension n_e.

As stressed in [Bishop *et al.*, 2001, Bishop, 2001, Majumdar *et al.*, 2001], the key point of such techniques is the possibility to estimate the reduction in forecast variance due to adaptive observations through the *signal variance*.

A.6 Time scheduling

We give in this section some (brief) insights in time scheduling, that is the appropriate choice of targeting times, in the particular case of atmospheric chemistry.

Atmospheric chemistry (chemical kinetics in general) is characterized by a wide range of timescales and one usually splits chemical species in slow species (long-lived) and fast-species (short-lived). We refer for instance to [Sportisse et Djouad, 2000] or [Sportisse et Djouad, 2002].

As noticed in [Fisher et Lary, 1995] or [Wang *et al.*, 2001], data assimilation or inverse modeling may be a difficult task for short-lived species. In this case, the impact of timescales for the accuracy of the data assimilation system is crucial ([Quélo *et al.*, 2002] for a deeper understanding of these aspects). In a variational context, the key point is the relation between the condition number of the Hessian matrix and the wide range of timescales in the forecast model.

As notices in another context in [Heineken *et al.*, 1967], the way times for adaptive observations have to be taken has to be strongly related with these remarks.

Conclusion

We have tried to summarize the different approaches currently used for network design and adaptive observations. Many sub-optimal techniques are now available with an evolution towards observation sensitivity approaches and probably a wider use of second-order adjoint methods in the future for 4D-var techniques.

Some benchmarks may be found in [Lorenz et Emanuel, 1998], [Morss, 1998] or [Hansen, 1998, Hansen et Smith, 2002] with the systematic use of a simplified 40-dimensional model proposed by Lorenz ([Lorenz et Emanuel, 1998]) for meteorological applications. The key issue is the systematic sensitivity of the data assimilation/forecast/observation system to all parameters (error statistics, number of observations, ...). The nonlinear effects are more particularly investigated in [Hansen, 1998] and one can wonder if the extension of such techniques for instance to other highly nonlinear geophysical flows (even if non chaotic, for instance : atmospheric chemistry, may not be quite difficult, since linear assumptions are often underlying assumptions.

Many works have been done in the context of estimation theory (as reflected in this survey paper) and the extension to the 4d-var context is largely to be done.

Appropriate use of numerical approaches of optimal design ([Fedorov et Hackl, 1997]) may be another way to propose new techniques.

Serial techniques, as proposed in [Majumdar *et al.*, 2001], are another challenging topics.

More generally, the use of some of these techniques in other application fields, is probably a key issue of the future of environmental monitoring networks, whatever the data are issued from (terrestrial sensors, sat data with the project Global Monitoring Environment System, GMES).

Acknowledgements

We thank François Xavier Le Dimet for fruitful discussions. Part of this work has been led in the framework of the INRIA Research Action COMODE (COupling MOdels and Data for Environment).

Nomenclature

– *General*
  x : n-dimensional state vector
  F : a n-dimensional nonlinear model
  M : the tangent linear model, linearized at the analysis trajectory
  $\tilde{M}$: the tangent linear model, linearized at the forecast trajectory
  $\epsilon^{(m)}$: random model error
  Q : model error covariance matrix
  x_f : n-dimensional forecast vector
  P_f : forecast error covariance matrix $(n \times n)$
  y : p-dimensional observation vector
  H : observation operator $(p \times n)$
  $\epsilon^{(o)}$: random observation error
  R : observation error covariance matrix $(p \times p)$
  x_a : n-dimensional analysis vector
  P_a : analysis error covariance matrix $(n \times n)$
  x_b : background state
  B : background error covariance matrix
  K : gain matrix $(n \times p)$
  d : innovation vector $(y - Kx_f)$
– *Adaptive observations*
  p_r : dimension of routine observations
  p_a : dimension of adaptive observations
  $H^{(r)}$: routine observation operator $(p_r \times n)$
  $H^{(tot)}$: total observation operator (including both routine and adaptive parts)
  n_t : number of target times
  t_0 : decision time
  t_1 : target time (if $n_t = 1$)
  $(t_{1,i})$: sequence of target times $(1 \leq i \leq n_t)$
  t_2 : verification time
  H_i : single observation operator for state variable i $((0 \ldots 0, 1, 0 \ldots, 0))$
– *Optimal design*
  Ψ : design criterion
  λ : design variable
  Λ : design space
  ξ : design probability
  I_F : Fisher information matrix
  ϕ : sensitivity design function
  V_H : design space for observation operator

– *Singular vectors*
  W_1 : metric at target time
  W_2 : metric at verification time
  $(\sigma_i^2)_W$: singular values for metrics W
  $(\mathbf{u_i})_W$: singular vectors for W
  $M^\dagger$: Moore-Penrose pseudo-inverse of M
– *Sensitivity analysis*
  h : forecast aspect (in $\mathbf{R}$)
  J : cost function for variational approach
  S : signal covariance matrix
  s : signal
– *Ensemble approaches*
  n_e : dimension of the ensemble
  $(x^{(k)})$: n_e-dimensional ensemble of forecasts
  $\bar{x}$: mean of the ensemble
  $\tilde{X}$: perturbation matrix of the ensemble $(n \times n_e)$
  $Z^{(q)}$: transformed perturbation matrix for the adaptive network q $(n \times n_e)$
  $T^{(q)}$ $n_e \times n_e$ transformation operator

Annexe B

Assimilation de données et modélisation inverse pour les modèles lent/rapide de chimie atmosphérique

Résumé

L'objet de ce chapitre est l'étude de l'impact de la dispersion des échelles de temps d'un modèle sur la qualité des résultats de l'assimilation de données. On envisage le cas des modèles dits "lent/rapide" caractéristiques du comportement dynamique de la cinétique chimique atmosphérique. Ces modèles ont notamment pour propriété de pouvoir être approchés en temps court par des modèles réduits ne conservant que la dynamique lente du système.

A l'aide de la théorie de la perturbation singulière, qui procure le cadre théorique adéquat classique pour l'étude de ces modèles, on étudie notamment les principales propriétés du hessien de la fonction coût qui donne une indication sur la qualité des résultats obtenus après assimilation.

Les principaux enseignements sont l'importance du moment d'observation et la nécessité de distinguer les espèces lentes et rapides.

Ce chapitre fait suite à [Quélo et al.*, 2002] et correspond à un article en cours de soumission ([Quélo et Sportisse, 2004]).*

Sommaire

B.1 Un modèle académique 186
B.2 Assimilation de données pour les modèles de chimie "lent/rapide" 188
B.2.1 Notations 188
B.2.2 Etude de la matrice hessienne pour le cas lent/rapide 191
B.2.3 Quelques remarques 192
B.3 Tests numériques 196
B.3.1 Conditions de calcul 196
B.3.2 Expériences de modélisation inverse 199

Introduction

Dans ce chapitre, nous nous intéressons à une catégorie de modèles chimiques que l'on rencontre en pratique : les systèmes dits "lent/rapide". Ils sont caractérisés par une dynamique transitoire rapide suivie d'une évolution lente [Sportisse et Djouad, 2002]. Les espèces chimiques peuvent alors être partitionnées en espèce *rapide* et espèce *lente*. Un bon exemple de tels systèmes est donné par la cinétique chimique atmosphérique et les flux d'émissions réactifs.

Comme mentionné dans des précédents travaux, il existe une relation étroite entre la qualité de l'assimilation de données et les propriétés dynamiques du système sous-jacent. Par exemple, dans [Wang *et al.*, 2001], à la page 2202, les auteurs affirment que : "the longer-lived species [...] are in fact a better fit than the shorter-lived species". Dans le même registre, dans le travail préliminaire de [Austin, 1992], une des conclusions est que "the assimilation technique [...] is invaluable in the analysis of short time scale species". Par ailleurs, le fait que la qualité des analyses de certains paramètres peut influencer les autres espèces et la relation avec les échelles de temps sont des questions sous-jacentes dans [Fisher et Lary, 1995].

L'objectif de ce chapitre est de déterminer l'influence de ce comportement dynamique sur l'assimilation de données. Le caractère "lent/rapide" de la cinétique chimique atmosphérique a plusieurs conséquences et son impact sur l'assimilation de données n'est qu'un point spécifique (voir [Sportisse, 2000] par exemple pour l'influence sur la résolution numérique en particulier les méthodes de séparation d'opérateurs).

Les questions posées sont les suivantes :

- Quelles conséquences pour l'assimilation de données (la modélisation inverse des conditions initiales) sont amenées par de tels systèmes ? Par exemple, il semble relativement intuitif qu'il est plus facile de retrouver la valeur d'une espèce chimique lente que celle d'une espèce chimique rapide.
- Quelle est l'importance du moment d'observation ? A priori, il parait évident qu'il faut observer les phénomènes quand ils se produisent.
- Quelles sont les implications pour la modélisation inverse des émissions ? De la même façon, il semble plus facile d'identifier les valeurs de paramètres physiques associés à la dynamique lente plutôt qu'à la dynamique rapide.

Ce chapitre est organisé de la manière suivante. La première section est consacrée à l'étude d'un modèle académique simple. Nous donnons le cadre théorique et une extension au cas général dans la section 2. Des illustrations numériques sont données dans la section 3 avec une application à une cinétique chimique troposphérique.

B.1 Un modèle académique

Considérons le modèle de boîte suivant :

$$\frac{dc}{dt} = -k\,c + E\ , \quad c(0) = c_0 \tag{B.1}$$

c correspond à une concentration chimique, k est une constante cinétique (ou l'inverse d'un temps de vie) et E est une émission. L'ordre de grandeur de k est relié aux échelles de temps du système : une petite (resp. grande) valeur de k correspond à une espèce *lente* (resp. *rapide*).

La solution s'obtient facilement :

$$c(t) = E/k + (c_0 - E/k)e^{-k\,t} \stackrel{\Delta}{=} F_t(c_0, k, E) \tag{B.2}$$

Au temps t fixé, $c(t)$ peut être vu comme une fonction des paramètres d'entrée (c_0, k, E) et nous notons $c(t) = F_t(c_0, k, E)$. Supposons à présent qu'un paramètre d'entrée n'est pas connu (ou connu avec une faible précision) et que nous voulons retrouver sa valeur réelle sur la base des observations par modélisation inverse. Le cas où le paramètre est k correspond à une expérience en chambre de réaction avec un ajustement des paramètres de cinétique chimique ; le cas où le paramètre est E est relié à la modélisation inverse des émissions dans le but d'améliorer un inventaire d'émission ; le cas où le paramètre est c_0 est le cas standard de l'assimilation de données pour la prévision de la qualité de l'air.

Nous avons un jeu d'observations disponibles *obs*. Par souci de clarté, nous supposons que nous ne disposons que d'une seule observation au temps t (l'unique temps d'observation). Si on définit les "vraies" valeurs des paramètres par $\bar{k}$, $\bar{E}$ et $\bar{c}_0$, l'observation est donnée par :

$$obs = \bar{E}/\bar{k} + (\bar{c}_0 - \bar{E}/\bar{k})e^{-\bar{k}\,t} + \eta \tag{B.3}$$

avec η une erreur d'observation supposée constante dans ce cas simplissime.

Par exemple, dans le cas de la modélisation inverse de k (E et c_0 sont supposés connus), nous obtenons une valeur de k telle que :

$$obs = F_t(\bar{c}_0, k, \bar{E}) \tag{B.4}$$

Nous avons tracé sur la figure B.1 le comportement du rapport $\Phi/\bar{\Phi}$, pour chaque paramètre d'entrée Φ, en fonction de $\bar{k}\,t$. Les données numériques sont les suivantes : $\bar{c}_0 = 1$, $\bar{E} = 1$, $t = 1$ (temps d'observation), $\eta = 0.01$ (l'erreur d'observation modélise un

seuil de détection de mesure) alors que $\bar{k}$ varie de 1 (espèce lente) à 100 (espèce rapide). Comme il existe une unique solution à l'équation (B.4), nous n'avons pas utilisé une procédure de minimisation classique et nous avons utilisé directement l'algorithme de Newton pour trouver la valeur inconnue du paramètre.

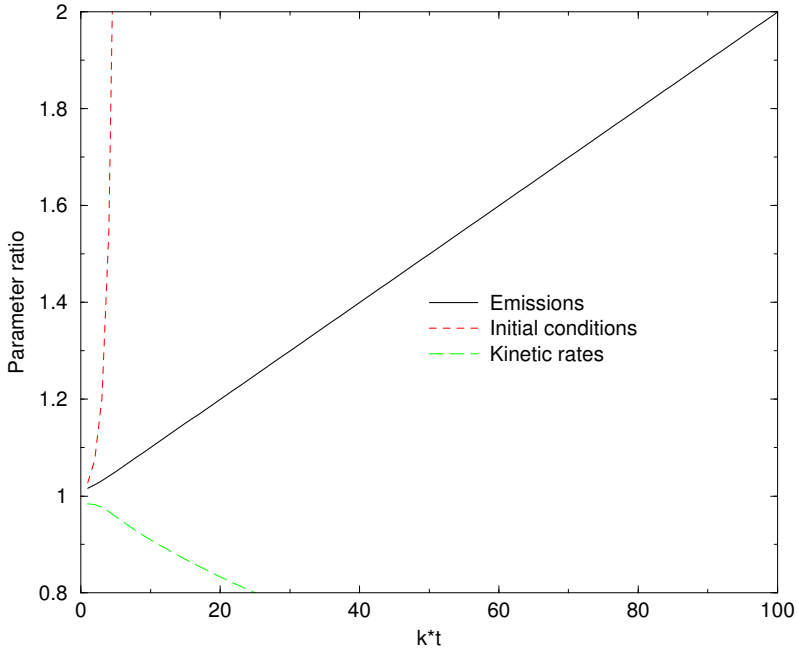

FIG. B.1 – Evolution du rapport de paramètre pour des espèces rapides et lentes.

La conclusion évidente est que plus l'espèce est rapide, plus l'erreur est grande. Remarquons que dans notre cas, le temps d'observation est beaucoup plus approprié pour des espèces rapides (t est de l'ordre de l'échelle de temps caractéristique d'une espèce lente). Ceci montre qu'il est impossible de retrouver des paramètres d'entrée associés aux espèces rapides si le temps d'observation n'est pas bien choisi.

Remarquons que nous avons une solution analytique pour les cas de E et c_0 :

$$c_0 = \bar{c}_0 + \eta\, e^{\bar{k}\,t} \;, \quad E = \bar{E} + \frac{\bar{k}}{1 - e^{-\bar{k}\,t}}\eta \tag{B.5}$$

et on peut vérifier que l'erreur est grande pour $\bar{k}\,t \gg 1$.

Nous concluons cet exemple simple avec quelques remarques concernant la nécessité d'inverser simultanément E et c_0 (k n'est pas traité par souci de clarté). En pratique, les émissions et les concentrations initiales peuvent être toutes deux connues de manière

imprécise et on peut s'interroger sur la sensibilité du résultat de modélisation inverse à un autre paramètre incertain.

Par exemple, recherchons la valeur de E en supposant que la condition initiale est c_0 (et non $\bar{c}_0$). On obtient facilement :

$$E = \bar{E} + \frac{\bar{k}}{1 - e^{-\bar{k}\,t}}(\eta + (\bar{c}_0 - c_0)e^{-\bar{k}\,t}) \tag{B.6}$$

Ceci montre que le résultat n'est pas très sensible à l'incertitude sur les conditions initiales et que, au premier ordre en $e^{-\bar{k}\,t}$, la sensibilité est une fonction décroissante du temps. De plus, on remarque que le manque de connaissances sur les conditions initiales a un impact plus grand sur les espèces lentes que sur les espèces rapides.

Une approche similaire peut être appliquée à la modélisation inverse de c_0 avec une incertitude sur E :

$$c_0 = \bar{c}_0 + \eta\, e^{\bar{k}\,t} + (\bar{E} - E)\frac{e^{\bar{k}\,t} - 1}{\bar{k}} \tag{B.7}$$

L'erreur est une fonction linéaire de $(\bar{E} - E)$ pour les espèces lentes alors qu'elle est contrôlée par ce terme pour les espèces rapides.

B.2 Assimilation de données pour les modèles de chimie "lent/rapide"

B.2.1 Notations

B.2.1.1 Cinétique chimique "lent/rapide"

L'objectif de ce paragraphe est d'étendre les principaux résultats obtenus avec le modèle académique. Dans la suite, on va supposer que les réactions chimiques peuvent être divisées en deux catégories selon leur temps caractéristique : certaines réactions chimiques peuvent être considérées comme des réactions *lentes* (resp. *rapides*) si leur échelle de temps est de l'ordre de 1 (resp. $\varepsilon \ll 1$).

Nous notons c le vecteur des concentrations chimiques de dimension n (n étant le nombre d'espèces chimiques). Afin de simplifier la présentation, on suppose que les mécanismes chimiques conduisent à un système **linéaire** d'équations différentielles

ordinaires. Ceci constitue bien sûr une hypothèse irréaliste en pratique vu que cela n'est rencontré que pour les réactions monomoléculaires, ce qui n'est pas le cas général de la chimie atmosphérique (exception faite des réactions photochimiques). Notre analyse peut cependant être généralisée à une version linéarisée de modèles plus sophistiqués et même à des modèles non-linéaires (voir la remarque ci-dessous).

Le modèle de cinétique chimique s'écrit alors sous la forme :

$$\frac{dc}{dt} = Mc + E \ , \quad c(0) = c_0 \tag{B.8}$$

où la matrice M contient les constantes cinétiques et les coefficients stoechiométriques et E est le vecteur des termes d'émissions. Sur la base du comportement lent/rapide, nous obtenons :

$$M = \frac{M_0}{\varepsilon} + M_1 \tag{B.9}$$

où M_0 (resp. M_1) est reliée aux réactions rapides (resp. lentes). Notons qu'il est possible d'ajouter un terme lent de lessivage humide et de dépôt dans M_1. Ce terme est en effet sous la forme $-\beta\, c$ avec β une matrice diagonale composée par les coefficients positifs de lessivage humide et de dépôt.

En supposant que M est une matrice inversible nous avons :

$$\begin{aligned} c(t) &= \exp(-Mt)c_0 + \varepsilon\left(\exp(Mt) - I\right) M^{-1}E \\ &= Ac_0 + BE \\ &\stackrel{\Delta}{=} F_t(c_0, E) \end{aligned} \tag{B.10}$$

ce qui définit les matrices A et B.

B.2.1.2 Base de partitionnement et modèle réduit

Il est cependant plus facile de travailler avec une base d'espèces plus appropriée. Le comportement lent/rapide des cinétiques chimiques est en effet relié à la propriété de partitionnement suivante ([Sportisse et Djouad, 2002] par exemple) : il existe une base (et donc une matrice inversible P) telle que

$$PMP^{-1} = \begin{bmatrix} \Lambda_0/\varepsilon & 0 \\ 0 & \Lambda_1 \end{bmatrix} \tag{B.11}$$

où Λ_0 est une matrice $n_f \times n_f$ strictement stable (une matrice inversible dont les valeurs propres ont une partie réelle strictement négative). n_f (resp. $n - n_f$) est le nombre de degrés de liberté rapide (resp. lent) du mécanisme chimique. La relation étroite entre P et le

"*lumping*" d'espèces ([Sportisse et Djouad, 2000]) ou les *familles* d'espèces ([Austin, 1992]) sera commentée par la suite. P peut être écrite sous la forme partitionnée :

$$P = \begin{bmatrix} P_f \\ P_s \end{bmatrix} \tag{B.12}$$

avec P_f (resp. P_s) une matrice $n_f \times n$ (resp. $(n - n_f) \times n$) telle que $c_f = P_f\, c$ (resp. $c_s = P_s\, c$) sont les degrés de liberté rapides (resp. lents).

On peut écrire à présent :

$$\begin{aligned} \tilde{B}(t) &= PBP^{-1} = \begin{bmatrix} \exp(t\Lambda_0/\varepsilon) & 0 \\ 0 & \exp(t\Lambda_1) \end{bmatrix} \\ \tilde{A}(t) &= PAP^{-1} = \begin{bmatrix} \varepsilon(\exp(t\Lambda_0/\varepsilon) - I)\Lambda_0^{-1} & 0 \\ 0 & (\exp(t\Lambda_1) - I)\Lambda_1^{-1} \end{bmatrix} \end{aligned} \tag{B.13}$$

Dans la nouvelle base, $\tilde{c} = Pc$ et $\tilde{E} = PE$ tel que :

$$\tilde{c}(t) = \tilde{B}\tilde{c}_0 + \tilde{A}\tilde{E} \tag{B.14}$$

Le point clé est qu'il existe alors un modèle réduit sous-jacent pour $t \simeq 1$:

$$\begin{aligned} \tilde{B}(t) &\simeq \begin{bmatrix} 0 & 0 \\ 0 & \exp(t\Lambda_1) \end{bmatrix} \\ \tilde{A}(t) &\simeq \begin{bmatrix} 0 & 0 \\ 0 & (\exp(t\Lambda_1) - I)\Lambda_1^{-1} \end{bmatrix} \end{aligned} \tag{B.15}$$

au premier ordre en ε. Ceci signifie en pratique que $\tilde{c}_f = 0$.

B.2.1.3 Assimilation de données

Précisons à présent le cadre de l'assimilation de données. On suppose que l'on dispose d'un ensemble I d'observations $(obs_i)_i$, données aux temps $(t_i)_i$ avec les matrices de covariance d'erreur d'observations $(R_i)_i$. L'opérateur d'observation est H_i (au temps t_i).

Dans un cadre variationnel, la fonction coût que l'on doit minimiser par rapport aux paramètres d'entrées (c_0 et/ou E en pratique) s'écrit :

$$J(c_0, E) = 1/2 \sum_{i=1}^{i=I} \left(obs_i - H_i c(t_i)\right)^T R_i^{-1} \left(obs_i - H_i c(t_i)\right) \;\; , \;\; c(t_i) = F_{t_i}(c_0, E) \tag{B.16}$$

Nous avons omis le terme d'ébauche qui peut aussi être pris en compte.

B.2.2 Etude de la matrice hessienne pour le cas lent/rapide

Un bon indicateur de la qualité de la procédure d'assimilation de données est l'étude du spectre de la matrice hessienne de J, Hess J, calculé par rapport aux paramètres d'entrée ([Thacker, 1989, Le Dimet *et al.*, 2002]). Par exemple, les petites valeurs propres de Hess J sont reliées aux paramètres qui sont difficiles à retrouver à cause du mauvais conditionnement (ceci constitue un point de vue numérique) ou pour lequel la confiance de l'analyse est faible (ceci est le point de vue des statistiques). Pour les modèles "lent/rapide", le mauvais conditionnement est fortement relié au comportement dynamique et on montre dans la suite que les paramètres mal conditionnés sont reliés aux processus rapides.

On obtient facilement :

$$\text{Hess J}_{c_0,E} = \sum_{i=1}^{i=I} [B_i \ A_i]^T [H_i^T R_i^{-1} H_i][B_i \ A_i] \tag{B.17}$$

où $B_i = B(t_i)$ et $A_i = A(t_i)$.

Dans la nouvelle base, la matrice hessienne s'écrit à présent :

$$\text{Hess J}_{\tilde{c}_0,\tilde{E}} = \sum_{i=1}^{i=I} [\tilde{B}_i \ \tilde{A}_i]^T [{P^{-1}}^T H_i^T R_i^{-1} H_i \, P^{-1}][\tilde{B}_i \ \tilde{A}_i] \tag{B.18}$$

On peut utiliser le modèle réduit pour les observations associées aux processus lent ($t \sim 1$, ce qui est le cas en pratique). Des calculs fastidieux mènent à la forme suivante partitionnée par blocs de la matrice hessienne :

$$\text{Hess}J_{\tilde{c}_0,\tilde{E}} = \sum_{i=1}^{i=I} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & * & 0 & * \\ 0 & 0 & 0 & 0 \\ 0 & * & 0 & * \end{bmatrix} \tag{B.19}$$

où le signe $*$ désigne un terme relié à la dynamique lente $\exp(t\Lambda_1)$ (omise ici pour plus de clarté). Précisons que la partition par blocs de la matrice hessienne est reliée au partitionnement des paramètres d'entrée dans l'ordre suivant : conditions initiales rapides (de dimension n_f), conditions initiales lentes (de dimension $n - n_f$), émissions des espèces rapides (de dimension n_f), émissions des espèces lentes (de dimension $n - n_f$). Cette forme approchée est une sorte de forme "réduite" de la matrice hessienne.

Ceci montre qu'il est impossible de retrouver les vraies valeurs des conditions initiales et les émissions des processus rapides, ce qui constitue l'extension des résultats obtenus avec le modèle académique.

B.2.3 Quelques remarques

B.2.3.1 Extension à un modèle de chimie–transport

L'analyse précédente a été effectuée avec l'hypothèse simplificatrice d'un modèle de boîte avec une chimie monomoléculaire. Ceci peut être facilement généralisé en prenant en compte le transport des espèces chimiques pour l'advection et la diffusion si on suppose que ce sont des processus lents (cas le plus courant en pratique). Pour être plus précis, intéressons-nous à la forme discrétisée du terme de transport, Tc, avec T un opérateur linéaire obtenu par exemple par différences finies. Le modèle s'écrit à présent :

$$\frac{dc}{dt} = Mc + E + Tc \ , \quad c(0) = c_0 \tag{B.20}$$

où c est le vecteur des espèces chimiques pour tout le maillage (et pas seulement dans une maille comme précédemment). Le résultat clé ([Sportisse et Djouad, 2000]) est que le modèle réduit sous-jacent et le partitionnement adéquat des espèces est toujours donné par le modèle de boîte $dc/dt = Mc$. La généralisation des résultats précédents est alors immédiate.

B.2.3.2 Généralisation au cas non linéaire

Une autre généralisation intéressante concerne bien sûr l'extension au cas non linéaire. Nous ne donnons ici qu'une indication en étudiant le comportement asymptotique de la fonction coût J (ce qui revient à étudier la forme "réduite" de J). Pour simplifier, on peut travailler avec un jeu d'observations continues $(obs(t))_{0 \leq t \leq T}$. La fonction coût s'écrit à présent :

$$J = 1/2 \int_0^T \left(obs(t) - Hc(t)\right)^T R^{-1} \left(obs(t) - Hc(t)\right) dt \tag{B.21}$$

Considérons le modèle sous sa forme générale (Il peut inclure la cinétique chimique, les émissions, les termes de transport) dans une base partitionnée $c = (c_f, c_s)$:

$$\begin{aligned} \varepsilon \frac{dc_f}{dt} &= g(c_f, c_s) \ , \quad c_f(0) = (c_f)_0 \\ \frac{dc_s}{dt} &= f(c_f, c_s) \ , \quad c_s(0) = (c_s)_0 \end{aligned} \tag{B.22}$$

De la même manière, on peut partitionner la matrice d'observations en $H = [H_f \ H_s]$.

Rappelons tout d'abord la définition d'un modèle adjoint. On considère le système dynamique $dc/dt = F(c) \ , \ c(0) = c_0$ qui définit $c(t) = F_t(c_0)$ pour $t \in [0, T]$. Notons la

fonction coût $\tilde{J}(c) = J(c_0)$. Le modèle adjoint est alors défini par :

$$\frac{d\hat{c}}{dt} = -\left(\frac{\partial F}{\partial c}\right)^T_{|c(t)} \hat{c} - \nabla \tilde{J} \ , \quad \hat{c}(T) = 0 \tag{B.23}$$

et le point clé est que l'intégration rétrograde donne $\nabla_{c_0} J = \hat{c}(0)$.

Dans notre cas, les équations continues adjointes sont donc :

$$\begin{cases} \varepsilon \dfrac{d\hat{c}_s}{dt} = -\left(\dfrac{\partial g}{\partial c_s}\right)^T_{|c} \hat{c}_f - \varepsilon \left(\left(\dfrac{\partial f}{\partial c_s}\right)^T_{|c} \hat{c}_s + H_s^T R^{-1}(obs - Hc) \right) \\ \hat{c}_s(T) = 0 \\ \varepsilon \dfrac{d\hat{c}_f}{dt} = -\left(\dfrac{\partial g}{\partial c_f}\right)^T_{|c} \hat{c}_f - \varepsilon \left(\left(\dfrac{\partial f}{\partial c_f}\right)^T_{|c} \hat{c}_s + H_f^T R^{-1}(obs - Hc) \right) \\ \hat{c}_f(T) = 0 \end{cases} \tag{B.24}$$

où $\hat{c}_f$ et $\hat{c}_s$ sont les *variables adjointes*. Les gradients de J par rapport aux conditions initiales sont obtenus en intégrant le système rétrograde :

$$\frac{\partial J}{\partial (c_f)_0} = \hat{c}_f(0) \ , \quad \frac{\partial J}{\partial (c_s)_0} = \hat{c}_s(0) \tag{B.25}$$

Après une période transitoire rapide ($t \gg \varepsilon$), le système adjoint peut être réduit à :

$$\begin{cases} \dfrac{d\hat{c}_s}{dt} = -\left(\dfrac{\partial f_{red}}{\partial c_s}\right)^T \hat{c}_s - \left(H_s - H_f \left(\dfrac{\partial g}{\partial c_f}\right)^{-1} \dfrac{\partial g}{\partial c_s} \right)^T R^{-1}(obs - Hc) \\ \hat{c}_s(T) = 0 \\ \hat{c}_f = 0 \end{cases} \tag{B.26}$$

avec $\dfrac{\partial f_{red}}{\partial c_s} = \left(\dfrac{\partial f}{\partial c_s} - \dfrac{\partial f}{\partial c_f} \left(\dfrac{\partial g}{\partial c_f}\right)^{-1} \dfrac{\partial g}{\partial c_s} \right)$.

Le point clé est que la sensibilité de la fonction coût par rapport aux conditions initiales des espèces rapides est nulle ($\hat{c}_f = 0$). Physiquement, cela signifie que l'assimilation de données des espèces rapides ne peut être menées à bien.

B.2.3.3 Assimilation de données pour les modèles réduits

Un autre produit dérivé de l'analyse précédente peut être représenté par le graphique suivant (une flèche verticale signifie *réduction*, une flèche horizontale *assimilation de données*) :

$$\begin{array}{lcl} (\text{exact model})_\varepsilon & \longrightarrow & J_\varepsilon, \nabla J_\varepsilon \\ \downarrow & & \downarrow \\ \text{reduced model} & \longrightarrow & J_{red}, \nabla J_{red} \end{array} \tag{B.27}$$

La forme réduite des équations adjointes associées au modèle exact est équivalente aux équations adjointes associées au modèle réduit :

$$0 = g(c_f, c_s) \ , \quad \frac{dc_s}{dt} = f(c_f, c_s) \ , \quad c_s(0) = (c_s)_0 \tag{B.28}$$

Les équations adjointes pour ce modèle sont :

$$\begin{aligned} \frac{d\hat{c}_s}{dt} &= -\left(\frac{\partial f_{red}}{\partial c_s}\right)^T \hat{c}_s - \left(H_s - H_f \left(\frac{\partial g}{\partial c_f}\right)^{-1} \frac{\partial g}{\partial c_s}\right)^T R^{-1}(obs - Hc) \\ \hat{c}_s(T) &= 0 \end{aligned} \tag{B.29}$$

et les seules différences avec (B.26) sont que les dérivées partielles sont calculées le long du modèle réduit (B.28) et non le long du modèle exact (B.22) (mais les solutions sont les mêmes au premier ordre en ε).

Ceci met en évidence l'utilité de travailler avec le modèle réduit, parce qu'on obtient le bon résultat pour les espèces lentes quel que soit le modèle (modèle exact ou modèle réduit). Le fait qu'il existe un modèle réduit sous-jacent a des conséquences négatives (les espèces rapides sont difficiles à inverser) mais aussi positives (leur comportement est forcé par les espèces lentes).

B.2.3.4 Lumping et préconditionnement

Il a été plus facile de travailler dans la base partitionnée définie par la matrice P. Nous allons à présent commenter la relation étroite avec les espèces lumpées ([Sportisse et Djouad, 2000, Djouad et Sportisse, 2002]) et les familles d'espèces chimiques ([Austin, 1992]). On définit habituellement une espèce lumpée par une combinaison linéaire d'espèces chimiques réelles $u_1 = L_1\, c$ telle que u_1 soit un degré de liberté lent (soit $L_1\, M_0 = 0$ ce qui implique que $du_1/dt = L_1\, M_1\, c$). On peut alors définir la matrice de partitionnement P de la manière suivante : chercher le jeu d'espèces indépendantes lumpées $u_1 \ldots u_{n-n_f}$ (tel que $P_s = [L_1^T \ldots L_{n-n_f}^T]^T$ est de rang plein) et compléter P avec des lignes définies par des vecteurs de base de $vect(L_1^T \ldots L_{n-n_f}^T)^\perp$. On se réfère à [Djouad et Sportisse, 2002] pour la génération automatique d'espèces lumpées pour la chimie atmosphérique multiphasique. L'algorithme bien connu consiste à construire les espèces lentes (par exemple $NO_x = NO + NO_2$ pour la chimie troposphérique) sur la base des espèces rapides (NO and NO_2).

La conséquence pour l'assimilation de données est que les conditions initiales et les paramètres d'émissions associés aux espèces lumpées (par exemple NO_x ou $NO_z = NO_3 + N_2O_5$) peuvent être inversés alors que c'est impossible pour des espèces rapides réelles (NO, NO_2). Cela signifie que les espèces rapides réelles agrégées dans un

lumping ne peuvent être dissociées.

En ce sens, l'utilisation d'espèces lumpées peut être vu comme une sorte de "préconditionnement physique" de la matrice hessienne pour le problème de l'assimilation de données.

B.2.3.5 Impact pour les concentrations

Le fait qu'il soit difficile de retrouver les conditions initiales rapides n'a pas d'impact sur la qualité des prévisions. En effet, l'évolution des espèces chimiques ne dépend pas des conditions initiales rapides. Par exemple, le modèle général donné par l'équation (B.22) peut être approché au premier ordre en ε après une phase transitoire par :

$$0 = g(c_f, c_s) \ , \quad \frac{dc_s}{dt} = f(c_f, c_s) \ , \quad c_s(0) = (c_s)_0 \tag{B.30}$$

qui ne dépend pas de $(c_f)_0$.

La situation est cependant différente pour la modélisation inverse des émissions vu qu'on est intéressé par l'amélioration l'inventaire d'émissions.

B.2.3.6 Chimie atmosphérique versus météorologie

Notons qu'il y a une différence fondamentale avec les modèles "lent/rapide" qui apparaissent en météorologie. A l'opposé de la météorologie, l'impact d'une précision faible pour l'analyse des conditions initiales rapides peut être négligé. Une analyse similaire peut être menée pour les modèles du type de ceux rencontrés en météorologie sur la base de modèles linéarisés : Cela consiste à remplacer $k \in \mathrm{R}$ dans notre modèle académique par $i\,k$ avec $i^2 = -1$ afin de décrire les ondes rapides de gravité (ou Λ_0 par $i\,\Lambda_0$ dans cette section). Des procédures spécifiques doivent être utilisées afin de "filtrer" l'impact des degrés de liberté rapides (on se réfère par exemple à [Kopell, 1985] et plus généralement à la bibliographie consacrée par exemple à l'initialisation des modes normaux).

B.2.3.7 Extension aux observations rapides

Une analyse similaire peut être menée pour les temps d'observation t_i associés aux phénomènes rapides, soit $t_i \simeq \varepsilon$ (ce qui n'est pas réaliste en pratique). Dans ce cas, on peut montrer de la même façon que précédemment que les conditions initiales rapides peuvent être retrouvées mais pas les émissions rapides ni les conditions initiales et les émissions lentes.

B.3 Tests numériques

B.3.1 Conditions de calcul

Nous avons effectué des expériences numériques pour illustrer le lien entre la modélisation inverse et le comportement dynamique lent/rapide de la chimie atmosphérique. Les tests ont été menés sur le modèle de boîte défini par :

$$\frac{\partial c}{\partial t} = \chi(c,t) + \frac{1}{H_{CL}}(E - v_{Dep}c)$$

où χ est le terme de production chimique, E est le flux d'émission, v_{dep} est la vitesse de dépôt sec et H_{CL} désigne la hauteur de mélange.

Tab. B.1 – Schéma chimique Ozone16. Les constantes cinétiques sont calculées par $k = a\,10^b$.

Nr	reaction	kinetic rate $a(b)$
1	$OD + O_2 + AIR \rightarrow O_3 + AIR$	6.02 (-34)
2	$O_3 + NO \rightarrow NO_2$	1.872 (-14)
3	$NO + HO_2 \rightarrow NO_2 + OH$	8.235 (-12)
4	$OH + NO_2 \rightarrow HNO_3$	1.1 (-11)
5	$NO_2 \rightarrow NO + OD$	8.88 (-3)
6	$RH + OH \rightarrow RO_2$	2.607 (-12)
7	$RCHO + OH \rightarrow RCO_3$	1.588 (-11)
8	$RCHO \rightarrow RO_2 + CO + HO_2$	3.18 (-6)
9	$NO + RO_2 \rightarrow RCHO + HO_2 + NO_2$	7.563 (-12)
10	$NO + RCO_3 \rightarrow NO_2 + RO_2 + CO_2$	7.563 (-12)
11	$NO_2 + RCO_3 \rightarrow PAN$	4.7 (-12)
12	$PAN \rightarrow RCO_3 + NO_2$	4.837 (-4)

Le schéma chimique utilisé, Ozone16, est présenté dans le tableau B.1. C'est une chimie simplifiée de l'ozone qui comporte 12 réactions chimiques mettant en jeu 16 espèces. L'ozone est formé par la réaction très rapide OD + O_2 + AIR → O_3 + AIR et a donc elle-même un comportement dynamique rapide. En comparaison, les réactions mettant en jeu NO, NO_2 et RH sont lentes. Les figures B.2 et B.3 représentent l'évolution temporelle[†1] des concentrations des espèces O3, NO, NO_2 et RH (composés organiques volatils). On remarquera que les dynamiques des espèces émises (NO, NO_2 et RH) ont le même temps caractéristique.

[†1] On notera la différence d'échelle temporelle.

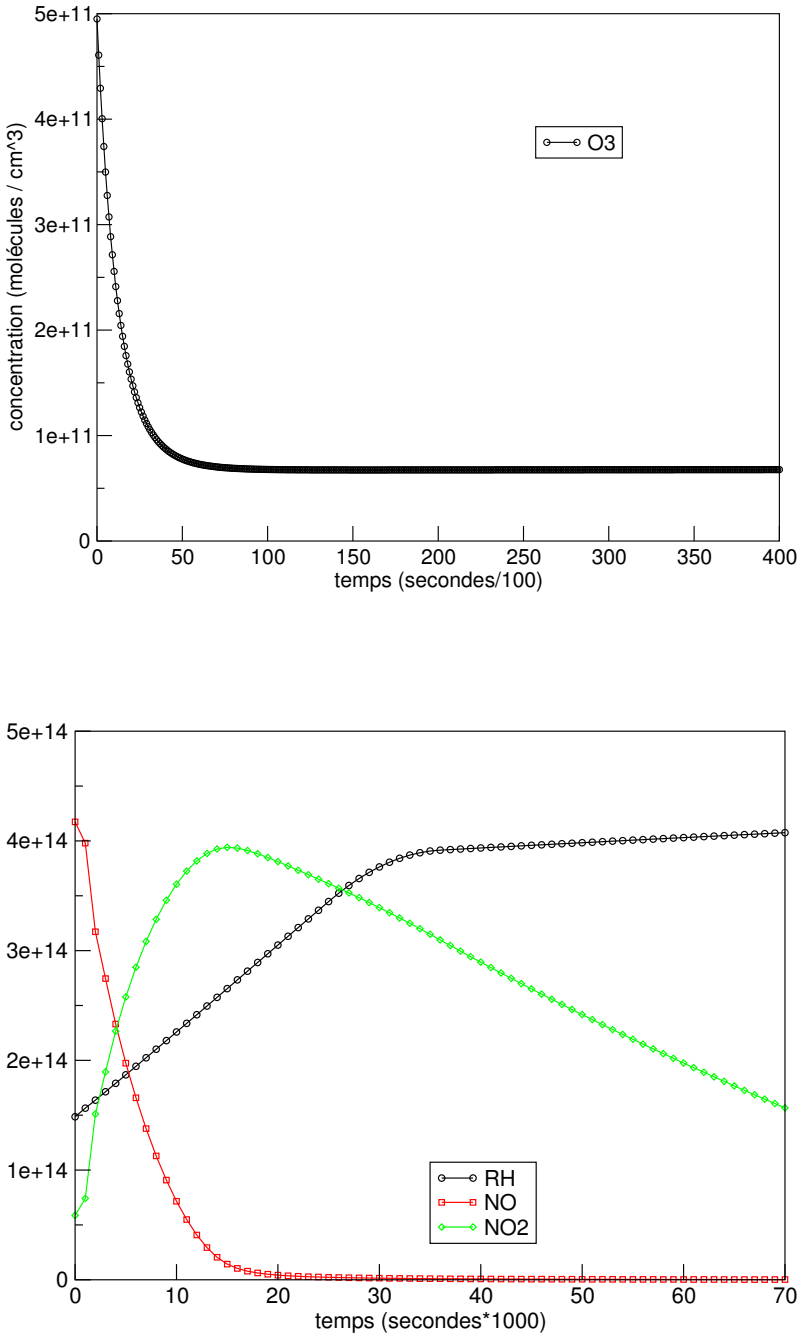

FIG. B.3 – Evolution temporelle des concentrations de NO, NO_2 et RH.

TAB. B.2 – Vitesse de dépôt sec ($\mathrm{cm\,s^{-1}}$)

species	O_3	NO_2	NO
v_{Dep}	0.6	0.6	0.1

TAB. B.3 – Flux d'émissions (molécules $\mathrm{cm}^{-2}\,\mathrm{s}^{-2}$)

species	RH	NO	NO_2
E	1.23E+13	1.764E+12	0.756E+12

TAB. B.4 – Conditions initiales (molécules cm^{-3})

species	Initial conditions
RH	$1.5E+14$
NO	$4.2.E+14$
NO_2	$0.6.E+14$
PAN	$4.60E+11$
$RCHO$	$9.98E+14$
O_3	$4.9E+11$
OH	$1.76E+07$
HO_2	$6.76E+07$
RCO_3	$5.98E+07$
RO_2	$6.78E+07$
OD	$1.12E+07$
O_2	$2.E+19$
AIR	$1.E+20$

Les valeurs des émissions et du dépôt sont regroupées dans les tableaux B.2 et B.3. Les conditions initiales (voir le tableau B.4) sont typiques des valeurs en milieu urbain (les conditions initiales de CO, HNO_3 et CO_2 qui ne sont que des espèces produites n'interviennent pas et ne sont donc pas renseignées).

B.3.2 Expériences de modélisation inverse

On présente d'abord le cadre commun de la modélisation inverse. On définit la fonction coût par :

$$J(p) = \sum_{l=1}^{n} \left(Hc^{(p)}(t_l) - obs(t_l) \right)^2 \tag{B.31}$$

où H désigne l'opérateur d'observation et p correspond aux conditions initiales, aux émissions ou aux constantes cinétiques. Les observations $obs(t_l)$ proviennent d'expériences jumelles (elles sont générées numériquement et non perturbées). On observe les espèces O_3, NO et NO_2 aux temps d'observations t_l (au nombre de 10) qui sont fixés toutes les dix minutes.

La minimisation de J se fait par l'algorithme itératif BFGS qui se base sur la connaissance du gradient de J par rapport à p. Le gradient a été obtenu par différenciation automatique du code direct par Odyssée [Faure et Papegay, 1997] qui génère directement le code adjoint à partir du code direct écrit en Fortran77.

Deux types d'expériences de modélisation inverse sont menées par rapport aux paramètres de contrôle suivants :

- conditions initiales de NO et O_3 ;
- constantes cinétiques des réactions 1 et 2 (notées k_1 et k_2).

B.3.2.1 Inversion des conditions initiales

L'assimilation de données est appliquée sur les conditions initiales de O_3 (rapide) et NO (lente). Les temps d'observations (toutes les 10 minutes) correspondent à des observations en temps long (relativement à la dynamique de l'ozone). Elles sont prises en dehors de la phase d'évolution très rapide de l'ozone. La figure B.4 montre que l'erreur relative sur les conditions initiales tend vers zéro seulement pour l'espèce lente NO alors qu'elle reste grande (à la valeur maximale définie par l'intervalle de variation des variables) pour l'espèce rapide O_3.

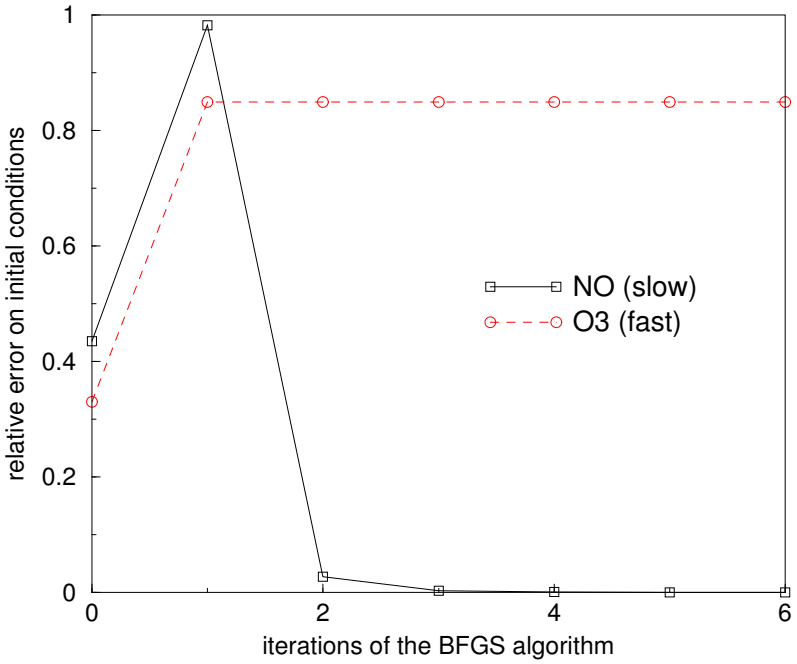

FIG. B.4 – Erreur relative sur les conditions initiales en fonction des itérations de BFGS.

B.3.2.2 Inversion des constantes cinétiques

La modélisation inverse conjointe des constantes cinétiques des réactions 1 et 2 du schéma Ozone16 [Djouad *et al.*, 2002b] (voir le tableau B.1) a été effectuée. L'erreur relative en fonction des itérations de BFGS est tracée sur la figure B.5. Le résultat obtenu est similaire à celui sur les conditions initiales : il est uniquement possible de retrouver la constante cinétique lente.

Conclusion

L'impact du comportement "lent-rapide" sur l'assimilation de données et la modélisation inverse des émissions a été mis en évidence par des résultats théoriques et des expériences numériques.

Pour un cas pratique, le point clé est qu'il est difficile de retrouver les conditions initiales des espèces rapides, ce qui n'est pas capital vu qu'elles n'ont pas d'influence sur le comportement à long terme du système.

Une question clé de l'assimilation de données concerne le choix des observations (contraintes techniques mises à part). Quelles espèces observées, à quels moments, en

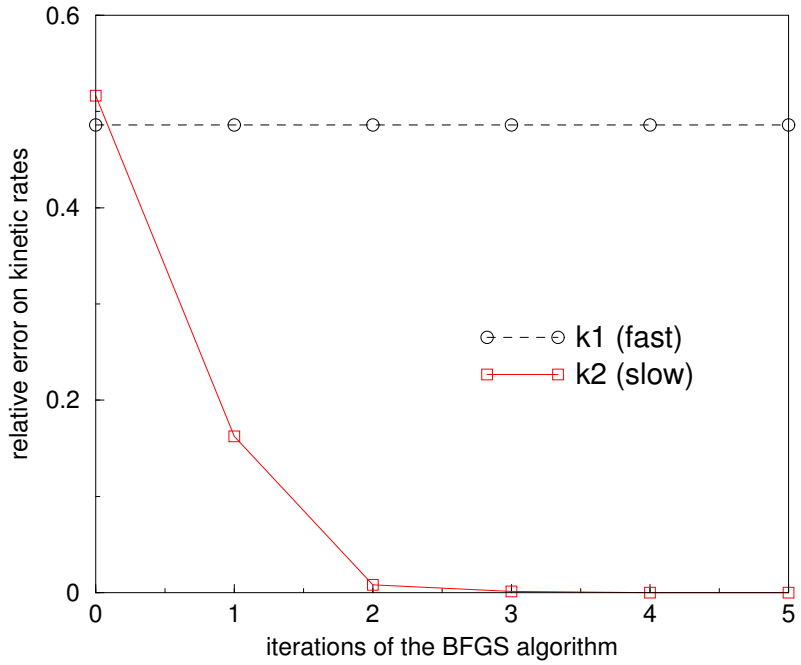

FIG. B.5 – Erreur relative sur les constantes cinétiques en fonction des itérations de BFGS.

quels lieux ? Cet exemple illustre la problématique du choix de l'espèce et du moment d'observation.

Annexe C

Assimilation de données pour la dispersion à petite échelle de radionucléides

Résumé

Dans le cas d'un incident ou d'un accident survenant dans une centrale nucléaire, le suivi de la dispersion des radionucléides rejetés dans l'atmosphère est bien entendu crucial. Pour un incident, la problématique est une problématique d'échelle locale (quelques kilomètres autour de la source) et les modèles utilisés dans une application opérationnelle sont typiquement des modèles de dispersion gaussienne, fondés sur des calculs analytiques.

La question de l'assimilation de données est bien entendu pertinente dans un contexte de prévision.

Ce chapitre est essentiellement basé sur un rapport à destination de l'Institut de Radioprotection et de Sûreté Nucléaire (IRSN) et un article en cours de publication dans Journal of Environmental Radioactivity. Il constitue une étude de faisabilité préliminaire et présente des calculs de second-ordre dans un contexte académique.

Sommaire

C.1 Problématique 205
C.1.1 Contexte industriel 205
C.1.2 L'intérêt de l'assimilation de données 206
C.1.3 Modélisation de la dispersion 206
C.1.4 Rappel sur les méthodes d'assimilation de données 209
C.1.5 Objectifs visés 214

C.2 Analyse de sensibilité 215
C.2.1 Méthodologie 215
C.2.2 Résultats 215
C.3 Expérience de modélisation inverse 221
C.3.1 Présentation 221
C.3.2 Aspects numériques 223
C.3.3 Recherche de l'ensemble des paramètres 223
C.3.4 Recherche des débits à météo et à paramètres de dépôt connus . 226
C.4 Analyse du second ordre et expériences de positionnement de capteurs 227
C.4.1 Présentation 227
C.4.2 Influence aux paramètres physiques 227
C.4.3 Influence aux paramètres d'assimilation 228
C.4.4 Sensibilité au réseau d'observation 228
C.5 Conclusion 230
C.5.1 Travail réalisé 230
C.5.2 Enseignements 230
C.5.3 Prospective 231

Introduction

L'objet de ce rapport est de faire un point sur les travaux de modélisation inverse et d'assimilation de données pour la dispersion de radionucléides à petite échelle, menés au sein du Pôle Air du Cereve pour le compte de l'IRSN. Ce travail correspond à l'étude de faisabilité telle que présentée dans la lettre de commande numéro IPSN/DPEA/SECRI/01-306.

On présente dans ce rapport :
- les objectifs recherchés, les modèles utilisés et la problématique de l'assimilation de données dans ce contexte (section C.1),
- le travail effectué à proprement parler avec des analyses de sensibilité (section C.2), des expériences de modélisation inverse (section C.3) et une étude préliminaire sur le positionnement optimal des capteurs de mesures (section C.4).

Cette étude de faisabilité, à la suite de [Chemith, 2001], montre la pertinence à utiliser les méthodes classiques d'assimilation de données dans le contexte de la dispersion à petite échelle des radionucléides et a permis de comparer diverses stratégies d'assimilation de données.

L'extension de ce travail dans un environnement à caractère plus opérationnel reste bien entendu à mener en parallèle avec une amélioration des modèles physiques. De plus, des points restent à approfondir dans une telle optique, notamment sur le positionnement optimal des capteurs ou, en retour d'expérience de campagnes de mesures, la construction d'une base de données précisant les statistiques d'erreur à prendre en compte.

C.1 Problématique

C.1.1 Contexte industriel

Lors d'un incident survenant sur une installation nucléaire, l'IRSN est chargé d'évaluer, en particulier via son Centre Technique de Crises (CTC) les conséquences à court terme d'un rejet de produits radioactifs dans l'environnement et d'apporter son conseil aux pouvoirs publics.

Pour assurer sa mission d'expertise en phase d'urgence, il dispose de deux sources d'information :

1. des *mesures* : Des mesures d'activité au sol sont réalisées dans la zone du rejet. L'interprétation de ces mesures est délicate car :
   - le nombre de mesures disponibles est limité.
   - l'incertitude sur les mesures peut être importante.

   Ces mesures peuvent être *a priori* fixes ou mobiles.

2. des *modèles* : L'IRSN dispose également d'un ensemble d'outils de prévision : la filière CONRAD (CONséquences RADiologiques). Celle-ci est composée pour partie d'outils rapides à précision limitée (ABAQUES) et de modèles de second niveau, fondés sur une description gaussienne de la dispersion atmosphérique (SIROCCO : Système Informatique Rapide et Opérationnel de Calcul des COnséquences, [Monfort, 1999]) qui permettent d'estimer rapidement (en conditions réelles opérationnelles) la dispersion des polluants dans l'atmosphère.

   Les données météorologiques (vent, température, classe de stabilité de l'atmosphère, pluviométrie) sont fournies en amont du modèle par Météo France.

   Enfin, en aval de cette chaîne, un module est plus orienté vers l'impact sanitaire (ASTRAL) et à partir des concentrations et des dépôts au niveau du sol permet d'évaluer notamment les doses reçues par inhalation et irradiation.

C.1.2 L'intérêt de l'assimilation de données

La validité de cette chaîne d'outils dépend bien entendu de plusieurs facteurs de nature radicalement différente, notamment :

- la qualité des données météorologiques fournies,
- la validité du modèle de dispersion gaussien à proprement parler,
- la qualité des données de rejet, estimées par des modèles d'évolution de réacteur.

De manière à présent classique en prévision environnementale, on se propose de *coupler les deux sources d'information (les mesures et l'outil de prévision)* de manière à :

- Réduire et quantifier les incertitudes de la prévision,
- Améliorer la connaissance de certains paramètres, notamment affiner la valeur des débits de rejet.

Dans les sous-sections qui suivent, on va d'abord rappeler la modélisation de la dispersion utilisée à ces échelles (modèles gaussiens, modèles à bouffée). Notre travail s'inscrit dans le contexte des méthodes d'assimilation de données, dont on rappelle les principes dans la section C.1.4.

Une nomenclature permet en annexe de préciser les diverses notations utilisées. Des éléments techniques (notamment sur les aspects numériques de l'assimilation de données) sont précisés en annexe.

C.1.3 Modélisation de la dispersion

C.1.3.1 Hypothèses simplificatrices

On utilise pour notre étude, un modèle de dispersion atmosphérique de type « panache »([Seinfeld, 1985]).

Il consiste à transporter des radio-éléments provenant :

– d'une source ponctuelle (la cheminée de l'installation),
– d'une source diffuse correspondant à une fuite dans l'enceinte de confinement. Cette partie diffuse est modélisée par un rejet ponctuel en amont de l'installation.

Dans cette étude de faisabilité, on ne considérera qu'*un seul* radioélément et l'on négligera les processus de filiation.

La dispersion à proprement dite est décrite par un profil gaussien caractérisé par des écarts types, donnés par les formules de Doury pour l'application de l'IRSN.

Lors du transfert dans l'atmosphère, le rejet est appauvri sous l'effet des dépôts secs et humides, auxquels s'ajoutent les processus de filiation en toute généralité.

C.1.3.2 Modèle utilisé

Ceci conduit à résoudre une expression analytique pour l'activité volumique au point spatial de coordonnées (x, y, z) de la forme :

$$act(x,y,z) = \frac{1}{2\pi u} f_{rad} f_{dep} f_y \big[\frac{Q_1}{\sigma_y(x)\sigma_z(x)} f_z(z_1) + \frac{Q_2}{\sigma_y(x-d)\sigma_z(x-d)} f_z(z_2) \big] \qquad \text{(C.1)}$$

où :

– $act(x, y, z)$ représente l'activité en un point spatial (en Bq/m^3), où x,y et z désignent les distances au point de rejet ponctuel.

– Q_1 et Q_2 sont des débits de rejet correspondant à deux sources ponctuelles. La source indicée 1 est le point de rejet de la cheminée tandis que celle indicée 2 est placée en amont du rejet à une distance d et modélise une source diffuse.

– u est la composante horizontale du vent, α l'angle du vent avec une direction fixée. Numériquement, on a choisi de prendre l'axe du maillage en x confondu avec la direction du vent, c'est à dire qu'en pratique, on a imposé au maillage une rotation d'angle $-\alpha$, centrée autour du point de rejet 1.

– $\sigma_y(x)$ (resp. $\sigma_z(x)$) est l'écart type de Doury pour la direction y (resp. z) de la forme $(A_y\, t)^{k_y}$ (resp. $(A_z\, t)^{k_z}$) avec (A_y, k_y) (resp. (A_z, k_z)) un jeu de constantes caractéristiques de la diffusion, qui varie en fonction de la stabilité de l'atmosphère, et $t = \frac{x}{u}$ le temps de transfert.
On se réfère plus spécifiquement à [Chemith, 2001] pour plus de précision sur le modèle direct, notamment pour des valeurs typiques des coefficients de Doury.
Notons enfin que la définition des écarts types est affinée dans le cadre du *modèle franco-allemand* ([Crabol et al, 1999]), notamment en fonction du site et des conditions météos.

– f_y et f_z ont pour expression :

$$f_y = \exp(-\frac{y^2}{2{\sigma_y}^2}) \tag{C.2}$$

$$f_z(z_i) = \exp(-\frac{(z-z_i)^2}{2{\sigma_z}^2}) + \exp(-\frac{(z+z_i)^2}{2{\sigma_z}^2}) \tag{C.3}$$

où z_i désigne la hauteur du rejet du rejet i.

– le terme d'appauvrissement f_{dep} est le produit des termes de dépôts secs f_{sec} et humides f_{hum} :

$$f_{sec} = \exp\Big(-\frac{v_d}{u}\sqrt{\frac{2}{\pi}}\int_0^{\frac{x}{u}}\frac{f_z(z_{ref})}{2{\sigma_z}^2}dt\Big) \tag{C.4}$$

$$f_{hum} = \exp(-\Lambda\, t_{pluie}) \tag{C.5}$$

où v_d désigne la vitesse de dépôt sec, z_{ref} correspond à une hauteur de référence à laquelle est calculée le flux de dépôt, Λ est le taux de lessivage par la pluie et t_{pluie} est la durée de la pluie.

– le terme f_{rad} correspond à un appauvrissement en fonction de la décroissance radioactive du radionucléide père et des processus éventuels de filiation.
Si l'on ne tient compte que de la décroissance radioactive :

$$f_{rad}^{(0)} = \exp(-\lambda_p\, x/u) \tag{C.6}$$

pour le radionucléide « père ».
Si l'on tient compte d'une filiation à un niveau (père/fils) :

$$f_{rad}^{(1)} = \frac{\lambda_f}{\lambda_p+\lambda_f}(\exp(-\lambda_p\, x/u) - \exp(-\lambda_f\, x/u)) \tag{C.7}$$

pour le radionucléide « fils ». λ_p (resp. λ_f) est la constante de décroissance radioactive du père (resp. du fils) en s^{-1}.

C.1.3.3 Relations entrées/sortie

Rappelons que dans la suite, on ne tiendra pas compte des termes de lessivage humide et des termes de filiation radioactive.

Pour synthétiser, le modèle utilisé calcule l'activité sur un maillage tridimensionnel cartésien (x, y, z) régulier en fonction des entrées suivantes :

– les paramètres météos : vitesse de vent u, angle du vent α, stabilité atmosphérique $stab$,
– les paramètres de rejet : débit Q_1 et Q_2, hauteurs z_1 et z_2, distance d du rejet diffus en amont du rejet ponctuel,

– les paramètres de dépôt sec : vitesse de dépôt sec v_d et hauteur de référence z_{ref}.

La figure C.1 représente un exemple de carte d'activité au sol adimensionnée par l'activité maximum (au lieu de rejet). Elle est tracée en échelles logarithmiques avec les paramètres numériques fournis dans l'annexe A.

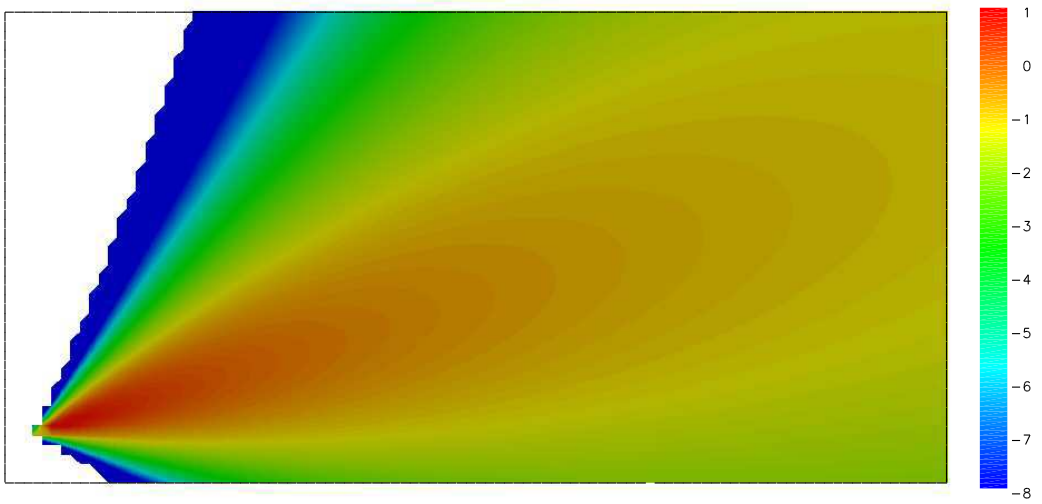

FIG. C.1 – Activité volumique au sol

C.1.4 Rappel sur les méthodes d'assimilation de données

C.1.4.1 Définition du problème de minimisation

On se réfère par exemple à [Tarantola, 1987] pour une présentation générale de ces techniques.

On cherche à prévoir un état de l'atmosphère, ici des composés traces de type radionucléides, dont la description est fournie par l'activité *act*.

On dispose à cette fin d'un modèle numérique f_{direct} (c'est ce que l'on appellera par la suite le modèle *direct*), décrivant l'évolution de la variable de sortie (*act*) en fonction des paramètres d'entrée (notés p) :

$$act = f_{direct}(p) \tag{C.8}$$

Le modèle direct a été décrit au paragraphe précédent par l'équation (C.1). p recoupe l'ensemble des paramètres déterminant les sorties :

– paramètres de débits : Q_1, Q_2, z_1, z_2, d.
– paramètres météos : u, α,
– paramètres de lessivages : v_d, z_{ref},

Les paramètres d'entrées sont connus avec plus ou moins de précision et l'utilisation de mesures (d'observations de la sortie) peuvent permettre de retrouver des valeurs de p (ou d'une partie de p) plus conformes à la réalité.

On suppose donc connu un jeu d'observations de la sortie, noté *obs*. On cherche à minimiser l'écart entre les sorties de modèles et les observations, par rapport aux paramètres d'entrée :

$$min_{p \in P_{adm}} J(p) = ||f_{direct}(p) - obs||^2 \tag{C.9}$$

où P_{adm} désigne l'espace des paramètres *admissibles* (par exemple, on cherche un débit positif). La fonction J est classiquement appelée *fonction coût* et donne une mesure de l'écart entre modèles et observations.

C.1.4.2 Définition de la fonction coût. Paramètres d'assimilation

Un élément important est le choix de la norme $||.||$ permettant d'estimer cet écart. Typiquement, on peut expliciter J selon :

$$J(p) = (p - p_b)^2/\sigma_b^2 + \sum_{i=1}^{i=n_{obs}} ((f_{direct}(p))_i - obs_i)^2/\sigma_i^2 \tag{C.10}$$

dans le cas d'un paramètre p scalaire. L'extension au cas vectoriel est immédiate.

Précisons les notations utilisées :

- Paramètre de connaissance *a priori* :
  - p_b désigne l'ébauche (*background* en anglais) pour le paramètre recherché, c'est à dire la connaissance *a priori* dont on dispose pour p.
  - σ_b est la variance de l'erreur d'ébauche. Autrement dit, pour une faible confiance accordée à l'ébauche, σ_b a une valeur élevée.
  - L'indiçage par i correspond à la i-ème composante des sorties du modèle (indicées par exemple par les positions spatiales discrètes sur le maillage de calcul).
- Paramètres d'observations
  - n_{obs} désigne le nombre d'observations.
  - obs_i correspond à une mesure de la i-ème composante de la sortie (c'est à dire $((f_{direct}(p)_i)$.
  - σ_i est la variance de l'erreur d'observation pour l'observation i. Notons que l'on a supposé implicitement que les erreurs d'observations étaient décorrélées entre elles et décorrélées de l'erreur d'ébauche.

Formellement, la fonction coût est donc à la fois une fonction des paramètres « physiques »(p) et des paramètres d'assimilation (ce qui détermine le choix de la norme $||.||$), que l'on notera p_{assim} :

$$p_{assim} = (p_b,\, \sigma_b,\, (\sigma_i)_i) \tag{C.11}$$

Pour marquer ces dépendances, on notera :

$$J(p, p_{assim}) \tag{C.12}$$

Il convient de mettre l'accent sur l'importance à bien connaître ces paramètres d'assimilation en situation réelle, c'est à dire en particulier les variances d'erreurs d'ébauche et d'observation. Pour les observations, ceci passe par le calcul de statistiques d'erreur : si act_r est la réalité, on a par définition de la matrice de variance de l'erreur d'observation :

$$\begin{aligned}(\sigma_{ij}^2) &= < (act_r - obs)^T (act_r - obs) > \\ &\simeq 1/n_{exp} \sum_{k=1}^{k=n_{exp}} (act_r^{(k)} - obs^{(k)})^T (act_r^{(k)} - obs^{(k)}) \end{aligned} \tag{C.13}$$

où $< . >$ correspond à une moyenne statistique, n_{exp} est le nombre de jeux d'observations disponibles (indicés par (k)). Avec l'hypothèse de décorrélation, nous avons supposé cette matrice diagonale et $\sigma_i = \sigma_{ii}$.

C.1.4.3 Résolution du problème de minimisation

La solution du problème de minimisation (C.9) est donnée par l'annulation du gradient de J par rapport à p et définit une valeur *optimale* du paramètre p, que l'on notera p^* :

$$\nabla_p J(p, p_{assim}) = 0 \quad \Leftrightarrow \quad p = p^*(p_{assim}) \tag{C.14}$$

Il convient de noter que cette valeur optimale dépend des valeurs des paramètres d'assimilation p_{assim}.

La résolution de (C.14) peut être effectuée :
- soit de manière directe, par résolution explicite, lorsque le modèle direct est linéaire (ou linéarisé). Ceci conduit alors typiquement aux algorithmes de type filtres de Kalman avec une expression analytique pour p^*.
- soit à une résolution itérative par convergence d'un algorithme de minimisation.

  p^* est alors obtenu comme limite d'une suite d'itérations. Par exemple, en utilisant un algorithme de type gradient :

  $$p_{k+1}^* = p_k^* - \beta_k \nabla_p J(p_k) \tag{C.15}$$

  avec β_k un paramètre numérique et k l'indice d'itération. On a alors :

  $$p^* = lim_{k \to \infty} p_k^* \tag{C.16}$$

  Typiquement, 10 à 15 itérations sont suffisantes pour obtenir la convergence avec des algorithmes de minimisation plus élaborés, comme la méthode BFGS ([H. Byrd et Zhu, 1995]), que nous utilisons en pratique. Pour le cas étudié, 3 itérations suffisent en général.

En général, le point de départ de ces algorithmes itératifs coïncident avec la valeur à priori p_b que l'on donne à ce paramètre.

C.1.4.4 Modèle adjoint et différenciation automatique

L'utilisation de méthodes itératives nécessite donc de pouvoir disposer du gradient de la fonction coût en fonction des paramètres, $\nabla_p J(p_n)$, et ce à chaque itération.

On note plus généralement $s(p)$ une sortie scalaire de modèle dont on veut calculer le gradient par rapport aux paramètres d'entrée : cela peut être la fonction coût J ou une composante des activités (act_i) pour le calcul de sensibilité.

Une première manière d'effectuer ce calcul de gradient est de procéder par différences finies : en perturbant le paramètre p_i de manière infinitésimale, soit δp_i, on calcule alors :

$$\frac{s_j(p_i + \delta p_i) - s_j(p_i)}{\delta p_i} \tag{C.17}$$

L'inconvénient de cette approche est d'une part le manque de précision (il y a une très forte sensibilité au choix de la perturbation), et d'autre part le coût calcul. Si on cherche à calculer une sensibilité pour n paramètres, pour une sortie de modèle, on a en effet besoin d'utiliser le modèle $n+1$ fois, ce qui peut être rapidement prohibitif, notamment dans un contexte opérationnel.

On a donc choisi d'utiliser une approche adjointe ([Le Dimet et Talagrand, 1986]), qui permet de calculer la sensibilité d'une sortie de modèle par rapport à l'ensemble des paramètres pour un coût calcul de l'ordre de deux fois le modèle direct (un peu plus en réalité). On se réfère à l'annexe B pour une présentation rapide de cette approche.

Le point clé est d'obtenir à partir du modèle (du code numérique) direct, donnant la dépendance $act(p)$, le modèle *adjoint*, permettant de calculer $\partial act_i / \partial p$.

Nous avons utilisé un différenciateur automatique, le logiciel ODYSSÉE, développé par l'INRIA ([Faure et Papegay, 1997]). Modulo une écriture « appropriée »du code initial, nous avons obtenu le code adjoint. Celui-ci a été validé par le test du gradient (annexe B).

C.1.4.5 Analyse du second ordre

Une fois calculé la solution optimale $p^*(p_{assim})$, on peut s'interroger sur la sensibilité du résultat aux paramètres de l'assimilation. Ceci recoupe :

- la sensibilité aux paramètres de positionnement du capteurs (c'est-à-dire le choix des composantes d'observation),
- la sensibilité aux observations elles-mêmes :

$$\frac{\partial p^*}{\partial obs} \tag{C.18}$$

- la sensibilité par rapport à la connaissance de l'ébauche :

$$\frac{\partial p^*}{\partial p_b} \tag{C.19}$$

– la sensibilité par rapport aux variances d'erreur :

$$\frac{\partial p^*}{\partial \sigma} \tag{C.20}$$

De manière analogue, si on suppose connue une partie des paramètres physiques (que l'on notera p_c : typiquement les champs météos), le modèle direct s'écrit :

$$act = f_{direct}(p, p_c) \tag{C.21}$$

et la fonction coût devient une fonction dont on écrit la dépendance selon :

$$J(p, p_c, p_{assim}) \tag{C.22}$$

La solution optimale dépend alors aussi des paramètres physiques connus :

$$p^*(p_c, p_{assim}) \tag{C.23}$$

On peut alors s'interroger sur la sensibilité des résultats à la valeur des paramètres physiques *supposés* connus (alors qu'ils ne le sont pas nécessairement en pratique) :

$$\frac{\partial p^*}{\partial p_c} \tag{C.24}$$

Toutes ces sensibilités de la solution optimale à divers paramètres définissent ce que l'on appellera une *analyse du second ordre* ([Wang *et al.*, 1992]). Ce calcul est habituellement omis alors qu'il est d'une importance cruciale puisqu'il donne une indication sur la robustesse des résultats.

Techniquement, on a aisément par différenciation, par exemple pour les paramètres physiques connus p_c :

$$\frac{\partial p^*}{\partial p_c} = -Hess(J)^{-1} \frac{\partial^2 J}{\partial p_c \partial p} \tag{C.25}$$

avec $Hess(J)$ le Hessien de J par rapport à p.

Une analyse du second ordre est donc coûteuse puisqu'elle nécessite de différencier le système d'optimalité obtenu.

On renvoie à l'annexe B pour un descriptif de la mise en œuvre de l'analyse du second ordre et les aspects à proprement parler numérique.

C.1.4.6 Conditions de calcul : expériences synthétiques

Dans la suite, on va effectuer des expériences synthétiques (appelées aussi « expériences jumelles »), c'est à dire que les observations sont générées numériquement puis perturbées selon des distributions gaussiennes centrées autour de la réalité (supposée connue) avec une variance correspondant à la variance d'erreur utilisée pour l'assimilation.

Notons qu'en toute généralité, on aurait pu distinguer la variance d'erreur utilisée pour l'assimilation de donnée (notée $\sigma^{(1)}$), qui reflète la connaissance *a priori*, de la variance d'erreur effective (notée $\sigma^{(2)}$). Cela aurait pu permettre d'évaluer les stratégies « frileuses »(resp. « confiantes ») définies par $\sigma^{(2)} \leq \sigma^{(1)}$ (resp. $\sigma^{(1)} \leq \sigma^{(2)}$).

Dans un premier temps, on suppose tous les paramètres d'entrée du modèle connus. On prend alors pour observations, les sorties d'une exécution du modèle, en supposant que celui-ci est parfait (Il n'introduit pas d'erreur de représentation). Ces observations peuvent aussi être bruitées. Pour la variable obs_j, on suppose connue la réalité act_{r_j}, avec une variance d'erreur σ_j. On tire alors pour valeur effective de obs_j :

$$obs_j = act_{r_j}(1 + \sigma_j w_j) \ , \quad w_j \sim N(0,1) \tag{C.26}$$

$N(0,1)$ correspondant à une distribution gaussienne de moyenne nulle, de variance 1.

On souhaite que les observations soient non biaisées, c'est à dire qu'elles correspondent en moyenne au modèle. En pratique, pour le tirage des $obs_j, j = 1, \ldots, ni_{obs}$, la dernière valeur $obs_{n_{obs}}$ est choisie telle que :

$$\sum_{j=1}^{n_{obs}} (obs_j - act_{r_j}) = 0 \tag{C.27}$$

ce qui conduit à :

$$obs_{n_{obs}} = act_{r_{n_{obs}}}(1 - \sum_{j=1}^{n_{obs}} act_{r_j}\sigma_j w_j) \tag{C.28}$$

Dans un second temps, on « oublie »la valeur de certains paramètres d'entrée et on essaye de les retrouver en ne tenant compte que des observations ainsi générées.

C.1.5 Objectifs visés

En pratique, il s'agit d'affiner la connaissance du débit à partir de mesures d'activité effectuées autour du lieu de rejet.

Dans les sections qui suivent, on présente successivement :

- l'étude de sensibilité des sorties du modèle par rapport aux principaux paramètres du modèle (section C.2).
  Cette analyse permet *a priori* d'avoir une estimation de la difficulté de mener à bien la modélisation inverse. Elle donne également des indications sur des positionnements de capteurs « préférables ».
- des expériences de modélisation inverse (section C.3) sur les paramètres déjà mentionnés : paramètres de débits ponctuels et diffus, paramètres météos, paramètres de dépôt.
  Les cartes de sensibilité et la modélisation inverse ont été faites à l'aide d'approches adjointes, qui ont nécessité la différenciation automatique du code (voir l'annexe B).

– des expériences portant sur la sensibilité des résultats de l'assimilation de données à l'égard de paramètres physiques (des paramètres non recherchés) et de paramètres définissant le réseau d'observation, notamment le nombre d'observation et leur position spatiale (section C.4).

Les tests ont été menés dans plusieurs types de régime (voir l'annexe A).

C.2 Analyse de sensibilité

C.2.1 Méthodologie

C.2.1.1 Paramètres de sensibilité

On cherche ici à quantifier l'importance respective des paramètres d'entrée du modèle pour la sortie (l'activité au sol en l'occurrence, act_j pour la position spatiale j). On note $s_j(p_1, \ldots, p_n)$ la dépendance de la sortie aux n paramètres d'entrées $p_1, \ldots, p_n$.

Les valeurs de sensibilité indiquent comment se propagent les incertitudes sur les paramètres. Ainsi, une perturbation δp_i du paramètre p_i conduit à une perturbation $\delta act = s_i \, \delta p_i$. On va donc chercher à calculer pour le paramètre p_i, la sensibilité sous la forme :

$$s_{ij} = \frac{\partial act_j}{\partial p_i} \tag{C.29}$$

Nos tests n'indiquent pas de différences notables entre les résultats pour les sensibilités absolues (définition ci-dessus) et les sensibilités relatives.

C.2.2 Résultats

Nous présentons les résultats sous la forme de *carte de sensibilité au sol* en échelles logarithmiques. Autrement dit, nous avons tracé le logarithme de la valeur absolue de la sensibilité de l'activité sur toutes les mailles au sol par rapport aux paramètres suivants :

– paramètres météo : vitesse du vent u et angle de vent α,
– paramètres du rejet : altitude z_1, débit Q_1 , débit du rejet diffus Q_2, positionnement du rejet diffus z_2 et d,
– paramètres de dépôt : vitesse de dépôt v_d et hauteur de référence z_{ref}.

Les paramètres de calculs sont fournis dans l'annexe A. Les résultats présentés dans la suite sont relatifs à cette situation.

C.2.2.1 Paramètres météo

La figure C.2 représente la sensibilité de l'activité au sol par rapport à la vitesse du vent. Elle montre l'existence de trois « panaches »de sensibilité : un dans l'axe du vent et

deux autres symétriques par rapport à cet axe. Il est possible de montrer l'existence de ces trois structures par un calcul analytique en considérant un profil de cette sensibilité à x fixé en $z = 0$. Un rapide calcul montre que ce profil suit une loi de la forme :

$$\frac{\partial act}{\partial u} = (f_1(u) - f_2(u)y^2)exp(-f_3(u)y^2) \tag{C.30}$$

où f_1, f_2 et f_3 sont trois fonctions de u positives, dont la valeur dépend des paramètres numériques choisis. Par exemple, pour $f_1(u) = f_2(u) = 1$ et $f_3(u) = 0.5$, la courbe obtenue est celle de la figure C.3.

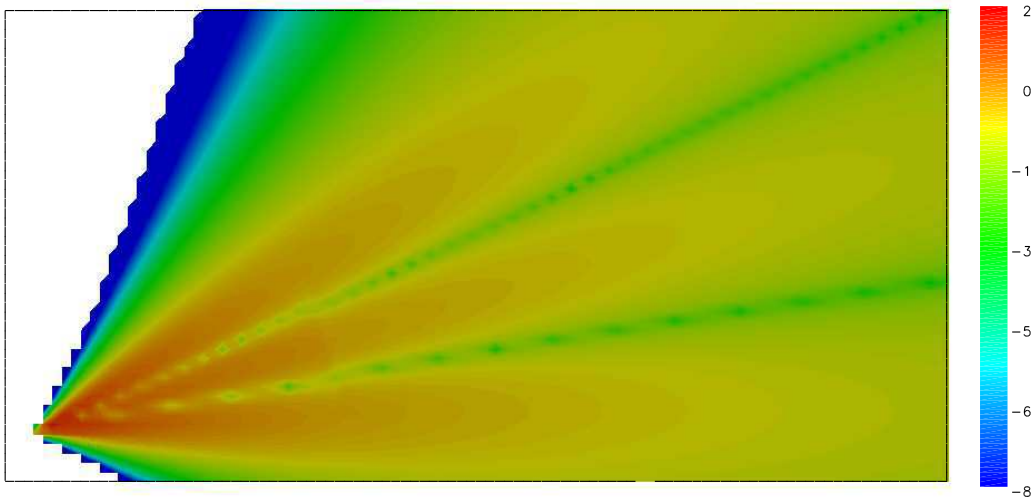

FIG. C.2 – Sensibilité au sol à la vitesse du vent

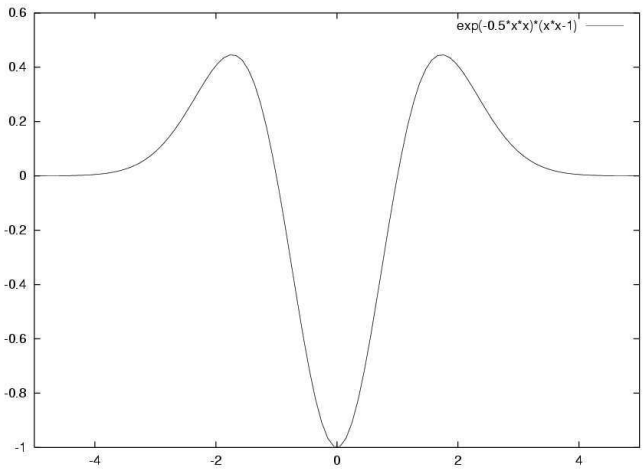

FIG. C.3 – Profil au sol à x fixé de la sensibilité à la vitesse du vent

La figure C.4 correspond à la sensibilité de l'activité par rapport à l'angle entre le vent et l'axe des abscisses. Elle indique que les valeurs de l'activité dans l'axe du rejet ne

dépendent pas (ou très peu) de ce paramètre pour ce jeu de donné (voir Annexe A). Par contre, on observe une forte sensibilité selon deux « panaches »symétriques par rapport à l'axe. Un profil à x fixé en $z = 0$ de cette sensibilité suit une fonction de la forme :

$$\frac{\partial act}{\partial \alpha} = (g_1(\alpha) + g_2(\alpha) * y^2) exp(-y^2 cos^2(\alpha) g_3(\alpha)) + g_4(\alpha) \tag{C.31}$$

où g_1, g_2, g_3 et g_4 sont des fonctions de α, dont la valeur dépend des paramètres numériques choisis. Il faut noter que le signe positif entre g_1 et g_2, conséquence de la dérivation de $-cos(\alpha)$. Par exemple, pour $g_1(\alpha) = 1$, $g_2(\alpha) = 10$, $cos(\alpha) * g_3(\alpha) = 1$ et $g_4(\alpha) = 0$ la courbe obtenue est celle de la figure C.5.

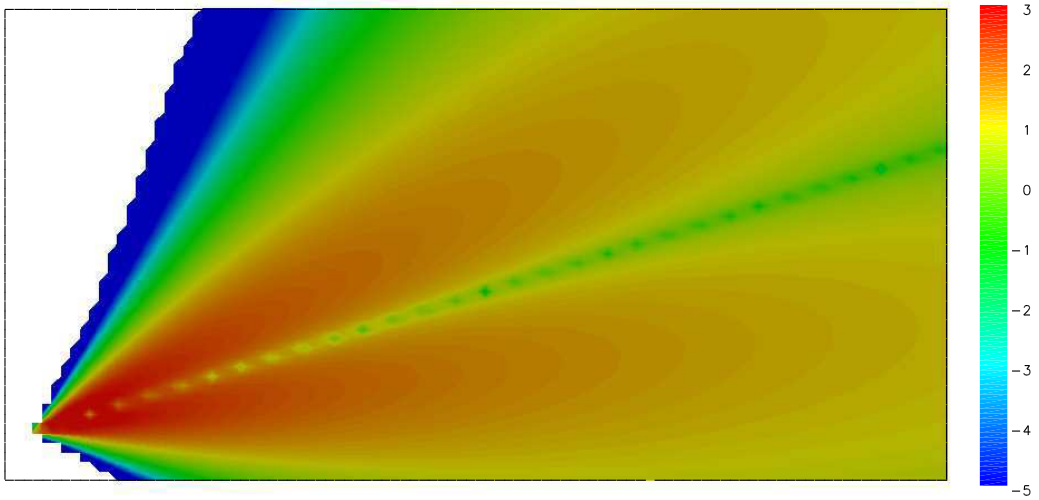

FIG. C.4 – Sensibilité à l'angle du vent

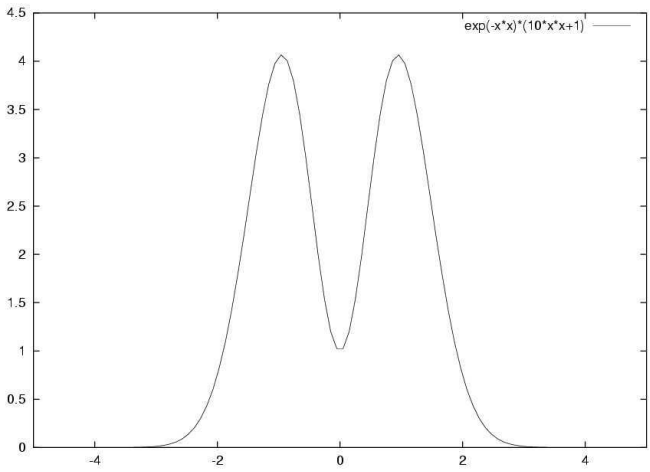

FIG. C.5 – Profil au sol à x fixé de la sensibilité à l'angle du vent

C.2.2.2 Paramètres de rejet

Les figures C.6 et C.7 représentent les sensibilités de l'activité au sol par rapport aux débits des rejets ponctuel et diffus. L'activité étant une fonction linéaire des débits, il est normal de constater que ces sensibilités sont directement proportionnelles à l'activité. Elles montrent une sensibilité très faible à ces deux paramètres.

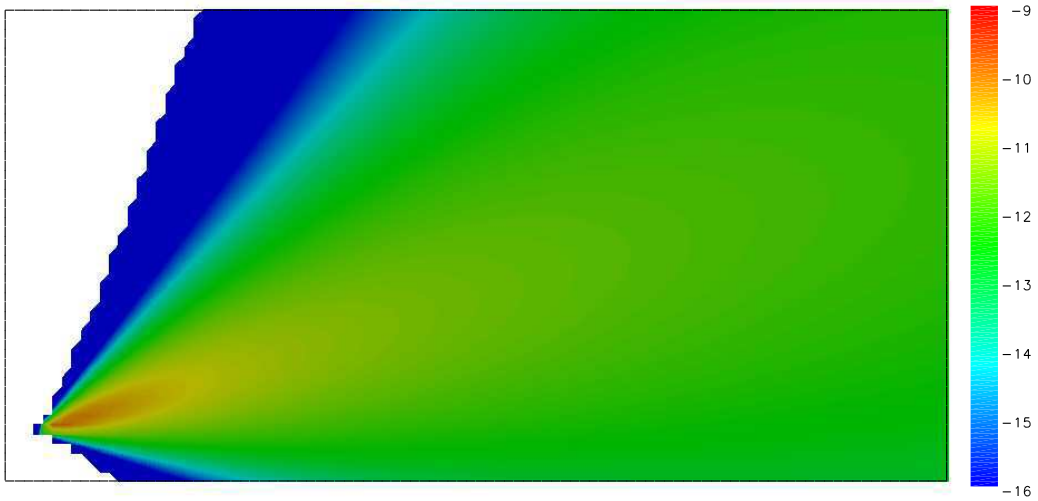

FIG. C.6 – Sensibilité au débit du rejet ponctuel

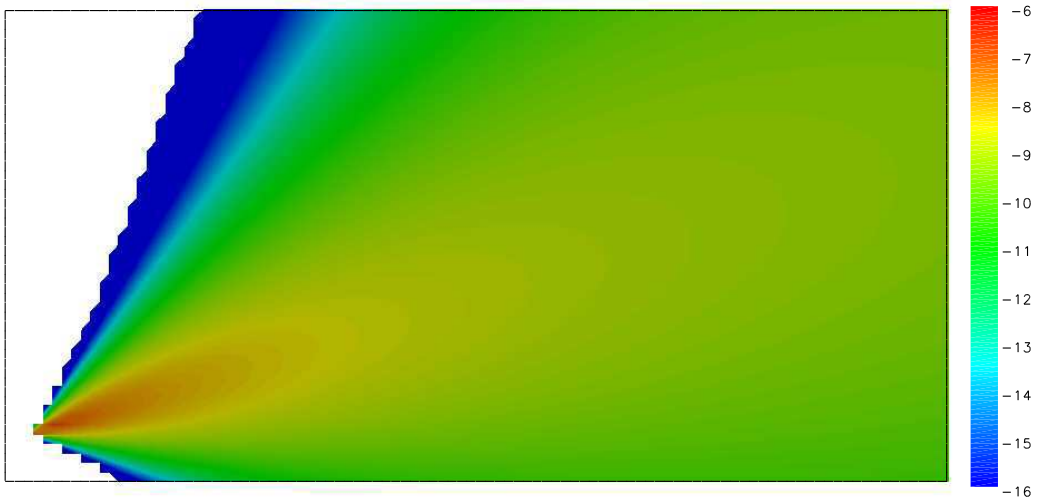

FIG. C.7 – Sensibilité au débit du rejet diffus

Les sensibilités de l'activité par rapport aux paramètres de positionnement spatial des rejets ($z1$, $z2$ et d) sont également faibles, comme on peut le voir sur les figures C.8, C.9 et C.10.

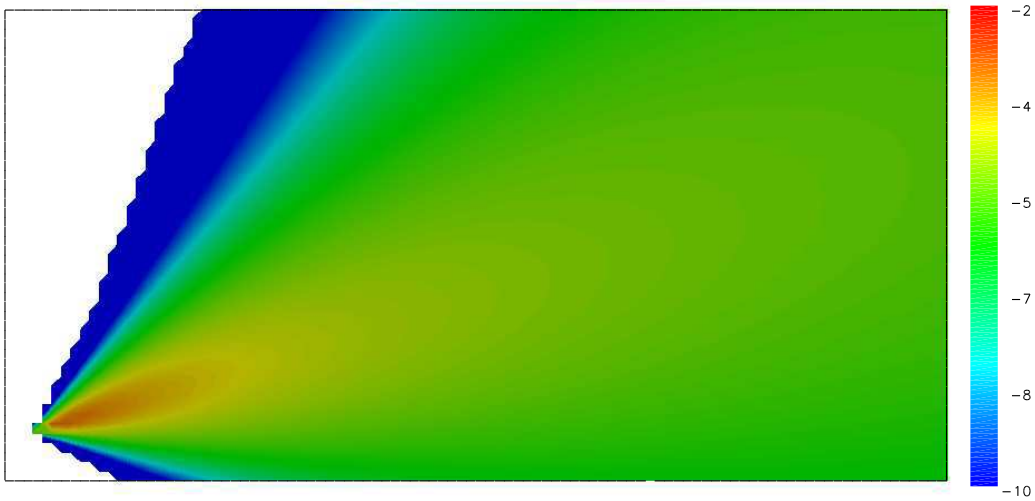

FIG. C.8 – Sensibilité à l'altitude du rejet ponctuel

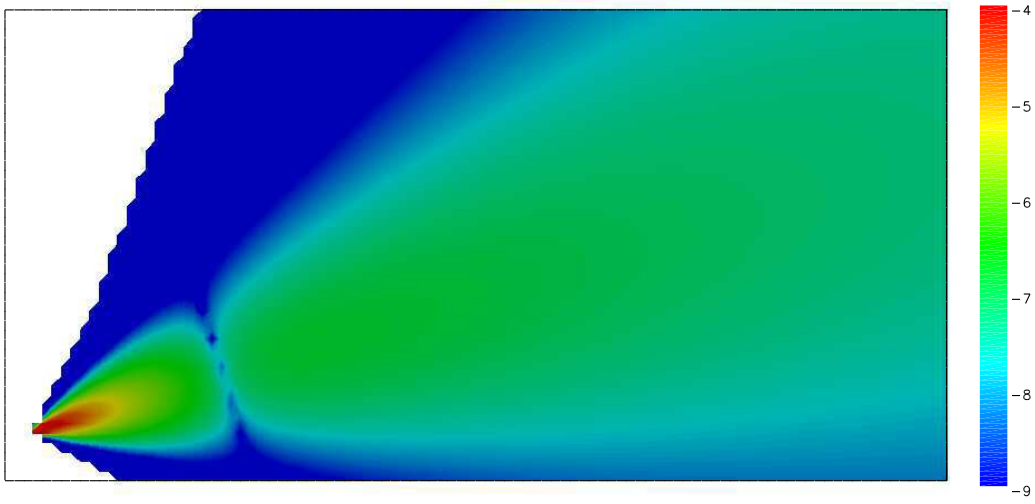

FIG. C.9 – Sensibilité à l'altitude du rejet diffus

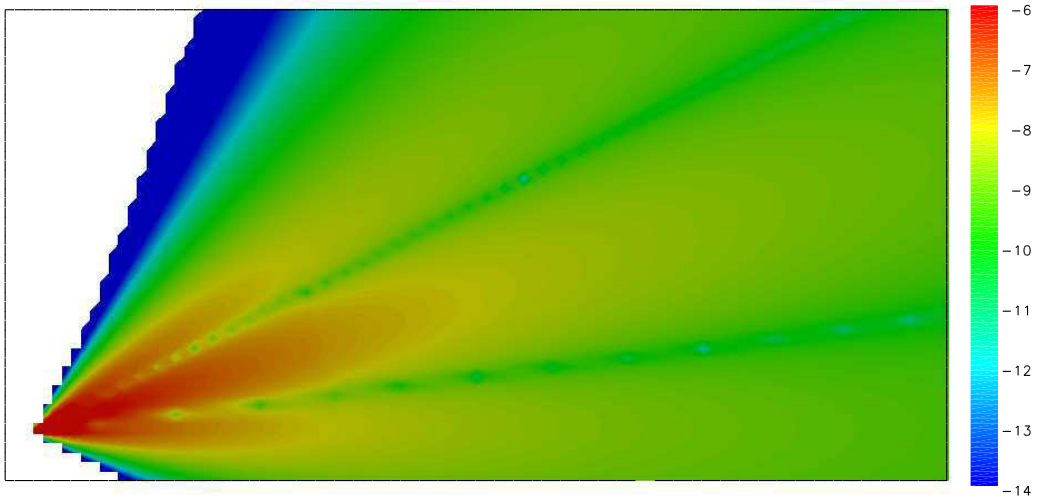

FIG. C.10 – Sensibilité à la distance entre les rejets ponctuel et diffus

C.2.2.3 Paramètres de dépôt

Les figures C.11 et C.12 représentent la sensibilité de l'activité au sol par rapport aux paramètres du dépôt sec. L'influence de la vitesse de dépôt est beaucoup plus locale que celle de la hauteur de référence.

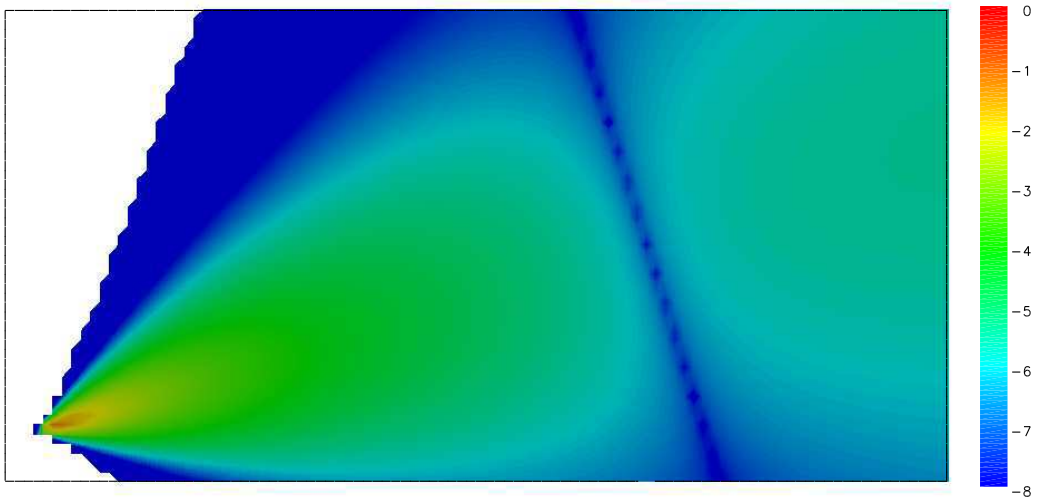

FIG. C.11 – Sensibilité à la vitesse de dépôt

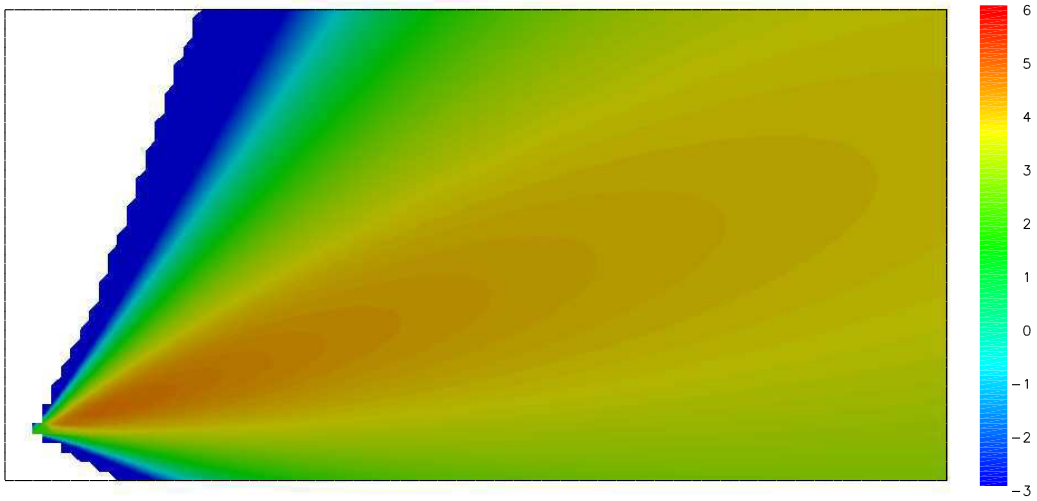

FIG. C.12 – Sensibilité à la hauteur de référence

C.2.2.4 Conclusion

L'étude de sensibilité nous a permis de « classer »les paramètres du modèle selon leur sensibilité. On rappelle que cette étude a été menée avec un seul jeu de paramètres et que les conclusions présentées ici sont relatives à cette situation particulière.

- Les paramètres météos et de dépôt sont les paramètres les plus importants du modèle pour le régime considéré, comme l'indiquent les grandes valeurs de sensibilité des figures C.2, C.4, C.11 et C.12. Il est indispensable par conséquent de bien les spécifier. Il faut noter que la vitesse de dépôt est un paramètre qui est sensible localement (à proximité du rejet uniquement).
- Les paramètres de rejet sont les moins sensibles. En particulier, la valeur des débits n'a pas beaucoup d'influence sur l'activité au sol. De plus, la zone de sensibilité est restreinte à proximité du rejet.

Elle a mis en évidence des régions à forte sensibilité situées à proximité du point du rejet. Leurs géométries diverses selon les paramètres montrent qu'une stratégie de mesures uniquement dans l'axe du rejet ne donnent pas nécessairement l'information la plus pertinente.

C.3 Expérience de modélisation inverse

C.3.1 Présentation

On cherche ici à retrouver des paramètres par modélisation inverse à partir d'observations synthétiques générées numériquement comme indiqué dans le paragraphe C.1.4.6. Le réseau d'observations est défini de la manière suivante :

– nombre d'observation : $n_{obs} = 30$,
– la fenêtre spatiale d'assimilation de donnée : les abscisses des observations sont réparties aléatoirement dans $[x_{min}, x_{max}]$. Typiquement, en mètres :

$$x_{min} = 1000 \ , \quad x_{max} = 5000 \tag{C.32}$$

– l'écart à l'axe du rejet est tiré selon une gaussienne de variance $\sigma_{axe} = 50$ mètres.
– des erreurs d'observation. $\sigma_o = 10\%$.

La figure C.13 représente les valeurs des observations en fonction de l'abscisse de l'axe de rejet. Les observations non perturbées sont aussi représentées.

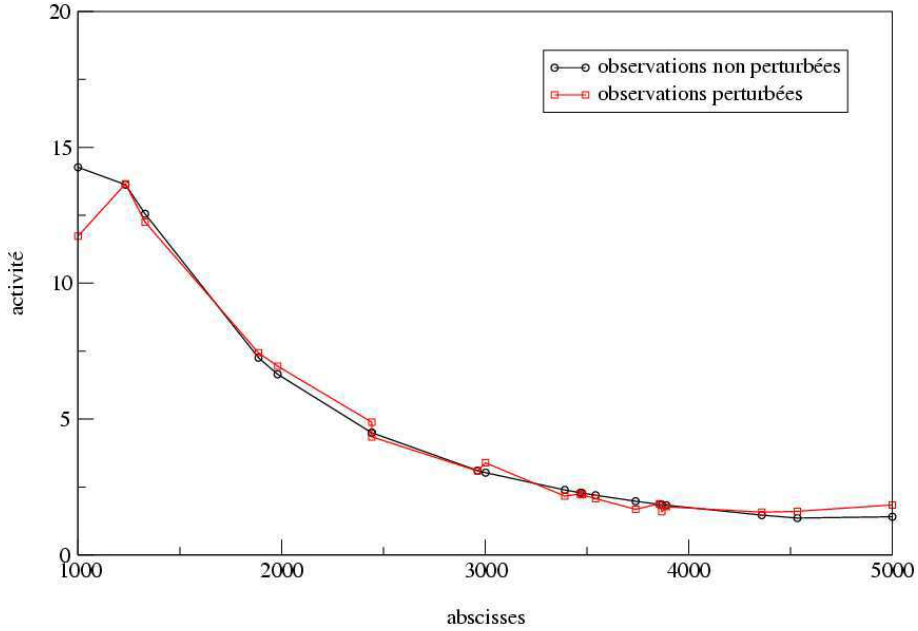

FIG. C.13 – Observations

On se réfère à l'annexe B pour un descriptif détaillé des paramètres de calcul. Notamment, le tableau C.3 de l'annexe B donne les valeurs utilisées dans le processus de modélisation inverse. Il contient :

– les incertitudes σ_b sur les paramètres permettent de choisir une valeur de point de départ (appelée « first guess »pour l'algorithme BFGS. On se place volontairement dans un cas défavorable, où l'erreur relative pour chacun des paramètres est d'environ $3\sigma_b$. On rappelle que la probabilité pour que l'erreur relative d'une variable suivant une loi gaussienne soit plus grande que $3\sigma_b$ est d'environ $0,1\%$. A partir de la valeur réelle du paramètre p_r et de l'incertitude qu'on accorde à cette valeur, on décide de prendre la connaissance à priori du paramètre $p_b = p_r\,(1 \mp 3\,\sigma_b)$.
– l'intervalle admissible, délimité par une borne inf et une borne sup dans lequel le minimiseur recherche la solution.

Dans la fonction coût, seul le terme correspondant à l'écart entre les observations et les sorties du modèle est pris en compte, pour les raisons évoquées dans l'annexe B.

L'algorithme itératif BFGS comporte une condition d'arrêt portant soit sur la valeur du gradient, soit sur la réduction de la valeur de la fonction coût entre deux itérations. En pratique, pour les calculs qui suivent, le test d'arrêt est pris le moins contraignant possible en le reliant directement à la précision de la machine. Dans une situation opérationnelle, il est évident que pour des raisons de coût de calcul, ce test sera plus contraignant. En effet, l'essentiel de la réduction de la fonction coût se fait dans les premières itérations, l'erreur relative sur les paramètres variant très peu par la suite.

C.3.2 Aspects numériques

Les résultats de la modélisation inverse sont particulièrement sensibles à la stratégie numérique utilisée. On présente dans l'annexe B rapidement :

- la problématique de préconditionnement par des matrices liées au Hessien,
- la sensibilité des résultats à la prise en compte ou non de l'ébauche dans la fonction coût.

Nos tests indiquent la pertinence à préconditionner le problème initial. L'algorithme de minimisation converge ainsi plus rapidement (gain de l'ordre de 2 à 3). Cette conclusion serait à affiner dans une application plus complexe.

Nos résultats présentent une grande sensibilité vis à vis de l'ébauche. Rappelons qu'un terme d'ébauche est habituellement utilisé notamment afin de rendre bien posé des problèmes sous-déterminés (plus de paramètres à inverser que d'observations disponibles). Dans notre cas, le problème est systématiquement sur-déterminé.

Nos tests montrent, pour des « faibles erreurs d'observation »(typiquement avec $\sigma_o \simeq 10\%$), la dégradation de la précision des résultats lorsque l'ébauche est prise en compte. On se réfère à l'annexe B pour des éléments de compréhension sur un cas simple.

Lorsque les erreurs d'observation deviennent « fortes »(typiquement 30 à 50 %), le terme d'ébauche permet de contraindre le problème et d'éviter de trop fortes erreurs.

La conclusion de cette étude est donc directement reliée aux statistiques d'erreurs d'observation dont nous pourrions disposer pour un cas pré-opérationnel.

C.3.3 Recherche de l'ensemble des paramètres

Lorsque l'on cherche à retrouver simultanément tous les paramètres, le processus de minimisation s'arrête en ayant réduit la fonction coût de l'ordre de 30%, ce qui est insuffisant. En adaptant le réseau d'observation (notamment l'écartement à l'axe des capteurs et la taille de la fenêtre spatiale d'observations), il est néanmoins possible de retrouver certains paramètres (les paramètres météos qui sont les plus sensibles).

Les problèmes rencontrés dans la minimisation sont souvent reliés à un mauvais conditionnement du Hessien. Il a la propriété de propager les erreurs faites sur les paramètres (cf annexe B). C'est le cas ici, comme le montre le spectre du Hessien représenté en échelles logarithmiques sur la figure C.14.

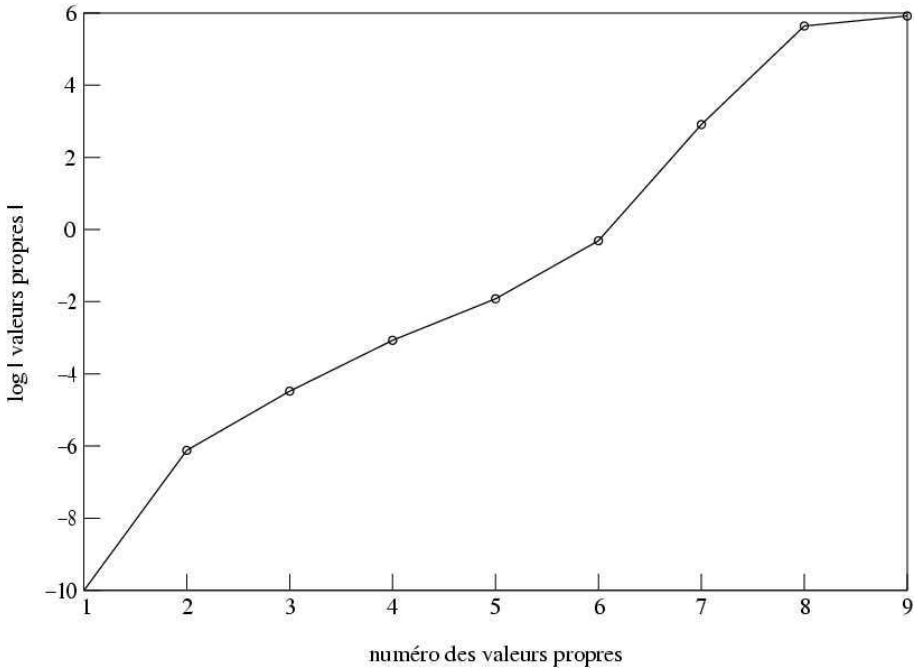

FIG. C.14 – Spectre du Hessien

Les vecteurs propres du Hessien associés aux grandes valeurs propres permettent de connaître les directions privilégiées dans lesquelles l'algorithme de minimisation va fonctionner. Ainsi, la projection des vecteurs de base ($\vec{e_i}$ pour le $i^{ème}$ paramètre) sur le sous-espace propre formé par les vecteurs propres associés aux plus grandes valeurs propres indiquent si le paramètre est facile ou non à retrouver par modélisation inverse. Sur notre exemple, on a reporté dans la figure C.15, la répartition des composantes des vecteurs propres selon les paramètres. Elle indique que les paramètres les plus faciles à retrouver par modélisation inverse sont les paramètres 1, 2 et 5 qui correspondent à u, α et v_d. En effet, les composantes des vecteurs propres associées aux 3 plus grandes valeurs propres sont portées par ces trois paramètres. Ces conclusions ont une valeur locale vu que le Hessien est calculé pour un seul jeu de paramètres (celui correspondant à la réalité). Ces résultats sont en adéquation avec l'analyse de sensibilité de la section 2.

En pratique, on va donc se donner des paramètres météos et de dépôt connus et se concentrer sur la recherche des paramètres de débit.

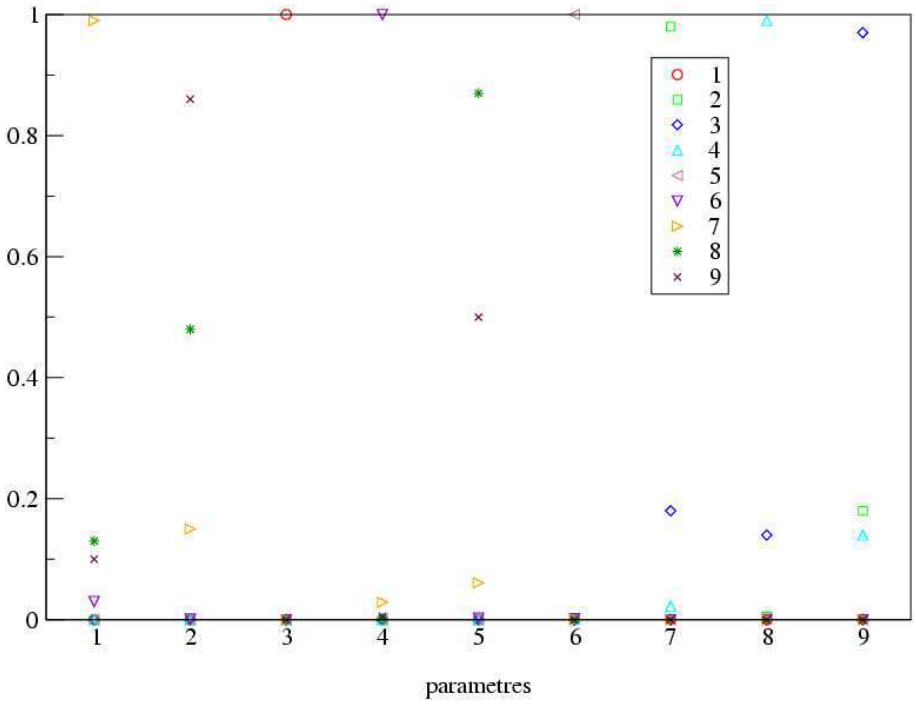

FIG. C.15 – Répartition des composantes des vecteurs propres en fonction des paramètres numérotés de 1 à 9 : u, α, Q_1, z_1, v_d, z_{ref}, Q_2, z_2, d

C.3.4 Recherche des débits à météo et à paramètres de dépôt connus

On cherche à présent uniquement à retrouver la valeur des débits ponctuel et diffus par modélisation inverse en supposant connus la vitesse et l'angle du vent d'une part et la vitesse de dépôt et la hauteur de référence d'autre part. De plus, la location spatiale des rejets est fixe. Le modèle considéré dans ce paragraphe ne contient donc plus que les débits comme paramètres d'entrée.

Les résultats sont conformes avec les calculs de sensibilité : le débit diffus Q_2 étant le paramètre le moins sensible, il est difficile de le retrouver. Quelle que soit la stratégie de connaissance *a priori* utilisée, la recherche de Q_2 est difficile. Nous avons notamment essayé (tests non reportés ici) :

- remplacer Q_2 par le rapport Q_1/Q_2,
- pénaliser par un terme de rappel la fonction coût (rajouter $||Q_1 - \beta\, Q_2||^2$ à la fonction coût, β fixé),

La seule stratégie qui soit efficace est celle consistant à se fixer le rapport $Q_2/Q_1 = 1/\beta$. Dans ces conditions, il n'y a plus qu'une seule variable à chercher (Q_1) et les résultats sont les meilleurs de tous ceux présentés ici.

On peut noter cependant que lorsque l'on multiplie la valeur de Q_2 par 100, il devient alors possible de le retrouver par modélisation inverse. On passe d'une erreur de 15% à 1.3%.

A l'inverse, le paramètre de rejet ponctuel est systématiquement retrouvé avec une stratégie d'observation adaptée (voir section C.4). Partant d'une erreur relative initiale de 9% sur la valeur du débit ponctuel, l'algorithme de minimisation permet en 3 itérations sur la configuration décrite dans l'annexe B, de réduire l'erreur sur ce paramètre à 1.2%. Le choix du point de départ de l'algorithme BFGS n'est ici pas déterminant et on obtient le même résultat en commençant la minimisation avec une erreur relative initiale de 50% par exemple.

Ces résultats s'expliquent par la forme du modèle considéré. Ici, on peut écrire au point spatial i :

$$act_i = a_i\, Q_1 + b_i\, Q_2 \tag{C.33}$$

avec $b_i \ll a_i$. Cet exemple linéaire est détaillé dans l'annexe B. On montre que l'erreur relative liée à la modélisation inverse de Q_1 est bornée par l'erreur d'observation alors que celle liée à la modélisation inverse de Q_2 est grande et est inversement proportionnelle à la valeur de Q_2.

C.4 Analyse du second ordre et expériences de positionnement de capteurs

C.4.1 Présentation

L'objet de cette section est d'étudier la sensibilité des résultats (p^*, ici Q_1^*) à l'ensemble des paramètres supposés connus :

- autres paramètres physiques (notamment la météo),
- paramètres d'assimilation de donnée,
- choix du réseau (nombre, position spatiale).

On se réfère à la section 1 pour un descriptif de l'approche utilisée. Les aspects à strictement parler algorithmiques (calcul de l'adjoint du second-ordre peuvent être trouvés en annexe B).

C.4.2 Influence aux paramètres physiques

Nous avons systématiquement calculé la sensibilité relative du paramètre recherché par rapport à la valeur des autres paramètres physiques supposés connus :

$$\frac{p_c}{p^*}\frac{\partial p^*}{\partial p_c} \tag{C.34}$$

Les paramètres auxquels la sensibilité est la plus forte sont les paramètres météos et la vitesse de dépôt (ce qui est en cohérence avec la sensibilité du modèle direct, voir section 1), comme le montre le tableau C.1. Il est à noter ici la grosse influence de la vitesse de dépôt sur la valeur du débit de rejet obtenu par modélisation inverse.

paramètre	$\frac{p_c}{p^*}\frac{\partial p^*}{\partial p_c}$
u	-2.4 E+5
α	-2.6 E+6
Q_1	1.
z_1	-2.1 E+4
v_d	-2.4 E+6
z_{ref}	-8.7
Q_2	1.7
z_2	-3. E+1
d	-4.7

TAB. C.1 – Sensibilité aux paramètres connus

C.4.3 Influence aux paramètres d'assimilation

Nous avons calculé la sensibilité relative des résultats à la valeur des observations :

$$\frac{obs}{p^*}\frac{\partial p^*}{\partial obs} = \text{constante} \tag{C.35}$$

On rappelle que nous avons fait l'hypothèse que les erreurs d'observations sont les mêmes en tout point spatial. Dans ce cas, la sensibilité des résultats à la valeur des observations n'est fonction que de la sensibilité du modèle à Q_1. Les erreurs de mesures ont donc d'autant plus d'impact qu'elles sont proches de l'endroit où le rejet atteint le sol et qu'on se place dans l'axe du rejet.

Notons que dans le cas où l'erreur d'observations n'est pas constante, la sensibilité à l'observation combine à la fois la sensibilité du modèle direct et les erreurs d'observation (Annexe B).

C.4.4 Sensibilité au réseau d'observation

Nous avons mené plusieurs expériences en modifiant la configuration du réseau d'observation. Les expériences présentées ont été menées sur une configuration qui permet de retrouver déjà correctement le débit de rejet ponctuel (1.2% d'erreur sur le paramètre après minimisation). Les résultats restent les mêmes pour une configuration moins favorable.

Pour cela, nous avons fait varier séparément ;

- le nombre d'observation n_{obs} (voir figure C.16),
- l'extension de la fenêtre d'observation en faisant varier la position du premier capteur vis à vis du rejet x_{min} à x_{max} fixé (voir figure C.17),
- l'éloignement à l'axe du rejet avec σ_{axe}, (voir figure C.18).

Ces tests montrent qu'un réseau d'observation efficace comporte :

- un nombre suffisant d'observations (Pour notre étude, $n_{obs} > 24$ donne de meilleurs résultats,
- une répartition des capteurs autour de l'axe de rejet et non pas uniquement sur l'axe.

Ces tests permettent de montrer la sensibilité des résultats au choix du réseau d'observation. Un « bon »réseau permet d'améliorer fortement les résultats. Il est difficile de mettre en relation les comportements des sensibilités avec les grandeurs physiques du modèle. Les paramètres numériques (Par exemple : le choix du point de départ de l'algorithme BFGS) ne peuvent pas être dissociés dans l'explication de ces résultats.

Nous n'avons pas balayé l'ensemble des configurations possibles d'observation et les réseaux obtenus ne sont que sous-optimaux. L'utilisation de méthodes automatiques reste à mener ([Sportisse et Quélo, 2002]).

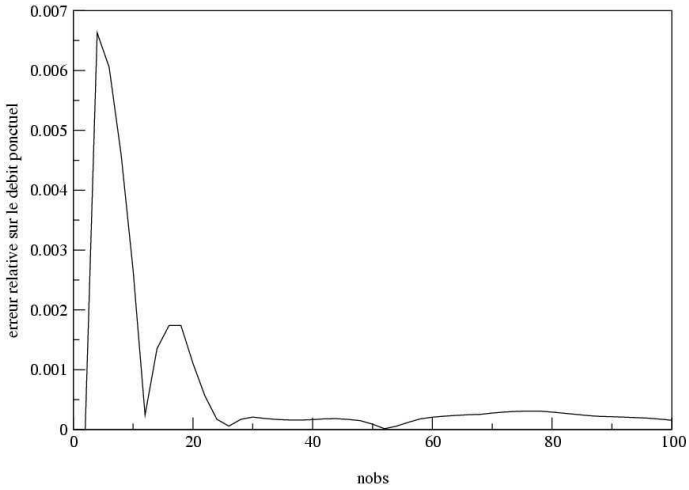

FIG. C.16 – Sensibilité au nombre d'observations

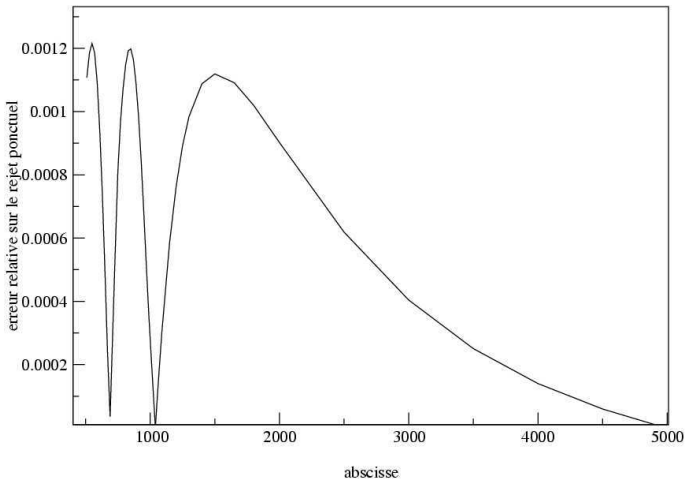

FIG. C.17 – Sensibilité à la taille de la fenêtre

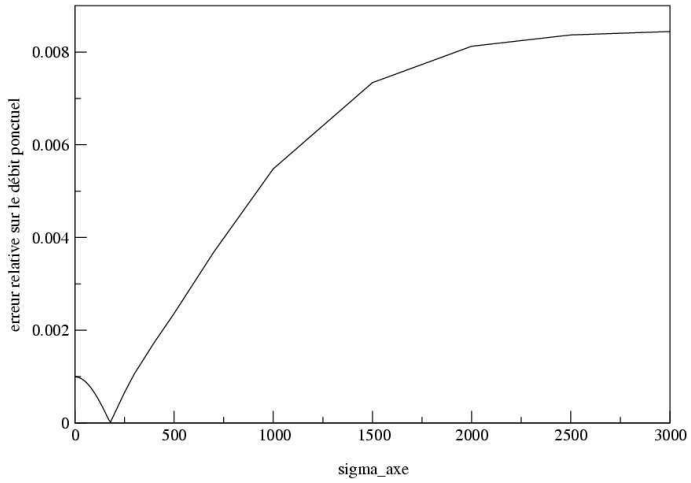

FIG. C.18 – Sensibilité à l'écart à l'axe du rejet

C.5 Conclusion

C.5.1 Travail réalisé

Ce travail a permis de mener à bien une étude de faisabilité de l'utilisation des méthodes d'assimilation de données pour la modélisation de la dispersion à petite échelle des radionucléides.

Le modèle utilisé est un modèle gaussien, une version simplifiée de la chaîne opérationnelle telle qu'elle est opérée par l'IRSN.

Un modèle adjoint a été construit à partir du code initial et a été utilisé pour diverses expériences d'assimilation de donnée.

Une *chaîne logiciel* de divers outils a ainsi pu être construite et est à présent disponible, permettant de :

- calculer des sensibilités des sorties aux divers paramètres d'entrée,
- effectuer la modélisation inverse de paramètres (comme les débits) en fonction d'observations synthétiques (générées numériquement) perturbées,
- étudier la sensibilité de second ordre aux paramètres physiques connus et aux paramètres d'assimilation (observations notamment),
- étudier la sensibilité à la définition du réseau de mesures.

C.5.2 Enseignements

Plusieurs enseignements peuvent être tirés de cette étude de faisabilité :

1. Les méthodes d'assimilation de données variationnelles peuvent être appliquées à ce contexte et permettent effectivement de retrouver des paramètres de débit.

   Le débit ponctuel est systématiquement retrouvé au contraire du débit diffus lorsque celui-ci joue un rôle faible (dans ce cas là, ceci n'a de toute manière pas d'impact sur la qualité de la chaîne de prévision).
2. A petite échelle, selon les configurations, la sensibilité des sorties aux entrées est parfois particulièrement faible. Dans ces cas là, il n'est pas possible de retrouver un jeu de paramètres trop important (combinant par exemple variables de débit et variables météos).
3. Les stratégies d'assimilation de données (préconditionnement, prise en compte ou non de terme d'ébauche) ont un impact fort sur la qualité du résultat. Comme le nombre de paramètres cherchés est faible et qu'aucune régularisation par un terme d'ébauche n'est nécessaire, il est préférable de ne tenir compte (dans ce jeu d'expériences synthétiques) que des observations. Ce point sera bien sûr à réexaminer dans un contexte plus « réel ».
4. Les informations liées aux analyses du second ordre sont riches et ont montré notamment la grande sensibilité des résultats aux conditions météos.
5. La qualité des résultats dépend fortement du réseau d'observation utilisé. Ce point est crucial pour la seconde phase et mérite d'être approfondi dans un contexte plus opérationnel.

C.5.3 Prospective

De manière prospective, on peut dégager plusieurs pistes pour l'utilisation de ces résultats :

1. Le modèle direct utilisé, un modèle gaussien, doit être amélioré afin que la comparaison à des données réelles puisse avoir un sens. On peut songer notamment à l'utilisation de modèles de dispersion plus élaborés et à une complexification de la physique prise en compte (notamment pour les aérosols).
2. Il convient de mener à bien une expérience avec des données réelles.
3. Avant de passer à une phase pré-opérationnelle, une connaissance plus fine des statistiques d'erreurs (notamment d'observation) est nécessaire. Ceci implique une mise à plat de l'existant en terme de mesures et de variances d'erreurs connues ou supposées.
4. Les expériences sur la définition du réseau de mesure doivent être poursuivies, du fait de la forte sensibilité des résultats aux caractéristiques (densité, positionnement) des capteurs de mesure.

Annexes

Paramètres physiques

Toutes les applications numériques ont été menées sur la même configuration d'étude :
- Nous avons effectué les calculs au sol uniquement, sur une grille cartésienne régulière 20000×10000 (en mètres) discrétisée par 100 points selon x et 50 points selon y.
- Les conditions météos sont un vent $u = 2\,m.s^{-1}$ faisant un angle $\alpha = 0,3$ rad avec l'axe des abscisses du maillage. On suppose être dans des conditions de diffusion normale.
- Le rejet ponctuel de débit $Q_1 = 1,1.10^6\,Bq.s{-}1$ s'effectue au point $(500, 1000, 80)$. La contribution du rejet diffus est de $Q_2 = 2.10^3\,Bq.s^{-1}$. Elle est située à la distance $d = 100\,m$ et à la hauteur $z_2 = 40\,m$.
- L'unique radionucléide considéré est un aérosol de vitesse de dépôt $v_d = 5.10^{-3}\,m.s^{-1}$. On a choisi la hauteur de référence $zref = 1\,m$.

Méthodes adjointes et aspects numériques

Adjointisation et différenciation automatique

L'utilisation d'un différenciateur automatique tel que Odyssée ([Faure et Papegay, 1997]) permet de générer automatiquement à partir d'un code existant un code dérivé permettant de calculer le gradient de sorties par rapport aux entrées du modèle.

Par exemple, pour le modèle simplissime suivant :

J=X*Y

pour lequel les paramètres d'entrées sont X et Y, on définit :

1. *le modèle linéaire tangent* qui associe à des perturbations de X et de Y (notées $dXtl$ et $dYtl$) la perturbation résultant pour J ($dJtl$) :

   dXtl= ...
   dYtl= ...
   dJtl(dXtl,dYtl)=X*dYtl+Y*dXtl

   Noter que le calcul de ∇J nécessite deux utilisations de ce code : d'abord $dJtl(1, 0)$ puis $dJtl(0, 1)$.

2. *le modèle adjoint* permet de calculer directement le gradient selon la procédure suivante :
   – on considère la relation :

$$\begin{bmatrix} dXtl \\ dYtl \\ dJtl \end{bmatrix} = M \begin{bmatrix} dXtl \\ dYtl \\ dJtl \end{bmatrix} \quad , \quad M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ Y & X & 0 \end{bmatrix} \tag{C.36}$$

   – puis on écrit pour les variables *adjointes* $d*cl$: :

$$\begin{bmatrix} dXcl \\ dYcl \\ dJcl \end{bmatrix} = M^T \begin{bmatrix} dXcl \\ dYcl \\ dJcl \end{bmatrix} \tag{C.37}$$

   soit encore :

   dXcl(dXcl,dYcl,dJcl)=dXcl+Y*dJcl
   dYcl(dXcl,dYcl,dJcl)=dYcl+X*dJcl
   dJcl=0

   Noter qu'à présent on a directement ∇J en prenant pour entrée $dJcl = 1$, les composantes du gradient étant à lire dans $(dXcl, dYcl)$.

L'extension de cette approche à un cas général (ie un code informatique avec une succession de lignes d'instructions) est relativement aisée. Un point important est l'inversion de la séquence des instructions :

$$(M_1\, M_2)^T = M_2^T\, M_1^T \tag{C.38}$$

Dans l'étape de validation du code dérivé, il faut vérifier la validité du calcul de toutes les dérivées directionnelles. Le test du gradient consiste à regarder la convergence du rapport de la dérivée partielle dans une direction obtenue par différences finies (relative à une certaine perturbation) sur le gradient provenant du code dérivé quand la perturbation est petite :

$$\frac{s_j(p_i + \delta p_i) - s_j(p_i)}{\delta p_i \dfrac{\partial act}{\partial p_i}} \tag{C.39}$$

Deux étapes sont à observer dans la convergence lorsque la perturbation tend vers zéro. Tout d'abord, ce rapport tend vers 1 quand la perturbation diminue, puis les erreurs d'arrondi prennent l'ascendant et la comparaison n'est plus valide.

Le tableau C.2 présente un exemple de test du gradient. Il s'agit de vérifier la validité de la dérivé $\frac{\partial act(x,y,z)}{\partial Q_1}$ au point spatial $(5000, 2300, 0)$.

perturbation relative	rapport
10E-1	0.994094073
10E-2	0.999361612
10E-3	0.999935578
10E-4	0.999993552
10E-5	0.999999355
10E-6	0.999999937
10E-7	0.999999997
10E-8	1.00000022
10E-9	1.00000306
10E-10	1.00010413

TAB. C.2 – Test du gradient

Choix des paramètres de modélisation inverse

Le tableau C.3 donne les valeurs utilisées dans le processus de modélisation inverse. Il contient :

- les incertitudes sur les paramètres qui ont permis de choisir une valeur de point de départ (appelée « first guess »pour l'algorithme BFGS. Nous avons choisi un cas défavorable en perturbant la valeur réelle des paramètres de 3 fois la valeur de l'incertitude,
- l'intervalle admissible, délimité par une borne inf et une borne sup dans lequel le minimiseur recherche la solution.

paramètre	valeur réelle	incertitude	first guess	borne inf	borne sup
u	2	0.03	2.2	0.01	12
α	0.05	0.03	0.055	-0.05	0.15
Q_1	1.1e6	0.03	1e6	1e5	1e7
z_1	80	0.01	81	120	40
v_d	5e-3	0.2	6e-3	5e-1	5e-5
z_{ref}	1	0.02	1.1	0.1	10
Q_2	2e3	0.1	1.8e3	20	2e5
z_2	40	0.05	42	60	20
d	100	0.01	110	50	150

TAB. C.3 – Paramètres de modélisation inverse

Préconditionnement

La qualité des résultats de la modélisation inverse est directement reliée au conditionnement du Hessien : les erreurs se propageront de manière préférentielle en effet le long

des directions associées aux plus petites valeurs propres du Hessien.

De plus, la convergence de l'algorithme itératif de minimisation est également reliée au conditionnement du Hessien.

De manière usuelle, on cherche donc à *préconditionner* le problème afin d'avoir un nombre de conditionnement aussi proche de 1 que possible.

Pour la minimisation de la fonction coût $J(p)$ ceci revient à minimiser par rapport à une variable préconditionnée $\tilde{p} = K_p\, p$ où K_p est une matrice de préconditionnement (diagonale en pratique) telle que le conditionnement du problème de minimisation de $\tilde{J}(\tilde{p}) = J(p)$ soit proche de 1.

Un conditionnement souvent utilisé ([Bouttier et Courtier, 1999]) est de prendre $K_p = L^{-1}$ où la matrice L est donnée par la décomposition de Choleski du Hessien selon $Hess = L\, L^T$.

Pour la modélisation inverse du débit, le préconditionnement permet de faire converger l'algorithme de minimisation en une itération (contre 3 sinon).

A propos de la prise en compte d'une ébauche

On précise ici le rôle joué par la prise en compte ou non de l'ébauche.

Pour simplifier, on considère un modèle linéaire que l'on notera $act = f \times p$ avec f un paramètre fixé, p un paramètre scalaire à estimer (typiquement le débit). On ne prend en compte qu'une observation.

Notons que dans notre cas d'application, le problème pour les débits est effectivement sur-déterminé (plus d'observations que de paramètres à estimer).

La fonction coût est alors en conservant les notations utilisées :

$$J(p) = (p - p_b)^2/\sigma_b^2 + (fp - obs)^2/\sigma_o^2 \tag{C.40}$$

On note p_r la valeur réelle du paramètre recherché. On écrit alors :

$$p_b = p_r(1 + \epsilon_b)\ ,\quad obs = f\, p_r(1 + \epsilon_o)\ ,\quad p^* = p_r(1 + \epsilon_p) \tag{C.41}$$

avec ϵ_b, ϵ_o et ϵ_p les erreurs (relatives) respectivement d'ébauche, d'observation et d'estimation du paramètre. Les erreurs ϵ_b et ϵ_o sont des données (non connues bien sûr) du problème d'assimilation et on cherche à réduire l'erreur d'estimation ϵ_p.

On a aisément par définition de p^* :

$$p^* = (p_b/\sigma_o^2 + f\, obs/\sigma_o^2)/Hess\ ,\quad Hess = 1/\sigma_b^2 + f^2/\sigma_o^2 \tag{C.42}$$

d'où une erreur d'estimation :

$$\epsilon_p = (\epsilon_b/\sigma_b^2 + f^2\,\epsilon_o/\sigma_o^2)/Hess\ , \tag{C.43}$$

On a directement :

$$\frac{\epsilon_p - \epsilon_o}{\epsilon_p - \epsilon_b} = -(\frac{f\,\sigma_b}{\sigma_o})^2 \leq 0 \tag{C.44}$$

ce qui implique que $min(\epsilon_o, \epsilon_b) \leq \epsilon_p \leq max(\epsilon_o, \epsilon_b)$.

Si les erreurs d'ébauche et d'observation vont dans le « même sens »(par exemple positives), on obtient donc une erreur dégradée par rapport à l'erreur de l'information la plus fiable (mais c'est en quelque sorte un pis aller puisque l'on ne connaît pas en réalité les erreurs bien sûr).

Cette indication n'est que parcellaire puisqu'il faut plutôt raisonner en terme de *cycle statistique* (ie sur l'ensemble des valeurs que peuvent prendre les variables stochastiques ϵ_b et ϵ_o).

Un rapide calcul montre que :

$$< \epsilon_p^2 > (a) = \sigma_o^2 \frac{a^2 + f^4}{(a^2 + f^2)^2}\ ,\quad a = \frac{\sigma_o}{\sigma_b} \tag{C.45}$$

où l'on a moyenné statistiquement en faisant l'hypothèse que $\epsilon_o \sim N(0, \sigma_o)$ et $\epsilon_b \sim N(0, \sigma_b)$.

Si la qualité de l'ébauche est faible, ie $a \ll 1$, alors on a effectivement fortement dégradé la précision des résultats puisque $< \epsilon_p^2 > \sim \sigma_o^2$.

Enfin un dernier élément qui va dans le sens où l'ébauche peut être omise dans notre problème est la dépendance à f (qui joue ici le rôle de la sensibilité du modèle aux paramètres d'entrée).

Pour des petites valeurs de f (notre cas en pratique, voir les tests de sensibilité), on a directement :

$$\epsilon_p \simeq \epsilon_b \tag{C.46}$$

c'est à dire que les observations n'amènent plus d'information et que la solution optimale est systématiquement la valeur d'ébauche (dans ces cas là, le calcul de l'équation (C.44) n'est plus licite). L'assimilation de donnée ne sert alors plus à rien.

Tous ces éléments nous ont conduit à ne pas prendre en compte l'ébauche, qui n'est pas nécessaire, le problème n'étant pas sous-déterminé.

Analyse du second ordre

Les analyses du second ordre nécessitent de calculer des dérivées secondes de la fonction coût $J(p, p_c, p_{assim})$ où l'on rappelle que p correspond aux paramètres physiques recherchés, p_c les paramètres physiques connus, p_{assim} les paramètres d'assimilation de données.

En pratique, nous avons utilisé un adjoint du second-ordre ([Wang *et al.*, 1992]), qui revient à calculer le modèle linéaire tangent du modèle adjoint. En effet, l'analyse du second ordre nécessite des évaluations Hessien×vecteur (voir section 1) :

Avec les notations utilisées, on a pour tout vecteur h ([Ngodock, 1996]) :

$$Hess(J)h = h/\sigma_b^2 + (\frac{\partial f}{\partial p})^T \frac{\partial f}{\partial p} h/\sigma_o^2 + (\frac{\partial^2 f}{\partial p^2} h)^T (f(p) - obs)/\sigma_o^2 \tag{C.47}$$

Le deuxième terme est calculable à l'aide du modèle adjoint avec en entrée une sortie du linéaire tangent. Le troisième terme est obtenu à l'aide d'un nouveau code (on a construit le linéaire tangent du code adjoint).

Cette formule est utilisable pour l'estimation de $\frac{\partial^2 J}{\partial p^2}$ et est généralisable dans le même esprit aux autres calculs de second ordre.

Sensibilité aux observations

Avec les notations utilisées, on a pour un modèle linéaire :

$$p^* = Hess^{-1}(p_b/\sigma_b^2 + \sum_{i=1}^{i=n_{obs}} (f_{direct})_i \, obs_i/\sigma_i^2) \tag{C.48}$$

D'où :

$$\frac{\partial p^*}{\partial obs_i} = Hess^{-1}(f_{direct})_i/\sigma_i^2 \tag{C.49}$$

et la sensibilité dépend donc à la fois de l'erreur d'observation et de la propagation par l'inverse du Hessien de la sensibilité $(f_{direct})_i$ (dans le cas général, c'est l'adjoint qui apparaît ici).

Erreur de modélisation inverse pour un paramètre linéaire

On considère le modèle linéaire $y_i = a_i \, p + b_i$ donnant l'activité au point i en fonction du paramètre scalaire p. a_i et b_i sont des constantes scalaires indépendantes de p.

La modélisation inverse de p consiste à minimiser la fonction coût :

$$J(p) = \frac{1}{2} \sum_{i=1}^{n_{obs}} (\frac{y_i - obs_i}{\sigma_o})^2 \tag{C.50}$$

à partir d'observations $obs_i = (a_i\, p_r + b_i)(1 + \sigma_o\, \epsilon_i)$ où $\epsilon_i \sim N(0,1)$ et p_r correspond à la réalité. On fait l'hypothèse supplémentaire que les observations sont non biaisées.

La valeur $p^\star$, solution du problème de modélisation inverse, réalise $\nabla J(p^\star) = 0$ et est donnée par : q

$$p^\star = p_r + \frac{\sum_{i=1}^{n_{obs}} a_i\,(a_i\, p_r + b_i)\,\epsilon_i}{\sum_{i=1}^{n_{obs}} a_i^2}\,\sigma_o \tag{C.51}$$

$$= p_r + Hess^{-1}\frac{\sum_{i=1}^{n_{obs}} a_i\,(a_i\, p_r + b_i)\,\epsilon_i}{\sigma_o} \tag{C.52}$$

où le Hessien est donné par :

$$Hess = \frac{1}{\sigma_o^2}\sum_{i=1}^{n_{obs}} a_i^2 \tag{C.53}$$

On remarque que si a_i est constant (ce qui correspond à une sensibilité uniforme sur le domaine considéré), on retrouve exactement $p^\star = p_r$, puisque les observations sont non biaisées.

L'erreur relative de modélisation inverse s'écrit dans le cas général :

$$\epsilon^\star = \frac{p^\star - p_r}{p_r} = \frac{\sum_{i=1}^{n_{obs}} a_i^2\,\epsilon_i}{\sum_{i=1}^{n_{obs}} a_i^2}\,\sigma_o + \frac{\sum_{i=1}^{n_{obs}} a_i\, b_i\,\epsilon_i}{\sum_{i=1}^{n_{obs}} a_i^2}\frac{\sigma_o}{p_r} \tag{C.54}$$

Elle évolue linéairement avec l'erreur d'observation σ_o. La première contribution de la somme donnant l'erreur relative est majorée par σ_o alors que la deuxième peut-être grande si b_i est prépondérant devant $a_i\, p_r$. Ce cas correspond à une sensibilité relative du modèle faible :

$$\frac{p_r}{y_{max}}\frac{\partial y_i}{\partial p} \simeq \frac{a_i\, p_r}{b_{max}} \ll 1 \tag{C.55}$$

On rappelle qu'en pratique il faut rajouter une erreur numérique à l'erreur relative de modélisation inverse en raison de l'inversion du Hessien lorsqu'il est petit (cas scalaire) ou mal conditionné (cas général).

Nomenclature

Paramètres physiques

$act(x, y, z)$: activité au point (x, y, z) (unité : Bq/m^3).

Q_1 : débit direct (unité : Bq/s).

z_1 resp. z_2) : hauteur de rejet du débit 1 (unité : m), resp. du débit 2.

Q_2 : débit diffus.

d : distance du rejet diffus en amont du rejet ponctuel.

u : composante horizontal du vent (unité : m/s).

α : angle du vent.

σ_y (resp. σ_z) : écart type de Doury dans la direction y (resp. z), unité : .

v_d : vitesse de dépôt sec (unité : m/s).

z_{ref} : hauteur de référence du dépôt sec.

Λ : taux de lessivage par la pluie (unité : s^{-1}).

t_{pluie} : durée de l'épisode de pluie (unité : s).

Paramètres de l'assimilation de donnée

J : fonction coût.

n_{obs} : nombre d'observation.

obs_i : i-ème observation.

p : paramètres physiques recherchés.

p_b : valeur de l'ébauche, p^* : valeur optimale, p_r : valeur réelle.

p_c : paramètres physiques connus.

σ_b : variance de l'erreur d'ébauche.

σ_i : variance de l'erreur d'observation pour l'observation i. (σ_o lorsqu'elles sont supposées constantes.

$Hess(J)$: Hessien de la fonction coût.

p_{assim} : ensemble des paramètres de l'assimilation $(p_b, \sigma_b, (\sigma_i)_i)$.

Bibliographie

[Audiffren *et al.*, 1998] Audiffren, N., Renard, M., Buisson, E. et Chaumerliac, N. (1998). Deviations from the henry's law equilibrium during cloud events : a numerical approach of the mass transfer between phases and its specific numerical effects. *Atmospheric Research*, 49(139-161).

[Austin, 1992] Austin, J. (1992). Towards the four dimensional assimilation of stratospheric chemical constituents. *J. Geophys. Res.*, 97(D2):2569–2588.

[Baker, 2000] Baker, N. (2000). *Observation adjoint sensitivity and the adaptive observation-targeting problem.* Thèse de doctorat, Naval Postgraduate School, Monterey.

[Baker et Daley, 2000] Baker, N. et Daley, R. (2000). Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Quart. J. Roy. Meteor. Soc.*, 126:1431–1554.

[Barkmeijer *et al.*, 1998] Barkmeijer, J., Van Gijzen, M. et Bouttier, F. (1998). Singular vectors and estimates of the analysis-error covariance metric. *Q.J.R.M. Meteorol.Soc.*, 124:1695–1713.

[Barrett et Berge, 1996] Barrett, K. et Berge, E. (1996). Transboundary air pollution in europe. msc-w status report 1/96. *EMEP Status Report.*

[Bennett, 1992] Bennett, A. (1992). *Inverse Methods in Physical Oceanography.* Cambridge University Press.

[Bensoussan, 1972] Bensoussan, A. (1972). Optimization of sensor's location in a distributed filtering problem. *In Stability of stochastic dynamical systems*, volume 294 de *Lecture notes in mathematics Springer*, pages 63–84.

[Bergamashi *et al.*, 2000] Bergamashi, P., Hein, R., Heinmann, M. et Crutzen, P. (2000). Inverse modeling of the global co cycle. 1. inversion of co mixing ratio. *J. Geophys. Res.*, 105(D2):1909–1927.

[Bergot, 2001] Bergot, T. (2001). Influence of the assimilation scheme on the efficiency of adaptative observations. *Quart. J. Roy. Meteor. Soc.*, 127:635–660.

[Bergot *et al.*, 1999] Bergot, T., Hello, G., Joly, A. et Malardel, S. (1999). Adaptive observations : a feasibility study. *Mon. Wea. Rev.*, 127(743-765).

[Berliner *et al.*, 1999] Berliner, L., Lu, Z. et Snyder, C. (1999). Statistical design for adaptive weather observations. *J. Atmos. Sci.*, 56:2536–2552.

[Bischof *et al.*, 1992] Bischof, C., Carle, A., Corliss, G., Griewank, A. et Hovland, P. (1992). ADIFOR generating derivative codes from FORTRAN programs. Rapport technique, Mathematics and Computer Science Division, Argonne National Laboratory.

[Bishop, 2001] Bishop, C. (2001). Estimation theory and adaptive observing techniques. *Quart. J. Roy. Meteor. Soc.*

[Bishop *et al.*, 2001] Bishop, C., Etherton, B. et Majumdar, S. (2001). Adaptive sampling with the ensemble transform Kalman filter. part 1 : Theoretical aspects. *Mon. Wea. Rev.*, 129:420–436.

[Bishop et Toth, 1999] Bishop, C. et Toth, Z. (1999). Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, 56:1748–1765.

[Blom et Roemer, 1997] Blom, J. et Roemer, M. (1997). Description of the 3d lotos model. part i : Dynamics. Rapport technique, CWI, Amsterdam.

[Blond, 2002] Blond, N. (2002). *Assimilation de données photochimiques et prévisions de la pollution troposphérique.* Thèse de doctorat, Ecole Polytechnique.

[Botchev, 2003] Botchev, M. (2003). Improving approximate matrix factorizations for implicit time integration in air pollution modelling. *J. Comp. Appl. Math*, 157:309–327.

[Bousquet *et al.*, 1999] Bousquet, P., Ciais, P., Peylin, P., Ramonet, M. et Monfray, P. (1999). Inverse modeling of annual atmospheric CO2 sources and sinks 1. method and control inversion. *J. of Geophys. Res.*, 104(D21):26,161–26,178.

[Boutahar, 2004] Boutahar, J. (2004). *Méthodes de réduction et de propagation d'incertitudes : Application à un modèle de Chimie-Transport pour la modélisation et la simulation des impacts.* Thèse de doctorat, Ecole Nationale des Ponts et Chaussées.

[Boutahar *et al.*, 2004] Boutahar, J., Lacour, S., Mallet, V., Quélo, D., Roustan, Y. et Sportisse, B. (2004). Development and validation of a fully modular platform for numerical modelling of air pollution : Polair3D. *Int. J. Env. and Pollution*, 22(1-2).

[Boutahar *et al.*, 2001] Boutahar, J., Sportisse, B., Musson-Genon, L., Quelo, D. et Plion, P. (2001). Description of a 3d chemistry-transport-model. Rapport technique, ENPC.

[Bouttier et Courtier, 1999] Bouttier, F. et Courtier, P. (1999). Data assimilation concepts and methods. Training course notes of ECMWF.

[Buizza et Montani, 1999] Buizza, R. et Montani, A. (1999). Targeting observations using singular vectors. *J. Atmos. Sci.*, 56(17):2965–2985.

[Buizza et Palmer, 1995] Buizza, R. et Palmer, T. (1995). The singular vector structure of the atmospheric global circulation. *Journal of the Atmospheric Sciences*, 52(9):1434–1456.

[Byun et Ching, 1999] Byun, D. et Ching, J. (1999). Science algorithms of the EPA models-3 Community Multiscale Air Quality (CMAQ) modeling system. Rapport technique, U.S. Environmental Protection Agency.

[Charpentier, 2000] Charpentier, I. (2000). The mesodif package for gradient computations with the atmospheric model meso nh. *Env. Mod. Soft.*

[Chemith, 2001] Chemith, C. (2001). Inversion d'un modèle opérationnel de dispersion atmosphérique. Rapport technique, ECL-IPSN. Rapport de stage.

[Cohn, 1997] Cohn, S. (1997). An introduction to estimation theory. *J. Meteor. Soc. of Japan*, 75:257–288.

[Crabol et al, 1999] Crabol, B. et al (1999). Modèle franco-allemand pour le calcul de la dispersion atmosphérique en cas d'accident nucléaire. Rapport technique DPEA/SECRI/121, IPSN.

[Dabberdt et Schlatter, 1996] Dabberdt, W. et Schlatter, T. (1996). Us weather research program : second prospectus development team. *Bull. Amer. Meteor. Soc.*, 77(2):305–323.

[Daescu et Carmichael, 2002] Daescu, D. et Carmichael, G. (2002). An adjoint sensitivity method for the adaptive location of the observations in air quality modeling. *J. Atmos. Sci.*

[Daley, 1991] Daley, R. (1991). *Atmospheric data analysis*. Cambridge University Press.

[Debry, 2004] Debry, E. (2004). *Numerical simulation of an atmospheric aerosol distribution*. Thèse de doctorat, ENPC, CEREA. In French.

[Djouad et Sportisse, 2002] Djouad, R. et Sportisse, B. (2002). Partitioning techniques for reduction in chemical kinetics. APLA : an Automatic Partitioning and Lumping Algorithm. *Appl. Numer. Math.*, 43(4):383 :398.

[Djouad *et al.*, 2002a] Djouad, R., Sportisse, B. et Audiffren, N. (2002a). Numerical simulation of aqueous-phase atmospheric models : use of a non-autonomous rosenbrock method. *Atmos. Env.*, 36:873–879.

[Djouad *et al.*, 2002b] Djouad, R., Sportisse, B., Audiffren, N. et Charpentier, I. (2002b). Modelling aqueous phase chemistry : numerical integration and sensitivity analysis. *In* Sportisse, B., éditeur : *APMS 2001*, Geosciences, pages 444–457. Springer.

[Durbiano, 2001] Durbiano, S. (2001). *Vecteurs caractéristiques de modèles océaniques pour la réduction d'ordre en assimilation de données*. Thèse de doctorat, UJF – Grenoble.

[Ebel *et al.*, 1989] Ebel, A., Geiss, H., Hass, H., Jakobs, H., Laube, M. et Memmesheimer, M. (1989). Transport and deposition of radionuclides in europe after the chernobyl accident studied with the european acid deposition model (eurad). *Österreichische Beiträge zur Meteorologie und Geophysik*, 1:119–128.

[Elbern et Schmidt, 1999] Elbern, H. et Schmidt (1999). A four dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling. *J. Geophys. Res.*, 104:18 583–18598.

[Elbern et Schmidt, 2001] Elbern, H. et Schmidt, H. (2001). Ozone episode analysis by four dimensional variational chemistry data assimilation. *J. Geophys. Res.*, 106:3569–3590.

[Elbern *et al.*, 1997] Elbern, H., Schmidt, H. et Ebel, A. (1997). Variational data assimilation for tropospheric chemistry modeling. *J. Geophys. Res.*, 102:15967 :15985.

[Elbern *et al.*, 2000] Elbern, H., Schmidt, H., Talagrand, O. et Ebel, E. (2000). 4D variational data assimilation with an adjoint air quality model for emission analysis. *Env. Mod. Soft.*, 15:539–548.

[Epstein, 1969] Epstein, E. (1969). Stochastic dynamic prediction. *Tellus*, 21:739–759.

[Eskes et Jeuken, 1998] Eskes, H. et Jeuken, A. (1998). Assimilation of tovs and gome total ozone with the tm3 model. *In Chemical Data Assimilation, SODA workshop*, page 77 :81. KNMI.

[Etherton, 2001] Etherton, B. (2001). *The use of ensemble based error statistics for data assimilation and targeted observations.* Thèse de doctorat, Penn. State Un.

[Fadale *et al.*, 1995] Fadale, T., Nenarokoumov, A. et Emery, A. (1995). Two approaches to optimal sensor locations. *Journal of Heat Transfer*, 117:373–379.

[Faure et Papegay, 1997] Faure, C. et Papegay, Y. (1997). Odyssée, version 1.6 : the language reference manual. Rapport technique RT-211, INRIA.

[Fedorov et Hackl, 1997] Fedorov, V. et Hackl, P. (1997). *Model-oriented design of experiments.* Numéro 125 de Lecture Notes in Statistics. Springer Verlag.

[Fiore *et al.*, 2003] Fiore, A., Jacob, D., Mathur, R. et Martin, R. (2003). Application of empirical orthogonal functions to evaluate ozone simulations with regional and global models. *J. of Geophys. Res.*, 108(D14).

[Fisher et Lary, 1995] Fisher, M. et Lary, D. (1995). Lagrangian four-dimensional variational data assimilation of chemical species. *Quart. J. Roy. Meteor. Soc.*, 121:1681–1704.

[Friedrich et Reis, 2004] Friedrich, R. et Reis, S., éditeurs (2004). *Emissions of Air Pollutants - Measurements, Calculations and Uncertainties.* Springer.

[Gelaro *et al.*, 1998] Gelaro, R., Buizza, B., Palmer, T. et Klinker, E. (1998). Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. *J. Atmos. Sci.*, 55:1012–1037.

[Gery *et al.*, 1989] Gery, M., Whitten, G., Killus, J. et Dodge, M. (1989). A photochemical kinetics mechanism for urban and regional scale computer modeling. *J. Geophys. Res.*, 94(D10):12925–12956.

[Giering et Kaminski, 1998] Giering, R. et Kaminski, T. (1998). Recipes for adjoint code construction. *ACM Trans. Math. Software*, 24(4):437–474.

[Golub et Van Loan, 1986] Golub, G. et Van Loan, C. (1986). *Matrix computations*. North Oxford academic.

[Gregoire et Chaumerliac, 1994] Gregoire, P. et Chaumerliac, N. (1994). Impact of cloud dynamics on tropospheric chemistry : advances in modeling the interactions between microphysical and chemical processes. *J.Atmos.Chem.*, 18:247–266.

[Griewank et Walther, 1999] Griewank, A. et Walther, A. (1999). Revolve : An implementation of checkpoint for the reverse or adjoint mode of computational differentiation. *ACM Trans. Math. Software*.

[H. Byrd et Zhu, 1995] H. Byrd, P. Lu, J. N. et Zhu, C. (1995). limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comp.*, 16(5):1190–1208.

[Hamill, 2002] Hamill, T. (2002). Ensemble-based data assimilation : a review. University of Colorado and NOAA-CIRES Climate Diagnostics Center Boulder.

[Hanna *et al.*, 2001] Hanna, S. R., Lu, Z., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M. et Hansen, D. A. (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.

[Hansen, 1998] Hansen, J. (1998). *Adaptive observations in spatially extended nonlinear dynamical systems*. Thèse de doctorat, Oxford.

[Hansen et Smith, 2002] Hansen, J. et Smith, L. (2002). The role of operational constraints in selecting supplementary observations. *Submitted.*

[Hauglustaine *et al.*, 1998] Hauglustaine, D., Brasseur, G., Walters, S., Rasch, P., Müller, J.-F., Emmons, L. et Carroll, M. (1998). Mozart, a global chemical transport model for ozone and related chemical tracers, 2, model results and evaluation. *J. of Geophys. Res.*, 103:28,291–28,335.

[He et al, 2000] He, S. et al (2000). Application of ADIFOR for air pollution model sensitivity studies. *Environmental Modeling and Software*, (15):475–489.

[Heineken *et al.*, 1967] Heineken, F., Tsuchyia et Aris, R. (1967). On the accuracy of determining rate constants in enzymatic reactions. *Mathematical biosciences*, 1:115–141.

[Heskes *et al.*, 1999] Heskes, H., Piters, A., Levelt, P., Allaart, M. et Kelder, H. (1999). Variational assimilation of gome total-column ozone satellite data in a 2d lat-lon tracer transport model. *J. Atmos. Sci.*, 56:3560 :3572.

[Hoelzemann *et al.*, 2004] Hoelzemann, J., Brasseur, G., Granier, C., Schultz, M. et Simon, M. (2004). The global wildland fire emission model GWEM : evaluating the use of global area burnt satellite data. *J. of Geophys. Res.*, 109(D14).

[Hoelzemann *et al.*, 2001] Hoelzemann, J., Elbern, H. et Ebel, A. (2001). Psas and 4d-var assimilation for chemical state analysis by urban and rural observation sites. *Physics and Chemistry of the Earth*, 26/10.

[Honoré, 2000] Honoré, C. (2000). *La Photochimie de l'Ozone à l'Echelle Urbaine, un Système Dynamique Non-Linéaire.* Thèse de doctorat, Université Paris 6.

[Hourdin et Issartel, 2000] Hourdin, F. et Issartel, J.-P. (2000). Sub-surface nuclear tests monitoring through the ctbt 133xe network. *Geophys. Res. Lett.*, 27:2245–2248.

[Houtemaker *et al.*, 1996] Houtemaker, P., Lefaivre, L., Derome, J., Ritchie, H. et Mitchell, H. (1996). A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, 124:1225–1241.

[Ide *et al.*, 1997] Ide, K., Courtier, P., Ghil, M. et Lorenc, A. (1997). Unified notation for data assimilation : operational, sequential and variational. *J. Meteor. Soc. of Japan*, 75:181–189.

[Issartel et Baverel, 2002] Issartel, J. et Baverel, J. (2002). Adjoint backtracking for the verification of the comprehensive test ban treaty. *Atmos. Chem. Phys. Discuss.*, 2:2133 :2150.

[Jazwinski, 1970] Jazwinski, A. (1970). *Stochastic processes and filtering theory.* Academic Press.

[Jeuken *et al.*, 1999] Jeuken, A., Eskes, H., Van Velthoven, P., Kelder, H. et Holm, E. (1999). Assimilation of total ozone satellite measurements in a three-dimensional tracer transport model. *J. Geophys. Res.*, 104:5551 :5563.

[Joly et al, 1997] Joly, A. et al (1997). The fronts and atlantic storm-track experiment (fastex) : Scientific objectives and experimental design. *Bull. Amer. Meteor. Soc.*, 78: 1917–1940.

[Kalnay, 2003] Kalnay, E. (2003). *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press.

[Kalnay et Pu, 1996] Kalnay, E. et Pu, Z. (1996). Using an adjoint model and/or a quasi-inverse linear model to target weather obervations. *In 7th conf of Mesoscale Processes*, pages 25–26. AMS.

[Kaminski, 1998] Kaminski, T. (1998). *On the benefit of the adjoint technique for inversion of the atmospheric transport employing carbon dioxide as an example of a passive tracer.* Thèse de doctorat, Un. Hamburg.

[Kern, 2003] Kern, M. (2003). *Problèmes inverses.* cours de l'Ecole Supérieure d'Ingénierie Léonard de Vinci.

[Khattatov et al, 1999] Khattatov, B. et al (1999). Assimilation of photochemically active species and a case analysis of uars data. *J. Geophys. Res.*, 104(18715-18737).

[Kopell, 1985] Kopell, N. (1985). Invariant manifolds and the initialization problem for some atmospheric equations. *Physica D*, 14:203–215.

[Langland et al, 1999] Langland, R. et al (1999). The north pacific experiment 98 (norpex 98) : targeted observations for improved north american weather forecasts. *Bull. Amer. Meteor. Soc.*, 80:1363–1384.

[Langland et Rohaly, 1996] Langland, R. et Rohaly, G. (1996). Adjoint-based targeting of observations for fastex cyclones. *In 7th conf of Mesoscale Processes*, pages 369–372. AMS.

[Le Dimet *et al.*, 2002] Le Dimet, F.-X., Navon, I. et Daescu, D. (2002). Second-order information in data assimilation. *Month. Weat. Rev*, 130:629–648.

[Le Dimet et Talagrand, 1986] Le Dimet, F.-X. et Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. *Tellus*, 38A:97–110.

[Lebelt *et al.*, 1996] Lebelt, P., Allaart, M. et Kelder, H. (1996). On the assimilation of total ozone data. 14:1111 :1118.

[Lelieveld et Crutzen, 1991] Lelieveld, J. et Crutzen, P. (1991). The role of clouds in tropospheric chemistry. *Journal of Atmospheric Chemistry*, 12:229–267.

[Levelt *et al.*, 1998] Levelt, P., Khattatov, B., Gille, J., Brasseur, G., Tie, X. et Waters, J. (1998). Assimilation of mls ozone measurements in the global 3d ctm rose. *Geophys. Res. Lett.*, 25(24):4493 :4496.

[Li et Navon, 2001] Li, Z. et Navon, I. (2001). Optimality of 4d-var and its relationship with the kalman filter and kalman smoother. *Q. J. R. Meteorol. Soc.*, 127(572):661–684.

[Lin *et al.*, 2002] Lin, C.-L., Chai, T. et Sun, J. (2002). On the smoothness constraints for four-dimensional data assimilation. *J. Comp. Phys.*, 181(2):430–453.

[Lorenz et Emanuel, 1998] Lorenz, E. et Emanuel, K. (1998). Optimal sites for supplementary weather observations : simulation with a small model. *J.Atm.Sc.*, 55:399–414.

[Louis *et al.*, 1982] Louis, J., Tiedke, M. et Geleyn, J. (1982). A short history of the operational PBL parameterization at ECMWF. workshop on planetary boundary layer parameterization at ECMWF. *In Workshop on planetary boundary layer parameterization*, pages 59–80, ECMWF.

[Majumdar *et al.*, 2001] Majumdar, S., Bishop, C., Szunyogh, I. et Toth, Z. (2001). Can an etkf predict the reduction in forecast error variance produced by targeted observations ? *Quart. J. Roy. Meteor. Soc.*

[Mallet, 2004] Mallet, V. (2004). Atmodata library : data processing and parameterizations in atmospheric sciences. Rapport technique 29, CEREA.

[Mallet, 2005] Mallet, V. (2005). *Prévision numérique de la qualité de l'air. Prévisions d'ensemble.* Thèse de doctorat, ENPC.

[Mallet et Sportisse, 2004] Mallet, V. et Sportisse, B. (2004). 3-d chemistry-transport model Polair numerical issues, validation and automatic-differentiation strategy. *Atmos. Chem. Phys. Discuss.*, 4:1371–1392.

[Massart, 2003] Massart, S. (2003). *Méthodologies De l'assimilation de données. Application à la chimie atmosphérique et à la mécanique des structures.* Thèse de doctorat, INPT-CERFACS.

[Mendoza-Dominguez et Russel, 2001] Mendoza-Dominguez, A. et Russel, A. (2001). Estimation of emission adjustments from the application of four-dimensional data assimilation to photochemical air quality modeling. *Atmos. Env.*, 35:2879–2894.

[Menut, 2003] Menut, L. (2003). Adjoint modelling for atmospheric pollution processes sensitivity at regional scale during the ESQUIF IOP2. *J. of Geophys. Res.*, 108(D17).

[Menut *et al.*, 2000] Menut, L., Vautard, R., Flamant, C., Abonnel, C., Beekmann, M., Chazette, P., Flamant, P. H., Gombert, D., Guedalia, D., Kley, D., Lefebvre, M., Lossec, B., Martin, D., Megie, G., Sicard, M., Perros, P. et Toupance, G. (2000). Atmospheric pollution over the paris area : The esquif project. *Annales Geophysicae*, 18:1467–1481.

[Middleton *et al.*, 1990] Middleton, P., Stockwell, W. R. et Carter, W. P. L. (1990). Aggregation and analysis of VOC emissions for regional modelling. *Atmos. Env.*, 24: 1107–1133.

[Molteni *et al.*, 1996] Molteni, F., Buizza, R., Palmer, T. et Petroliagis, T. (1996). The ecmwf ensemble prediction system : methodology and validation. *Quart. J. Roy. Meteor. Soc.*, 122:73–120.

[Monfort, 1999] Monfort, M. (1999). Présentation du modèle sirocco. Rapport technique DPEA/SECRI/121, IPSN.

[Morss, 1998] Morss, R. (1998). *Adaptive observations : idealized sampling strategies for improving numerical weather prediction.* Thèse de doctorat, MIT.

[Ngodock, 1996] Ngodock, H. E. (1996). *Assimilation de données et Analyse de sensibilité : Une application à la circulation océanique.* Thèse de doctorat, Un. Joseph Fourier, Grenoble I.

[Ntziachristos *et al.*, 2002] Ntziachristos, L., Tourlou, P. M., Samaras, Z., Geivanidis, S. et Andrias, A. (2002). National and central estimates for air emissions from road transport. Rapport technique, EEA.

[Omatu et Seinfeld, 1983] Omatu, S. et Seinfeld, J. (1983). Optimization of sensor an actuator locations in a distributed parameter system. *J. of the Franklin Institute*, 315:407–421.

[Omatu et Seinfeld, 1989] Omatu, S. et Seinfeld, J. (1989). *Distributed parameter systems. Theory and applications.* Oxford Mathematical monographs.

[Palmer *et al.*, 1998] Palmer, T., Gelaro, R., Barkmeijer, J. et Buizza, R. (1998). Singular vector, metrics and adaptive observations. *J. Atmos. Sci.*, 55(633-653).

[Pham, 1996] Pham, D. (1996). A singular evolutive interpolated kalman filter for data assimilation in oceanography. Rapport technique RT 163, LMC/IMAG.

[Pham *et al.*, 1998] Pham, D., Verron, J. et Roubaud, M. C. (1998). A singular evolutive extended kalman filter for data assimilation in oceanography. *J. Marine Systems*, 16:323–340.

[Planet et al, 1984] Planet, W. et al (1984). Determination of total ozone amount from tiros radiance measurements. *J. Applied Meteor.*, 23(208).

[Pu et Kalnay, 1999] Pu, Z. et Kalnay, E. (1999). Targeting observations with the quasi-inverse linear and adjoint ncep global models : performance during fastex. *Quart. J. Roy. Meteor. Soc.*, 125:3329–3337.

[Pu *et al.*, 1997] Pu, Z., Kalnay, E., Sela, J. et Szunyogh, I. (1997). Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method. *Mon. Wea. Rev.*, 125:2479–2503.

[Quélo *et al.*, 2004] Quélo, D., Lagache, R. et Sportisse, B. (2004). Etude de l'impact qualité de l'air des scénarios PDUs sur Lille à l'aide du modèle de chimie-transport POLAIR3D. Rapport technique 28, CEREA/ENPC.

[Quélo et Sportisse, 2004] Quélo, D. et Sportisse, B. (2004). Assimilation de données – 2ème partie : Implémentation dans polair3d. Rapport technique 25, CEREA/ENPC.

[Quélo *et al.*, 2002] Quélo, D., Sportisse, B., Berroir, J. et Charpentier, I. (2002). Some remarks concerning inverse modeling and data assimilation for slow-fast atmospheric chemical kinetics. *In* Sportisse, B., éditeur : *APMS 2001*, Geosciences, pages 499–513. Springer.

[Quélo *et al.*, 2005] Quélo, D., Sportisse, B. et Isnard, O. (2005). Data assimilation for short range atmospheric dispersion of radionuclides : a case study of second-order sensitivity. *J. of Env. Radioact.*, 84:393–408.

[Quélo et Sportisse, 2002] Quélo, D. et Sportisse, B. (2002). Assimilation de données pour la qualification d'un rejet radioactif. Rapport technique Cereve 2002-7, ENPC. Rapport de contrat IRSN.

[Quélo et Sportisse, 2004] Quélo, D. et Sportisse, B. (2004). The influence of timescales for data assimilation of chemistry. *Quart. J. Roy. Meteor. Soc.* En cours de soumission.

[Rabier *et al.*, 1996] Rabier, F., Klinker, E., Courtier, P. et Hollingsworth, A. (1996). Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, 122:121–150.

[Reynolds *et al.*, 2000] Reynolds, C., Gelaro, R. et Palmer, T. (2000). Examination of targeting methods in a simplified seting. *Tellus*, 52A:391–411.

[Riishojgaard, 1996] Riishojgaard, L. (1996). On four-dimensional variational assimilation of ozone data in weather prediction models. *Quart. J. Roy. Meteor. Soc.*, 122:1545–1571.

[Risken, 1984] Risken, H. (1984). *The Fokker-Planck Equation.* Springer.

[Roustan, 2005] Roustan, Y. (2005). Modélisation du mercure à l'échelle du continent européen. I. paramétrisations. Rapport technique 43, CEREA.

[Roustan et Bocquet, 2006] Roustan, Y. et Bocquet, M. (2006). Sensitivity analysis for mercury over Europe. *J. of Geophys. Res.*, 111(D14304).

[Sartelet, 2004] Sartelet, K. (2004). Modal aerosol model MAM – technical description. Rapport technique 22, CEREA.

[Sartelet *et al.*, 2002] Sartelet, K. N., Boutahar, J., Quélo, D., Coll, I., Plion, P. et Sportisse, B. (2002). Development and validation of a 3d chemistry-transport model, POLAIR3D, by comparison with data from ESQUIF campaign. *In Proceedings of the 6th Gloream workshop : Global and regional atmospheric modelling*, pages 140–146, Aveiro, Portugal.

[Schmidt et Martin, 2003] Schmidt, H. et Martin, D. (2003). Adjoint sensitivity of episodic ozone in the Paris area to emissions on the continental scale. *J. of Geophys. Res.*, 108.

[Schwartz, 1986] Schwartz, S. (1986). Mass-transport considerations pertinent to aqueous phase reactions of gases in liquid-water clouds. NATO ASI Series, Vol. G6, Springer Verlag. Chemistry of multiphase atmospheric systems.

[Segers, 2002] Segers, A. (2002). *Data assimilation in atmospheric chemistry models using Kalman filtering.* Thèse de doctorat, TU Delft.

[Sei et Symes, 1995] Sei, A. et Symes, W. (1995). A note on consistency and adjointness for numerical schemes. Rapport technique TR95-04, Department of Computational and Applied Mathematics, Rice University and Center for Research in Parallel Computation.

[Seigneur *et al.*, 1982] Seigneur, C., Stephanopoulos, G. et Carr Jr, R. (1982). Dynamic sensitivity analysis of chemical reaction systems. *Chem. Eng. Sci.*, 37(6):845–853.

[Seinfeld, 1985] Seinfeld, J. (1985). *Atmospheric Physics and Chemistry of Air Pollution.* Wiley.

[Simpson *et al.*, 1999] Simpson, D., Winiwarter, W., Börjesson, G., Cinderby, S., Ferreiro, A., Guenther, A., Hewitt, N. C., Janson, R., Khalil, M. A. K., Owen, S., Pierce, T. E., Puxbaum, H., Shearer, M., Skiba, U., Steinbrecher, R., Tarrasón, L. et Oquist, M. G. (1999). Inventorying emissions from nature in europe. *J. of Geophys. Res.*

[Sirkes et Tziperman, 1997] Sirkes, Z. et Tziperman, E. (1997). Finite difference of adjoint or adjoint of finite difference ? *Mon. Wea. Rev.*, 125:3373–3378.

[Spee, 1998] Spee, E. (1998). *Numerical methods in global transport-chemistry models*. Thèse de doctorat, Univ. Amsterdam.

[Sportisse, 1999] Sportisse, B. (1999). *Contribution à la modélisation des écoulements en milieu réactif : réduction des modèles de cinétique chimique et simulation de la pollution atmosphérique*. Thèse de doctorat, École Polytechnique.

[Sportisse, 2000] Sportisse, B. (2000). An analysis of operator splitting techniques in the stiff case. *J. Comp. Phys.*, 161:140–168.

[Sportisse *et al.*, 2002] Sportisse, B., Boutahar, J., Debry, E., Quélo, D. et Sartelet, K. (2002). Some tracks in air pollution modeling. *RACSAM (Journal of the Spanish Science Academy, Real Academia de Ciencias de España)*, 96(3):507–528.

[Sportisse et Djouad, 2000] Sportisse, B. et Djouad, R. (2000). Reduction of chemical kinetics in air pollution modelling. *J.Comp.Phys.*, 164:354–376.

[Sportisse et Djouad, 2002] Sportisse, B. et Djouad, R. (2002). *Some aspects of multi-timescales issues for the numerical modeling of atmospheric chemistry*, volume 130, chapitre 4, page 39 :61. IMA Volume in Mathematics and its applications. Springer New York.

[Sportisse et Quélo, 2002] Sportisse, B. et Quélo, D. (2002). Observational network design, adaptive observations and targeting strategies in atmospheric data assimilation : sketch of a methodological review. *In INRIA-CEA-EDF Summer School on Data Assimilation for Geophysical flows.*

[Sportisse et Quélo, 2003] Sportisse, B. et Quélo, D. (2003). Data assimilation and inverse modeling of atmospheric chemistry. *Proc. of Indian National Science Academy. Part A Physical Sciences*, 69.

[Sportisse et Quélo, 2004] Sportisse, B. et Quélo, D. (2004). Assimilation de données – 1ère partie : Eléments théoriques. Rapport technique 24, CEREA/ENPC.

[Stauffer et Bao, 1993] Stauffer, D. et Bao, J. (1993). Optimal determination of nudging coefficients using the adjoint equations. *Tellus*, 45A:358–369.

[Stephenson, 1997] Stephenson, D. (1997). Correlation of spatial climate/weather maps and the advantage of using the mahalanobis metric in predictions. *Tellus*, 49 A:513–527.

[Stockwell *et al.*, 1997a] Stockwell, W., Kirchner, F., Kuhn, M. et Seefeld, S. (1997a). A new mechanism for Regional Atmospheric Chemistry Modeling. *J. Geophys. Res.*, 95(D10):16343 :16367.

[Stockwell *et al.*, 1990] Stockwell, W., Middleton, P., Chang, J. et Tang, X. (1990). The second regional acid deposition model chemical mechanism for regional air quality modeling. *J. Geophys. Res.*, 95(D10):16343–16367.

[Stockwell *et al.*, 1997b] Stockwell, W. R., Kirchner, F., Kuhn, M. et Seefeld, S. (1997b). A new mechanism for regional atmospheric chemistry modeling. *J. of Geophys. Res.*, 102:25,847–25,879.

[Szunyogh et al, 2000] Szunyogh, I. et al (2000). The effect of targeted dropsonde observations during the 1999 winter storm reconnaissance program. *Mon. Wea. Rev.*, 128:3520–3537.

[Taghavi *et al.*, 2004] Taghavi, M., Musson-Genon, L. et Sportisse, B. (2004). Impact des rejets de la centrale thermique de martigues sur la qualité de l'air. modélisation avec polair3d. Rapport technique, ENPC-CEREA. Rapport pour la Mission Thermique EDF.

[Talagrand, 1997] Talagrand, O. (1997). Assimilation of observations, an introduction. *J. Meteor. Soc. Japan*, 75:191–209.

[Tarantola, 1987] Tarantola, A. (1987). *Inverse Problem Theory*. Elsevier Amsterdam.

[Thacker, 1989] Thacker, W. (1989). The role of the hessian matrix in fitting models to measurements. *J. Geophys. Res.*, 94(C5):6177 :6196.

[Tilmes et Zimmermann, 1998] Tilmes, S. et Zimmermann, J. (1998). Investigation on the spatial scales of the variability in measured near-ground ozone mixing ratios. *Geophys. Res. Lett.*, 25:3827–3830.

[Toth et Kalnay, 1997] Toth, Z. et Kalnay, E. (1997). Ensemble forecasting at nmc and the breeding method. *Mon. Wea. Rev.*, 125:3297–3319.

[Tulet, 2000] Tulet, P. (2000). *Modélisation physico-chimique de la pollution régionale : impacts des divers processus de transport des polluants gazeux en situations complexes et tests de validations*. Thèse de doctorat, Université Paul Sabatier (Toulouse III).

[van Loon *et al.*, 2000] van Loon, M., Builtjes, P. et Segers, A. (2000). Data assimilation of ozone in the atmospheric transport chemistry model LOTOS. *Env. Mod. and Soft.*, 15:603–609.

[Vautard *et al.*, 2001] Vautard, R., Beekmann, M., Roux, J. et Gombert, D. (2001). Validation of a deterministic forecasting system for the ozone concentrations over the Paris area. *Atmos. Env.*, 35:2449–2461.

[Verlaan, 1998] Verlaan, M. (1998). *Efficient Kalman filtering algorithms for hydrodynamic models*. Thèse de doctorat, TU Delft.

[Verwer *et al.*, 1998] Verwer, J., Hundsdorfer, W. et Blom, J. (1998). Numerical time integration for air pollution models. *In Proceedings of the Conference APMS'98*. ENPC-INRIA.

[Verwer *et al.*, 1999] Verwer, J., Spee, E., Blom, J. et Hundsdorfer, W. (1999). A second order Rosenbrock method applied to photochemical dispersion problem. *SIAM J. Sci. Comp.*, 20(4):1456–1480.

[Wang *et al.*, 2001] Wang, K., Lary, D., Shallcross, D., Hall, S. et Pyle, J. (2001). A review on the use of the adjoint method in four-dimensional atmospheric chemistry data assimilation. *Quart. J. Roy. Meteor. Soc.*, 127:2181–2294.

[Wang *et al.*, 1992] Wang, Z., Navon, I., Le Dimet, F. et Zou, X. (1992). The second order adjoint analysis : theory and applications. *Meteor.Atmos.Phys.*, (50):3 :20.

[Wesely, 1989] Wesely, M. L. (1989). Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmos. Env.*, 23:1293–1304.

[Zou *et al.*, 1992] Zou, X., Navon, I. et Le Dimet, F.-X. (1992). An optimal nudging data assimilation scheme using parameter estimation. *Q. J. R. Meteorol. Soc.*, 118:1163–1186.