



Assimilation de données photochimiques et prévision de la pollution troposphérique

Nadège Blond

► **To cite this version:**

Nadège Blond. Assimilation de données photochimiques et prévision de la pollution troposphérique. Physique [physics]. Ecole Polytechnique X, 2002. Français. <tel-00003337>

HAL Id: tel-00003337

<https://pastel.archives-ouvertes.fr/tel-00003337>

Submitted on 10 Sep 2003

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.

THÈSE DE DOCTORAT DE L'ÉCOLE POLYTECHNIQUE

présentée pour obtenir le grade de

Docteur de l'École Polytechnique

Spécialité : Physique et Chimie de l'atmosphère

par

Nadège Blond

*Assimilation de données photochimiques
et prévision de la pollution troposphérique*

Soutenue le 20 décembre 2002 devant le jury composé de :

Dr. Hendrick Elbern	<i>Rapporteur</i>
Pr. Daniel Guédalia	<i>Rapporteur</i>
Pr. Gérard Mégie	<i>Examineur</i>
Dr. Vincent-Henri Peuch	<i>Examineur</i>
Dr. Robert Vautard	<i>Examineur</i>
Dr. Hans Wackernagel	<i>Examineur</i>
Dr. Christian Elichegaray	<i>Invité</i>
Dr. Jean-Paul Boch	<i>Invité</i>

Thèse préparée au Laboratoire de Météorologie Dynamique,
Centre National de la Recherche Scientifique, École polytechnique,
financée par l'Agence de l'Environnement et De Maitrise de l'Énergie et TotalFinaElf.

Ces trois années de thèse se sont déroulées au Laboratoire de Météorologie Dynamique. Je remercie Claude Basdevant, directeur à l'époque, pour m'y avoir accueilli dans de très bonnes conditions. Je le remercie aussi pour sa présence auprès des doctorants et son écoute. Je remercie également la nouvelle direction qui, malgré la forte crise du logement, a accepté de me laisser *seule* dans un grand bureau pour finir ma thèse. Je remercie au passage Michel Desbois de m'avoir laissé son bureau.

Je tiens à remercier les membres du jury et ses invités. Je remercie notamment Hendrik Elbern, un des pionniers de l'assimilation de données photochimiques, et Daniel Guédalia, spécialiste confirmé en pollution atmosphérique pour avoir accepté d'être les rapporteurs de cette thèse. Je remercie Hans Wackernagel et Vincent-Henri Peuch pour avoir participé au jury et s'être intéressés à ce travail pendant ces trois années. Je remercie Gérard Mégie pour m'avoir fait l'honneur de présider ce jury. Je remercie l'ADEME et Total-Final-Elf pour m'avoir permis de réaliser ce travail en le finançant. Je voudrais spécialement remercier Christian Elichegaray et Nathalie Poisson du département de l'air de l'ADEME pour leur accueil et l'intérêt qu'ils ont porté à ce travail. Je remercie Christian Elichegaray et Jean-Paul Boch pour avoir accepté l'invitation à la table du jury.

Je remercie Robert Vautard, qui m'a fait partager son savoir et son enthousiasme pendant ces trois années. Je lui suis profondément reconnaissante de m'avoir accordée une grande liberté dans mon travail, tout en me guidant efficacement pendant ces trois années. Il m'a toujours offert de précieux conseils et restera pour moi un modèle d'efficacité.

Je remercie Laurent Menut et Hauke Schmidt qui m'ont accompagnée au cours de cette thèse. Je les remercie pour leur aide, leur disponibilité et leur soutiens. J'ai profondément apprécié les innombrables discussions que nous avons eues, qui m'ont permis d'avancer dans mes travaux bien sûr, mais aussi d'apprécier leur recul, leur rigueur et leur expertise dans le domaine de la pollution atmosphérique et de l'assimilation de données. Je tiens à remercier Lyliane Bel qui m'a fait découvrir le monde des statisticiens, et Olivier Talagrand qui m'a ouvert celui des *assimilateurs*.

Je remercie les associations pour la surveillance de la qualité de l'air, qui grâce à leur active collaboration, ont permis la mise en place d'un système de cartographie en temps quasi réel à l'Échelle Européenne et Régionale. Je les remercie d'avoir rapidement et gratuitement mis à disposition leurs données d'observations. Je remercie en particulier Mireille Lattuati pour avoir concrétisé une partie de ce travail à AIRPARIF, et Alexandre Armengaud (AIRMARAIX) pour son accueil et les informations qu'il m'a fournies.

Le soutien logistique des informaticiens du laboratoire a aussi permis à ce travail de s'accomplir. En particulier, je voudrais remercier Patrick Rabéranteau, Christophe Boitel et Jean-Louis Monge pour leur précieuse aide informatique et leur capacité à supporter mes colères quand Moebius, Gauss ou Wincenter me faisaient des caprices.

De même, j'ai beaucoup apprécié l'aide d'Éliane Rier, Martine Roux, France-Lise Robin et Stéphane Sportouch dans mes démarches administratives. Merci à Marie-Claire Lanceau de m'avoir souvent aidé dans mes recherches bibliographiques. Je remercie également le Bureau d'Étude pour leur aide diverse et variée. Je remercie toutes ces personnes pour leur permanente bonne humeur.

Un grand merci à tous les gens du LMD, et de l'IPSL : Jean Roux, Alma Hodzic, Albert Hertzog, Chiara Cagnazzo, François Vial, Olivier Chomette, Fadoua Eddounia, Michel Capederou, Robert Kandel, Michel Viollier, Isabelle Jobard, Guillem Candille, Herman Moldovan, Cécile Honoré, Fabio d'Andréa, Matthias Beekmann, Claude Derognat, Benjamin Sultan, Serge Janicot, André Szantai, Philippe Drobinsky, Alain Chédin, Federico Fierli ... J'ai beaucoup apprécié leur compagnie, et, de manière générale, l'ambiance chaleureuse et détendue régnant au sein du Laboratoire. Je remercie tout spécialement : Remy Roca, pour tous les bons conseils qu'il m'a donnés et tous ces agréables moments passés dans les bistros de Paris, Abdoulaye Deme, pour les discussions et les fous rires que nous avons eus.

Je voudrais également remercier les élèves étrangers de l'École Polytechnique, que j'ai encadré pendant ma thèse en tant que monitrice. Je garderai un très bon souvenir de ces premières années d'enseignement. Je remercie notamment Omid Amini, Sunyoto Sumantri, Sebastian Zepeda Garcia, Reynaldo Sandoval Lopez, Draltan Marin, Vo Ngoc Minh.

Je remercie enfin toutes les personnes qui m'ont fait oublier cette thèse quand il le fallait : Gilbert et Marie-Reine Blond (mes parents), Gaëlle Blond (ma soeur), Jérémie Léonard, Sébastien Chupin, Christophe Winisdörffer, Fabien Locher, Stéphanie Marco, Lydiane Crouzel, Pierre Philippe, Nadège et Benoit Courjal et bien d'autres déjà cités ou impardonnablement oubliés.

Table des matières

1	Introduction : étude et prévision de la pollution photochimique atmosphérique	13
1.1	Mise en place d'un réseau de surveillance	13
1.1.1	Prise de conscience des problèmes liés à la pollution atmosphérique	13
1.1.2	La loi sur l'air	14
1.1.3	Mesure stratégique	14
1.2	La modélisation numérique et la prévision de la pollution photochimique	15
1.2.1	Les modèles numériques de Chimie-Transport (CTM)	16
1.2.2	La prévision de la pollution	17
1.3	Pourquoi cette thèse? Les objectifs	18
1.4	Assimilation de données en météorologie et en océanologie	20
1.5	Les méthodes d'assimilation de données	23
1.5.1	Les méthodes séquentielles	24
1.5.2	Les méthodes variationnelles	27
1.6	Les méthodes géostatistiques	29
1.7	Application à la chimie atmosphérique	29
1.8	Choix d'une méthode d'assimilation de données	31
1.8.1	Coûts numériques des méthodes d'assimilation de données	31
1.8.2	Différences majeures entre les méthodes d'assimilation de données	32
1.8.3	Conclusions	32
1.9	Plan de la thèse	32
2	Krigeage Ordinaire, Interpolation Statistique, formulations mathématiques	35
2.1	Une illustration simple d'estimation statistique	35
2.2	Le Krigeage	38
2.2.1	Principe du Krigeage	38
2.2.2	Expression des poids et de l'erreur de Krigeage	38
2.3	Interpolation Statistique : utilisation d'un modèle de chimie-transport	40
2.3.1	Les erreurs de modélisation et d'observation	40
2.3.2	Principe de l'Interpolation Statistique	40
2.3.3	Interpolation Statistique, Krigeage des innovations	41
2.4	Nécessité d'une modélisation des biais et des covariances	42
2.4.1	Problème des biais	42
2.4.2	Les covariances d'erreur	42
2.5	Modélisation des covariances	43
2.5.1	Choix d'un ensemble statistique pour calculer les covariances	44

2.5.2	Moyennes temporelles : utilisation d'une climatologie	44
2.5.3	Moyennes spatiales	47
2.6	Quelques exemples simples d'Interpolations Statistiques	48
2.6.1	Cas d'une seule observation	48
2.6.2	Cas de deux observations	49
2.7	Extension à des champs de variances non définies : Krigeage Intrinsèque	50
2.8	Conclusions	53
3	Évaluation du modèle CHIMERE à l'aide des observations	55
3.1	Le modèle CHIMERE multi-échelles	55
3.1.1	Généralités	55
3.1.2	Description du modèle	56
3.2	Les observations	61
3.2.1	Problèmes de représentativité	61
3.2.2	Les observations de surface	62
3.2.3	Observations en altitude	66
3.3	Outils statistiques d'évaluation d'un modèle	68
3.4	Évaluation statistique de la qualité du modèle CHIMERE continental à l'aide des observations de surface	68
3.5	Évaluation statistique de la qualité de CHIMERE régional à l'aide des observations de surface	74
3.5.1	Cas de la région d'Île-de-France	74
3.5.2	Un terrain complexe : Bassins de Marseille et de Berre	82
3.6	Évaluation statistique de la qualité de CHIMERE à l'aide des données aéroportées de la campagne ESQUIF	86
3.7	Conclusions	90
4	Analyse des champs de concentrations d'ozone sur la région de l'Île-de-France	91
4.1	Mise en place des Interpolations Statistiques	91
4.1.1	Traitement des biais	92
4.1.2	Modélisation de la fonction de covariance	94
4.1.3	Analyses 2D ou 3D	98
4.2	Mise en place de méthodes de Krigeage	99
4.2.1	Krigeage sur les observations, méthode OBK	100
4.2.2	Krigeage sur les innovations, méthode INK	101
4.3	Validation et comparaison des méthodes d'analyse de données, représentation des champs d'ozone de surface	102
4.3.1	Méthode objective de validation : le <i>Leave-one-out</i>	102
4.3.2	Confrontation des méthodes ISI et ASI	103
4.3.3	Confrontation des méthodes ISI, ASI, OBK, et INK	105
4.4	Exemple d'analyses de champs de concentration d'ozone (étés 1999 et 2001)	107
4.4.1	Cartes de concentrations	107
4.4.2	Séries temporelles	109
4.4.3	Cartes d'erreurs d'analyse	112
4.5	Validation en altitude des analyses obtenues par la méthode ASI	114
4.6	Conclusions	115

5	Analyse des champs de concentrations d’ozone à l’échelle continentale et prévision de la pollution	117
5.1	Mise en place d’une Interpolation Statistique	118
5.1.1	Problème des biais d’observation et de modélisation	118
5.1.2	Choix des stations	118
5.1.3	Modélisation de la fonction de covariance	118
5.2	Qualité des analyses des champs de concentrations d’ozone	118
5.2.1	Utilisation des observations de surface	119
5.2.2	Utilisation des observations d’altitude de la campagne ESQUIF	123
5.3	Application à la prévision	125
5.3.1	Une série d’expériences en temps réel	125
5.3.2	Cartographie en quasi temps réel	129
5.3.3	Impact des analyses sur la prévision	135
5.4	Conclusions	143
6	Autres applications des méthodes d’analyse de données ASI et INK	145
6.1	Les limites du modèle CHIMERE	145
6.1.1	Le dioxyde d’azote	145
6.1.2	La région de Berre et de Marseille	146
6.2	Représentation des champs de concentrations du dioxyde d’azote sur l’Île-de-France	147
6.2.1	Méthodes d’analyse de données	147
6.2.2	Validation objective des méthodes d’analyse	147
6.2.3	Profils temporels et cycles diurnes	149
6.3	Étude de la pollution sur les bassins de Berre et de Marseille	151
6.3.1	Cas de l’ozone	151
6.3.2	Cas du dioxyde d’azote	155
6.4	Conclusions	156
7	Conclusions et Perspectives	159
7.1	Synthèse des travaux entrepris	159
7.1.1	Diagnostic des erreurs de modélisation du modèle CHIMERE	160
7.1.2	La cartographie	160
7.1.3	La prévision	163
7.2	Perspectives	163
A	La pollution photochimique troposphérique	167
A.1	Quelques espèces chimiques	167
A.1.1	L’ozone	167
A.1.2	Les oxydes d’azote	168
A.1.3	Les composés organiques volatils	168
A.2	Équilibre photochimique	168
A.3	Production nette d’ozone	169
A.3.1	Cycle des radicaux	169
A.3.2	Puits et sources de radicaux	170
A.3.3	Régimes chimiques	171
A.4	Conditions favorables aux épisodes de pollution photochimique	172
A.5	Les niveaux d’alerte	172
A.5.1	Le niveau d’information et de recommandation	172

A.5.2 Le niveau d'alerte	172
B Rappels de notions statistiques	175
B.1 L'espérance mathématique	175
B.2 Approximation de l'espérance mathématique	175
B.3 Variance et covariance	176
C Réseau européen de surveillance de la qualité de l'air	177
D Analyses météorologiques	181

Notations et unités

Unités des concentrations

Les concentrations des polluants sont données, dans ce document, en parties par billion (ppb). Une partie par billion correspond à une entité (atome, radical, molécule) pour un milliard de molécules d'air. Pour convertir cette unité en microgrammes par mètre cube, il suffit de multiplier les concentrations par le facteur k suivant :

$$k = \frac{pM}{RT} \times 10^{-3}$$

où p est la pression atmosphérique (Pa), T la température (K), M la masse molaire de l'entité étudiée (g.mol^{-1}) et R la constante des gaz parfaits ($\text{J.K}^{-1}.\text{mol}^{-1}$). Les valeurs de k pour l'ozone (O_3) et le dioxyde d'azote (NO_2) sont respectivement de 1.997, et de 1.914 dans les conditions de l'atmosphère standard.

$$R = 8.9149 \text{ J.K}^{-1}.\text{mol}^{-1}, M(\text{O}_3) = 47.998 \text{ g.mol}^{-1}, M(\text{NO}_2) = 46.006 \text{ g.mol}^{-1}.$$

Le temps universel

Le temps universel coordonné, que l'on note UTC, est l'heure locale à Greenwich, une ville d'Angleterre qui se trouve sur le méridien de longitude 0° . Il est midi en UTC lorsque le Soleil passe au méridien de Greenwich.

Le temps atomique international (TAI) est calculé à partir des données de plus de 200 horloges atomiques situées dans des instituts de métrologie ou des observatoires de plus de 30 pays. C'est une échelle uniforme et stable ; il ne suit pas par conséquent les légères irrégularités du mouvement de rotation de la Terre. Or pour certains besoins généraux et pratiques, il est nécessaire d'avoir une échelle qui en tienne compte au long terme ; c'est ce que fait le Temps universel coordonné qui est identique au TAI à cela près qu'on y ajoute de temps en temps une seconde intercalaire de telle sorte que le Soleil passe au méridien de Greenwich, à midi de UTC, à moins de 0,9 seconde près, en moyenne annuelle.

Liste des acronymes

Acronymes

Un certain nombre d'acronymes sont utilisés dans ce document et explicités ci-dessous :

Les organismes

CEPMMT	Centre Européen de Prévision Météorologique à Moyen Terme
ADEME	Agence De l'Environnement et la Maîtrise de l'Énergie
IPSL	Institut Pierre Simon Laplace
INSU	Institut National des Sciences de l'Univers
PNCA	Programme National de Chimie Atmosphérique
AASQA	Associations Agréées pour la Surveillance de la Qualité de l'Air
CITEPA	Centre Interprofessionnel Technique d'Étude de la Pollution Atmosphérique

Les Laboratoires

LMD	Laboratoire de Météorologie Dynamique
LISA	Laboratoire Interuniversitaire des Systèmes Atmosphériques
LA	Laboratoire d'Aérodynamique
SA	Service d'Aéronomie

Les campagnes de mesures

ESQUIF	Étude et Simulation de la Qualité de l'air en Île-de-France
--------	---

Les projets ou expériences

PIONEER	Prévisibilité et Incertitudes de l'Ozone à l'Échelle Européenne et Régionale
---------	--

Les modèles

CTM	<i>Chemistry Transport Model</i>
MELCHIOR	Modèle d'Étude Lagrangienne de la Chimie de l'Ozone à l'échelle Régionale
RADM2	<i>Regional Acid Deposition Model 2</i>
EURAD	EUROpean Air Pollution Dispersion

Les polluants

COV	Composés Organiques Volatils
-----	------------------------------

Autres

POI	Période d'Observation Intensive
ARAT	Avion de Recherche Atmosphérique et de Télédétection de l'INSU
PPM	<i>Parabolic Piecewise Method</i>

Chapitre 1

Introduction : étude et prévision de la pollution photochimique atmosphérique

”Il y a pollution de l’air lorsque la présence d’une substance étrangère ou une variation importante dans la proportion de ses constituants est susceptible de provoquer un effet nuisible compte tenu des connaissances scientifiques du moment, ou de créer une gêne.” Conseil de l’Europe, 1987.

Un polluant atmosphérique n’est pas toujours une espèce étrangère à l’atmosphère. En effet, les composés chimiques introduits dans l’air par les diverses activités humaines (transport, combustions diverses, agriculture ...) sont souvent de même nature que les espèces émises par des processus naturels. Ils deviennent polluants si la modification de leur concentration montre un impact négatif sur les êtres vivants et/ou sur l’environnement.

Les espèces émises dans l’atmosphère vont parfois réagir entre elles et donner naissance à de nouveaux constituants. Si la réaction nécessite un apport d’énergie lumineuse, la pollution résultante est appelée *pollution photochimique*. Les polluants émis directement dans l’atmosphère sont dits *primaires* tandis que ceux produits par des réactions chimiques sont dits *secondaires*. L’objectif de la présente introduction est de présenter les contextes social et scientifique qui motivent la démarche mise en œuvre au cours de cette thèse.

1.1 Mise en place d’un réseau de surveillance

1.1.1 Prise de conscience des problèmes liés à la pollution atmosphérique

La pollution atmosphérique n’est pas un phénomène nouveau. Depuis la révolution industrielle du 19^{ème} siècle, caractérisée par une utilisation massive des combustibles fossiles, un certain nombre d’épisodes tragiques ont marqué l’histoire. Un des plus célèbres est celui du smog de Londres de l’hiver 1952 (smog est la contraction de *smoke* et de *fog*). Une augmentation brutale des concentrations de dioxyde de soufre due à des conditions météorologiques particulièrement stagnantes, causa près de 4000 morts [Brimblecombe, 1987].

Les premiers pics de pollution photochimique apparaissent dans les années 40. Ils sont observés à Los Angeles au cours de périodes de fort ensoleillement (en été) [Leighton, 1961]. Cette pollution est caractérisée par l'augmentation du pouvoir oxydant de l'atmosphère et l'apparition de polluants secondaires tels que l'ozone.

Les années 70 marquent le début de la prise de conscience par la population des effets de la pollution atmosphérique sur la végétation. Ce sont notamment les Norvégiens et les Suédois qui attirent l'attention sur l'acidification des précipitations, qui se traduit alors par une acidification des eaux de surface. Ce problème donne lieu à de nombreuses études scientifiques entre 1970 et 1980, et c'est au début des années 1980 que le phénomène de dépérissement des forêts est constaté dans les régions industrialisées. La pollution atmosphérique, précédemment considérée comme un problème local, prend une *dimension régionale*, voire *continentale*.

Ce sont malheureusement ces effets néfastes pour l'environnement, ajoutés à des accidents industriels, tels que l'accident de Seveso en Italie (1976, pollution par la dioxine), celui à Bhopal en Inde (1984, pollution par le méthylisocyanide) et la catastrophe de Chernobyl (1986), qui ont favorisé la prise de conscience des gouvernements, des industriels et du public aux problèmes liés à la pollution atmosphérique via une forte médiatisation.

1.1.2 La loi sur l'air

Au niveau européen, des directives (valeurs guides assorties d'un temps d'exposition) émergent dans les années 80 pour un certain nombre de polluants : le dioxyde de soufre SO_2 (1980), le plomb Pb (1982), le dioxyde d'azote NO_2 (1985). Elles aboutissent surtout à des bilans de la qualité de l'air.

La directive Européenne de 1992 concernant l'ozone a un impact important dans l'histoire de la pollution atmosphérique en Europe. Plus novatrice et opérationnelle, elle implique la *création d'un réseau de surveillance* et d'une procédure d'information de la population, et d'alerte (le seuil d'information et de recommandation est, à cette date, de $180\mu g/m^3$ et le seuil d'alerte de $360\mu g/m^3$ en moyenne horaire). L'obligation d'action est aussi renforcée : en cas de risque de dépassement, les états membres doivent établir des plans d'action indiquant les mesures à prendre à court terme afin de réduire le risque et d'en limiter la durée. Cette directive est à l'origine de la **loi sur l'air et l'utilisation rationnelle de l'énergie, votée le 30 décembre 1996 en France**. Depuis cette date, un certain nombre d'Associations Agréées de Surveillance de la Qualité de l'Air (les AASQA) se sont créées, couvrant ainsi l'ensemble du territoire français. Regroupées au sein du réseau ATMO, ces associations sont chargées de surveiller la qualité de l'air, d'informer les autorités et les citoyens, de comprendre et prévoir les épisodes de pollution, et enfin d'évaluer l'impact des mesures de réduction des émissions. La loi sur l'air définit notamment les zones prioritaires à surveiller, et vise à terme une surveillance de l'ensemble du territoire.

1.1.3 Mesure stratégique

La pollution par l'ozone est surveillée depuis plusieurs décennies mais, avant la loi sur l'air, les capteurs étaient principalement placés dans ou au voisinage des grandes agglomérations. La loi sur l'air a permis le *développement du réseau des petites villes et mêmes des régions rurales éloignées*.

Il est également important de rappeler que les connaissances en matière de pollution atmosphérique ont fortement progressé depuis le vote de cette loi. A cette époque, par exemple, une "bulle" de pollution était présumée se développer autour des grandes agglomérations, comme Paris, mais ne pouvait être démontrée à cause du manque de mesures. Aujourd'hui le terme de panache urbain est totalement banalisé [ESQUIF, 2001].

Nous savons aussi aujourd'hui que la combinaison de réactions chimiques rapides avec la dynamique du transport conduit à une variabilité notable à court terme (sub-horaire) et à de fortes hétérogénéités spatiales des espèces réactives. Ainsi, en fonction de la réactivité de l'espèce mesurée, *les capteurs donnent une information plus ou moins localisée sur l'état de pollution de l'atmosphère*. Une observation d'ozone en milieu rural est souvent, par exemple, plus représentative des quelques kilomètres qui l'entoure qu'une observation en milieu urbain, plus soumise aux phénomènes locaux (réaction avec le monoxyde d'azote NO, turbulence ...). Pour diagnostiquer correctement l'état de pollution de l'atmosphère, l'idéal serait d'installer autant de stations de mesure qu'il est nécessaire pour identifier les structures de polluants. Ceci n'est pas concevable car le dispositif est très coûteux. Les capteurs sont alors placés à des endroits stratégiques : sous le vent des agglomérations (dans le panache urbain), près du trafic ...

La question essentielle qui reste alors en suspens et à laquelle il est important de répondre est la suivante : *A quels niveaux de concentration de polluants les êtres vivants sont-ils exposés, eux qui vivent précisément entre les stations de mesures ? Celle-ci nous amène à réfléchir sur les observations du réseau. Sont-elles suffisamment représentatives pour permettre une estimation sur toute une région ? Un capteur isolé, par exemple placé en milieu rural, qui mesure une valeur forte de polluant, donne-t-il une information sur un phénomène local ou sur une situation généralisée à toute une région (i.e. détecte-t-il une petite bulle de pollution ou un panache de quelques kilomètres d'extension) ? De quel domaine la mesure est-elle représentative ? Comment estimer quantitativement sa représentativité ?*

La modélisation numérique peut nous aider à répondre en partie à ces questions puisqu'un modèle peut donner une estimation de l'état de pollution hors des zones couvertes par la mesure, cohérente avec notre connaissance des lois d'évolution du système. Cependant, vu la complexité des phénomènes, qui sont étudiés, les modèles numériques ne sont que de représentations approximatives de la réalité, et contiennent ainsi des erreurs.

1.2 La modélisation numérique et la prévision de la pollution photochimique

Les besoins ressentis dans le domaine de la qualité de l'air (estimer la pollution hors des zones couvertes par la mesure, définir des mesures de réduction des émissions de polluants en cas de dépassement des seuils d'alerte et prévoir la pollution), et la complexité des phénomènes de pollution (en particulier due à la non linéarité du système chimique), ont incité les scientifiques à *développer des modèles déterministes de simulation de la qualité de l'air*. En outre, la prévision de la pollution est d'un intérêt scientifique majeur dans la mesure où une bonne prévision est le résultat d'une bonne compréhension des processus chimiques et physiques mis en jeu.

1.2.1 Les modèles numériques de Chimie-Transport (CTM)

Le premier modèle de simulation de la pollution photochimique urbaine a été réalisé par *Reynolds et al.* [1973]. Ce modèle était dédié à l'étude de la pollution à Los Angeles. De tels modèles, il en existe aujourd'hui une large variété de par le monde. Ceux-ci couvrent différentes échelles spatiales :

- l'échelle régionale : SMOG (Los Angeles) [*Lu et Turco, 1997*]; CALGRID (Milan) [*Silibello et al., 1998*]; Méso-NH-C (Toulouse) [*Tulet et al., 2000*]; CHIMERE (Paris) [*Vautard et al., 2000*].
- l'échelle continentale :
 - USA : [*Peters et al., 1995*]
 - EUROPE : EURAD [*Jakobs et al., 1995*]; [*EMEP, 1998*]; [*Olendrzynski, 1999*]; LOTOS [*Builtjes, 1992*]; CHIMERE [*Schmidt et al., 2001*]
- l'échelle globale : MOZART [*Hauglustaine et al., 1998*]; TM3 [*Lelieveld et Dentener, 2000*]; MOCAGE [*Cathala et al., 2002*]

Les modèles dédiés à l'étude de la pollution urbaine (modèles régionaux) ont plusieurs fonctions. Ils aident au problème de spatialisation évoqué dans le paragraphe 1.1.3. Ils permettent aussi d'interpréter les observations, dans la mesure où des études de sensibilité sur des paramètres peuvent être menées afin de comprendre les phénomènes d'épisodes passés. En retour, bien sûr les observations permettent de mieux comprendre les incertitudes de modélisation. Enfin, ils permettent de prévoir l'évolution future des concentrations des polluants et sont ainsi des outils d'aide à la prise de décision dans des plans d'action de réduction des émissions.

Le développement de tels modèles de chimie-transport sophistiqués a permis de considérables avancées dans le domaine de la pollution atmosphérique. Il a été notamment très rapidement établi que l'échelle spatiale des épisodes de pollution photochimique, dépassant largement l'échelle régionale, il était nécessaire de disposer d'outils de simulation de plus grande échelle. En effet si les pointes de pollution dans les basses couches de l'atmosphère sont généralement observées à proximité des sources majeures de polluants primaires (pôles industriels, agglomérations urbaines), une part importante de la pollution photochimique est généralement due au transport à longue distance de polluants (ozone, précurseurs) à laquelle vient s'ajouter la pollution locale [*Vautard et al., 2000*]. C'est précisément pour ces raisons que le modèle de simulation numérique CHIMERE, qui couvre l'Europe de l'Ouest tout en laissant la possibilité de "zoomer" les simulations sur des régions urbanisées, a été développé à l'IPSL.

Les modèles de chimie-transport permettent aujourd'hui de simuler de façon *assez réaliste* le transport et la chimie des différents polluants photochimiques. Cependant, la complexité des phénomènes étudiés, la variabilité spatiale et temporelle des émissions de polluants sont telles que, quel que soit le modèle, il ne pourra jamais reproduire parfaitement les concentrations. En effet, le modèle numérique est la traduction de l'état de connaissance scientifique du moment, via la modélisation physique et chimique des processus mis en jeu et la description de paramètres d'entrée comme par exemple les émissions, le type de sol, les concentrations au bord du domaine de modélisation. Lorsqu'un paramètre interne ou d'entrée du modèle est mal décrit, une erreur importante peut être observée certains jours sur les champs de concentration des polluants. Certains de ces paramètres peuvent même être intentionnellement éloignés de leur valeur réelle (nombre de polluants et/ou d'équations chimiques mis en jeu, résolution verticale et/ou horizontale ...) pour

réduire le temps de calcul, ce qui devient crucial si l'objectif du modèle est de faire de la prévision.

Ainsi, il est évident qu'avant même d'utiliser un modèle pour réaliser des simulations, études de scénarios ou des prévisions, il est nécessaire d'identifier ses limites (erreurs, incertitudes). Ceci ne peut être fait qu'à l'aide des observations. C'est une étape indispensable dans le développement d'un modèle.

1.2.2 La prévision de la pollution

La prévision des épisodes de pollution répond à une forte attente du public et des autorités. Elle permet de prendre, à court et à moyen terme, des mesures de réduction des émissions afin de limiter l'exposition de la population à des niveaux de pollution qui pourraient être nocifs pour la santé des personnes les plus sensibles, voire pour la santé de l'ensemble de la population.

La prévision des pics de pollution est essentiellement effectuée de part le monde à l'aide de modèles statistiques [*Simpson et Layton*, 1983; *Robeson et Steyn*, 1990] qui ont un succès incontestable. Ces modèles utilisent des méthodes de régression, de classification, d'arbre de décision [*Burrows et al.*, 1995] et des réseaux de neurones [*Gardner et Dorling*, 1998]. L'inconvénient de ces méthodes statistiques est que, plus un épisode a été rare par le passé, plus un épisode de même type sera difficile à prévoir. En outre, ils ne permettent pas de comprendre l'origine des pics de pollution prévus, puisqu'ils s'appuient sur des variables corrélées avec la pollution mais non directement causales. C'est principalement le cas pour la prévision des pics d'ozone avec la température.

Les modèles déterministes tridimensionnels de simulation de la qualité de l'air sont mieux adaptés à la compréhension de la pollution atmosphérique que les modèles statistiques, mais sont en général extrêmement détaillés dans leur formulation physico-chimique, et s'avèrent lourds de manipulation. Ainsi l'utilisation de *la plupart de ces modèles* ne peut être envisagée que dans le cadre de simulations courtes, de quelques jours. Ils ne peuvent donc pas être testés sur un nombre suffisamment grand de conditions météorologiques différentes pour juger de leurs performances. Cependant, quelques études montrent des comparaisons à plus longues échéances. Pour l'Est des USA, *Sirois et al.* [1999] présentent une comparaison sur deux mois de l'été 1995. En Europe, *Simpson* [1991] compare les concentrations simulées par un modèle lagrangien avec des observations pour les périodes d'été de 1985 à 1986. Une évaluation du même modèle est donnée de 1989 à 1996 par EMEP [1998]. De telles études permettent d'espérer aujourd'hui la réalisation de prévisions à courte échéance des épisodes de pollution photo-oxydante à l'échelle européenne. *Tilmes et al.* [2002] comparent les prévisions d'ozone pour 5 modèles eulériens pour *un été* (l'été 1999) sur les données allemandes. *Schmidt et al.* [2001] montrent l'efficacité du modèle CHIMERE à représenter les concentrations d'ozone à l'échelle européenne sur l'été 1998. Il semble maintenant envisageable de pouvoir mettre à disposition des décideurs et des associations chargées de la surveillance de la qualité de l'air en France des prévisions de qualité de l'air à deux ou trois jours d'échéance. Cette perspective est particulièrement importante car des délais de cet ordre sont nécessaires pour décider de mesures préventives, en informer le public et éventuellement infléchir les épisodes de pollution prévus. Il semble aussi établi que les prévisions des modèles de chimie-transport sont bien plus sensibles aux émissions, aux paramètres météorologiques, aux concentrations importées dans la région d'étude (conditions aux limites), et au dépôt, qu'aux conditions initiales. En effet, le système photochimique troposphérique ne semble pas être chaotique [*Honoré*, 2000]. En tout cas, aucune étude n'a mis en évidence la présence de chaos.

Le modèle CHIMERE a été construit pour être un outil diagnostique et aussi pronostique à coût numérique modéré. Il permet ainsi d'effectuer des simulations longues (plusieurs années) et est testé depuis maintenant trois ans en temps réel en mode de prévision dans le cadre du projet PIONEER (Prévisibilité et Incertitudes de l'Ozone à l'Échelle Européenne et Régionale). PIONEER est un projet destiné à étudier la prévisibilité de la pollution photo-oxydante aux échelles régionales/continentales, ainsi qu'à tester plusieurs méthodologies d'estimation des incertitudes de cette prévision, au moyen de deux modèles numériques de chimie-transport, CHIMERE et MOCAGE (Météo-France). Les expériences de prévision en temps réel, effectuées pendant l'été 2001 dans le cadre de ce projet, montrent que le système a une faible mémoire des conditions initiales (deux à trois jours). En outre, elles ont mis en évidence une difficulté : le transport d'erreurs de prévision. En effet, des erreurs commises à un endroit pour des raisons variées (problème dans les émissions, erreurs météorologiques, ...) peuvent éventuellement se propager pendant quelques jours et participer à accentuer l'erreur à un instant supérieur vers une autre région. Ainsi, même si, contrairement à la prévision météorologique, la prévision de la pollution atmosphérique n'est pas un problème de conditions initiales, pour éviter la propagation d'erreur il est nécessaire de "soigner" les conditions initiales d'une prévision à court terme. [Elbern et Schmidt, 2001] montrent notamment qu'une meilleure définition des conditions initiales améliore les prévisions jusqu'à 12 h après l'initialisation.

1.3 Pourquoi cette thèse ? Les objectifs

Le développement des modèles de chimie-transport sophistiqués ainsi que l'extension du réseau opérationnel de surveillance de la qualité de l'air ont permis d'améliorer notre compréhension du système photochimique. Cependant, cette compréhension, comme notre capacité à prévoir la pollution atmosphérique, sont limitées par un certain nombre de problèmes.

Premièrement, le nombre de mesures disponibles sur des polluants autres que l'ozone ou le dioxyde d'azote, est faible. Les stations de mesures de dioxyde d'azote sont, en outre, principalement localisées à l'intérieur des agglomérations, ce qui rend difficile l'évaluation des simulations de ce polluant à grande échelle. Deuxièmement, il n'y a pas de mesures de routine en altitude. Seules quelques campagnes de mesures nous permettent de contrôler la qualité des simulations dans les couches supérieures à la surface. Ainsi, il n'est pas possible aujourd'hui de valider complètement les simulations d'un modèle numérique. Cependant, une des façons de s'assurer qu'un modèle est performant, est de comparer ses sorties avec les observations disponibles sur des périodes temps suffisamment longues pour regrouper différentes situations.

Le premier objectif de cette thèse a été précisément de confronter les simulations du modèle numérique CHIMERE et les observations pour diagnostiquer les erreurs du modèle. Ces erreurs sont de différents types : erreur sur les valeurs météorologiques analysées et prévues, erreur sur les émissions de polluants, erreur sur les paramétrisations et le traitement des processus sous-maille, erreur due aux résolutions horizontale et verticale, erreur sur les concentrations initiales de la prévision. Dans la mesure où les modèles de chimie-transport s'appuient sur les forçages météorologiques, le premier type d'erreur est une erreur extérieure qu'il n'est pas susceptible de pouvoir corriger. L'incertitude sur les émissions est évidente vu la complexité des processus d'émissions et la multitude des différentes sources anthropiques et naturelles. Les erreurs de modélisation sont liées à la très difficile représentation de processus essentiels, comme l'impact de la nébulosité ou

des aérosols sur la photolyse, ainsi qu'à la représentation nécessairement condensée de la chimie.

Les méthodologies permettant de traiter, ou tout au moins d'aborder les deux dernières sources d'erreurs existent. Le manque de résolution d'un modèle d'échelle européenne peut être pallié par une approche de modèles de grille emboîtés sur des régions-clé comme les zones fortement urbanisées ou industrielles, ou les régions côtières. Enfin, les erreurs de prévisions dues à une mauvaise initialisation du modèle, qui s'accumulent et se transportent, peuvent être réduites grâce aux *méthodes d'assimilation de données*. Le deuxième objectif a donc été de développer de telles méthodes. Il s'agit de corriger l'*erreur globale* commise sur les simulations des champs de polluants afin d'offrir des cartes de concentrations de polluants, plus réalistes et cohérentes avec notre connaissance de la physique des phénomènes mis en jeu. Des techniques de *Krigeage* sont aussi testées. Le dernier objectif a été de répondre à la question suivante : est-ce que les champs de polluants, ainsi corrigés peuvent *améliorer la prévision de la pollution photochimique*, s'ils sont utilisés pour initialiser le modèle de prévision ?

Ces objectifs sont pratiquement ceux que se sont fixés les pionniers de la météorologie. Aujourd'hui nous bénéficions bien sûr de tout le savoir et l'expérience acquise par les météorologues. La problématique de la pollution est cependant très différente, car le problème des conditions initiales, qui a guidé le travail des météorologues, n'est pas primordial dans le domaine de la qualité de l'air. Même si la préparation de l'état initial d'une prévision est importante pour éviter la propagation d'erreurs, ce travail de thèse a surtout été motivé par la nécessité de disposer d'une *base de données réalistes de champs de polluants*.

Cette thèse s'est organisée en plusieurs étapes. Le travail a consisté à :

- **Évaluer les limites du modèle de chimie-transport CHIMERE**

Le modèle CHIMERE est constitué de plusieurs versions : CHIMERE continental, qui couvre l'Europe de l'Ouest avec une résolution spatiale assez faible ($0.5^\circ \times 0.5^\circ$) ; des modèles régionaux mieux résolus (Île-de-France, Régions de Berre et de Marseille, Alsace, Région toulousaine). Nous évaluons, dans un premier temps, la capacité du modèle CHIMERE à simuler les concentrations des polluants photochimiques à l'échelle régionale et à l'échelle continentale. A l'échelle régionale, nous focalisons sur les deux polluants les plus documentés, c'est-à-dire l'ozone et le dioxyde d'azote. Le modèle est testé sur deux régions particulièrement différentes en terme d'émissions, d'orographie et de phénomènes météorologiques : l'Île-de-France, et la région de Berre et de Marseille.

La région d'Île-de-France est une région éloignée des côtes, à la topographie très faible. Tandis que la région de Berre et de Marseille est soumise à des facteurs locaux importants : brises, orographie caractérisée principalement par les vallées de la Durance et du Rhône et les reliefs (Alpilles, Chaîne de l'Estaque, de la Ste Victoire, de la Ste Baume, du Lubéron ...). La région de Berre et de Marseille est aussi une région de fortes émissions, notamment sur la zone industrielle de l'étang de Berre.

Nous présentons de façon détaillée les résultats de comparaison des concentrations d'ozone et de dioxyde d'azote, simulées et observées en surface, pendant les étés 1999 ou 2001. Ces étés correspondent à des périodes, dont les conditions météorologiques ont été particulièrement propices à la formation d'épisodes de pollution photochimique. L'été 2000 a été, en revanche, marqué par des

conditions nuageuses, limitant l'établissement d'une telle pollution.

- **Appliquer des méthodes d'assimilation de données, déjà existantes dans d'autres domaines, les adapter au cas de la pollution photochimique, et valider ces méthodes pour corriger les simulations du modèle CHIMERE en utilisant des observations**

Afin d'obtenir de meilleures représentations spatiales des champs de concentration de polluants que celles des simulations, nous introduisons des observations de surface dans le modèle. C'est précisément ce que proposent les *méthodes d'assimilation de données*, déjà utilisées en météorologie et en océanographie. Ce travail de thèse s'est focalisé ainsi sur le développement, l'application, et la validation de certaines de ces méthodes. Pour pallier à des inconvénients particuliers au domaine de la qualité de l'air, nous appliquons d'autres méthodes couramment utilisées dans les domaines minier et pétrolier, et appelées les méthodes de Krigeage. Toutes les méthodes sont testées à l'échelle continentale, comme à l'échelle régionale pour produire des champs de concentrations d'ozone ou de dioxyde d'azote, bidimensionnels ou tridimensionnels. Les champs corrigés résultants, appelés *analyses*, offrent alors une vision synoptique de l'état de pollution, directement utilisable pour repérer rapidement les défaillances du modèle numérique.

Nous cherchons également à estimer *objectivement* la qualité des analyses ainsi produites. Nous utilisons, pour cela, une méthode appelée *leave-one-out*. Le principe est de comparer les analyses avec des observations indépendantes (non utilisées dans l'algorithme de correction). Sur la région d'Île-de-France, nous utilisons en plus les observations aéroportées, effectuées pendant la campagne ESQUIF (Étude et Simulation de la QUalité de l'air en Île-de-France) pour estimer la qualité des analyses tridimensionnelles.

- **Quantifier l'effet d'une initialisation du modèle CHIMERE avec les champs de polluants corrigés sur ses prévisions à court terme**

La question est de savoir si les analyses, utilisées pour initialiser le modèle de prévision, au lieu des prévisions de la veille, peuvent améliorer les prévisions à court terme. Nous présentons les résultats d'une expérience de prévision de l'ozone, menée d'abord en temps réel dans le cadre du projet EUROPOLLUX pendant l'été 2000, puis reconduite dans le cadre du projet PIONEER pendant les étés 2001 et 2002. Ces expériences ont permis d'initier la mise en place d'une plateforme de prévision des concentrations d'ozone en temps réel, comprenant l'acquisition en temps réel de données d'observation, et l'assimilation de ces données.

Les prochains paragraphes proposent une revue bibliographique de l'assimilation de données et un rappel des méthodes qui sont actuellement utilisées en météorologie, en océanologie et en géostatistique. Le but de ces sections est de justifier le choix des méthodes mises en œuvre au cours de cette thèse.

1.4 Assimilation de données en météorologie et en océanologie

L'objectif des méthodes d'assimilation de données, développées dans le domaine de la météorologie et de l'océanologie, est de combiner des observations et des simulations d'un modèle numérique. Ceci a pour but d'estimer de façon la plus réaliste possible *la valeur réelle moyenne par maille du modèle d'une variable* et d'obtenir des cartes de cette variable cohérentes avec les observations et

les simulations. Le résultat final, qui correspond à une simulation corrigée par les observations, est appelé une *analyse*.

L'analyse subjective

Les méthodes d'assimilation de données trouvent leur origine dans la seconde moitié du XIX^{ème} siècle, à l'époque où les premières cartes météorologiques ont été réalisées. Ces cartes étaient le résultat d'une simple interpolation, faite à la main, des observations. En Europe, *Le Verrier* (directeur de l'observatoire de Paris) fut le premier à construire de telles cartes, avec lesquelles il observa le développement et l'évolution de la terrible tempête du 14 novembre 1854, qui détruisit la flotte française engagée dans la guerre de la Crimée [*Locher*, Thèse en cours]. Il collecta des observations sur l'Europe pour la période précédant la tempête. La figure 1.1 présente un fac-similé de la carte qu'il présenta à l'Académie des Sciences le 31 décembre 1855. Pour *Le Verrier*, les lignes isobariques indiquent que la tempête aurait pu être prévue quelques jours auparavant, "l'onde atmosphérique" étant d'abord passée au-dessus de la France. Ses travaux ont été le point de départ de l'organisation d'un service de prévision météorologique à l'observatoire de Paris. Ces cartes furent adoptées ensuite par les météorologues, qui espéraient, grâce à elles, établir des lois de l'écoulement de l'air atmosphérique et faire de la prévision météorologique.

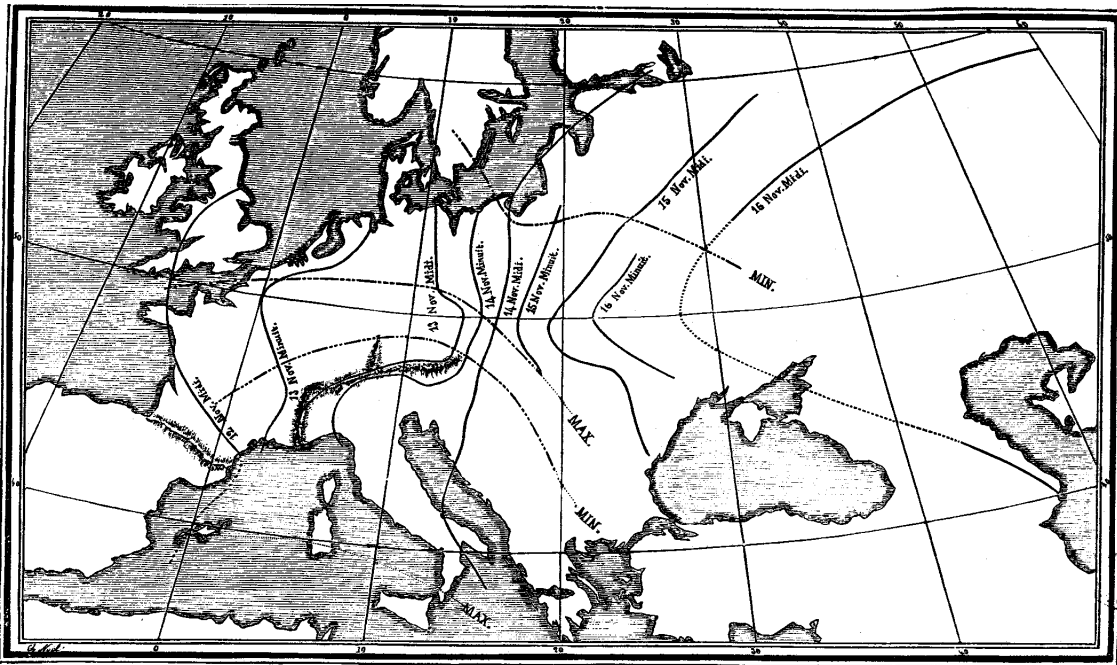


FIG. 1.1 – Fac-similé de la carte présentée à l'Académie des Sciences par Urbain Le Verrier, le 31 décembre 1855. Les lignes isobariques ayant traversé l'Europe du 12 au 15 novembre 1854 y sont tracées. Source : [*Locher*, Thèse en cours]

Comme les cartes ou *analyses* étaient le résultat de leur propre jugement, ces analyses sont

maintenant dites *analyses subjectives* pour les distinguer des *analyses objectives* utilisées actuellement.

L'analyse objective

La première analyse objective est produite par *Panofsky* [1949] pour remplacer l'interpolation graphique manuelle. Il utilise un polynôme pour interpoler les observations. Les poids de l'interpolation sont déterminés par la méthode des moindres carrés. Ils dépendent de la qualité de l'observation. *Gilchrist et Cressman* [1954] introduisent ensuite le concept de *rayon d'influence*, qui définit la distance à partir de laquelle les observations n'apportent plus d'information. Ils remarquent qu'il serait plus efficace d'utiliser dans l'analyse les sorties d'un modèle numérique.

Utilisation d'un modèle numérique

L'idée de combiner des observations et un modèle n'est pas nouvelle. En effet, *Gauss* [1809] dans son ouvrage sur *La théorie du mouvement des corps célestes* décrit des méthodes de moindre carré pour optimiser son estimation de l'orbite des comètes et des planètes à partir de ses observations et des lois de la mécanique Newtonienne.

Il écrit : "*If the astronomical observations and other quantities on which the computations of orbits is based were absolutely correct, the elements also, whether deduced from three or four observations, would be strictly accurate (so far indeed as the motion is supposed to take place exactly according to the laws of Kepler), and therefore, if other observations were used, they might be confirmed, but not corrected. But since our measurements and observations are nothing more approximations to truth, the same must be true of all calculations resting upon them, and the highest aim of all computation made concerning concrete phenomena must be to approximate, as nearly as practicable to the truth. But this can be accomplished in no other way than by a suitable combination of more observations than the number absolutely requisite for the determination of the unknown quantities. This problem can only be properly undertaken when an approximate knowledge of the orbit has been already obtained, which is afterward to be corrected, so as to satisfy all of the observations in the most accurate manner possible.*"

Ce texte contient un certain nombre d'idées clefs [*Daley, 1997*] : (1) tous les modèles et toutes les observations sont des *approximations de l'état réel*, (2) le résultat de la combinaison de ces deux estimations est aussi une *approximation*, (3) les observations doivent être incluses dans le problème de manière *optimale*, (4) Plus le nombre d'observations est grand, meilleure est la détermination du problème, (5) le modèle peut être utilisé comme *première estimation*.

L'arrivée des ordinateurs donne naissance à la prévision numérique. *Bergthorsson et Doos* [1955] et *Thompson* [1961] sont les premiers à montrer que les analyses sont améliorées si un modèle de prévision est utilisé en plus des observations disponibles. Ils posent ainsi les premières pierres de l'assimilation de données. Depuis, les méthodes n'ont pas beaucoup évolué. Dès les années 1960, les analyses objectives sont produites en temps réel dans la plupart des centres nationaux du temps. *Cressman* [1959] décrit le système américain basé sur les travaux de *Bergthorsson et Döös* [1955]. Les poids de l'interpolation linéaire sont inversement proportionnels à la distance entre l'observation et le point d'analyse. Leur forme et le rayon d'influence sont choisis arbitrairement. Loin des observations l'analyse est relaxée vers la simulation.

Les méthodes statistiques

Eliassen [1954] donne une forme statistique aux méthodes d'assimilation de données et définit ce que nous appelons maintenant l'*Interpolation Statistique* en météorologie. Les observations sont considérées comme des réalisations d'un champ de la variable étudiée (vent, concentration d'un polluant, ...) qui a une certaine structure de covariance spatiale stationnaire dans le temps. Les poids de l'interpolation sont déterminés statistiquement en étudiant cette structure de covariance sur une *base de données climatologique* (correspondant à un ensemble d'information sur l'état passé du système, c'est-à-dire sur la distribution spatio-temporelle des variables étudiées). Toutes les méthodes, qui ont succédé, sont basées sur celle-ci.

Thompson [1961] montre que les corrections en un temps et un lieu donné permettent d'améliorer les prévisions en un autre endroit. Il introduit ainsi le concept important de *propagation de la correction*.

Gandin [1963] étudie théoriquement les analyses multi-variables. Cependant il a conscience que les méthodes qu'il développe sont difficilement applicables car les corrélations croisées ne sont pas encore physiquement comprises. A l'époque les ordinateurs ne sont pas assez puissants pour analyser deux variables simultanément. Le développement informatique rend possible les études des corrélations croisées. Ces études montrent qu'analyser plusieurs variables en même temps améliore la qualité des analyses [*Kluge*, 1970; *Shlatter*, 1975; *Bergman*, 1979]. *Rutherford* [1973] montre qu'une ré-initialisation du modèle avec des analyses améliore les prévisions météorologiques jusqu'à 12 h après l'initialisation.

Depuis ces dernières années, d'autres méthodes d'assimilation de données ont été développées. Ces algorithmes sont classés en deux types : les *méthodes séquentielles* et les *méthodes variationnelles* [*Talagrand*, 1997].

Parallèlement aux travaux de *Gandin*, *Matheron* [1963] décrit une technique géostatistique, appelée le Krigeage. Cette méthode, utilisée dans l'industrie minière est assez similaire à l'*Interpolation Statistique*, à la différence près qu'elle ne nécessite pas de connaissance *a priori* de l'état passé du système.

1.5 Les méthodes d'assimilation de données

Les méthodes d'assimilation de données donnent la possibilité d'estimer de façon la plus réaliste possible le champ d'une variable (température, vent, concentrations de polluants, ...) à partir (1) d'une climatologie (2) d'un modèle numérique (qui permet d'établir les liens entre les variables observées et les composés non observés, et qui donne leur évolution) (3) des observations. Les principales méthodes d'assimilation existant actuellement sont décrites ci-dessous. Elles diffèrent essentiellement par leur coût numérique et leur optimalité (qualité du résultat obtenu).

Par souci de cohérence avec les méthodes géostatistiques, que nous utilisons également, nous avons choisi un système de notations, inspiré de [*Bertino*, 2001], qui diffère de celui utilisé habituellement en météorologie ou en océanologie. Les concepts utilisés ci-dessous, tels que l'espérance mathématique, la covariance, la variance, la corrélation, sont définis dans l'Annexe B, dédiée à des rappels de statistique.

Objectifs et définitions

Un modèle est limité par sa résolution, il ne peut pas simuler l'état réel "fin" de l'atmosphère. L'objectif des méthodes d'assimilation de données, à défaut d'avoir accès à l'état réel, est donc d'estimer de façon la plus réaliste possible l'état réel moyen par maille du modèle de l'atmosphère. Cet état réel moyen est représenté par le vecteur \mathbf{Z}^t de dimension n , le nombre de points de grille du modèle. (L'exposant t signifie *true* en anglais.)

À un instant donné, une simulation (un champ de vent, de température, un champ de concentration de polluants ...) est corrigée sur un certain domaine de simulation, en utilisant un jeu de p observations. Cette simulation, appelée couramment *ébauche*, est représentée par un vecteur \mathbf{Z}^b (b pour *background* en anglais), aussi de dimension n , qui regroupe les valeurs simulées en chaque point de grille du modèle. Nous définissons $\epsilon^b = \mathbf{Z}^b - \mathbf{Z}^t$, l'erreur du modèle.

Les observations, disponibles uniquement aux *points de mesure* sont, quant à elles, représentées par un vecteur \mathbf{Y}^o , de dimension p . Comme les points de mesure ne correspondent que très rarement à un point de grille du modèle, les simulations sont, dans un premier temps, interpolées aux points de mesure. Cette interpolation est le plus souvent effectuée à l'aide d'une combinaison linéaire des valeurs simulées \mathbf{Z}^b , et représentée par un opérateur, noté \mathbf{H} , appelé *opérateur d'observation*. Nous définissons également les *erreurs d'observation* par $\epsilon^o = \mathbf{Y}^o - \mathbf{H}\mathbf{Z}^b$.

Le résultat de la correction de l'ébauche par les observations s'appelle l'*analyse*. Cette analyse est représentée par le vecteur \mathbf{Z}^a de dimension n . L'erreur d'analyse est définie par $\epsilon^a = \mathbf{Z}^a - \mathbf{Z}^t$.

1.5.1 Les méthodes séquentielles

L'Interpolation Statistique (IS)

L'Interpolation Statistique [Daley, 1991] a été la méthode la plus utilisée dans les principaux centres de prévision météorologique pour produire des analyses météorologiques. Elle présente le grand avantage de fournir une procédure relativement facile à mettre en œuvre qui permet de traiter un ensemble quelconque d'observations.

Les différences entre les observations et les simulations prises aux points de mesure $\mathbf{Y}^o - \mathbf{H}\mathbf{Z}^b$ sont interpolées par l'intermédiaire d'un opérateur linéaire \mathbf{K} , appelée *matrice de gain*. Le résultat de l'interpolation, qui définit la correction à appliquer sur le domaine étudié, est ensuite ajouté à l'ébauche \mathbf{Z}^b pour donner finalement une "simulation corrigée", c'est à dire l'*analyse*. Sous certaines hypothèses qui seront énoncées au chapitre 2, où nous décrivons la méthode de façon plus détaillée, l'expression générale de l'analyse est donc définie par :

$$\mathbf{Z}^a = \mathbf{Z}^b + \mathbf{K}[\mathbf{Y}^o - \mathbf{H}\mathbf{Z}^b] \quad (1.1)$$

En outre, on montre que, sous les contraintes définies ci-dessous, la matrice de gain s'écrit :

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1} \quad (1.2)$$

\mathbf{K} est en fait la matrice regroupant les poids de l'interpolation linéaire. Ces derniers sont déterminés de telle façon que l'analyse est statistiquement la meilleure possible, c'est à dire que sa

variance d'erreur $E[\epsilon^a(\mathbf{s})^2]$ est minimale en tout point de grille \mathbf{s} . $\mathbf{B} = E(\epsilon^b(\epsilon^b)^T)$ est la matrice des covariances des erreurs du modèle, $\mathbf{R} = E(\epsilon^o(\epsilon^o)^T)$, celle des covariances des erreurs d'observation. \mathbf{HBH}^T correspond à une réduction de la matrice \mathbf{B} aux points de mesures. \mathbf{BH}^T est la matrice des covariances d'erreur du modèle entre les points de grille et les points de mesure.

Ainsi, les poids dépendent des covariances des erreurs du modèle et des observations, leur covariances croisées étant supposées nulles. Si ces covariances sont connues, les poids de l'interpolation peuvent être calculés. Or, la difficulté des méthodes d'assimilation de données est précisément liée au fait que ces covariances d'erreur ne sont pas accessibles directement, et qu'il est nécessaire de les modéliser. Une des principales hypothèses faite dans le cadre d'une Interpolation Statistique est que les covariances des erreurs du modèle sont constantes dans le temps (stationnaires).

Optimisés, les poids d'interpolation permettent de calculer la matrice de covariance d'erreur d'analyse $\mathbf{A} = E(\epsilon^a(\epsilon^a)^T)$, qui est donnée par :

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B} \quad (1.3)$$

Les éléments diagonaux de cette matrice correspondent aux variances d'erreur d'analyse. Elles permettent d'estimer la qualité de l'analyse obtenue.

Il est souvent fait l'hypothèse que seules quelques observations sont nécessaires pour déterminer la correction à effectuée en un point. Lorsqu'une telle hypothèse est faite, la méthode prend le nom d'*Interpolation Optimale (OI)*.

Filtre de Kalman linéaire (KF)

Le filtre de Kalman [Kalman, 1960; Kalman et Bucy, 1961; Ghil, 1989; Lacarra et Talagrand, 1988; Ghil et Malanotte-Rizzoli, 1991; Errico et al., 1993] est une extension de la méthode d'Interpolation Statistique au cas d'une *assimilation en continu des observations*. Le modèle numérique, linéaire dans ce cas, est corrigé au fur et à mesure que les observations sont disponibles. Cette méthode est souvent utilisée dans des systèmes de prévision. L'ébauche est alors une prévision, et est notée Z^f (f de *forecast* en anglais). Elle provient d'une simulation qui a été initialisée par une précédente analyse. Dans l'*erreur de prévision*, notée $\epsilon^f = \mathbf{Z}^f - \mathbf{Z}^t$, on distingue alors l'erreur de modélisation (due à une erreur de paramétrisation et/ou erreur dans les entrées du modèle : prévisions météorologiques, émissions ...), de l'erreur sur les conditions initiales (erreur commise sur la précédente analyse).

Les covariances d'erreur de prévision, notées $\mathbf{P}_f = E(\epsilon^f(\epsilon^f)^T)$, ne sont plus supposées être stationnaires dans le temps, comme dans le cas de l'Interpolation Statistique. En effet, le Filtre de Kalman propose précisément une loi d'évolution de ces covariances.

Sous les mêmes hypothèses que l'Interpolation Statistique, la meilleure façon d'assimiler des observations est définie par récurrence sur les temps d'observation i :

$$\mathbf{Z}^f(i+1) = \mathbf{M}_{i \rightarrow i+1} \mathbf{Z}^a(i) \quad (1.4)$$

$$\mathbf{P}^f(i+1) = \mathbf{M}_{i \rightarrow i+1} \mathbf{P}^a(i) \mathbf{M}_{i \rightarrow i+1}^T + \mathbf{Q}(i) \quad (1.5)$$

$$\mathbf{K}(i) = \mathbf{P}^f(i) \mathbf{H}^T(i) [\mathbf{H}(i) \mathbf{P}^f(i) \mathbf{H}^T(i) + \mathbf{R}(i)]^{-1} \quad (1.6)$$

$$\mathbf{Z}^a(i) = \mathbf{Z}^f(i) + \mathbf{K}(i) [\mathbf{Y}^o(i) - \mathbf{H}(i) \mathbf{Z}^f(i)] \quad (1.7)$$

$$\mathbf{P}^a(i) = (\mathbf{I} - \mathbf{K}(i) \mathbf{H}(i)) \mathbf{P}^f(i) \quad (1.8)$$

La première équation traduit juste le fait que le modèle est utilisé pour prévoir l'état de l'atmosphère aux instants supérieurs en partant de l'analyse précédente $\mathbf{Z}^a(i)$. \mathbf{M} est l'*opérateur linéaire de prévision* associé à ce modèle.

La deuxième équation est l'équation de *propagation des covariances de l'erreur* de prévision. Elle montre que $\mathbf{P}^f(i+1)$ est la somme de deux termes : les covariances de l'erreur d'analyse $\mathbf{P}^a(i)$ propagées par le modèle et les covariances de l'erreur du modèle $\mathbf{Q}(i)$. Cette équation est obtenue en soustrayant l'état réel $\mathbf{Z}^t(i+1)$ à l'équation précédente et en utilisant la linéarité de l'opérateur de prévision \mathbf{M} ,

$$\mathbf{Z}^f(i+1) - \mathbf{Z}^t(i+1) = \mathbf{M}_{i \rightarrow i+1} [\mathbf{Z}^a(i) - \mathbf{Z}^t(i)] + \mathbf{M}_{i \rightarrow i+1} \mathbf{Z}^t(i) - \mathbf{Z}^t(i+1) \quad (1.9)$$

puis en multipliant cette équation par sa transposée, et en prenant l'espérance du résultat. Les termes de corrélation entre les erreurs d'analyses et les erreurs modèles sont supposées être nuls. $\mathbf{M}_{i \rightarrow i+1} \mathbf{Z}^t(i) - \mathbf{Z}^t(i+1)$ est l'erreur de modélisation.

Les trois dernières équations sont exactement les mêmes que celles de l'Interpolation Statistique.

Comme son nom l'indique, le Filtre de Kalman linéaire, est un algorithme, utilisable pour assimiler en continu des observations dans un modèle *linéaire*. Le point fort de cette méthode est la modélisation de l'évolution des matrices de covariances d'erreur de prévision. Son point faible est son coût numérique. Ceci est discuté dans le paragraphe 1.8.

Filtre de Kalman Étendu (EKF)

La plupart des modèles en météorologie et en océanologie sont non linéaires. L'algorithme du Filtre de Kalman linéaire, présenté dans le paragraphe précédent, peut être généralisé à des opérateurs d'observation H et des modèles M non linéaires. On parle, dans ce cas, de Filtre de Kalman Étendu. Les opérateurs sont alors linéarisés au voisinage de la variable à analyser, c'est à dire qu'ils sont supposés être linéaires au voisinage de n'importe quel état \mathbf{Z} . Par exemple, pour tout $\mathbf{Z}(i)$ proche de $\mathbf{Z}^a(i)$, $M_{i \rightarrow i+1} \mathbf{Z}^a(i) - M_{i \rightarrow i+1} \mathbf{Z}(i) \simeq M_{i \rightarrow i+1} [\mathbf{Z}^a(i) - \mathbf{Z}(i)]$. On dit qu'ils sont alors définis par leur *modèle linéaire tangent* \mathbf{M} (i.e leur différentiel).

Si les non linéarités sont très importantes, il peut être nécessaire d'introduire des termes de correction dans l'équation précédente ou d'utiliser une méthode *stochastique* plus générale telle que le *Filtre de Kalman d'Ensemble*.

Filtre de Kalman d'Ensemble (EnKF)

Le Filtre de Kalman d'Ensemble est une méthode encore relativement peu utilisée, qui a été initialement introduite par *Evensen* [1994]. L'objectif est, comme dans le cas du Filtre de Kalman, de prévoir l'évolution des matrices de covariances d'erreur de prévision. Cependant, le problème est abordé de manière *stochastique*. L'état initial, provenant d'une analyse, est perturbé compte tenu de la variance d'erreur déterminée sur cet état pour créer un *ensemble d'états initiaux*. Le modèle (en général non linéaire) est ensuite intégré à partir de chacun de ces états initiaux, créant ainsi un *ensemble de prévisions*. Une erreur modèle aléatoire est introduite à chaque pas d'intégration.

À tout instant, les covariances d'erreur de prévision peuvent être calculées sur l'ensemble statistique ainsi construit : l'erreur de prévision, qui inclut l'erreur modèle et l'erreur commise sur l'état initial, est alors approximée en prenant l'écart d'une simulation à la moyenne de l'ensemble. Les observations sont ensuite assimilées de la même manière qu'une Interpolation Statistique. En outre, il est nécessaire de prendre en compte l'erreur d'observation pour éviter que l'ensemble des prévisions ne converge pas trop vite vers la même analyse. C'est pourquoi, avant de les assimiler dans chaque membre de l'ensemble, les observations sont perturbées en fonction de la variance d'erreur d'observation supposée.

1.5.2 Les méthodes variationnelles

Analyse variationnelle tridimensionnelle (3D-VAR)

Il y a deux façons de définir le problème de l'Interpolation Statistique [*Lorenc*, 1986] : (1) supposer que les matrices de covariances d'erreur sont connues et dériver les équations de l'analyse en cherchant le minimum de variance de l'erreur d'analyse, c'est la méthode que nous avons exposée dans le paragraphe 1.5.1, ou alors (2) supposer que les erreurs du modèle et des observations sont gaussiennes et dériver les équations en cherchant l'état le plus probable; c'est alors la méthode du 3D-VAR [*Parrish et Derber*, 1992; *Courtier et al.*, 1998].

La distribution de probabilité $P(\mathbf{Z})$ pour la réalisation d'un état réel donné \mathbf{Z} s'exprime en fonction des distributions de probabilité, supposées gaussiennes, pour les erreurs d'observation et les erreurs du modèle. L'état analysé est alors défini comme le maximum de $P(\mathbf{Z})$, c'est à dire la valeur la plus probable. Une *fonction de coût* est définie par $J(\mathbf{Z}) = -\ln(P(\mathbf{Z}))$. *Lorenc* [1986] montre que $J(\mathbf{Z})$ s'écrit :

$$J(\mathbf{Z}) = (\mathbf{Z} - \mathbf{Z}^b)^T \mathbf{B}^{-1} (\mathbf{Z} - \mathbf{Z}^b) + (\mathbf{Y}^o - \mathbf{H}\mathbf{Z})^T \mathbf{R}^{-1} (\mathbf{Y}^o - \mathbf{H}\mathbf{Z}) \quad (1.10)$$

L'analyse, alors définie comme le minimum de J , est approchée en faisant plusieurs évaluations de cette fonction et de son gradient. La principale hypothèse faite est qu'un nombre limité d'itérations est suffisant pour atteindre le minimum. Les covariances d'erreur du modèle ne sont pas directement calculées, il suffit de leur donner une forme, et d'optimiser les paramètres de leur expression.

Le principe du 3D-VAR est strictement équivalent à l'Interpolation Statistique si l'opérateur d'observation est linéaire et si les distributions des erreurs du modèle et des observations sont effectivement gaussiennes.

Assimilation variationnelle quadridimensionnelle (4D-VAR)

La méthode du 4D-VAR [Lewis et Derber, 1985; Talagrand et Courtier, 1987; Lacarra et Talagrand, 1988; Thépaut et Courtier, 1991; Rabier et Courtier, 1992; Errico et al., 1993] est une généralisation du 3D-VAR pour des observations *distribuées dans le temps*. Son principe est également d'analyser l'état de l'atmosphère à *un instant donné*, mais en tenant compte d'observations *antérieures et postérieures* au temps de l'analyse. Autrement dit, il s'agit là d'estimer un état en cherchant à minimiser les écarts entre le modèle et les observations sur une certaine période de temps, i.e. sur une *fenêtre d'assimilation*. Cette méthode est particulièrement utile pour faire des *ré-analyses* d'épisodes passés.

À l'instant de mesure i , \mathbf{Y}_i^o , \mathbf{Z}_i et \mathbf{Z}_i^t désignent respectivement les observations, les simulations et l'état réel. L'algorithme du 4D-VAR est défini comme la minimisation de la fonction de coût :

$$J(\mathbf{Z}) = (\mathbf{Z} - \mathbf{Z}^b)^T \mathbf{B}^{-1} (\mathbf{Z} - \mathbf{Z}^b) + \sum_{i=0}^n (\mathbf{Y}_i^o - \mathbf{H}_i \mathbf{Z}_i)^T \mathbf{R}_i^{-1} (\mathbf{Y}_i^o - \mathbf{H}_i \mathbf{Z}_i) \quad (1.11)$$

\mathbf{R}_i est la matrice de covariances d'erreurs d'observation. La matrice de covariance des erreurs du modèle \mathbf{B} est seulement définie à l'instant de l'analyse. Les équations sont les mêmes que le 3D-VAR, avec cela en plus que l'opérateur d'observation inclut le modèle de prévision, ce qui permet de comparer les observations \mathbf{Y}_i^o et les simulations \mathbf{Z}_i au de temps de la mesure. En outre, le problème de minimisation est contraint par le fait que la séquence des états du modèle doivent être une solution des équations du modèle :

$$\forall i, \mathbf{Z}_i = M_{0 \rightarrow i}(\mathbf{Z}) \quad (1.12)$$

où $M_{0 \rightarrow i}(\mathbf{Z})$ est l'opérateur du modèle *non linéaire* de prévision de l'instant initial jusqu'au temps t_i . Le 4D-VAR est ainsi un problème d'optimisation à contrainte non linéaire très difficile à résoudre sous sa forme générale. Heureusement, il peut être simplifié en faisant deux hypothèses. La première est que le modèle peut être linéarisé, la deuxième est qu'il peut être exprimé comme le produit d'étapes *linéaires* de prévision. Ainsi toute simulation \mathbf{Z}_i s'écrit

$$\mathbf{Z}_i = \mathbf{M}_i \mathbf{M}_{i-1} \dots \mathbf{M}_1 \mathbf{Z} \quad (1.13)$$

Cette équation reflète la causalité de la nature. Les modèles \mathbf{M}_i sont les *modèles linéaires tangent d'étape de prévision*. Le premier terme de la fonction de coût $J^b(\mathbf{Z}) = (\mathbf{Z} - \mathbf{Z}^b)^T \mathbf{B}^{-1} (\mathbf{Z} - \mathbf{Z}^b)$, dit *fonction de coût de l'ébauche*, est le même que pour le 3D-VAR. L'évaluation du second terme $J_o = \sum_{i=0}^n (\mathbf{Y}_i^o - \mathbf{H}_i \mathbf{Z}_i)^T \mathbf{R}_i^{-1} (\mathbf{Y}_i^o - \mathbf{H}_i \mathbf{Z}_i)$, dit *fonction de coût d'observation* et de son gradient est lourde et nécessite un algorithme basé sur l'intégration des transposés des modèles linéaires tangents des modèles étapes de prévision : \mathbf{M}_i^T (appelés *modèles adjoints*) [Talagrand et Courtier, 1987].

L'inconvénient majeur du 4D-VAR est la méconnaissance des erreurs du modèle, qui interviennent dans les termes J^b et J^o . C'est pourquoi, la plupart du temps, le modèle est supposé parfait (i.e. pas d'erreur de modélisation). Le terme J^b devrait alors disparaître de l'expression de J . Cependant, pour que l'analyse soit au pire égale à l'estimation provenant d'un 3D-VAR, le terme J^b est conservé.

1.6 Les méthodes géostatistiques

Les méthodes géostatistiques sont souvent utilisées pour analyser le champ d'une variable à partir uniquement de mesures distribuées sur une certaine région de l'espace. En fonction des hypothèses, qui sont faites sur la variable à analyser, les méthodes de Krigeage prennent un nom différent (Krigeage Simple, Ordinaire, Intrinsèque, Universel ...). Elles sont d'un intérêt majeur pour l'estimation de *covariances spatiales*.

L'Interpolation statistique est très proche d'un Krigeage, les différences apparaissant seulement dans la détermination des poids de l'interpolation. Dans le cas du Krigeage, ces poids sont le produit de statistiques faites sur des données distribuées *dans l'espace*, tandis que dans le cas de l'Interpolation Statistique, ils sont déterminées à partir de statistiques faites sur des données distribuées *dans l'espace et dans le temps*.

Par rapport aux méthodes d'assimilation de données, qui utilisent un modèle numérique apportant une information sur les contraintes physiques spatio-temporelles, les méthodes de Krigeage ne permettent la prise en compte que de contraintes simples (linéaires [Lajaunie, 1997] ou géométriques). La géostatistique spatio-temporelle est abordée par Matheron [1970], cependant à cause de problèmes d'anisotropie entre la dimension spatiale et temporelle, il est rare que les théories géostatistiques soient appliquées à l'espace-temps [Bertino, 2001].

1.7 Application à la chimie atmosphérique

Depuis le milieu des années 1980, les méthodes précédemment décrites sont appliquées au domaine de la pollution chimique.

Étude de la chimie stratosphérique

L'assimilation d'observations satellitaires de gaz trace dans les modèles atmosphériques est l'objet de nombreuses publications. Fisher et Lary [1995] sont les premiers à introduire un processus d'assimilation variationnelle pour étudier la chimie atmosphérique. Ils utilisent un mécanisme chimique très réduit (6 espèces et 19 réactions) et montrent que le mécanisme de Chapman avec le cycle catalytique des NO_x , les observations d'ozone et de dioxyde d'azote suffisent à analyser toutes les autres espèces. Levelt et al. [1996] assimilent des colonnes d'ozone dans un modèle d'advection de traceur. Lyster et al. [1997] utilisent pour la première fois un Filtre de Kalman pour assimiler les mesures de méthane du satellite UARS dans un modèle bidimensionnel global de transport atmosphérique. Ils mettent en évidence les difficultés quant à appliquer un Filtre de Kalman complet. Récemment Ménard et al. [2000], Ménard et Chang [2000] proposent une paramétrisation simplifiée de l'évolution des covariances des erreurs. Khattatov et al. [2000] étendent la méthode à un modèle tridimensionnel. Ekes et al. [1999] construisent des analyses bidimensionnelles d'ozone à l'aide d'observations satellitaires et d'une méthode variationnelle. Jeuken et al. [1999] utilisent une Interpolation Optimale pour assimiler des colonnes d'ozone dans le modèle tridimensionnel de transport TM3.

Étude de la chimie troposphérique

En chimie troposphérique, plusieurs études cherchent à estimer la précipitation acide à l'aide d'un Krigeage [Finkelstein, 1984 ; Fedorov, 1989].

Les premiers essais pour estimer la pollution urbaine photochimique hors des zones couvertes par la mesure ont été réalisés en interpolant les observations avec des techniques mathématiques plus ou moins sophistiquées. Par exemple, Ionescu *et al.* [2000] utilisent une interpolation par des *splines plats*, qui est équivalent à un Krigeage Intrinsèque d'ordre 1 [Cressie, 1993], pour estimer les champs de concentration de dioxyde d'azote sur la ville de Paris. Le principal inconvénient de telles interpolations est que le résultat dépend beaucoup de la capacité du réseau à couvrir la zone étudiée et peut être incohérent avec la physique du problème. En effet loin des observations, aucune information n'est disponible pour guider l'interpolateur vers une valeur réaliste de l'état de l'atmosphère. Dans une étude analytique des concentrations d'acide nitrique, Fuentes [2001] utilise un modèle numérique pour déterminer les poids de Krigeage et pour pallier au manque de données d'observation.

Elbern *et al.* [1997] montrent, par une suite d'expériences *jumelles* idéalisées, le bénéfice apporté par l'assimilation d'observations d'ozone synthétiques dans la représentation des concentrations d'ozone et des oxydes d'azotes (NO, NO₂, HNO₂, HNO₄, PAN), des radicaux (OH, HO₂), et du formaldéhyde (HCHO). Ces expériences sont effectuées avec le mécanisme chimique Regional Acid Deposition Model 2 (RADM2), associé à un 4D-VAR. Ils montrent également les limites de la méthode d'assimilation lorsque des observations d'ozone sont trop espacées dans le temps (5 heures), la qualité des analyses des espèces non observées est alors complètement contrôlée par celle de l'ébauche. Elbern *et Schmidt* [1999] étendent l'algorithme d'assimilation variationnel précédent à un modèle de chimie-transport incluant les opérateurs d'advection horizontale et verticale, de diffusion verticale, et le module chimique RADM2. Ils montrent que l'assimilation d'observation d'ozone disponibles toutes les demi-heures à *chaque point de grille du modèle* améliore la représentation des champs des oxydes d'azote NO_x. Ils montrent aussi que l'assimilation supplémentaire d'observations de dioxyde d'azote améliore l'estimation du monoxyde d'azote mais pas systématiquement celle des Composés Organiques Volatiles. La prise en compte d'une contrainte sur la composition relative des COV améliore les résultats [Elbern *et al.*, 2000].

Elbern *et Schmidt* [2001] appliquent pour la première fois la méthode de 4D-VAR à des observations réelles d'ozone et d'oxydes d'azote. Ils assimilent des observations d'ozone sur la fenêtre d'assimilation [6 h UTC, 12 h UTC], et quantifient l'impact de cette assimilation sur les prévisions à deux jours du modèle EURAD (EUROpean Air Pollution Dispersion). Ils améliorent les prévisions jusqu'à 12 h après la fin de l'assimilation inculant l'heure du pic du jour même.

van Loon *et Heemink* [1997] présentent un Filtre de Kalman Étendu réduit pour assimiler des observations dans un modèle simple de chimie atmosphérique. van Loon *et al.* [2000] couplent le modèle de chimie-transport LOTOS à un Filtre de Kalman d'ensemble. Ils montrent que l'assimilation de concentration d'ozone avec une perturbation des émissions de NO_x, SO_x, VOC, et CO améliore les prévisions et permettent d'indiquer des corrections de facteurs d'émissions. Ils suggèrent qu'il est nécessaire de perturber d'autres paramètres que seulement les émissions : dépôt, diffusion verticale, données météorologiques.

1.8 Choix d'une méthode d'assimilation de données

Comme le modèle CHIMERE a pour objectif d'être un modèle de prévision opérationnelle et d'offrir la possibilité de faire des simulations longues, les efforts sont maintenus pour que les temps de calculs restent faibles. Or les méthodes d'assimilation de données ont précisément des coûts numériques très variables.

1.8.1 Coûts numériques des méthodes d'assimilation de données

La dimension de l'état \mathbf{Z}^b du modèle CHIMERE continental est de $n = [65 \times 33 \text{ points de grille horizontaux}] \times [5 \text{ niveaux verticaux}] \times [44 \text{ espèces}] \sim 5.10^5$. La dimension du vecteur d'observation \mathbf{Z}^o est de $p \sim 2.10^2$ (sachant que seules les observations d'ozone sont disponibles). Le problème est donc très largement sous déterminé (bien que par endroit il peut être surdéterminé quand par exemple la densité d'observations est grande devant la résolution du modèle). La méthode d'Interpolation Statistique nécessite en principe la connaissance des matrices \mathbf{B} et \mathbf{R} qui contiennent respectivement $n^2/2$ et $p^2/2$ coefficients distincts (matrices symétriques). La détermination de \mathbf{K} nécessite l'inversion d'une matrice de taille $p \times p$. Dans le cas du 3D-VAR, la minimisation exacte de la fonction de coût J nécessite en principe $n + 1$ évaluations de la fonction de coût et de son gradient en supposant que J est quadratique et qu'il n'y a pas d'erreurs numériques.

Le coût numérique du Filtre de Kalman est celui de l'Interpolation Statistique, plus celui nécessaire à l'estimation des covariances de l'erreur d'analyse, plus celui de l'équation de propagation de la matrice d'erreur de prévision. Cette dernière opération nécessite $2n$ prévisions du modèle. De plus les coûts de stockage sont très grands car chaque matrice \mathbf{P}^f , qui est une matrice de dimension $n \times n$ (même si la moitié suffit car les matrices sont symétriques), comme les matrices \mathbf{KH} , doivent être conservées.

Dans le 4D-VAR, le plus coûteux numériquement est l'évaluation de la fonction de coût d'observation (la fonction de coût de l'ébauche est identique à celle du 3D-VAR). *A priori*, cette évaluation nécessite n intégrations du modèle de prévision depuis l'analyse à tous les temps t_i d'observation et même plus pour le calcul du gradient. Cependant, cet algorithme peut être arrangé de façon à ce qu'il soit plus efficace en temps de calcul. En effet l'évaluation de la fonction de coût d'observation et de son gradient nécessite en fait une seule intégration directe du modèle de t_0 à t_n et une *intégration adjointe*. Cette dernière intégration consiste à l'intégration des transposés des modèles linéaires tangents des modèles adjoints.

Finalement, le 4D-VAR peut être implémenté plus facilement dans un système de prévision en temps réel qu'un Filtre de Kalman, beaucoup plus coûteux en temps de calcul. D'ailleurs, lorsqu'un Filtre de Kalman est utilisé en prévision, il est toujours simplifié. Par exemple, l'équation de propagation de l'erreur de prévision a longtemps été remplacée par une simple multiplication des variances d'erreur d'analyse par un coefficient [Talagrand, 1997]. En météorologie, beaucoup d'études traitent de la simplification de cette équation [Cohn et Parrish, 1991; Dee, 1991; Bouttier, 1994]. Dans le domaine de la qualité de l'air, Chenevez et Jensen [2001] négligent totalement les termes de corrélation de l'erreur de prévision dans le système opérationnel de prévision de la pollution photochimique DACFOS.

1.8.2 Différences majeures entre les méthodes d'assimilation de données

Mis à part les problèmes numériques, les algorithmes d'assimilation de données montrent des avantages et des inconvénients. Par exemple, par rapport à un 3D-VAR ou à un Filtre de Kalman, l'inconvénient majeur du 4D-VAR est que le modèle est supposé être parfait (ce qui consiste à prendre $\mathbf{Q} = \mathbf{0}$ dans l'algorithme du KF). De plus, en temps réel, il est nécessaire d'attendre que toutes les observations soient arrivées dans la fenêtre d'assimilation pour pouvoir les assimiler. Avec les autres algorithmes, les observations peuvent être intégrées très rapidement après leur arrivée.

En revanche, si l'état analysé déterminé par la méthode du 4D-VAR est utilisé comme état initial pour une prévision, alors la prévision sera complètement cohérente avec les équations du modèle et la distribution quadridimensionnelle des observations jusqu'à la fin de la fenêtre d'assimilation.

1.8.3 Conclusions

Les objectifs du travail, qui sont principalement la représentation des champs de polluants et la prévision ont bien sûr guidé le choix de la méthode d'assimilation de données à développer. Ce choix s'est décidé sur plusieurs critères : (1) mettre en place le premier système d'assimilation de données avec CHIMERE (2) respecter le coût numérique pour faire de la prévision en temps réel avec assimilation de données.

Ainsi, nous avons choisi de mettre en place un système d'assimilation de données *rapide* utilisant des techniques telles que l'Interpolation Statistique ou le Krigeage. Nous avons cherché à développer des méthodes permettant de construire en un temps raisonnable des champs de polluants plus réalistes que les simulations du modèle CHIMERE. L'idée était, bien sûr, d'utiliser ces champs de polluant pour améliorer les prévisions du modèle. De telles cartes sont maintenant produites en temps quasi-réel (chaque jour pour la veille) et sont très utiles pour diagnostiquer rapidement les erreurs commises par le modèle.

1.9 Plan de la thèse

Ce travail est organisé en sept chapitres et quatre annexes. Le premier chapitre consiste en une introduction générale de la problématique actuelle de la pollution de l'air. (L'annexe A donne quelques éléments de compréhension de la formation d'épisodes de fortes concentrations d'ozone). Il apparaît que deux besoins majeurs restent à combler : la *cartographie* des champs de concentration de polluants atmosphériques et la prévision. Ces sujets sont le fil conducteur de cette thèse. Dans ce premier chapitre, nous proposons également une description rapide des principales méthodes d'assimilation de données, qui sont actuellement utilisées pour produire des champs cohérents de données environnementales (données météorologiques, océanologiques, et géostatistiques en général) à partir de simulations et d'observations.

Le chapitre 2 est dédié à la présentation théorique détaillée des méthodes, sur lesquelles est basé notre travail : l'Interpolation Statistique et le Krigeage. Quelques rappels de statistique sont exposés dans l'Annexe B.

Le chapitre 3 concerne la description du modèle CHIMERE et de la base de données d'observation de deux polluants : l'ozone et le dioxyde d'azote. Les résultats d'une comparaison entre les sorties du modèle (dans ses versions continentale et régionale) et les observations y sont aussi présentées. L'annexe C donne la liste des organisations qui nous ont fourni les observations.

Le chapitre 4 expose la mise en place des méthodes exposées au chapitre 2 pour produire des champs de concentrations d'ozone sur l'Île-de-France. Ces champs sont produits soit à partir, seulement d'observations du polluant en surface, soit à partir de ces mêmes observations et des sorties du modèle régional. De nouvelles méthodes, adaptées au cas de la pollution atmosphérique, y sont présentées. Ce chapitre donne également les résultats d'une validation objective et d'une comparaison des différentes méthodes. Des exemples frappants mettent en évidence les points forts et les inconvénients de chaque méthode.

Le chapitre 5 présente l'application des méthodes d'assimilation à la prévision. Les analyses de champs de concentrations d'ozone produites à l'échelle européenne sont d'abord présentées puis ensuite utilisées pour initialiser le modèle de prévision. L'annexe D donne quelques exemples de cartes météorologiques nécessaires à la compréhension de cas de pollution.

Le chapitre 6 est dédié à la mise en œuvre de la méthodologie d'analyse pour évaluer la représentation des champs de concentration du dioxyde d'azote sur la région Île-de-France et, de l'ozone et du dioxyde d'azote, sur la région de Berre et de Marseille.

Enfin, le dernier chapitre est composé des conclusions et des perspectives de ce travail.

Chapitre 2

Krigeage Ordinaire, Interpolation Statistique, formulations mathématiques

Ce chapitre est dédié à la présentation mathématique des méthodes d'assimilation de données et de Krigeage sur lesquelles ces travaux sont basés : l'Interpolation Statistique et le Krigeage Intrinsèque. Les concepts présentés étant parfois difficiles à appréhender, nous proposons d'étudier tout d'abord un exemple très simple d'estimation statistique, avant de décrire des méthodes plus élaborées. Quelques notions statistiques sont rappelées dans l'Annexe B.

2.1 Une illustration simple d'estimation statistique

Nous cherchons à estimer la concentration Z^t d'un gaz dans l'atmosphère. Supposons que deux mesures de cette quantité Y_1^o et Y_2^o sont effectuées au même instant et au même endroit par deux instruments différents. Si nous considérons que les deux mesures sont de même qualité, il est naturel de prendre comme estimation de Z^t , notée Z^a , la moyenne de Y_1^o et Y_2^o . En revanche, si nous savons qu'un instrument est plus performant qu'un autre, il devient plus logique que la mesure faite par l'instrument le plus fiable soit favorisée dans l'estimation. Il est alors préférable de prendre une combinaison linéaire des deux mesures avec des poids différents.

$$Z^a = w_1 Y_1^o + w_2 Y_2^o \quad (2.1)$$

L'objectif est alors de déterminer w_1 et w_2 pour que l'estimation Z^a soit la meilleure possible. Supposons dans un premier temps que les mesures ne contiennent pas d'erreur systématique, c'est-à-dire que si nous répétons les mesures plusieurs fois, la moyenne des mesures est égale à la moyenne de la vraie valeur Z^t . En termes mathématiques, cela s'écrit $E(Y_1^o - Z^t) = E(Y_2^o - Z^t) = 0$, et nous disons que les mesures ne sont pas *biaisées*. $E(\cdot)$ est l'opérateur d'espérance mathématique, défini en Annexe B. De même, nous ne voulons pas que Z^a soit biaisée, c'est-à-dire que nous cherchons une estimation Z^a telle que $E(Z^a - Z^t) = 0$.

Or,

$$Z^a - Z^t = w_1 Y_1^o + w_2 Y_2^o - Z^t = w_1(Y_1^o - Z^t) + w_2(Y_2^o - Z^t) + Z^t(w_1 + w_2 - 1) \quad (2.2)$$

En prenant l'espérance de l'équation précédente et en utilisant l'hypothèse que les mesures ne sont pas biaisées $E(Y_1^o - Z^t) = E(Y_2^o - Z^t) = 0$, l'espérance de l'erreur d'estimation $Z^a - Z^t$ s'écrit :

$$E(Z^a - Z^t) = E(Z^t)(w_1 + w_2 - 1) \quad (2.3)$$

Ainsi l'estimation est une estimation linéaire non biaisée si la somme des poids est égale à 1. Introduisons maintenant $\epsilon_1 = Y_1^o - Z^t$, $\epsilon_2 = Y_2^o - Z^t$, les erreurs aléatoires commises respectivement sur la mesure 1 et sur la mesure 2. En considérant que la somme des poids est égale à 1 dans l'équation 2.2, l'erreur d'estimation $\epsilon_a = Z^a - Z^t$ s'écrit :

$$\epsilon_a = w_1 \epsilon_1 + w_2 \epsilon_2 \quad (2.4)$$

Dans le cas non biaisé, Z^a est la meilleure estimation possible si la variance de son erreur $E(\epsilon_a^2)$ est minimale. Cette variance s'écrit :

$$E(\epsilon_a^2) = E[(w_1 \epsilon_1 + w_2 \epsilon_2)^2] \quad (2.5)$$

$$= w_1^2 E(\epsilon_1^2) + w_2^2 E(\epsilon_2^2) + 2w_1 w_2 E(\epsilon_1 \epsilon_2) \quad (2.6)$$

Posons $\sigma_a^2 = E(\epsilon_a^2)$, $\sigma_1^2 = E(\epsilon_1^2)$ et $\sigma_2^2 = E(\epsilon_2^2)$ respectivement les variances d'erreur de Z^a , Y_1^o et Y_2^o . Les dernières sont normalement données avec les caractéristiques des instruments. Supposons que les erreurs d'observations sont décorréliées, ce qui est raisonnable si les mesures sont effectuées avec deux instruments différents. Alors :

$$E(\epsilon_1 \epsilon_2) = 0 \quad (2.7)$$

et la variance de l'estimation s'écrit :

$$\sigma_a^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 \quad (2.8)$$

Pour minimiser cette variance, en respectant la contrainte qui est que la somme des poids soit égale à 1, nous utilisons la méthode des multiplicateurs de Lagrange, qui consiste à construire une fonction J , et à minimiser cette fonction par rapport aux poids w_1 et w_2 .

$$J = \sigma_a^2 + \mu(1 - (w_1 + w_2)) \quad (2.9)$$

$$= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + \mu(1 - (w_1 + w_2)) \quad (2.10)$$

En différenciant J et en annulant les dérivées partielles, nous obtenons le système de deux équations :

$$\frac{\partial J}{\partial w_1} = 2w_1 \sigma_1^2 - \mu = 0 \quad (2.11)$$

$$\frac{\partial J}{\partial w_2} = 2w_2 \sigma_2^2 - \mu = 0 \quad (2.12)$$

d'où

$$w_1 = \frac{\mu}{2\sigma_1^2}, w_2 = \frac{\mu}{2\sigma_2^2} \quad (2.13)$$

La somme des poids doit être égale à 1, donc

$$\mu = \frac{2}{\sigma_1^{-2} + \sigma_2^{-2}} \quad (2.14)$$

et

$$w_1 = \frac{\sigma_1^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}}, w_2 = \frac{\sigma_2^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}} \quad (2.15)$$

L'estimation s'écrit finalement :

$$Z^a = \frac{\sigma_1^{-2}Y_1^o}{\sigma_1^{-2} + \sigma_2^{-2}} + \frac{\sigma_2^{-2}Y_2^o}{\sigma_1^{-2} + \sigma_2^{-2}} \quad (2.16)$$

Nous constatons que plus σ_1 est faible (i.e. σ_1^{-2} est fort), plus l'estimation Z^a sera proche de Y_1^o . A la limite, dans le cas de l'observation parfaite, si σ_1 est nulle, l'estimation Z^a sera strictement égale à Y_1^o . Ce qui est bien ce que nous recherchons : plus une observation sera fiable (variance d'erreur faible), plus l'analyse sera proche de cette observation. Introduisons les poids ainsi calculés dans l'équation 2.8, la variance d'erreur d'estimation s'écrit :

$$\sigma_a^2 = \frac{1}{\sigma_1^{-2} + \sigma_2^{-2}} \quad (2.17)$$

Ces résultats sont généralisables à plus de deux mesures. Pour N mesures, nous avons :

$$\sigma_a^2 = \frac{1}{\sum_{n=1}^N \sigma_n^{-2}} \quad (2.18)$$

Il en ressort que toute nouvelle observation, *quelle que soit sa qualité*, va toujours *réduire l'erreur de l'estimation*. Considérons que la mesure Y_1^o a l'incertitude la plus petite, alors pour toute observation $Y_{i>1}^o$, $\sigma_i^2 \geq \sigma_1^2$, et

$$1/\sigma_a^2 = \sum_{n=1}^N \sigma_n^{-2} > \sigma_1^{-2} \quad (2.19)$$

Finalement,

$$\sigma_a^2 < \sigma_1^2 \quad (2.20)$$

La variance d'estimation est plus petite que la plus faible des variances des observations. L'estimation est statistiquement meilleure que chacune des observations prises individuellement. Cette remarque est capitale : elle justifie le travail effectué au cours de cette thèse. Le problème exposé ci-dessus est étendu dans la suite à un problème multidimensionnel. En effet, nous allons chercher à estimer une valeur d'une variable en un point de l'espace, alors que celle-ci sera observée en d'autres points.

2.2 Le Krigeage

Estimer, à un instant donné un champ d'une variable, représentée par un vecteur \mathbf{Z}^t , à partir d'un jeu d'observations est un problème classique traité par les méthodes de Krigeage. La difficulté supplémentaire par rapport à l'exemple donné dans le paragraphe 2.1, est que nous cherchons maintenant à estimer une variable $Z^t(\mathbf{s}_a)$ en un point \mathbf{s}_a de l'espace alors que celle-ci est *a priori* observée en d'autres points \mathbf{s}_k .

2.2.1 Principe du Krigeage

Considérons un champ bidimensionnel d'une concentration d'un gaz dans l'atmosphère sur une région donnée. Nous disposons de p observations de \mathbf{Z}^t , distribuées sur cette région, que nous regroupons dans un vecteur \mathbf{Y}^o . Le Krigeage consiste à estimer \mathbf{Z}^t , en n'importe quel point de la région d'étude, par une combinaison linéaire des observations. L'expression générale de cette estimation en un point \mathbf{s}_a est notée :

$$Z^a(\mathbf{s}_a) = \sum_{k=1}^p w_k(\mathbf{s}_a) Y^o(\mathbf{s}_k) \quad (2.21)$$

où $w_k(\mathbf{s}_a)$ est le poids de l'observation $Y^o(\mathbf{s}_k)$, positionnée en \mathbf{s}_k , dans l'estimation de \mathbf{Z}^t au point analysé \mathbf{s}_a .

2.2.2 Expression des poids et de l'erreur de Krigeage

Supposons que la moyenne de \mathbf{Z}^t , $E(\mathbf{Z}^t)$, est uniforme et égale à m . Autrement dit, nous recherchons un champ de concentration uniforme en moyenne sur la région étudiée, mais qui prend des valeurs aléatoires en chaque point de cette région. Supposons également que les observations ne sont pas biaisées, i.e. $E(Y^o(\mathbf{s}_k)) = E(Z^t(\mathbf{s}_k)) = m$. La moyenne de l'erreur d'estimation $\epsilon_a(\mathbf{s}_a)$ s'écrit :

$$E[\epsilon^a(\mathbf{s}_a)] = E[Z^a(\mathbf{s}_a)] - E[Z^t(\mathbf{s}_a)] = \sum_{k=1}^p w_k(\mathbf{s}_a) E[Y^o(\mathbf{s}_k)] - E[Z^t(\mathbf{s}_a)] \quad (2.22)$$

$$= \left(\sum_{k=1}^p w_k(\mathbf{s}_a) - 1 \right) m \quad (2.23)$$

L'estimation $Z^a(\mathbf{s}_a)$ de $Z^t(\mathbf{s}_a)$ est alors non biaisée si la somme des poids est égale à 1. Nous allons donc chercher, comme dans le paragraphe 2.1, à minimiser en chaque point analysé, la variance de l'erreur d'estimation en contraignant la somme des poids à être égale à 1. Dans le cas d'une estimation non biaisée, cette variance d'erreur s'écrit :

$$\sigma_a^2(\mathbf{s}_a) = E\left[\sum_{k=1}^p w_k(\mathbf{s}_a) (Y^o(\mathbf{s}_k) - Z^t(\mathbf{s}_a)) \right]^2 \quad (2.24)$$

$$= \sum_{k=1}^p \sum_{l=1}^p w_l(\mathbf{s}_a) w_k(\mathbf{s}_a) E[(Y^o(\mathbf{s}_k) - Z^t(\mathbf{s}_a))(Y^o(\mathbf{s}_l) - Z^t(\mathbf{s}_a))] \quad (2.25)$$

L'équation 2.25 fait intervenir trois types de quantités : les covariances entre observations $E[(Y^o(\mathbf{s}_k) - m)(Y^o(\mathbf{s}_l) - m)]$, les covariances entre les observations et le point analysé $E[(Y^o(\mathbf{s}_k) - m)(Z^t(\mathbf{s}_a) - m)]$, et enfin la variance du champ réel $E[(Z^t(\mathbf{s}_a) - m)^2]$. En faisant intervenir les erreurs d'observation $\epsilon^o(\mathbf{s}_k)$, *non corrélées au champ réel Z^t* , ces termes se réécrivent de la façon suivante :

$$E[(Y^o(\mathbf{s}_k) - m)(Y^o(\mathbf{s}_l) - m)] = E[(Z^t(\mathbf{s}_k) - m)(Z^t(\mathbf{s}_l) - m)] + E[\epsilon^o(\mathbf{s}_k)\epsilon^o(\mathbf{s}_l)] \quad (2.26)$$

$$E[(Y^o(\mathbf{s}_k) - m)(Z^t(\mathbf{s}_a) - m)] = E[(Z^t(\mathbf{s}_k) - m)(Z^t(\mathbf{s}_a) - m)] \quad (2.27)$$

Nous définissons $C_{k,l} = E[(Z^t(\mathbf{s}_k) - m)(Z^t(\mathbf{s}_l) - m)]$ les *covariances du champ réel* entre deux points quelconques \mathbf{s}_k et \mathbf{s}_l , et $R_{k,l} = E[\epsilon^o(\mathbf{s}_k)\epsilon^o(\mathbf{s}_l)]$ les covariances entre les erreurs d'observation. Comme les observations sont effectuées avec des instruments différents, les erreurs d'observation ne sont pas corrélées entre elles, c'est à dire que $R_{k,l} = R_{k,k}\delta_{kl}$. En tenant compte de ces définitions, la variance d'erreur d'estimation devient :

$$\sigma_a^2(\mathbf{s}_a) = \sum_{k=1}^p \sum_{l=1}^p w_l(\mathbf{s}_a)w_k(\mathbf{s}_a)[C_{k,l} + R_{k,k}\delta_{kl} - C_{a,k} - C_{a,l} + C_{a,a}] \quad (2.28)$$

$$= C_{a,a} - 2 \sum_{k=1}^p w_k(\mathbf{s}_a)C_{a,k} + \sum_{k=1}^p \sum_{l=1}^p w_k(\mathbf{s}_a)w_l(\mathbf{s}_a)(C_{k,l} + R_{k,k}\delta_{kl}) \quad (2.29)$$

La fonction J à minimiser s'écrit

$$J(\mathbf{s}_a) = \sigma_a^2(\mathbf{s}_a) + 2\mu \left(\sum_{k=1}^p w_k(\mathbf{s}_a) - 1 \right) \quad (2.30)$$

Elle est minimisée par rapport aux p poids w_k et au paramètre μ . Cette minimisation conduit, en plus de la contrainte sur la somme des poids, au système linéaire suivant :

$$\forall k = 1, \dots, p, \quad \sum_{l=1}^p w_l(\mathbf{s}_a)[C_{k,l} + R_{k,k}\delta_{kl}] + \mu = C_{k,a} \quad (2.31)$$

Introduisons les poids optimisés dans l'expression de la variance d'erreur d'analyse, donnée par l'équation 2.29,

$$\sigma_a^2(\mathbf{s}_a) = C_{a,a} - \sum_{k=1}^p w_k(\mathbf{s}_a)C_{a,k} - \mu \quad (2.32)$$

Le système 2.31 montre que les poids ne peuvent être calculés que si les covariances du champ réel entre deux points quelconques sont connues. Or la difficulté est que nous n'avons précisément pas d'accès direct à ces covariances, nous devons donc *faire des hypothèses pour les modéliser*. Le paragraphe 2.5 est entièrement consacré à ce problème fondamental. En outre, l'équation 2.32 montre que la qualité de l'estimation des covariances détermine celle de l'analyse résultante.

2.3 Interpolation Statistique : utilisation d'un modèle de chimie-transport

Supposons maintenant que nous disposons aussi d'un modèle numérique capable de simuler les concentrations que nous voulons estimer. Le champ simulé est noté \mathbf{Z}^b (*b* de *background*). D'après les résultats obtenus dans le paragraphe 2.1, le modèle devrait, comme une observation supplémentaire, améliorer l'estimation de ce champ. Il est donc naturel de penser que les erreurs du modèle vont aussi intervenir dans l'estimation de \mathbf{Z}^t . Ainsi avant de décrire l'interpolation statistique, nous proposons de détailler les erreurs commises sur les simulations.

2.3.1 Les erreurs de modélisation et d'observation

Les erreurs de modélisation

Les erreurs de modélisation peuvent être liées à la conception du modèle (description imparfaite des processus), à la mise en œuvre pratique (calage de paramètres, résolution ...) ou à des erreurs dans les paramètres d'entrée (émissions, données météorologiques). Posons $\epsilon_b = \mathbf{Z}^b - \mathbf{Z}^t$ l'erreur du modèle.

Les erreurs d'observations : erreurs instrumentales et erreurs de représentativité

De manière générale, un modèle numérique est limité par sa résolution. Si une observation est contrôlée par des processus de trop petite échelle par rapport à la maille du modèle, le modèle ne les simulera pas ou mal. Ces erreurs, dites *erreurs de représentativité* sont, en pratique, toujours considérées comme des erreurs d'observation dans les algorithmes d'assimilation de données. Ce transfert d'erreur évite que l'analyse ne se rapproche d'une information trop localisée, mais qu'elle tende vers l'*état réel moyen par maille du modèle*. Les erreurs commises sur les observations sont donc de deux types. Elles incluent les erreurs instrumentales, liées aux caractéristiques des instruments de mesure utilisés, et les erreurs de représentativité.

2.3.2 Principe de l'Interpolation Statistique

L'objectif est de corriger chacune des valeurs simulées par le modèle en interpolant les différences constatées entre les observations et les simulations et en ajoutant le champ de correction à la simulation initiale. L'expression générale de la valeur du champ analysé final en un point \mathbf{s}_a est :

$$Z^a(\mathbf{s}_a) = Z^b(\mathbf{s}_a) + \sum_{k=1}^p w_k(\mathbf{s}_a)(Y^o(\mathbf{s}_k) - Z^b(\mathbf{s}_k)) \quad (2.33)$$

où $w_k(\mathbf{s}_a)$ est le poids de l'observation $Y^o(\mathbf{s}_k)$, positionnée au point \mathbf{s}_k , dans l'estimation de \mathbf{Z}^t au point analysé \mathbf{s}_a . Nous ne tenons pas compte du fait que les points de mesure ne correspondent que très rarement à un point de grille du modèle, et qu'il est ainsi nécessaire pour comparer les simulations avec les observations de les interpoler au points de mesures. Nous supposons ici que cette opération a déjà été effectuée. Ainsi $Z^b(\mathbf{s}_k)$ est une valeur simulée interpolée au point de mesure \mathbf{s}_k . Cela revient à définir de nouveaux points de grille aux points d'observation.

2.3.3 Interpolation Statistique, Krigeage des innovations

L'équation générale de l'Interpolation Statistique 2.33 peut être regardée comme une équation de Krigeage, dans laquelle la variable à estimer est le champ d'erreur du modèle $\mathbf{X}^t = \mathbf{Z}^t - \mathbf{Z}^b = -\epsilon^b$. Les "observations" $X^o(\mathbf{s}_k)$ de ce champ sont les différences $Y^o(\mathbf{s}_k) - Z^b(\mathbf{s}_k)$ entre les observations réelles et les simulations prises aux points de mesures \mathbf{s}_k . Elles sont appelées les *innovations*. L'estimation, $X^a(\mathbf{s}_a) = Z^a(\mathbf{s}_a) - Z^b(\mathbf{s}_a)$, de \mathbf{X}^t s'écrit alors :

$$X^a(\mathbf{s}_a) = \sum_{k=1}^p w_k(\mathbf{s}_a) X^o(\mathbf{s}_k) \quad (2.34)$$

Supposons temporairement, que ni les observations, ni les simulations ne sont biaisées, c'est-à-dire que $E[Z^b(\mathbf{s}_a) - Z^t(\mathbf{s}_a)] = E[Y^o(\mathbf{s}_k) - Z^t(\mathbf{s}_k)] = 0$. Cette hypothèse équivaut à supposer que $E[X^o(\mathbf{s}_k)] = 0$, l'estimation \mathbf{X}^a n'est alors pas biaisée. Il n'est donc plus nécessaire d'imposer à la somme des poids d'être égale à 1. Les équations du Krigeage 2.31 sont, dans ce cas, valables en prenant $\mu = 0$. A condition que \mathbf{X}^t soit indépendant de "l'erreur d'observation" commise sur \mathbf{X}^o , c'est à dire que les erreurs d'observations et les erreurs du modèle soient décorréelées (ce qui est le cas car les observations et les simulations sont des estimations indépendantes), les poids sont solutions du système suivant :

$$\forall k = 1, \dots, p, \quad \sum_{l=1}^p w_l(\mathbf{s}_a) [B_{k,l} + R_{k,k} \delta_{kl}] = B_{k,a} \quad (2.35)$$

avec

$$B_{k,l} = E[(X^t(\mathbf{s}_k) X^t(\mathbf{s}_l))] = E[\epsilon^b(\mathbf{s}_k) \epsilon^b(\mathbf{s}_l)] \quad (2.36)$$

Les $B_{k,l}$ sont les covariances des erreurs du modèle, qui ne sont *a priori* pas nuls. En effet, la variabilité spatiale de l'erreur du modèle n'est pas purement aléatoire : si le modèle commet une erreur due à une erreur de conception ou si il y a une erreur dans les entrées du modèle, il y a une forte probabilité pour que celle-ci se répète sur plusieurs points de grille. Plus deux points de grille sont rapprochés l'un de l'autre, plus les erreurs commises par le modèle en ces deux points ont des chances d'être proches (il faut cependant rester prudent sur ce genre de considération : nous montrons précisément dans ce document que le rapprochement n'est pas toujours un bon critère dans l'étude de la chimie atmosphérique). L'important est de retenir qu'il y a une corrélation dans la distribution spatiale des erreurs du modèle. C'est-à-dire que les erreurs ont une certaine *structure de corrélation spatiale*. L'équation 2.35 est, à l'interpolateur \mathbf{H} près, l'équation de l'Interpolation Statistique 1.2 qui définit la matrice de gain introduite au chapitre 1. La variance de l'erreur d'estimation s'écrit :

$$\sigma_a^2(\mathbf{s}_a) = B_{a,a} - \sum_{k=1}^p w_k(\mathbf{s}_a) B_{a,k} \quad (2.37)$$

Cette équation donne les éléments diagonaux de la matrice \mathbf{A} de l'équation 1.3. Nous appellerons par abus de langage, *erreur d'analyse*, la racine carré de cette variance (i.e. son écart type). A chaque analyse, nous pourrions donc associer une carte d'erreur d'analyse.

2.4 Nécessité d'une modélisation des biais et des covariances

Les méthodes décrites ci-dessus montrent que la détermination des covariances (du champ réel ou du champ des erreurs du modèle) est primordiale puisque ces covariances sont à la base du calcul des poids de l'interpolation. Si les covariances ne sont pas bien estimées, les poids ne sont pas optimaux, et l'erreur d'analyse, minimisée, est systématiquement sous-estimée. Par ailleurs, si les observations ou les simulations sont biaisées, les analyses résultantes le sont aussi.

2.4.1 Problème des biais

Nous avons supposé jusqu'à maintenant que ni les observations ni les simulations ne sont biaisées. Si c'est le cas, il faut retirer aux observations et au modèle leur biais respectif avant d'interpoler les valeurs. Cependant comment estimer ces biais? Les seules informations disponibles, qui puissent apporter de l'information sur les erreurs, et en particulier sur les biais, sont les innovations. Leur moyenne s'écrit :

$$E[Y^o(\mathbf{s}_k) - Z^b(\mathbf{s}_k)] = E[\epsilon^o(\mathbf{s}_k)] - E[\epsilon^b(\mathbf{s}_k)] \quad (2.38)$$

Des innovations biaisées $E[Y^o(\mathbf{s}_k) - Z^b(\mathbf{s}_k)] \neq 0$ sont le signe d'un biais dans les observations ou le modèle ($E[\epsilon^o(\mathbf{s}_k)] \neq 0$ ou $E[\epsilon^b(\mathbf{s}_k)] \neq 0$). Mais comment faire la part entre l'erreur d'observation et l'erreur due au modèle?

Si les innovations ne sont pas biaisées, cela peut vouloir dire que ni les observations, ni le modèle ne sont biaisés, mais aussi qu'il y a compensation entre le biais du modèle et le biais opposé des observations. Le problème n'est pas simple à régler et est rarement abordé dans la littérature. Dans le domaine de la qualité de l'air, ces biais sont en partie dus à des problèmes de représentativité, et ne sont pas négligeables. *Elbern et al.* [2001] proposent de prendre un biais de représentativité uniforme sur toutes les stations urbaines et égal à 10% de la valeur mesurée. Nous proposons dans le chapitre 4 une modélisation du biais de modélisation en fonction de la concentration moyenne simulée, le biais résiduel étant pris comme le biais d'observation.

2.4.2 Les covariances d'erreur

La variance des erreurs instrumentales est normalement donnée par le constructeur de l'instrument. En revanche, nous n'avons aucun accès direct aux erreurs de représentativité, et aux erreurs du modèle.

Les covariances d'erreur d'observation

Les covariances d'erreur instrumentale d'observation sont souvent supposées nulles, dans les algorithmes d'assimilation de données. Ceci est justifié si les mesures sont effectuées avec des instruments différents. Ce n'est pas le cas, quand les observations proviennent par exemple de mesures aéroportées ou de mesures satellitaires.

Dans le cas d'un réseau d'observation dense, les erreurs de représentativité ont, quant à elles, de fortes chances d'être corrélées puisqu'elles sont liées à la résolution du modèle. La présence de covariances d'erreur d'observation non négligeables va avoir pour effet de diminuer les poids en moyenne, et de favoriser les valeurs observées les plus corrélées. En effet, supposons que nous

cherchons à estimer une variable en un point et que nous disposons de trois observations effectuées en trois autres points équidistant du point d'analyse. Deux points de mesures sont proches l'un de l'autre, tandis que le dernier est à l'opposé par rapport au point d'analyse. Il est logique d'accorder plus de poids à la mesure isolée, qu'à chacune des deux mesures rapprochées spatialement (probablement plus corrélées), qui peuvent apporter deux fois la même information. En pratique, il est plus facile de minimiser ces covariances d'erreur que de les estimer. C'est pourquoi, les observations les plus représentatives sont, souvent, les plus utilisées. Tandis que les observations, les moins représentatives sont soit rejetées, soit regroupées dans une "super-observation" si elles donnent une information sur les concentrations dans la même maille du modèle [Daley, 1991].

Si les erreurs d'observation sont décorrélées, alors les variances d'erreur d'observation et du modèle peuvent être simplement déterminées par la méthode de [Hollingsworth et Lönnberg, 1986] décrite ci-dessous. Le rapport de la variance des erreurs du modèle sur celle des erreurs d'observation contrôle la distance entre les observations et l'analyse. Si ces variances ne sont pas bien spécifiées, les corrections des simulations seront soit trop petites soit trop fortes.

Les covariances d'erreur du modèle

Les covariances d'erreur du modèle sont essentielles pour plusieurs raisons. D'abord, si les observations sont peu nombreuses, ce sont ces termes qui déterminent comment la simulation loin des points de mesure va être modifiée par l'introduction des observations. Autrement dit, ces termes définissent l'influence des observations dans la correction des simulations. Si le réseau d'observation est dense, suivant la structure des covariances, l'analyse va être plus ou moins lissée en fonction de l'influence qu'on accorde à une observation. Enfin ces covariances apportent dans l'algorithme d'Interpolation Statistique de l'information sur la physique du problème. Si les erreurs d'observation sont décorrélées, les covariances d'erreur du modèle peuvent être également déterminées par la méthode de [Hollingsworth et Lönnberg, 1986].

2.5 Modélisation des covariances

De manière générale, nous cherchons à modéliser $Cov(X^t(\mathbf{s}_k), X^t(\mathbf{s}_l))$, les covariances du champ réel (les covariances $C_{k,l}$ du champ de concentration ou celle du champ des erreurs du modèle $B_{k,l}$) entre n'importe quels points \mathbf{s}_k et \mathbf{s}_l . Pour cela, nous allons utiliser les seules informations, dont nous disposons, à savoir les observations ou les innovations, notées $X^o(\mathbf{s}_k)$. En effet, nous avons déjà montré que

$$Cov(X^o(\mathbf{s}_k), X^o(\mathbf{s}_l)) = Cov(X^t(\mathbf{s}_k), X^t(\mathbf{s}_l)) + R_{k,k} \delta_{k,l} \quad (2.39)$$

D'après cette équation, tant que les erreurs d'observations ne sont pas corrélées, pour tout $k \neq l$, $Cov(X^o(\mathbf{s}_k), X^o(\mathbf{s}_l))$ est une estimation directe de $Cov(X^t(\mathbf{s}_k), X^t(\mathbf{s}_l))$. Ainsi, pour tout couple de points de mesure distincts, les covariances $C_{k,l}$ ou $B_{k,l}$ des équations 2.31 et 2.35 peuvent être calculées directement. Il reste néanmoins à choisir correctement l'ensemble statistique de données sur lequel ces covariances vont être calculées (en théorie, il faudrait un "nombre infini" de données).

En revanche, dans les mêmes équations 2.31 et 2.35, les termes $B_{k,k}$ ou $C_{k,k}$ de variances du champ à estimer, comme les covariances $B_{a,k}$ ou $C_{a,k}$ entre les points de mesure et des points non systématiquement observés, restent indéterminés.

Nous sommes donc face à deux problèmes : (1) définir un ensemble sur lequel vont être calculées les covariances (2) modéliser les covariances pour pouvoir les déterminer *entre n'importe quels points du domaine étudié*.

2.5.1 Choix d'un ensemble statistique pour calculer les covariances

Pour calculer les covariances, il y a deux manières de procéder selon que les covariances sont constantes ou non dans le temps. Si elles sont stationnaires, alors la covariance entre deux points peut être estimée par la *moyenne temporelle* des produits des "observations" (observations réelles ou innovations) à condition, bien sûr, d'avoir accumulé suffisamment de données (observations et/ou simulations) par le passé. Si les covariances ne sont pas constantes dans le temps, alors elles doivent être calculées uniquement sur les observations réalisées au temps de l'*analyse* à l'aide de *moyennes spatiales*.

2.5.2 Moyennes temporelles : utilisation d'une climatologie

Supposons dans un premier temps que les covariances sont stationnaires. Elles peuvent être alors approchées par une moyenne temporelle calculée sur une base de données, i.e. une climatologie. Notons C_{kl}^o les covariances évaluées à partir des données "observées".

$$Cov(X^o(\mathbf{s}_k), X^o(\mathbf{s}_l)) \sim C_{kl}^o = \frac{1}{N} \sum_{n=1}^N (X_n^o(\mathbf{s}_k) - m(\mathbf{s}_k))(X_n^o(\mathbf{s}_l) - m(\mathbf{s}_l)) \quad (2.40)$$

N étant le nombre de jours considérés et $m(\mathbf{s}_k) = \frac{1}{N} \sum_{n=1}^N X_n^o(\mathbf{s}_k)$. Pour pouvoir estimer les covariances $C_{k,l}$ en toutes positions \mathbf{s}_k et \mathbf{s}_l , une méthode classique consiste à construire une *fonction de covariance homogène et isotrope*.

Modélisation de covariances isotropes et homogènes

Avant de décrire la méthode, il convient de rappeler quelques définitions :

- ▷ **Condition d'homogénéité.** La fonction $C(\mathbf{s}_k, \mathbf{s}_l)$ est homogène si elle ne dépend que du déplacement relatif entre les vecteurs \mathbf{s}_k et \mathbf{s}_l et non des positions absolues \mathbf{s}_k et \mathbf{s}_l , i.e. $C(\mathbf{s}_k, \mathbf{s}_l) = C(\mathbf{s}_k - \mathbf{s}_l)$.
- ▷ **Condition d'isotropie.** Dans un repère centré sur une position \mathbf{s}_o , la fonction $C(\mathbf{s}_o, \mathbf{s}_l)$ est isotrope si elle ne dépend que du module de \mathbf{s}_l et non de sa direction. Cette fonction peut dépendre aussi de la position absolue de \mathbf{s}_o , dans ce cas, la fonction de covariance est isotrope mais pas homogène.
- ▷ **Conditions d'homogénéité et d'isotropie réunies.** La fonction $C(\mathbf{s}_k, \mathbf{s}_l)$ est homogène et isotrope si elle ne dépend que de la distance qui sépare \mathbf{s}_k et \mathbf{s}_l , i.e. $C(\mathbf{s}_k, \mathbf{s}_l) = f_B(h)$ en posant $h = |\mathbf{s}_k - \mathbf{s}_l|$.

Les conditions d'homogénéité et d'isotropie consistent à supposer que les variances sont homogènes, et que les covariances des erreurs du modèle ne dépendent que de la distance qui sépare

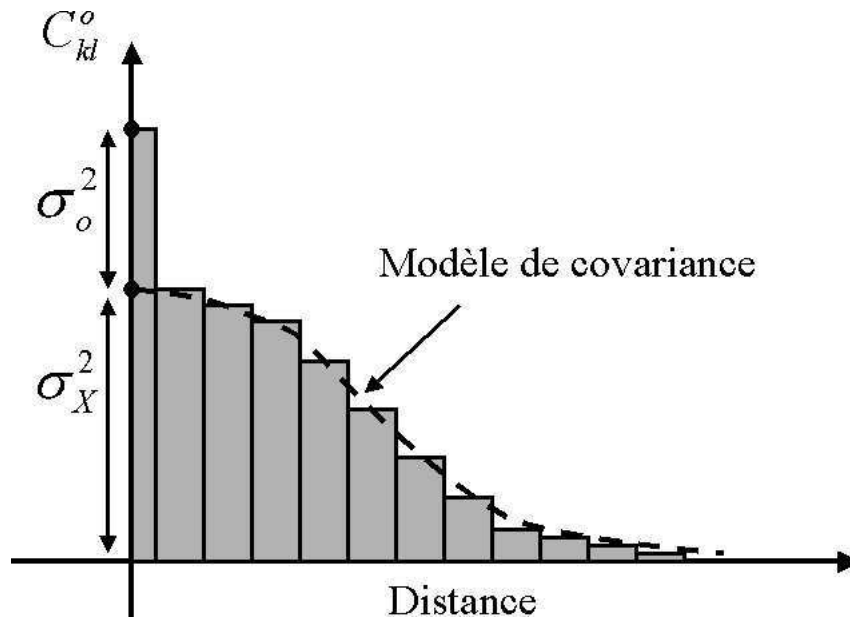


FIG. 2.1 – Représentation schématique d'un covariogramme. Courbe en pointillée : modèle de covariance C .

les positions entre lesquelles elles sont étudiées. Posons σ_o^2 la variance homogène de l'erreur d'observation et σ_X^2 celle de la variable X^t (dans le cas où X^t est l'erreur du modèle, σ_X^2 est la variance de l'erreur du modèle).

Pour déterminer le modèle de covariance C , le principe de la méthode consiste à tracer les covariances $C_{k,l}^o$ en fonction de la distance qui sépare les points entre lesquels elles sont calculées. Le graphe résultant est appelé *covariogramme*. La figure 2.1 montre une représentation schématique d'un tel covariogramme.

D'après les équations 2.39 et 2.40, tant que les observations sont décorréliées, c'est à dire tant qu'on considère deux observations distinctes, $C_{k,l}^o$ est une estimation de $C_{k,l}$. A la distance zéro, les erreurs d'observation contribuent à augmenter la variance et sont responsables de ce qu'on appelle couramment *l'effet pépité*. Moyennant l'hypothèse que les observations sont non corrélées, les erreurs d'observations n'apportent leur contribution qu'à cette distance nulle. Ainsi sans prendre en compte les variances, la variation des covariances des observations en fonction de la distance permet de définir la fonction de covariance C . $C(h)$ prend pour valeur σ_X^2 en $h = 0$, et nous avons une estimation de la variance de X^t .

En pratique, si les variances des observations ne sont pas tout à fait homogènes, il suffit de prendre leur moyenne. A partir de cette moyenne, qui est une estimation de $\sigma_X^2 + \sigma_o^2$, et de $\sigma_X^2 = C(0)$, il est possible de calculer la variance de l'erreur d'observation σ_o^2 .

Dans le cas homogène, les covariances $C_{k,l}^o$ sont généralement normalisées par la variance $\sigma_X^2 + \sigma_o^2$, c'est-à-dire qu'elles sont remplacées par les corrélations. L'algorithme n'est pas modifié, il suffit juste de noter qu'à la distance nulle, la fonction d'ajustement f_C prend la valeur :

$$f_C(0) = \frac{\sigma_X^2}{\sigma_o^2 + \sigma_X^2} \quad (2.41)$$

Elle n'est donc pas égale à 1 à cause des erreurs d'observation. La fonction de covariance C s'obtient finalement en normalisant f_C par $f_C(0)$. Elle s'écrit ainsi :

$$C(h) = \sigma_X^2 * \frac{f_C(h)}{f_C(0)} = [\sigma_o^2 + \sigma_X^2] f_C(h) \quad (2.42)$$

Dans le cas hétérogène, une telle séparation n'est possible que si la variance a une variabilité spatiale d'échelle caractéristique plus faible que celles des covariances. Si c'est le cas, les covariances ont alors la même forme que les corrélations. Ceci permet de *modéliser* séparément les corrélations et les variances.

Dans le cas où X^t est l'erreur du modèle, cette méthode correspond à la *méthode des observations*, couramment utilisée en météorologie [Hollingsworth et Lönnberg, 1986]. Plusieurs fonctions ont été utilisées par le passé en météorologie pour modéliser la fonction de corrélation des erreurs des modèles numériques [Buell, 1972; Julian et Thiebaut, 1975]. Ces fonctions doivent être telles que les matrices de corrélation d'erreur dérivées de ces fonctions soient définies positives ([Gaspari et Cohn, 1999], annexe B). La fonction isotrope et homogène la plus utilisée est une fonction décrite par [Balgovind et al., 1983] :

$$f_B(h) = \left(1 + \frac{h}{L}\right) \exp\left(-\frac{h}{L}\right) \quad (2.43)$$

où L est appelée *longueur de corrélation*.

Modèles de covariance d'erreur anisotropes et hétérogènes

Il existe aussi des modèles de covariances hétérogènes et anisotropes. En météorologie, Desrozières [1997] construit par exemple des fonctions de corrélation anisotropes pour analyser les champs de vent et de température près des fronts. La longueur de corrélation est plus petite le long du front, que dans la direction perpendiculaire. D'autres exemples de fonctions de covariances anisotropiques peuvent être trouvés dans [Thiebaut, 1976, 1977; Hollingsworth, 1987]. Riishøgaard [1998], qui étudie l'assimilation de colonnes d'ozone dans un modèle de transport, propose une fonction de covariance qui *dépend du champ simulé de la variable étudiée* \mathbf{Z}^b . Il définit ainsi un modèle hétérogène, anisotrope et, de plus, non stationnaire dans le temps :

$$B_{k,l} = \sigma_b^2(\mathbf{s}_k) \sigma_b^2(\mathbf{s}_l) \left[1 + \frac{1}{2} \left(\frac{|\mathbf{s}_k - \mathbf{s}_l|}{L} \right)^2 \right]^{-1} \exp \left[- \left(\frac{Z^b(\mathbf{s}_k) - Z^b(\mathbf{s}_l)}{T} \right)^2 \right] \quad (2.44)$$

Dans le domaine de la qualité de l'air, *Hoelzemann et al.* [2001] proposent une fonction basée sur une classification des sites en fonction du régime chimique :

$$B_{k,l} = \sigma_b^2(1 - \alpha|u_k - u_l|)f_B(|\mathbf{s}_k - \mathbf{s}_l|) \quad (2.45)$$

où f_B est la fonction de Balgovind et u_k sont des *indices du régime chimique*. Ces indices sont calculés à partir des profils de concentration d'ozone observés [*Elbern et al.*, 1996] et à partir des taux d'émission [*Memmesheimer et al.*, 1995]. Ils varient entre 0 et 1, $u = 0$ étant caractéristique d'un milieu rural et $u = 1$ d'un milieu urbain. De plus, ils limitent la longueur de corrélation en fonction du nombre d'observations qui entourent la station de mesure.

Nous proposons au chapitre 4 de modéliser la fonction de covariance en fonction de la covariance des concentrations simulées par le modèle, les covariances des innovations étant tracées en fonction de la covariances des concentrations observées. Ceci nous permet de conférer au modèle de covariance un caractère hétérogène et anisotrope que les observations mettent clairement en évidence. Cette modélisation nous permet de nous affranchir de toute classification difficile et laborieuse des sites de pollution. Elle nous permet aussi de prendre en compte de la présence de barrière dynamique. En effet, cette méthode est basé sur la simple intuition que plus des concentrations sont corrélées, plus les erreurs commises sur ces concentrations doivent être corrélées. La présence de sources d'émission (une voie rapide, une région de fortes émissions ...), de vallées et de massifs montagneux, qui pourraient éventuellement décorréler les erreurs commises sur deux régions, devrait être indiquée par la structure de corrélation du champ de concentrations des polluants.

2.5.3 Moyennes spatiales

Lorsque les covariances ne sont pas stationnaires dans le temps, ou tout simplement si aucune climatologie n'est disponible, seules les observations collectées au temps de l'analyse peuvent être utilisées pour calculer les covariances. L'*ensemble statistique* est alors fortement réduit, et il est nécessaire de faire des hypothèses sur le champ réel recherché.

Le Krigeage ordinaire

Le *Krigeage Ordinaire* est une méthode qui suppose que la variable à analyser $X^t(\mathbf{s})$ est une fonction aléatoire *stationnaire d'ordre 2*. Cela signifie que sa moyenne et sa covariance sont invariantes par translation dans l'espace :

$$\forall \mathbf{s}, E(X^t(\mathbf{s})) = m \quad (2.46)$$

$$\forall (\mathbf{s}, \mathbf{h}), Cov(X^t(\mathbf{s}), X^t(\mathbf{s} + \mathbf{h})) = C(\mathbf{h}) \quad (2.47)$$

La moyenne m , inconnue *a priori*, est juste supposée uniforme. Nous avons déjà fait cette hypothèse lorsque les équations du Krigeage ont été écrites. La contrainte sur la somme des poids, que nous imposons égale à 1, assure qu'avec une telle hypothèse le champ analysé ne sera pas biaisé. En pratique, si cette moyenne n'est pas uniforme, des méthodes plus sophistiquées peuvent être utilisées (Krigeage Universel)[*Chauvet*, 2000].

Modélisation des covariances : le Covariogramme

La covariance $C_{k,l}^o$ est la moyenne des produits des observations, calculée entre tous les points de mesure séparés par le *même vecteur* $\mathbf{s}_k - \mathbf{s}_l$. Si la fonction de covariance recherchée C est supposée isotrope, la covariance $C_{k,l}^o$ est la moyenne des produits des observations, calculée entre tous les points de mesure séparés par la *même distance* $|\mathbf{s}_k - \mathbf{s}_l|$.

Plaçons nous dans le cas d'une fonction de covariance isotrope. Des classes de distances h_i sont alors définies, et les covariances d'observation C_i^o sont calculées sur l'ensemble des stations k et l séparées par la distance $|\mathbf{s}_k - \mathbf{s}_l| = h_i$. Comme les observations sont souvent trop peu nombreuses dans chaque classe pour obtenir une estimation suffisamment correcte de ces covariances, une tolérance sur les distances est accordée.

Soit N_j la classe de distance j , nous écrivons :

$$N_j = \left\{ (s_k, s_l); \left| |s_k - s_l| - (2j - 1)\frac{L}{2} \right| < \frac{L}{2} + \tau \right\} \quad (2.48)$$

Cette classe regroupe l'ensemble des couples d'observations tels que la distance qui les sépare est comprise dans l'intervalle $[(j - 1)L, jL]$, à plus ou moins τ , la tolérance, près. A chaque classe est associée une distance moyenne,

$$h_j = \frac{1}{|N_j|} \sum_{N_j} |s_k - s_l| \quad (2.49)$$

et une covariance, dite *expérimentale*,

$$C^o(h_j) = \frac{1}{|N_j|} \sum_{N_j} (Y^o(s_k) - m(s_k))(Y^o(s_l) - m(s_l)) \quad (2.50)$$

où $|N_j|$ est le nombre d'éléments de la classe, et $m(s_k) = \frac{1}{|N_j|} \sum_{N_j} Y^o(s_k)$. De la même manière que dans le paragraphe 2.5.2, un covariogramme peut être construit à partir des valeurs h_j et $C^o(h_j)$. Celui-ci est ensuite utilisé pour estimer les covariances $C_{k,l}$ entre n'importe quel couple de points (s_k, s_l) , nécessaires au calcul des poids de l'interpolation.

2.6 Quelques exemples simples d'Interpolations Statistiques

Dans ce paragraphe, nous nous intéressons à quelques exemples simples d'interpolation statistique dans le cas unidimensionnel.

2.6.1 Cas d'une seule observation

Supposons que nous cherchons à estimer une concentration Z^t en un point x_a , à partir d'une observation non biaisée Y^o positionnée en x_0 , de variance d'erreur σ_o^2 , et d'une simulation non biaisée Z^p , de variance d'erreur σ_b^2 . D'après les équations développées précédemment, nous estimons Z^t de la meilleure manière par $Z^a(x_a) = Z^b(x_a) + w(Y^o(x_0) - Z^b(x_0))$ en prenant :

$$w = \frac{B_{0a}}{\sigma_o^2 + \sigma_b^2} \quad (2.51)$$

où B_{0a} est bien sûr la covariance des erreurs du modèle entre le point d'analyse x_a et le point de mesure x_0 . La variance d'erreur de l'estimation s'écrit :

$$\sigma_a^2 = B_{0a} - \frac{B_{0a}^2}{\sigma_o^2 + \sigma_b^2} \quad (2.52)$$

Dans le cas où $x_a = x_0$, l'estimation s'écrit $Z^a = Z^b + w(Y^o - Z^b)$ et $B_{0a} = \sigma_b^2$. Si l'observation est parfaite, alors $\sigma_o^2 = 0$, le poids w est égal à 1, et l'estimation est égale à l'observation ($Z^a(x_a) = Y^o(x_a)$). Au contraire si la simulation est parfaite, le poids est nul, et l'estimation est égale à la valeur simulée ($Z^a(x_a) = Z^b(x_a)$). Dans les deux cas, l'erreur d'estimation est nulle.

2.6.2 Cas de deux observations

Supposons maintenant que nous cherchons à estimer une concentration Z^t au point x_a à partir de deux observations non biaisées, Y_1^o et Y_2^o , données aux points de mesure x_1 et x_2 et d'une simulation non biaisée Z^b . Les erreurs d'observations sont supposées être décorréliées, ce qui n'est pas le cas des erreurs du modèle. Nous désignons par σ_{oi}^2 les variances des erreurs d'observation aux points de mesure x_i . La variance d'erreur du modèle est supposée homogène et égale à σ_b^2 . L'estimation est alors nettement plus compliquée :

$$Z^a(x_a) = Z^b(x_a) + \frac{B_{1,a}(\sigma_b^2 + \sigma_{o2}^2) - B_{2,a}B_{1,2}}{(\sigma_b^2 + \sigma_{o1}^2)(\sigma_b^2 + \sigma_{o2}^2) - B_{1,2}^2}(Y_1^o - Z^b) \quad (2.53)$$

$$+ \frac{B_{2,a}(\sigma_b^2 + \sigma_{o1}^2) - B_{1,a}B_{1,2}}{(\sigma_b^2 + \sigma_{o1}^2)(\sigma_b^2 + \sigma_{o2}^2) - B_{1,2}^2}(Y_2^o - Z^b) \quad (2.54)$$

L'erreur commise sur l'estimation s'écrit :

$$\sigma_a^2(x_a) = \sigma_b^2(x_a) - \frac{B_{1,a}^2(\sigma_b^2 + \sigma_{o2}^2) + B_{2,a}^2(\sigma_b^2 + \sigma_{o1}^2) - 2B_{1,a}B_{2,a}B_{1,2}}{(\sigma_b^2 + \sigma_{o2}^2)(\sigma_b^2 + \sigma_{o1}^2) - B_{1,2}^2} \quad (2.55)$$

Prenons $\sigma_{o1}^2 = \sigma_{o2}^2 = 0.25\sigma_b^2$, et utilisons le modèle de corrélation homogène et isotrope de [Balgovind et al., 1983] décrit par l'équation 2.43. La figure 2.2 montre l'évolution de l'erreur d'analyse et des poids d'interpolation des deux observations dans l'estimation de la variable Z_a^t positionnée en $x_a/L = 0$ en fonction de la position du point de mesure x_2/L . L'autre observation est fixe et située au point $x_1/L = -2.0$.

Lorsque l'observation 2 est à l'infini, son poids tend vers 0, tandis que celui de l'observation 1 tend vers une constante. Cette constante correspond au poids d'une observation seule, déterminé par l'équation 2.51. Lorsque l'observation 2 est positionnée au même endroit que l'observation 1, i.e $x_2 = -2.0$, les poids sont égaux. C'est aussi le cas quand l'observation 2 est positionnée à l'opposé du point d'analyse et à la même distance que l'observation 1, en $x_2 = +2.0$. Mais les poids sont alors un peu plus forts car la covariance des erreurs du modèle entre les deux points est plus faible. De plus, nous remarquons que l'erreur d'analyse est plus faible, montrant qu'à même distance du point analysé, deux observations distinctes apportent plus d'information que deux observations proches. Nous constatons également sur cette figure que les poids peuvent être négatifs.

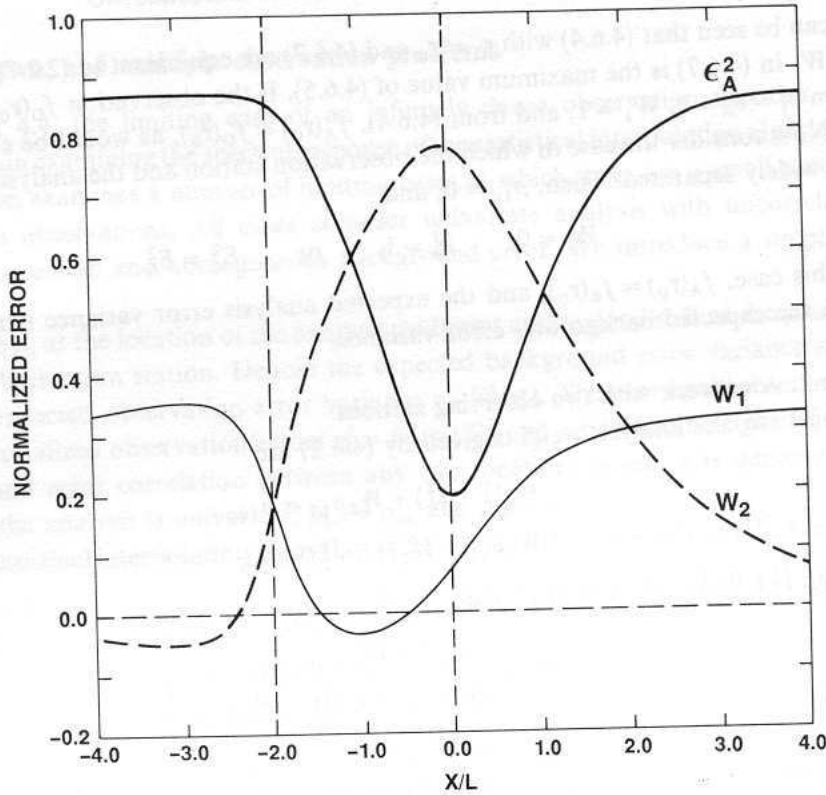


FIG. 2.2 – Évolution de l’erreur d’analyse et des poids d’interpolation de deux observations dans l’estimation d’une variable quelconque positionnée en $x_a/L = 0$. Une des observations est fixe et située au point $x_1/L = -2.0$ tandis que la position de l’autre observation varie entre $x_2/L = \pm\infty$. Source : [Daley, 1991].

2.7 Extension à des champs de variances non définies : Krigeage Intrinsèque

L’hypothèse de stationnarité d’ordre 2 du Krigeage Ordinaire est une hypothèse forte. Elle suppose que la fonction $X^t(\mathbf{s})$ a une variance finie. En effet, l’équation 2.47 suppose qu’à distance nulle, la variance est égale à une constante : $Var(X^t(\mathbf{s})) = C(0)$. Nous pouvons toujours calculer expérimentalement une telle variance, mais si celle-ci croît avec l’extension de la région d’étude, cela signifie qu’elle n’est bien approchée, c’est-à-dire que l’ensemble statistique n’est pas représentatif, ou pire, qu’elle n’existe pas. Il est alors nécessaire d’utiliser des modèles de fonction aléatoire à ”variance infinie”.

Krigeage Intrinsèque

Le Krigeage Intrinsèque [Cressie, 1993] suppose que la variable à analyser, $X^t(\mathbf{s})$, est *intrinsèquement stationnaire*. C'est-à-dire que :

$$\forall(\mathbf{s}_1, \mathbf{s}_2), E(X^t(\mathbf{s}_1) - X^t(\mathbf{s}_2)) = 0 \quad (2.56)$$

$$\forall(\mathbf{s}_1, \mathbf{s}_2), \frac{1}{2}Var(X^t(\mathbf{s}_1) - X^t(\mathbf{s}_2)) = \gamma(\mathbf{s}_1 - \mathbf{s}_2) \quad (2.57)$$

où $\gamma(\mathbf{s}_1 - \mathbf{s}_2)$ est la *semivariance* des différences entre deux valeurs de la variable \mathbf{X}^t à analyser. Ces différences sont appelées *les accroissements*. La méthode consiste en fait à supposer que ces accroissements sont stationnaires d'ordre 2.

Ces hypothèses sont moins fortes que les hypothèses énoncées dans le cas du Krigeage Ordinaire. En effet, ne faisant aucune hypothèse sur leur variance (mais sur la variance des accroissements), le Krigeage Intrinsèque permet de traiter des variables de plus grande variabilité, et en particulier des fonctions aléatoires à "variance infinie", ainsi que l'illustre l'exemple suivant.

Exemple de fonction non stationnaire d'ordre 2, intrinsèquement stationnaire

Prenons le cas d'une marche aléatoire : un tirage à pile ou face. Chaque tirage x_i est indépendant et prend la valeur 1 (face) ou la valeur -1 (pile) avec la probabilité $\frac{1}{2}$. Après chaque tirage n , nous construisons la fonction aléatoire sur les instants précédents $Y_n = \sum_{i=1}^n X_i$. La moyenne de X est $E(X) = 1 * \frac{1}{2} + (-1) * \frac{1}{2} = 0$, sa variance $E(X^2) = 1^2 * \frac{1}{2} + (-1)^2 * \frac{1}{2} = 1$. La moyenne de Y s'écrit : $E(Y) = \sum_{i=1}^n E(X_i) = 0$, sa variance est $E(Y) = E(\sum_{i=1}^n X_i)^2 = \sum_{i=1}^n \sum_{j=1}^n E(X_i X_j) = \sum_{i=1}^n E(X_i^2) = n$ car $\forall i \neq j E(X_i X_j) = 0$. La "variance" de Y tend vers l'infini quand n tend donc vers l'infini. La covariance de Y s'écrit $Cov(Y_n, Y_{n+k}) = E(\sum_{i=1}^n X_i \sum_{j=1}^{n+k} X_j) = \sum_{i=1}^n E(X_i^2) = n$. Cette covariance tend elle-aussi vers l'infini quand n tend vers l'infini. La fonction aléatoire n'est pas stationnaire d'ordre 2.

Considérons maintenant les accroissements $Y_{n+k} - Y_n = \sum_{i=n+1}^{n+k} X_i$, leur moyenne est égale à $E(Y_{n+k} - Y_n) = \sum_{i=n+1}^{n+k} E(X_i) = 0$, et leur variance $E(Y_{n+k} - Y_n)^2 = \sum_{i=n+1}^{n+k} \sum_{j=n+1}^{n+k} E(X_i X_j) = \sum_{j=n+1}^{n+k} E(X_i)^2 = k$. Cette variance dépend de l'accroissement k . Les accroissements de la marche aléatoire sont donc stationnaires d'ordre 2. Finalement, la fonction aléatoire, que nous avons considérée n'est pas stationnaire d'ordre 2 mais est *intrinsèquement stationnaire*. Notons qu'une fonction stationnaire d'ordre 2 est toujours une fonction intrinsèquement stationnaire.

Expression des poids du Krigeage Intrinsèque

Les équations du Krigeage Ordinaire sont légèrement modifiées dans le cas du Krigeage Intrinsèque. Posons $2\gamma_{k,l} = Var(X^t(\mathbf{s}_k) - X^t(\mathbf{s}_l))$ la semi-variance du champ \mathbf{X}^t à analyser. En tenant compte de cette nouvelle définition, les termes qui apparaissent dans l'équation 2.25 peuvent être écrits de manière générale sous la forme suivante :

$$\begin{aligned} E[(X^o(\mathbf{s}_k) - X^t(\mathbf{s}_a))(X^o(\mathbf{s}_l) - X^t(\mathbf{s}_a))] &= \frac{1}{2}[E[(X^o(\mathbf{s}_k) - X^t(\mathbf{s}_a))^2] + E[(X^o(\mathbf{s}_l) - X^t(\mathbf{s}_a))^2] \\ &\quad - E[(X^o(\mathbf{s}_l) - X^o(\mathbf{s}_k))^2]] \\ &= \gamma_{a,k} + \gamma_{a,l} - \gamma_{k,l} + R_{k,k} \delta_{k,l} \end{aligned}$$

Les équations du Krigeage Intrinsèque s'obtiennent de la même façon que pour le Krigeage Ordinaire et s'écrivent :

$$\forall k = 1, \dots, K, \sum_{l=1}^p w_l(\mathbf{s}_a) [\gamma_{k,l} - R_{k,k} \delta_{kl}] - \mu = \gamma_{k,a} \quad (2.58)$$

avec la condition que la somme des poids doit être égale à 1. L'erreur d'analyse est donnée par :

$$\sigma_a^2(\mathbf{s}_a) = \sum_{k=1}^p w_k(\mathbf{s}_a) \gamma_{a,k} - \gamma_{a,a} - \mu \quad (2.59)$$

Modélisation des semivariances : le variogramme

La démarche adoptée pour modéliser les semivariances $\gamma_{k,l}$ consiste à construire un *variogramme* à partir des observations. Ce variogramme est obtenu de la même manière que dans le cas du Krigeage Ordinaire, excepté qu'au lieu de tracer les covariances des observations, nous traçons les semivariances des différences spatiales de ces mêmes observations.

Ainsi, si la fonction de covariance γ est isotrope, sur la classe N_j , définie par l'ensemble des couples de stations éloignées de la distance h_j (à une tolérance près), nous calculons les variances sur les J classes :

$$2\gamma^o(h_j) = \frac{1}{|N_j|} \sum_{N_j} (X^o(\mathbf{s}_k) - X^o(\mathbf{s}_l))^2 \quad j = 1, \dots, J \quad (2.60)$$

Quelque soit la classe de distance, ces variances contiennent toujours les variances d'erreur d'observation. En effet $2\gamma_{k,l}^o = E[(X^o(\mathbf{s}_k) - X^o(\mathbf{s}_l))^2] = 2\gamma_{k,l} + \frac{1}{2}(R_{k,k} + R_{l,l} - 2R_{l,k} \delta_{kl})$. Supposons que les variances d'erreur d'observation sont homogènes et égales à σ_o^2 . A distance nulle, $\gamma_{k,l}^o = \gamma_{k,l} = 0$. En revanche, à distance non nulle, $\gamma_{k,l}^o = \gamma_{k,l} + \sigma_o^2$: le variogramme expérimental donne une estimation de $\gamma_{k,l}$ à σ_o^2 près. C'est pourquoi, le variogramme est modélisé par une fonction qui dépend de la variance de l'erreur d'observation. Dans ce document, nous avons choisi d'utiliser la fonction exponentielle suivante :

$$\gamma(h) = c \left(1 - \exp\left(-\frac{h}{a}\right) \right) + \sigma_o^2 \quad (2.61)$$

Cette fonction est définie négative [Cressie, 1993]. Avec seulement deux paramètres a et b , elle permet de modéliser un grand nombre de variogrammes, allant du variogramme plat $\gamma(h) = \sigma_o^2$ (a très petit ou c nul) au variogramme linéaire (a très grand). Si la variance de la variable existe, un variogramme est relié à un covariogramme par la relation $\gamma(h) = C(0) - C(h)$.

Un variogramme plat est caractéristique d'une variable sans structure spatiale, c'est-à-dire d'un *bruit blanc*. Un variogramme exponentiel est caractéristique d'une variable qui a une structure de covariance spatiale. Si le variogramme est linéaire, cela signifie que la semivariance des accroissements peut tendre vers l'infini quand la distance augmente. Cependant la taille de la région n'est pas suffisamment grande pour le démontrer. Dans ce cas, à grande distance des observations,

l'erreur d'analyse, donnée par l'équation 2.59 peut prendre de grandes valeurs : la qualité de l'analyse est dégradée. Le fait qu'un variogramme tende vers l'infini aux grandes distances peut avoir plusieurs raisons : cela notamment peut être du à la présence d'une structure de grande échelle, il faut alors élargir le domaine d'étude, ou la présence d'une anisotropie, il faut alors faire des variogrammes dans plusieurs directions pour le vérifier.

2.8 Conclusions

Nous avons décrit dans ce chapitre le principe des différentes méthodes, permettant d'interpoler des données. Nous appliquons ces méthodes dans la suite, pour produire des analyses d'ozone ou de dioxyde d'azote.

Nous avons successivement décrit le Krigeage Ordinaire, l'Interpolation statistique, et le Krigeage Intrinsèque. Nous avons montré que l'Interpolation Statistique est très proche d'un Krigeage appliqué aux différences entre les observations et les simulations prises aux points de mesure. La principale différence apparaît dans la détermination des poids de l'interpolation, qui sont reliés à des valeurs statistiques : les covariances des erreurs du modèle numérique et des observations, ou, les semi-variances des accroissements de la variable que l'on cherche à analyser. Ces quantités non accessibles directement doivent être modélisées. Comme nous l'avons vu, il y a plusieurs manières de procéder.

La première méthode consiste à considérer que les poids sont stationnaires dans le temps, comme dans le cas de l'Interpolation Statistique. Les poids sont alors déterminés à partir d'une climatologie, c'est à dire d'une longue série temporelle de simulations et d'observations. L'estimation des covariances des erreurs de modélisation et d'observation est basée sur le calcul de *moyennes temporelles*, les covariances des innovations.

La deuxième méthode consiste à considérer que les poids de l'interpolation sont variables dans le temps. Cela nécessite donc de limiter l'ensemble statistique sur lesquels vont être calculés ces poids. C'est ainsi que seules les données disponibles (observations ou innovations) au temps de l'analyse sont utilisées. Les méthodes de Krigeage proposent une méthodologie pour construire des ensembles statistiques réduits, basés sur la notion de classes de distance. Les poids résultent alors du calcul de *moyennes spatiales*.

Quatre méthodes ont été testées dans la suite. La première est une Interpolation Statistique basée sur un modèle de covariance des erreurs de modélisation homogène et isotrope. Mis à part la modélisation particulière du biais de modélisation, cette méthode est classique, et est, dans ce document, utilisée comme référence. Basée sur la précédente, la deuxième méthode est plus originale. En effet, nous avons cherché à construire un modèle de covariance hétérogène et anisotrope. La troisième consiste en un Krigeage Intrinsèque que nous appliquons directement sur les observations. C'est aussi une méthode classique couramment utilisée dans le domaine pétrolier. Cette méthode est perfectionnée en utilisant le modèle CHIMERE, i.e en appliquant un Krigeage Intrinsèque sur les différences simulations-observations, c'est ce qui fait toute l'originalité de cette dernière démarche. La mise en place, la validation et la comparaison des quatre méthodes sont présentées dans les prochains chapitres.

Chapitre 3

Évaluation du modèle CHIMERE à l'aide des observations

L'évaluation d'un modèle, faite en comparant ses sorties à des observations est une étape indispensable. Elle permet de définir des axes futurs de développement et d'évaluer les marges d'incertitudes à prendre en compte sur tout résultat fourni par le modèle. C'est aussi la première étape avant la mise en place d'un algorithme d'assimilation de données. En effet tout algorithme de correction d'un modèle nécessite d'abord une étude statistique de ses erreurs et de celles commises sur les observations.

Ce chapitre est consacré à une description du modèle CHIMERE multi-échelles ainsi qu'à une présentation des mesures, utilisées dans le cadre de cette étude. Nous confrontons également les concentrations de deux polluants photochimiques, l'ozone et le dioxyde d'azote, simulées par le modèle aux valeurs observées. Cette comparaison, effectuée sur les données des étés 1999 et 2001, à l'échelle continentale ou à l'échelle régionale, permet de détailler différents types d'erreurs que nous regrouperons ultérieurement dans une *erreur globale* dans les algorithmes d'assimilation de données.

3.1 Le modèle CHIMERE multi-échelles

Nous utilisons la version 200108 du modèle, correspondant à son état d'avancement en août 2001. Cette version est accessible sur le site internet <http://euler.lmd.polytechnique.fr/chimere/> et décrite de façon détaillée dans [Schmidt et al., 2001].

3.1.1 Généralités

L'objectif de ce modèle est de simuler les concentrations de gaz liés à la photochimie de la basse troposphère. Le modèle de simulation numérique, baptisé CHIMERE continental couvre l'Europe de l'Ouest avec une résolution spatiale assez faible, de sorte qu'avec cet outil, il n'est pas possible de simuler précisément des concentrations urbaines ni même le panache d'une agglomération (car les concentrations sont trop diluées). En revanche, les concentrations fournies par ce modèle

continental sont utilisées comme conditions aux limites pour alimenter des versions régionales plus résolues (approche appelée "nesting").

La principale contrainte imposée est la rapidité d'utilisation du modèle. Pour garantir une telle souplesse, deux conditions sont respectées : la rapidité du calcul numérique et l'accessibilité des données d'entrée. Les différentes versions de CHIMERE utilisées répondent à ces deux critères : d'une part, le temps de simulation est raisonnable, et d'autre part, le modèle est alimenté par les données météorologiques, issues quotidiennement du Centre Européen de Prévision Météorologique à Moyen Terme (CEPMMT), ce qui évite de recalculer les champs météorologiques par un modèle de type " méso-échelle ". Le prix à payer pour ce dernier choix est que l'on ne peut pas espérer simuler correctement le transport des polluants dans les régions fortement soumises à des effets météorologiques "locaux" comme les circulations dans les massifs montagneux ou les brises en régions côtières. En revanche, en l'absence de ces phénomènes (ce qui est vrai pour une grande partie de l'Europe, notamment l'Europe du Nord), il est raisonnable de supposer que les champs météorologiques simulés par le modèle du CEPMMT sont suffisamment réalistes et peuvent ainsi être utilisés pour calculer le transport et la chimie de la pollution photo-oxydante.

La version continentale décrite ici nécessite un temps de simulation, sur une station de travail (PC sous LINUX), d'environ 6 minutes CPU par jour simulé. Ceci permet d'effectuer rapidement (en moins de 24h) une simulation sur une saison entière, et d'obtenir ainsi un grand nombre de résultats pouvant être analysés de façon statistique.

3.1.2 Description du modèle

CHIMERE est un modèle de chimie-transport simulant les variations spatio-temporelles concentrations d'une quarantaine d'espèces gazeuses dans une centaine de réactions chimiques. Il est basé sur l'équation de continuité des différentes espèces chimiques de concentration \mathbf{c} :

$$\left(\frac{\partial \mathbf{c}}{\partial t}\right) + \nabla(\mathbf{u}\mathbf{c}) = \nabla(K\nabla\mathbf{c}) + P - L \quad (3.1)$$

Dans cette équation, représentant une approche Eulérienne [Seinfeld and Pandis, 1998], \mathbf{c} est un vecteur contenant les concentrations de toutes les espèces chimiques à tous les points de grille du modèle, \mathbf{u} est le vecteur vitesse du vent, K le tenseur des diffusivités turbulentes. P et L représentent respectivement les termes de production et de pertes (dûes aux réactions chimiques, aux émissions et au dépôt sec).

La géométrie du modèle continental

Horizontalement, le modèle décrit les champs de concentrations des polluants avec un maillage de 1/2 degré (soit environ 50 km à notre latitude). C'est à peu près la résolution effective du modèle de prévision météorologique du CEPMMT. Comme le montre la figure 3.1, le domaine couvre une grande partie de l'Europe de l'Ouest (-10W ; 22E ; 40N ; 57N). Le maillage est constitué de 65x33 cellules. La résolution choisie correspond également approximativement à celle de la base de données d'émissions (voir plus loin).

Verticalement, le modèle décrit les concentrations dans 5 couches couvrant l'ensemble de la couche limite, de la surface à 750 hPa (\approx 3000 mètres), et la basse troposphère libre. La première

couche a une hauteur d'environ 50 mètres, et les autres sont fixées en " niveaux hybrides sigma-p ", afin d'être mieux adaptées aux niveaux du modèle météorologique du CEPMMT.

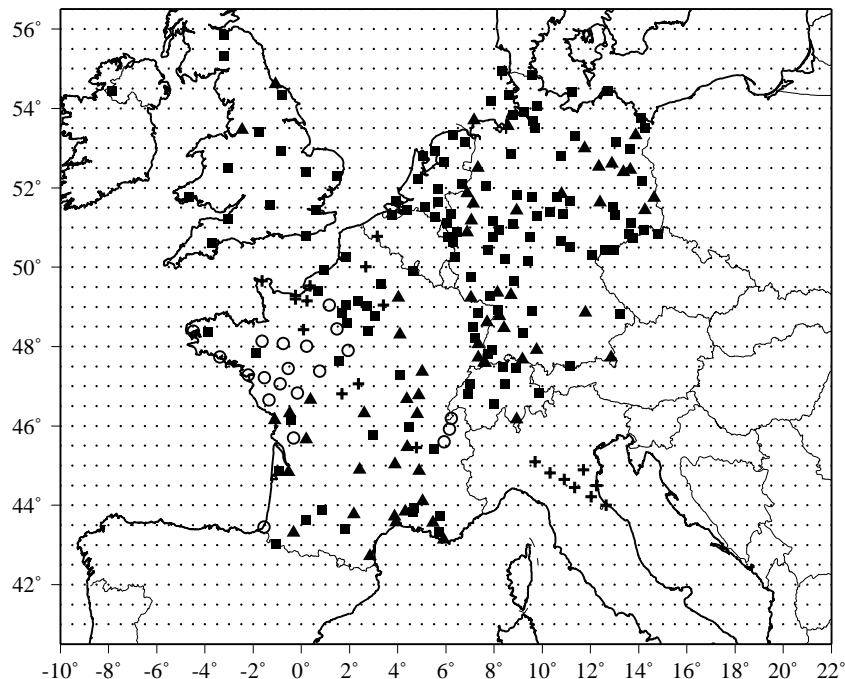


FIG. 3.1 – Grille horizontale du modèle CHIMERE continental et positions des 232 stations mesurants l’ozone en 2001 et utilisées lors de sa validation. Les cercles, les carrés et les triangles pleins repèrent respectivement l’emplacement des stations urbaines, rurales, et péri-urbaines. Les croix représentent les stations dont la typologie est inconnu.

Données météorologiques d’entrée

Le modèle nécessite, en entrée les variables standard des modèles numériques de prévision du temps en chacun des noeuds du maillage : vent horizontal, température, humidité spécifique, pression de surface, couverture nuageuse de trois différents types (les nuages bas, nuages de moyenne altitude et nuages hauts).

Dans ce document, les données météorologiques du CEPMMT "premières ébauches" (non assimilées, i.e non revues par les observations) sont utilisées. Elles correspondent aux prévisions à 6 heures. Elles ont une résolution horizontale de 0.5 degré. Il y a dix huit niveaux verticaux entre 0 et 3000 m. Les données sont moyennées sur les niveaux du modèles CHIMERE, sauf pour la première couche (de 50 mètres) où les données du premier niveau du modèle du CEPMMT sont utilisées directement. Ces données, d’une résolution temporelle de 3 heures, sont interpolées linéairement toutes les dix minutes.

Les émissions

Le modèle CHIMERE-continental et sa chimie MELCHIOR réduite requierent en entrée les émissions de 16 espèces représentatives des composés essentiels mis en jeu dans la photochimie.

Les émissions anthropiques portent sur 5 classes de composés : NO_x, SO₂, CO, CH₄ et les Composés Organiques Volatiles Non Méthaniques (Ethane, n-Butane, Ethène, Propène, Isoprène, o-Xylène, Formaldéhyde, Acétaldéhyde, Méthyle-Ethyle-Cétone). Elles proviennent de la base de données EMEP pour l'année 1998, et sont délivrées sous forme de totaux annuels. Afin de transformer ces données brutes en grandeurs utilisables par le modèle CHIMERE, c'est à dire en données horaires, variant en fonction du mois, du jour de la semaine, et de l'heure de la journée, un travail d'interface est nécessaire. Cette interface procède en deux temps : d'abord les données annuelles sont réparties temporellement en utilisant une modulation mensuelle, hebdomadaire puis horaire. Ensuite les COVNM sont regroupés en 32 classes de composés de réactivité similaire. Finalement, ces classes de composés sont agrégées dans les composés "modèle" décrits ci-dessus, en utilisant la procédure proposée par *Middleton et ses collaborateurs* [1990]. Les émissions anthropiques de NO par les sols, dues à la transformation micro biologique de l'épandage des engrais durant les mois d'avril à août, sont aussi prises en compte, et dépendent fortement de la température. Pour ces flux d'émissions, la méthodologie de [*Stohl et al.*, 1996] est utilisée. Le fait que la partie anthropique des émissions corresponde à l'année 1998, alors que les années simulées sont 1999 et 2001, s'explique par la non disponibilité des données d'émissions pour ces années.

Les émissions biogéniques de COV concernent principalement l'isoprène et de terpène (α -pinène). Les émissions d'isoprène sont essentiellement issues de certaines espèces de chênes et des épicéas. Les émetteurs de terpènes sont essentiellement les conifères et en particulier les épicéas. Les émissions de terpènes concernent aussi une grande variété d'espèces chimiques, mais dont la dégradation photochimique n'est, en général, pas bien connue. Pour cette raison, toutes les émissions de terpènes sont représentées dans le mécanisme chimique par le α -pinène, souvent émis de façon majoritaire et dont la dégradation photochimique est relativement bien comprise. Ces émissions biogéniques sont paramétrisées par des flux [*Günther*, 1997], qui dépendent du potentiel d'émission de l'espèce et de la densité foliaire. Ces deux valeurs dépendent elles-mêmes du type de sol et ont été estimées selon *Simpson et ses collaborateurs* [1999]. La classification des sols est celle d'EMEP. Elle prend en compte 9 type de sols : forêt d'érables, herbe, céréales, plan d'eau, surface urbaine, océan, forêt de permanents, forêt de caduques, autres. Un facteur correctif est appliqué à ces flux pour tenir compte de la dépendance des émissions en température et en insolation [*Günther et al.*, 1994; *Günther*, 1997; *Simpson et al.*, 1999].

La chimie

Une version réduite du mécanisme chimique MELCHIOR (Modèle d'Etude Lagrangienne de la CHimie de l'Ozone à l'échelle Régionale) est utilisée. Développé au Service d'Aéronomie du CNRS [*Lattuati*, 1997] à partir d'un mécanisme antérieur dû à *Hov et ses collaborateurs* [1985], elle correspond à un ensemble de réactions chimiques se produisant dans l'atmosphère. MELCHIOR est en fait un modèle plus général destiné à élucider le rapport entre émissions d'oxydes d'azote, composés organiques volatils et pollution photo oxydante à l'échelle régionale. En conséquence, le mécanisme chimique doit être applicable à la fois à une atmosphère très polluée en précurseurs d'ozone (NO_x et COV) près des sources d'émission et à une atmosphère relativement propre, n'ayant pas subie d'apports majeurs en polluants depuis environ une semaine.

La chimie des espèces hydrocarbonées est assez similaire au mécanisme en phase gazeuse d'EMEP [Simpson, 1992]. Les constantes de vitesses ont été actualisées selon [Atkinson et al., 1997], et [De Moore et al., 1994]. La version réduite de ce mécanisme contient un total de 44 espèces chimiques condensées et 116 réactions chimiques. La réduction suit le concept d'*opérateurs chimiques* [Carter, 1990]. Il est très efficace en termes de temps de calcul, car il est très condensé et a montré un excellent accord avec le mécanisme initial.

Le traitement horizontal du transport

Le transport horizontal est traité au moyen du schéma numérique " PPM " d'ordre 3 (*Parabolic Piecewise Method*), développé par Colella et Woodward [1984], qui a montré de meilleures performances que d'autres schémas, et qui a l'avantage de conserver la masse et d'être précis [Carpenter et al., 1990]. Toutefois il est assez coûteux en temps, c'est pourquoi seules les espèces lentes (20 sur les 44) sont transportées avec ce schéma, les autres étant transportées avec le schéma classique "amont" d'ordre 1, ce qui a un impact négligeable sur les concentrations d'ozone.

Le transport vertical et la diffusion

Le transport vertical est calculé de sorte à compenser la convergence ou la divergence des flux de masse horizontaux. Si par exemple, dans la couche de surface, le bilan de masse est positif (plus de matière entrante que sortante), un flux vertical ascendant est créé entre la couche de surface et la couche suivante. Ce nouveau transport vertical est intégré au bilan de masse de la couche suivante et le processus est répété jusqu'au sommet du modèle, où les polluants sont définitivement évacués ou bien importés depuis la troposphère libre. Cette méthode présente l'avantage évident de conserver la masse et de créer un transport vertical consistant avec le transport horizontal. Les vents verticaux obtenus sont en accord avec ceux du CEPMMT, qui ne sont pas utilisables directement car le maillage vertical de ces données est différent de celui de CHIMERE.

La diffusion verticale est paramétrisée par une diffusivité turbulente. La diffusion horizontale est négligée comme il est couramment fait dans les modèles méso-échelles [Chang et al., 1987]. Pour calculer la diffusivité turbulente verticale K , la méthode d'O'Brien [1970] est mise en œuvre. La hauteur de la couche limite est définie comme étant l'altitude à laquelle le nombre de Richardson de couche atteint la valeur critique de 0.05. Pour éviter aussi de trop faibles hauteurs de mélange, la valeur minimale est fixée à 50 m.

Les conditions aux limites et initiales

Pour minimiser les effets de bord, il est nécessaire de connaître les concentrations importées loin du domaine d'intérêt [Chang et al., 1987]. Cependant ceci ne peut pas être totalement respecté, vu les critères de rapidité imposés au modèle. Il n'est donc pas possible d'utiliser un modèle global pour forcer les concentrations aux limites. Dans CHIMERE, les concentrations de 13 espèces lentes importées sont estimées à partir du modèle climatologique MOZART (Model for OZone And Related chemical Tracers) [Hauglustaine et al., 1998] à chaque point de grille du domaine limite. Les concentrations des autres espèces sont prises à 10^{-10} ppb. La simulation est lancée dix jours avant la période intéressante, afin de laisser le temps au modèle de se mettre à l'équilibre.

Le rayonnement

Le rayonnement est pris en compte dans les réactions de photolyse. Les valeurs des taux de photolyse sont tabulées en fonction de l'angle zénithal, en utilisant différentes sources pour ces cal-

culs. Toutefois, ces tabulations ne sont valables qu'en ciel clair. Afin de paramétriser l'atténuation due à la présence de nuages, nous utilisons un modèle statistique prenant en compte les nébulosités moyennes et hautes, et les profils d'humidité dans les basses couches. Sa description figure dans le rapport du projet ESQUIF [ESQUIF, 2001].

Le dépôt sec

Le dépôt sec est pris en compte et paramétré par des flux sortant de la couche inférieure du modèle. La vitesse de dépôt est estimée par analogie à une résistance électrique [Wesely and Hicks, 1989]. Ce dépôt sec consiste en un transfert de molécules de gaz depuis l'atmosphère jusqu'à la surface de la terre. Les processus de dépôt sec peuvent se décomposer en trois étapes. La première est le transfert des gaz de la couche limite de surface vers la couche directement adjacente à la surface (ou couche laminaire) et constitue la composante aérodynamique du transfert. La seconde correspond à la diffusion des gaz à travers la couche laminaire. Enfin, la troisième, déterminée par la solubilité, la réactivité du gaz et l'état de surface, correspond au piégeage (potentiel) des gaz par la surface.

Au sein de CHIMERE, la résistance aérodynamique est définie comme l'inverse du coefficient de traînée et du module du vent dans la plus basse couche du modèle [Wesely and Hicks, 1977]. Le coefficient de traînée est donné par la méthode de Louis [1979]. La résistance à travers la couche laminaire est quant à elle approchée par la formule de [Hicks et al., 1987]. La paramétrisation de la vitesse de friction est celle de [Louis, 1979]. Le calcul de la résistance de surface suit les travaux de [Erisman et al., 1994]. Il repose sur la solubilité ou le pouvoir oxydant de l'espèce chimique considérée, intervenant dans la paramétrisation des processus de transfert à travers les stomates, les mésophylles et les cuticules (toujours par analogie à des résistances). Pour tenir compte du type de sol lors du calcul des vitesses de dépôt, nous utilisons la base de données d'utilisation des sols *pan european data base* construit par RIVM sur une grille d'un dixième de degré de résolution [Van de Velde et al., 1994].

Approche numérique

La méthode numérique pour résoudre le système raide d'équations différentielles (3.1) est une adaptation de l'algorithme du second ordre TWOSTEP proposé par [Verwer, 1994]. Le système non-linéaire obtenu par ce schéma est résolu par la méthode Gauss-Seidel à chaque itération. Dans le cas du modèle continental, un pas de temps de 10 minutes est choisi, ce qui est un compromis raisonnable entre temps de calcul et précision. A chaque pas de temps, tous les processus physiques et chimiques sont actualisés simultanément. La technique de *séparation des opérateurs*, méthode standard pour résoudre des équations chimie-transport 3D [Mc Rae et al., 1982], rapide mais potentiellement imprécise, est évitée. L'utilisation d'une telle technique est possible parce que les processus physiques et chimiques, sont tous formulés par des flux entrants ou sortants dans chaque boîte du modèle.

Les modèles régionaux emboîtés

Les modèles régionaux sont des copies du modèle continental à certains détails près : géométrie, temps d'intégration (de 2.5 minutes au lieu de 10 minutes), résolution, émissions et conditions aux limites. La résolution est rapportée à celle des émissions qui sont fournies par les Associations Agréées pour la Surveillance de la Qualité de l'Air (AASQA). Elle est de 6 km sur l'Ile-de-France et sur

la région de Berre et de Marseille. L'extension de ces modèles est montrée sur les figures 3.2 sur 3.3.

Les variables météorologiques d'entrée sont interpolées à partir des données continentales sur la grille du modèle régional. Les émissions biogéniques sont également interpolées. En revanche, les émissions anthropiques proviennent de sources différentes. Pour l'Ile-de-France, nous utilisons les données fournies par AIRPARIF, valables pour un " jour moyen ". Elles sont modifiées de telle sorte que leurs totaux annuels correspondent à ceux donnés par le Centre Interprofessionnel Technique d'Étude de la Pollution Atmosphériques (CITEPA), par grands secteurs d'activités. Les profils saisonniers et les profils de spéciation des Composés Organiques Volatiles (COV) sont pris égaux à ceux fournis par l'université de Stuttgart (programme EUROTRAC/GENEMIS). Pour la région méditerranéenne, nous utilisons l'inventaire fourni par le projet ESCOMPTE.

Enfin, les modèles régionaux prennent comme conditions aux limites les concentrations de 14 espèces simulées par le modèle continental, qui sont interpolées sur les bornes du domaine. Ces espèces comprennent notamment les espèces lentes, l'ozone et ses précurseurs (hydrocarbures et NO_x). En fonction de la direction du vent, ces concentrations aux bornes du domaine sont advectées à l'intérieur par le schéma de transport PPM.

3.2 Les observations

3.2.1 Problèmes de représentativité

Avec un modèle de résolution de quelques dizaines de kilomètres, il n'est pas possible de simuler correctement la structure à petite échelle des panaches des villes ou des sources ponctuelles [Mc Nair et al., 1996]. L'objectif d'un tel modèle est donc, à défaut de donner accès à l'état réel de la pollution de l'air, de simuler de façon la plus réaliste possible son *état réel moyen par maille du modèle*.

La comparaison de simulations avec des observations influencées par des effets locaux met ainsi souvent en évidence de fortes erreurs liées au manque de résolution du modèle. Ces erreurs qui devraient être associées au modèle, sont pourtant toujours regroupées dans l'erreur d'observation dans les algorithmes d'assimilation de données. Ceci est fait pour forcer l'analyse, à ne pas trop se rapprocher des observations mais à tendre vers les niveaux de concentrations moyens par maille du modèle. Ces erreurs sont appelées les *erreurs de représentativité*. Elles sont contrôlées par les phénomènes dynamiques locaux et les émissions, et dépendent de la nature des surfaces qui environnent les stations.

Puisque les observations sont souvent localisées à proximité des villes où les sources locales peuvent avoir une influence, leur *représentativité* est toujours discutable [Tilmes et Zimmermann, 1998]. Ceci constitue l'un des principaux problèmes rencontrés lors de l'évaluation d'un modèle avec des observations. Ainsi, la première démarche à adapter pour se donner les moyens d'analyser correctement les résultats d'une comparaison simulations-observations est donc de se renseigner sur la localisation des stations auprès du personnel du réseau de surveillance en charge de leur maintenance. Une observation d'ozone en milieu rural ou en altitude est souvent, par exemple, plus représentative des quelques kilomètres qui l'entourent qu'une observation en milieu urbain, plus soumise aux phénomènes locaux (titration par NO, turbulence ...).

Pour ce qui est des précurseurs de l'ozone, les problèmes de représentativité sont encore plus

présents. En effet, la durée de vie de polluants tels que le monoxyde d'azote ou le dioxyde d'azote, est de l'ordre de quelques secondes à quelques heures. Ces polluants restent donc très localisés autour de leur source.

3.2.2 Les observations de surface

Les observations de surface utilisées dans cette étude proviennent de différentes associations pour la surveillance de la qualité de l'air en Europe. Les stations sont classifiées selon des critères européens, qui tiennent compte de leur localisation. Les stations, avec leur caractéristiques sont listées dans l'Annexe C dans les tableaux C.1 et C.2, dans lesquels les organismes propriétaires des données sont également mentionnés.

Typologie des stations

Les stations de mesure automatiques sont souvent implantées et équipées pour répondre en priorité à une préoccupation de santé publique. Il en existe six types, qui sont définis sur la base de critères précis et chiffrés au niveau européen. Cette coordination européenne répond à un besoin de comparaison des niveaux de pollution, ainsi qu'à un besoin d'estimer leur évolution d'une agglomération européenne à une autre. Dans cette étude, le type de la station est utilisé comme une première indication sur son environnement.

- ▷ **Les stations urbaines** (densité de population minimum 4000 habitants par km^2 dans un rayon de 1 km autour de la station) et **les stations périurbaines** (sous l'influence directe de l'agglomération) sont situées loin des sources directes de pollution (industrielle ou automobile) : parcs, cours d'école, zones d'habitation ... Elles permettent d'estimer la pollution à laquelle est soumise la population, quelle que soit son activité au cours de la journée. Elles mesurent en priorité le dioxyde de soufre, les oxydes d'azote, l'ozone, et les poussières fines.
- ▷ **Les stations trafic** sont situées près des voies de circulation, sur le trottoir, à deux ou trois mètres des pots d'échappement. Elles permettent d'évaluer le risque d'exposition maximal auquel sont soumis les piétons ou les automobilistes dans leur voiture. L'exposition du public en situation de proximité automobile est en général de courte durée mais les niveaux de pollution y sont élevés. Les polluants surveillés sont uniquement d'origine automobile : oxydes d'azote, monoxyde de carbone, poussières en suspension, hydrocarbures ...
- ▷ **Les stations industrielles** sont localisées au voisinage d'installations telles que les centrales thermiques ou les unités d'incinération des ordures ménagères. Situées dans des zones fortement peuplées, elles détectent les pollutions atmosphériques provoquées lors de retombées de panaches industriels. La population plus ou moins avoisinante, suivant la force, la direction du vent ainsi que la hauteur des cheminées, peut être exposée très occasionnellement à ce type de pollution. Ces stations mesurent principalement le dioxyde de soufre, les oxydes d'azote et parfois les poussières.
- ▷ **Les stations rurales** régionales sont installées à 50 km en moyenne des agglomérations où elles permettent de suivre les phénomènes de transfert de pollution par l'action du vent. Elles mesurent l'impact de certains polluants, créés par réactions chimiques à partir de polluants émis directement par les véhicules de l'agglomération (oxydes d'azote, hydrocarbures...).

Ces stations surveillent essentiellement l’ozone.

- ▷ **Les stations d’observation** ne répondent pas aux critères des stations précédentes, mais sont utilisées pour compléter le réseau et essentielles à la compréhension des phénomènes de pollution.

Choix des stations de mesure utilisées

A cause des problèmes de représentativité, évoqués ci-dessus, toutes les observations d’ozone disponibles durant l’été 1999 et 2001 n’ont pas été utilisées dans l’évaluation de CHIMERE continental et dans son algorithme de correction.

Le choix des 232 stations, dont la localisation est représentée sur la figure 3.1, a été effectué en fonction de la disponibilité en temps réel des données lors de l’expérience PIONEER 2001 et du caractère "le plus rural possible" (éloignement de grandes sources urbaines ou industrielles) de la typologie de la station, afin qu’elle ait une représentativité la plus proche de celle du modèle. Le choix, finalement subjectif, a largement été guidé par une discussion avec le personnel du réseau de surveillance en charge de la maintenance de la station. Il faut noter que certaines grandes régions ne possèdent pas ou très peu de stations de mesure d’ozone hors des zones urbaines. Dans ce cas, nous avons généralement opté pour des stations péri-urbaines, ou si ce n’était pas possible, les stations urbaines qui donnaient les valeurs moyennes observées les plus fortes possibles.

A l’échelle régionale, nous utilisons les mesures de toutes les stations en fonctionnement durant les étés 1999 et 2001. La distribution des sites mesurant l’ozone et le dioxyde d’azote en 2001 en Île-de-France est montrée sur la figure 3.2. Il existe douze stations urbaines mesurant l’ozone dans Paris et sa banlieue, quatre pré-urbaines et six stations rurales réparties autour de l’agglomération à environ 50 km. Les stations mesurant le dioxyde d’azote sont principalement localisées au coeur de l’agglomération, là où les concentrations sont les plus élevées.

Sur la région de Berre et de Marseille, les stations sont regroupées autour d’Avignon, de Marseille, de Toulon (ces stations sont gérées par l’AASQA AIRMARAIX), de l’étang de Berre (AIR-FOBEP) et de Montpellier (AIR LANGDOC ROUSSILLON). Parmi les stations mesurant l’ozone, quatre stations sont rurales et donnent une information sur les niveaux d’ozone en Provence, quatre autres sont industrielles. Pour le reste, la plus grande partie de ces stations sont urbaines (Cf. figure 3.3).

Incertitudes sur les mesures d’oxydes d’azote et d’ozone

La mesure de l’ozone est effectuée par absorption dans l’ultraviolet (253.7 nm). L’air échantillonné est envoyé dans une chambre optique soit directement, soit après passage sur un filtre éliminant l’ozone. La mesure de l’absorption due à l’ozone est déterminée par différence entre l’absorption UV de l’échantillon et l’absorption UV de l’échantillon exempt d’ozone.

La méthode de mesure du monoxyde d’azote (NO) se fait par chimiluminescence avec l’ozone. Un échantillon d’air est mélangé à l’ozone dans une chambre optique. L’ozone réagit avec le monoxyde d’azote pour former du dioxyde d’azote sous une forme excitée (Cf. équation (3.2)). En se désexcitant, le dioxyde d’azote émet un rayonnement (Cf. équation (3.3)). Le signal émis est mesuré par un photomultiplicateur et permet de calculer le taux de NO présent.



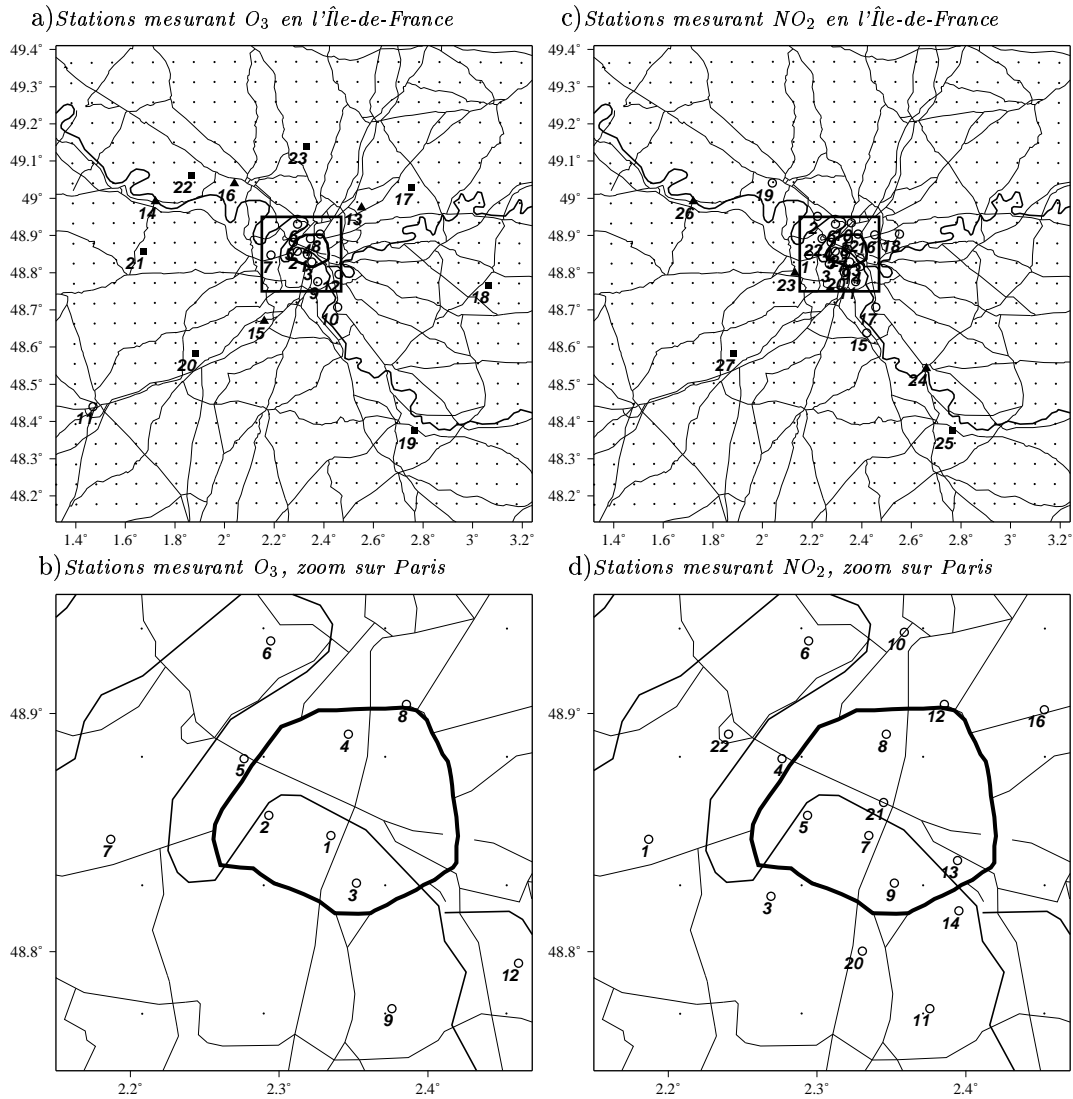


FIG. 3.2 – Paris est au centre de la carte. La ligne tracée en gras représente le périphérique autour de la ville. Les cercles, les carrés et les triangles pleins repèrent respectivement l'emplacement des stations urbaines, rurales, et péri-urbaines. Les noms des stations mesurant l'ozone sont : 1. PARIS06, 2. PARIS07, 3. PARIS13, 4. PARIS18, 5. NEUILLY, 6. GENNEVILLIERS, 7. GARCHES, 8. AUBERVILLIERS, 9. VITRY, 10. MONTGERON, 11. CHARTRES-FULBERT, 12. CRETEIL, 13. TREMBLAY, 14. MANTES, 15. LES-ULIS, 16. CERGY, 17. MONTGERON, 18. TREMBLAY, 19. CERGY, 20. CACHAN, 21. PARIS01, 22. LA-DEFENCE, 23. VERSAILLES. Les noms des stations mesurant le dioxyde d'azote sont : 1. GARCHES, 2. ARGENTEUIL, 3. ISSY, 4. NEUILLY, 5. PARIS07, 6. GENNEVILLIERS, 7. PARIS06, 8. PARIS18, 9. PARIS13, 10. ST-DENIS, 11. VITRY, 12. AUBERVILLIERS, 13. PARIS12, 14. IVRY, 15. EVRY, 16. BOBIGNY, 17. MONTGERON, 18. TREMBLAY, 19. CERGY, 20. CACHAN, 21. PARIS01, 22. LA-DEFENCE, 23. VERSAILLES, 24. MELUN, 25. FONTAINEBLEAU, 26. MANTES, 27. RAMBOUILLET

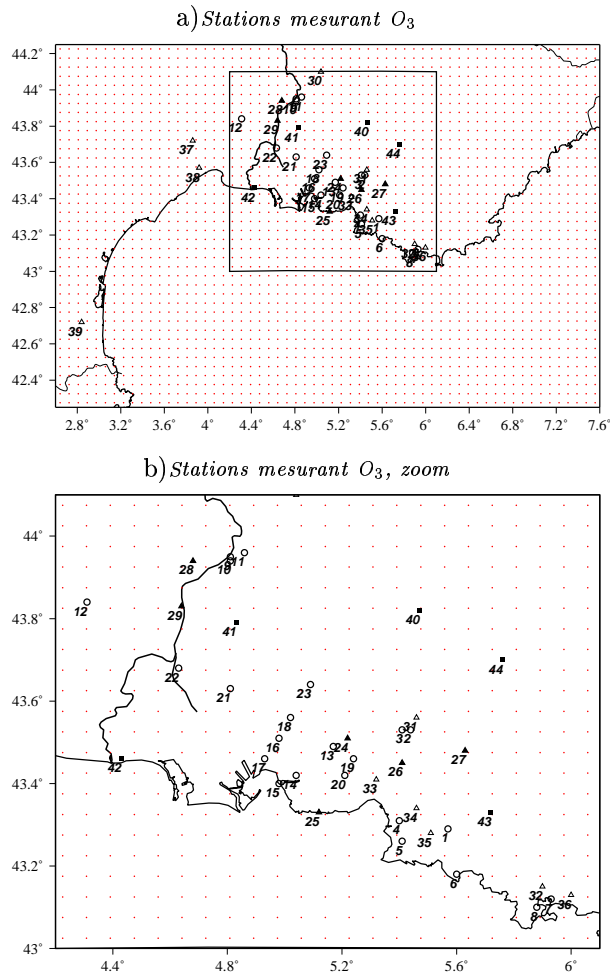


FIG. 3.3 – Distribution des stations mesurant l’ozone dans la région de Berre et de Marseille. Les cercles, les carrés et les triangles pleins montrent respectivement l’emplacement des stations urbaines, rurales, et péri-urbaines. Les triangles vides représentent les stations industrielles. Les noms des stations mesurant l’ozone sont : 1. AUBAGNE-PENITENTS, 2. AIX-CENTRE-ART, 3. AIX-OUEST, 4. MARSEILLE-5AVENUES, 5. MARSEILLE-MARGUERITE, 6. LA-CIOTAT, 7. TOULON-ARSENAL, 8. LA-SEYNE-SUR-MERS, 9. AVIGNON-MAIRIE, 10. AVIGNON-ARROUSAIRE, 11. LE-PONTET-MARGUERITE, 12. NIMES-GARDIES, 13. BERRE, 14. PAGODE, 15. PORT-DE-BOUC, 16. ISTRES, 17. FOS-Carabins, 18. MIRAMAS, 19. VITROLLES, 20. MARGNANE-Ville, 21. ST-MARTIN-de-CRAU, 22. ARLES, 23. SALON, 24. ROGNAC, 25. SAUSSET, 26. BOUC-BEL-AIR, 27. ROUSSET, 28. SAZE, 29. VALLABREGUES, 30. COMTAT-VENAISSIN, 31. AIX-PLATANES, 32. TOULON-OLIVE, 33. PENNES-MIRABEAU, 34. CUQUES-ALLAUCH, 35. VALLEE-HUVEAUNE, 36. LA-VALETTE-LA-GARDE, 37. MONTPELLIER-NORD, 38. MONTPELLIER-SUD, 39. ST-ESTEVE, 40. AURIBEAU, 41. SAINT-REMY, 42. ST-MARIE-de-la-MER, 43. AUPS-STE-BAUME, 44. CADARACHE.

Le dioxyde d'azote NO_2 n'est pas mesuré directement. Après mesure du contenu en NO de l'échantillon d'air, le NO_2 est converti en NO grâce à un four à catalyse. Le NO est alors mesuré une seconde fois par chimiluminescence avec l'ozone. Cette mesure donne la concentration totale en oxydes d'azote ($NO_x = NO + NO_2$) de l'échantillon initial. La concentration de NO_2 est alors déduite de la différence entre les deux mesures de NO.

L'incertitude de la mesure liée aux analyseurs est résumée dans le tableau 3.1 pour les polluants : O_3 , NO_2 , NO ou NO_x . L'incertitude de la mesure de NO_2 est la combinaison de l'incertitude de la mesure de NO et de celle de NO_x .

Concentration	Incertitudes		
	O_3	NO ou NO_x	NO_2
10	± 3.1	± 3.4	± 6.8
30	± 3.1	± 3.4	± 6.8
50	± 3.6	± 3.6	± 7.2
70	± 4.2	± 4.1	± 8.2
90	± 4.9	± 4.7	± 9.4
100	± 5.3	± 5.0	± 10.0
150	± 7.6	± 6.9	± 13.8
200	± 10.0	± 9.0	± 18.0
250	± 12.5	± 11.2	± 22.4

TAB. 3.1 – Incertitudes (en ppb) des mesures des oxydes d'azote et de l'ozone, effectuées par le réseau AIRPARIF, en fonction de leur concentration (en ppb) [ESQUIF, 2001]

Il est très important de signaler que la réduction catalytique n'est pas une réaction spécifique au dioxyde d'azote. En effet, d'autres composés azotés, comme par exemple le PAN, le HONO, le HNO_3 , et certains nitrates organiques, peuvent s'ils sont présents dans l'échantillon d'air, être réduits en monoxyde d'azote. Les concentrations de dioxyde d'azote sont alors souvent surestimées. Ceci est particulièrement le cas sur les zones rurales parce qu'au cours de son transport, le dioxyde d'azote peut parfois réagir avec des radicaux RO2 (produits soit par oxydation de composés organiques, soit par photolyse) pour former d'autres composés azotés : les peroxy nitrates. C'est ce qui est couramment appelé le *vieillessement chimique* des masses d'air. Parmi les peroxy nitrates, les plus stables sont les peroxyacétylnitrates, et parmi ces derniers le plus important d'entre eux est le peroxyacétylnitrate (PAN) produit à partir de l'acétaldéhyde.

3.2.3 Observations en altitude

La campagne ESQUIF

La campagne ESQUIF (Étude et Simulation de la Qualité de l'air en Île-de-France), réalisée par l'Institut Pierre Simon Laplace (IPSL), le Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA), Météo-France, et Laboratoire d'Aérodologie (LA), et le Forschungszentrum Jülich, visait à améliorer la compréhension de la formation des épisodes de pollution photochimique sur la région d'Île-de-France. Afin de mieux documenter la structure tridimensionnelle des champs de polluants, de nombreuses mesures aéroportées de composés chimiques ont été réalisées [Menut et al., 2000; ESQUIF, 2001]. Ces mesures d'altitude ont permis d'évaluer les modèles numériques, qui sont, depuis, utilisés pour, par exemple, quantifier les émissions [Vautard et al., 2002; Kley et al., 2002] ou étudier la production d'ozone dans le panache [Perros et al., 2002].

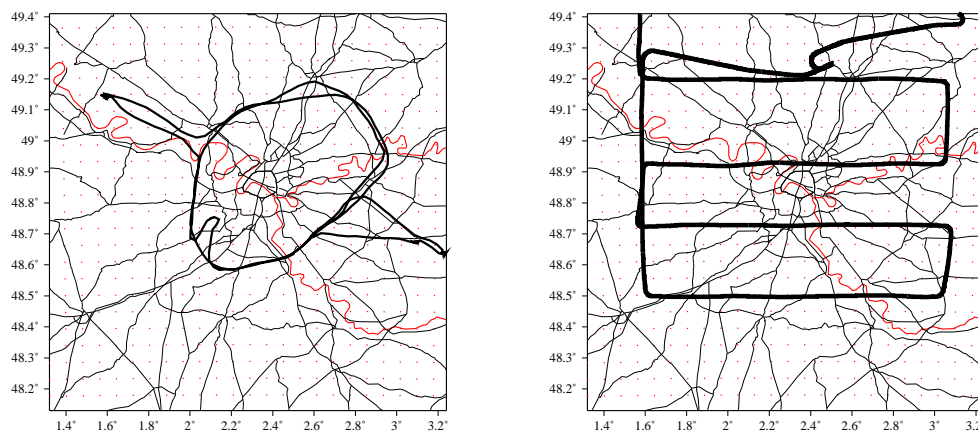


FIG. 3.4 – Trajectoire du DIMONA le 16 Juillet 1999 entre 12 :00 and 15 :40 UTC, et celle de l'ARAT le 30 Juillet 1999 entre 12 :10 and 14 :45 UTC. La grille du modèle CHIMERE est représentée par les points régulièrement répartis.

En plus des mesures de routine, des mesures intensives ont été effectuées lors d'épisodes de pollution. Six Périodes d'Observations Intensives (POI) ont été déclenchées pendant l'été 1999. Nous utilisons dans cette étude les mesures effectuées durant quatre de ces POI (POI 4, 5, 6 et 8 [ESQUIF, 2001]) par deux avions, le DIMONA (de Metair) et l'ARAT (Avion de Recherche Atmosphérique et de Télédétection de l'INSU). Les figures 3.4 montrent deux trajectoires types de ces avions.

Les vols DIMONA sont des vols approximativement circulaires effectués autour de l'agglomération parisienne, à altitude quasi-constante. Certains de ces vols offrent aussi des profils verticaux. L'objectif de ces mesures circulaires autour de Paris est d'examiner le contenu en polluants photochimiques des masses d'air entrant et sortant de l'agglomération. Les mesures de l'avion ARAT sont plus utilisées pour étudier les *régimes chimiques* de la région (Cf. Annexe A, [Honoré, 2000]). Le tableau 3.2 présente les données aéroportées utilisées dans le présent document (provenance des données, dates des POI, temps des vols, niveau du modèle associé à l'altitude des vols).

Avion	POI	Date	Heures de vol [UTC]	Altitude ¹	Niveau ²
DIMONA	4	25/06/1999	12.32 - 16.38	270	2
		02/07/1999	07.52 - 09.90	295	2
	6	16/07/1999	11.57 - 15.63	299	2
		17/07/1999	07.87 - 10.10	262	2
		17/07/1999	14.98 - 18.40	322	2
	18/07/1999	07.98 - 10.38	270	2	
ARAT		17/07/1999	11.24 - 13.97	847	3
			15.18 - 17.82	906	3
	8	07/29/1999	11.10 - 13.63	830	3
			14.86 - 17.15	781	3
			12.12 - 14.72	849	3

¹ : Altitude moyenne du vol (en mètre)

² : Niveau du modèle CHIMERE dans lequel évolue l'avion

TAB. 3.2 – Présentation des vols ESQUIF.

3.3 Outils statistiques d'évaluation d'un modèle

Supposons temporairement que les observations ne contiennent pas d'erreur systématique (biais instrumental ou de représentativité). Les différences constatées entre les simulations et les observations sont alors associées aux erreurs de modélisation. Plusieurs variables statistiques, dites *scores*, sont utilisées pour évaluer la qualité du modèle :

- ▷ Le *biais* mesure l'écart moyen entre les valeurs simulées Z_n^b et les valeurs observées Y_n^o . S'il est positif, le modèle surestime en moyenne les concentrations, s'il est négatif, le modèle les sous-estime. Ce biais peut être nul, ce qui signifie non pas que le modèle ne commet pas d'erreur, mais que le modèle ne commet pas d'*erreur systématique* (i.e. qui se répètent). Il s'écrit :

$$BIAIS = \langle Z_n^b - Y_n^o \rangle \quad (3.4)$$

où $\langle . \rangle$ est l'opérateur moyenne sur un échantillon de N valeurs (N=123 pour une validation du modèle sur un été, correspondant à la comparaison d'une valeur horaire par jour du 1^{er} mai au 31 août) :

$$\langle . \rangle = \frac{1}{N} \sum_{n=1}^N . \quad (3.5)$$

- ▷ La *moyenne quadratique* des différences entre concentrations simulées et observées, dite "RMS" (pour Root Mean Square, moyenne quadratique) permet d'estimer l'erreur moyenne commise en valeur absolue. Elle s'écrit :

$$RMS = \sqrt{\langle (Z_n^b - Y_n^o)^2 \rangle} \quad (3.6)$$

- ▷ La *corrélation* donne une indication relative de la bonne simulation de la variabilité indépendamment de l'amplitude des phénomènes. Cette valeur statistique permet donc de mieux repérer les mesures et simulations dont la variabilité des fluctuations est très différente.

$$COR = \frac{\langle (Z_n^b - \langle Z_n^b \rangle)(Y_n^o - \langle Y_n^o \rangle) \rangle}{\langle (Z_n^b - \langle Z_n^b \rangle)^2 \rangle \langle (Y_n^o - \langle Y_n^o \rangle)^2 \rangle} \quad (3.7)$$

- ▷ Le pourcentage d'erreurs "raisonnables" est calculé comme le pourcentage, sur la saison, d'erreurs absolues inférieures à un certain seuil que nous nous fixons. Ce dernier est pris à 15 ppb pour la comparaison des pics journaliers d'ozone ou de dioxyde d'azote. Plus ce pourcentage, que nous appellerons désormais E15, est grand, plus le nombre de "grosses erreurs" commises par le modèle est faible.

3.4 Évaluation statistique de la qualité du modèle CHIMERE continental à l'aide des observations de surface

Nous comparons dans ce paragraphe les concentrations d'ozone simulées par CHIMERE continental avec celles observées pendant l'été 2001 (de mai à août inclus). La distribution de ces stations est montrée sur la figure 3.1. Les observations de surface sont en majeure partie confrontées aux valeurs simulées dans la première couche du modèle. Cependant, la résolution du modèle est telle

que la topographie est lissée. Ainsi, les stations de montagne sont souvent plus élevées dans la réalité que dans le modèle, et la pertinence d'une comparaison des mesures de ces stations avec les concentrations simulées une certaine couche est toujours discutable. Nous comparons généralement les observations d'altitude avec les concentrations simulées au deuxième ou au troisième niveau du modèle.

Pour donner une vue synthétique des niveaux d'ozone observés et simulés pendant cet été, la figure 3.5 présente des séries temporelles de concentration à 15 h UTC (heure qui correspond à peu près au pic journalier) pour cinq sites localisés en France : à Nantes, Fremainville (Nord-Ouest de Paris), Saints (Nord-Est de Paris), dans les Vosges, et au Pays-bas (NL-232). Les mesures montrent une grande variabilité d'un jour à l'autre, résultant principalement de la variabilité naturelle des conditions météorologiques. En effet les émissions anthropiques (dues à l'activité humaine) de polluants ont un cycle diurne assez stable d'un jour à l'autre.

La période estivale 2001 a été marquée par des conditions météorologiques (températures souvent élevées, supérieures à $30^{\circ}C$), telles que de nombreux dépassements d'information et de recommandation (90 ppb) pour l'ozone ont été enregistrés. Ils sont regroupés en trois long épisodes de pollution continentale : 22-26 Juin avec développement d'un panache très concentré autour de l'agglomération de Paris (comme sur la station FREMAINVILLE où les niveaux d'ozone atteignent 100 ppb), 25-31 Juillet, 22-26 août. Il est intéressant de noter que de fortes concentrations d'ozone (proches de 80 ppb) ont été observées le 2 août sur la station d'altitude VOSGES-NORD, alors que sur les autres stations les concentrations sont plus faibles (inférieures à 60 ppb). Ces fortes concentrations sont probablement la conséquence d'un transport en altitude de l'ozone produit dans le nord de l'Europe quelque jours auparavant.

Ces séries temporelles suggèrent que le modèle est généralement capable de suivre les événements épisodiques de fortes concentrations d'ozone, non seulement sur des sites proches de grandes villes comme FREMAINVILLE mais aussi dans des villes comme Nantes, ou encore sur des sites moins soumis aux émissions comme Vosges-Nord. Autrement dit, il respecte non seulement la variabilité temporelle mais aussi la variabilité spatiale des concentrations d'ozone. Néanmoins, les profils observés et simulés sur Nantes-Jardin révèlent une des principales erreurs du modèle. En effet la titration locale de l'ozone par le monoxyde d'azote ne peut pas être correctement reproduite par le modèle continentale vu sa résolution. Les émissions de polluants primaires sont trop diluées.

La figure 3.6 regroupe, dans des histogrammes, les valeurs statistiques de vérification (biais, RMS, corrélation, erreurs raisonnables), définies dans le paragraphe 3.3. Ces dernières portent sur les concentrations journalières d'ozone à 15 h UTC pour chacune des 232 stations. Les histogrammes sont construits de la façon suivante : nous comptons tout simplement le nombre de stations, dont la valeur statistique associée est dans un certain intervalle. Les tableaux 3.3 et 3.4 donnent le détail de toutes ces statistiques, station par station, et en moyenne sur l'ensemble des stations.

Le biais ne dépasse généralement pas 7 ppb et sa distribution est centrée sur une valeur moyenne positive de 3.2 ppb, ce qui est le signe d'une tendance du modèle à la surestimation. Il s'agit d'une caractéristique récurrente du modèle qui peut être imputable aux problèmes de résolution. En effet, la faible résolution horizontale ne permet pas de comparer les concentrations simulées avec celles observées à proximité de sources majeures, ces sources étant "diluées" dans de grandes mailles du

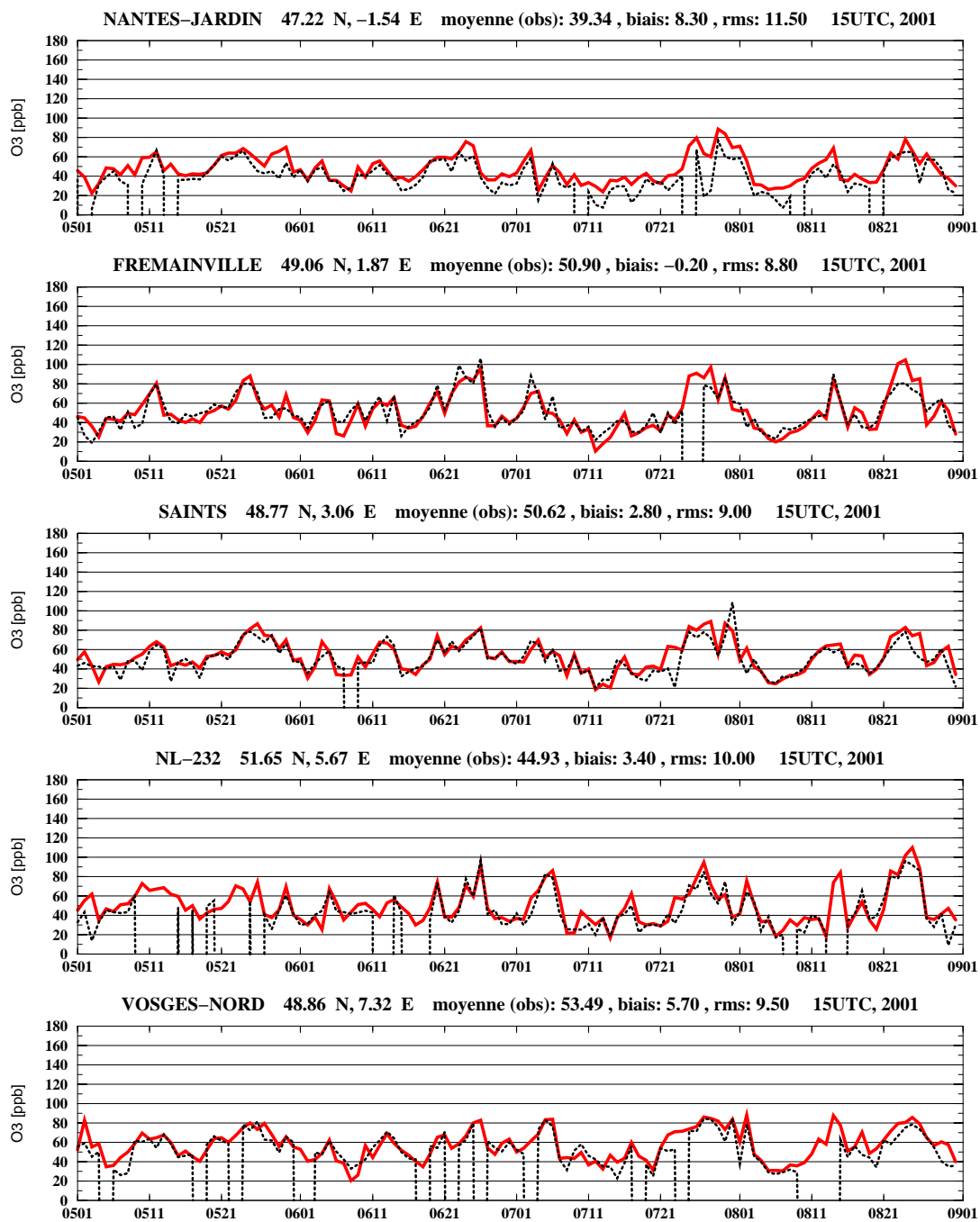


FIG. 3.5 – Séries temporelles des concentrations d’ozone observées (courbe noire pointillée) et simulées (courbe grise) par CHIMERE continental à 15 h UTC du 1^{er} Mai au 31 août 2001. L’unité des concentrations est en ppb.

3.4 Évaluation statistique de la qualité du modèle CHIMERE continental à l'aide des observations de surface

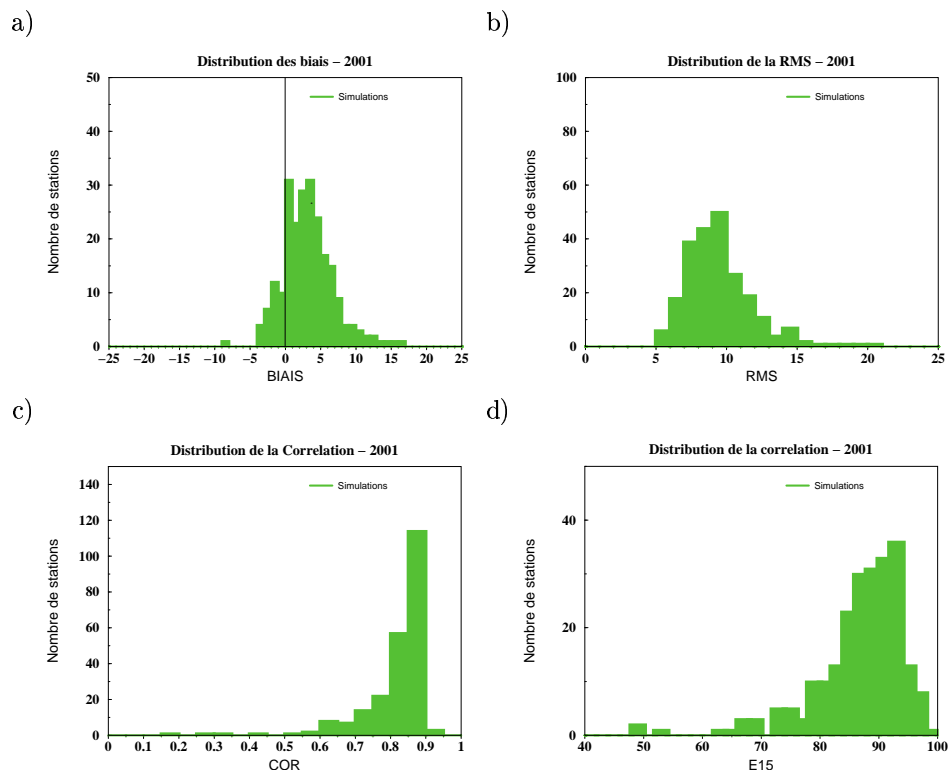


FIG. 3.6 – Statistiques calculées sur les différences entre les concentrations d’ozone observées et simulées par CHIMERE continental à 15 h UTC sur 232 stations de mesure européennes pendant l’été 2001. Les statistiques sont regroupées sous forme d’histogrammes : a) Biais b) RMS c) Corrélation d) Erreurs raisonnables. L’unité des concentrations est en ppb.

modèle. Dans ces cas, les concentrations du monoxyde d’azote sont sous-estimées, ce qui s’accompagne (par un effet de sous-estimation de la titration) d’une surestimation de l’ozone. L’effet de dilution peut également modifier le régime chimique en le faisant passer de ”sensible aux COV” à ”sensible aux NOx”. Cela conduit également à une surestimation générale des concentrations d’ozone, surtout au voisinage des grandes agglomérations.

La RMS de l’erreur est généralement située entre 5 et 12 ppb avec une moyenne générale de 9.3 ppb. L’analyse détaillée des stations pour lesquelles la RMS est élevée (supérieure à 9 ppb) montre que celles-ci correspondent à des stations situées dans les régions de terrains complexes (Alpes, côte Méditerranéenne, vallées du Rhône ou de la Saône). Les faibles résolutions verticale (accrue dans les versions ultérieures du modèle) et horizontales ne permettent pas de représenter de façon précise les écoulements de méso-échelle, écoulements complexes le long de côtes, dans les zones montagneuses, etc. La figure 3.7 donne un exemple de profil observé et simulé par CHIMERE continental sur une station méditerranéenne. Les dépassements du seuil d’information et de recommandation ne sont pas reproduits par le modèle.

Nous comparons ici des valeurs horaires (à 15 h UTC). Les scores sont en général moins bons

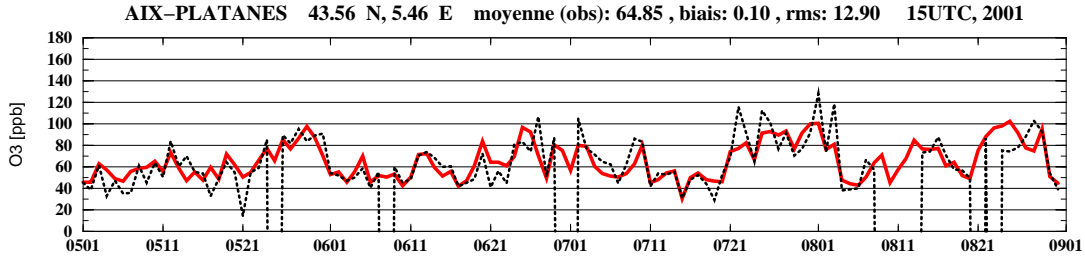


FIG. 3.7 – Série temporelle des concentrations d’ozone observées (courbe noire pointillée) et simulées (courbe grise) par CHIMERE continental à 15 h UTC du 1^{er} Mai au 31 août 2001 sur une station du réseau AIRMARAIX. L’unité des concentrations est en ppb.

que ceux calculés pour les pics journaliers. Cependant, ces scores se comparent bien avec les scores de simulation/prévision obtenus sur les pics journaliers d’ozone par d’autres groupes européens de modélisation/prévision. *Tilmes et ses collaborateurs* [2000] comparent les prévisions des pics d’ozone pour cinq modèles eulériens pour l’été 1999 sur les données allemandes. *Schmidt et et ses collaborateurs* [2001] donnent des scores un peu moins bons (autour de 10 ppb en moyenne) sur l’été 1998. Cette différence peut s’expliquer par une variabilité différente des situations météorologiques ou une amélioration des entrées météorologiques du modèle (le CEPMMT a augmenté les résolutions horizontales et verticale de leur modèle météorologique).

La corrélation est généralement comprise entre 0.8 et 0.9. Seules quelques stations ont une corrélation plus faible. Ces stations sont, en général, situées près des côtes (Mer du Nord, Océan atlantique), ou dans des régions à la topographie variable (Pyrénées, Alpes, Massif Central). Sur les stations d’altitude, la variabilité est faible et essentiellement due aux fluctuations de la concentration d’ozone dans la troposphère libre qui n’est pas prise en compte dans le modèle. En effet, la faible extension verticale du modèle ne permet pas de représenter des fluctuations comme les intrusions stratosphériques enrichissant la troposphère en ozone. De même les phénomènes convectifs profonds (situations orageuses ou frontales) sont mal pris en compte, les échanges entre le sommet du modèle et la troposphère supérieure se limitant à un simple coefficient d’échange paramétrisé en fonction de la précipitation convective. Des corrélations assez faibles sont aussi constatées sur des stations situées en bordure de panache, c’est le cas de la station MAGADINO.

Le pourcentage d’erreurs ”raisonnables” E15 se situe généralement au dessus de 85%, montrant le faible nombre de ”grosses erreurs” commises sur les pics. Les quelques stations ayant un taux important de ”grosses erreurs” (E15 inférieur à 80%) sont celles fréquemment soumises à des niveaux très élevés (notamment en Italie du Nord) sous l’effet probable des panaches urbains ou industriels. Là encore, la faible résolution permet d’expliquer la sous-estimation des extrêmes, car les panaches concentrés sont de taille généralement bien plus faible que celle de la maille du modèle.

Finalement, pour donner une idée des performances du modèle au cours de la journée, nous montrons sur la figure 3.8 la variation diurne des concentrations d’ozone observées et simulées. Le comportement de l’ozone est bien respecté mais une surestimation moyenne est constatée de 5 à 10 ppb la nuit. Cette surestimation provient de problèmes liés aux résolutions horizontales et verticale du modèle. Nous remarquons qu’elle est bien plus marquée sur les stations urbaines

TAB. 3.4 – Suite du tableau 3.3.

Station	Niv.	Long. [° N]	Lat. [° E]	BIAIS	RMS	COR	E15	Station	Niv.	Long. [° N]	Lat. [° E]	BIAIS	RMS	COR	E15
FRANCE								ROYAUME UNI							
CHARAVINES	1	5.5	45.4	6.4	11.1	0.80	83.6	BOTTESFORD	1	-0.8	52.9	-3.8	9.6	0.81	90.0
ST-ETIENNE-DIEME	1	4.4	45.5	2.4	13.5	0.74	81.0	ROCHESTER	1	0.6	51.5	-3.0	11.9	0.52	79.3
	1	4.5	46.0	2.4	11.5	0.78	80.8	REDCAR	1	-1.1	54.6	0.7	8.8	0.61	90.8
HOLLANDE															
ANNEMASSE-G	1	6.2	46.2	3.2	10.8	0.84	83.5	NL-934	1	6.3	53.3	3.1	8.1	0.75	93.8
ANNECY-NOVE	1	6.1	45.9	6.6	10.5	0.85	84.5	NL-538	1	5.0	52.8	2.4	7.5	0.81	96.6
CHAMBERY-LE	1	5.9	45.6	6.6	12.3	0.79	82.6	NL-107	1	6.0	51.1	4.4	10.1	0.89	93.3
ROCHES-DE-C	1	4.8	45.5	2.4	9.4	0.86	85.1	NL-918	1	5.6	52.9	3.9	8.3	0.85	97.1
ALENCON	1	0.1	48.4	3.0	8.0	0.86	93.4	NL-913	1	6.8	53.1	4.5	8.1	0.82	93.2
LISIEUX	1	0.2	49.1	12.2	14.8	0.85	70.0	NL-318	1	3.8	51.3	-1.2	9.1	0.80	89.6
OUISTREHAM	1	-0.2	49.3	16.6	19.5	0.62	54.5	NL-235	1	4.4	51.4	6.1	11.3	0.84	86.1
RANVILLE	1	-0.2	49.2	14.9	17.6	0.70	64.2	NL-230	1	5.2	51.5	6.8	12.7	0.84	84.6
CHERBOURG	1	-1.6	49.6	5.8	11.2	0.33	92.7	NL-227	1	5.6	51.3	5.3	11.0	0.87	88.8
ALLEMAGNE															
FORST	1	14.6	51.8	4.5	8.2	0.86	92.2	NL-722	1	6.7	52.1	6.5	11.6	0.82	90.3
WITTENBERGE	1	11.8	53.0	2.1	5.9	0.89	99.2	NL-724	1	5.7	52.0	4.4	9.2	0.88	91.6
PREMNITZ	1	12.3	52.5	0.4	6.0	0.89	99.2	NL-818	1	5.9	52.6	4.3	9.9	0.79	93.3
PRENZLAU	1	13.9	53.3	0.8	6.0	0.86	96.4	NL-301	1	3.9	51.7	-8.2	14.0	0.74	76.1
NAUEN	1	12.9	52.6	6.3	11.1	0.72	86.0	NL-232	1	5.7	51.6	3.4	10.0	0.87	92.2
BERLIN-MARI	1	13.4	52.4	2.0	6.9	0.88	95.5	NL-633	1	4.8	52.2	0.3	9.7	0.76	92.2
SUISSE															
BERLIN-FR.H	1	13.7	52.5	4.7	8.6	0.84	89.9	PAYERNE	1	6.9	46.8	2.0	7.4	0.87	94.6
WIESLOCH	1	8.7	49.3	8.3	11.5	0.89	81.3	CHAUMONT	3	7.0	47.0	4.4	9.1	0.85	92.4
WEIL-AM-RHE	1	7.6	47.6	3.7	13.2	0.76	74.6	JUNGFRAUJOC	5	8.0	46.5	-0.0	6.8	0.63	92.9
WELZHEIMER-SCHWARZWAL-FREUDENSTAD	1	9.6	48.9	1.9	9.6	0.87	87.1	LAEGEREN	1	8.3	47.5	6.2	9.3	0.90	90.2
KONSTANZ	1	9.2	47.7	7.3	11.7	0.83	80.2	RIGI	2	8.5	47.1	3.7	9.9	0.80	88.2
BADEN-BADEN	1	8.2	48.8	-1.9	9.5	0.90	88.3	TAENIKON	1	8.9	47.5	3.5	9.2	0.86	86.8
								ITALIE							
SCHWABEBISCH	1	9.2	48.4	-1.3	8.0	0.86	92.5	MAGADINO	1	8.9	46.2	9.6	16.1	0.65	67.0
BAD-WALDSEE	1	9.8	47.9	4.6	8.6	0.86	91.1	DAVOS	3	9.9	46.8	8.0	10.8	0.64	83.0
NEUSTADT-A	1	11.8	48.9	0.7	8.1	0.85	90.1								
BAD-REICHEN	1	12.9	47.7	11.5	14.2	0.80	69.9	PARMA-CITTA	1	10.3	44.8	10.6	15.6	0.64	71.3
GARM.-PART.	3	11.1	47.5	5.4	8.9	0.78	88.7	MODENA-20-S	1	10.9	44.6	2.5	12.6	0.77	75.4
BREMERHAVEN	1	8.6	53.6	7.8	11.6	0.80	83.9	BOLOGNA-CUC	1	11.3	44.5	13.4	20.1	0.59	50.5
GREBENAU	1	9.5	50.8	0.5	7.3	0.87	94.8	GHERARDI	1	11.7	44.9	-1.6	15.4	0.43	77.6
WITZENHAUSE	1	9.8	51.3	-3.1	8.1	0.89	94.6	COTIGNOLA	1	12.2	44.5	2.5	14.1	0.66	74.6
MOYENNES				3.2	9.3	0.80	89.6	FORLI-GIARD	1	12.0	44.2	4.7	10.1	0.81	89.6
								RIMINI-RICC	1	12.7	44.0	10.7	14.8	0.59	75.2
								CASTEL-SAN-	1	9.7	45.1	6.8	13.8	0.72	76.6

que sur les stations rurales, et notamment pendant les heures de fortes émissions d'oxyde d'azote correspondant aux pointes de trafic. Ces différences sont le reflet du manque de représentativité des stations urbaines, plus proches de sources d'émission, dont l'intensité est diluée dans les mailles de 50 km du modèle. Par rapport au jour, l'ozone est titré par le monoxyde d'azote et cette titration n'est pas compensée par la photolyse du dioxyde d'azote. Cette surestimation est en plus amplifiée par le manque de mélange. La couche mélangée peut être très fine (une dizaine de mètres) alors que l'épaisseur de la première couche est de 50 m, ce qui accentue encore la titration éventuelle, mais aussi les effets de dépôt sec. Par ailleurs, lorsque la couche limite est mince, les résultats du modèle sont plus sensibles à tous les paramètres, et il n'est pas étonnant que les erreurs résultantes soient accentuées.

3.5 Évaluation statistique de la qualité de CHIMERE régional à l'aide des observations de surface

3.5.1 Cas de la région d'Île-de-France

Nous présentons dans ce paragraphe une comparaison des concentrations d'ozone et de dioxyde d'azote simulées par CHIMERE régional, centré sur l'Île-de-France, avec les valeurs observées par les stations du réseau AIRPARIF pendant l'été 2001 (de mai à août inclus). La localisation des

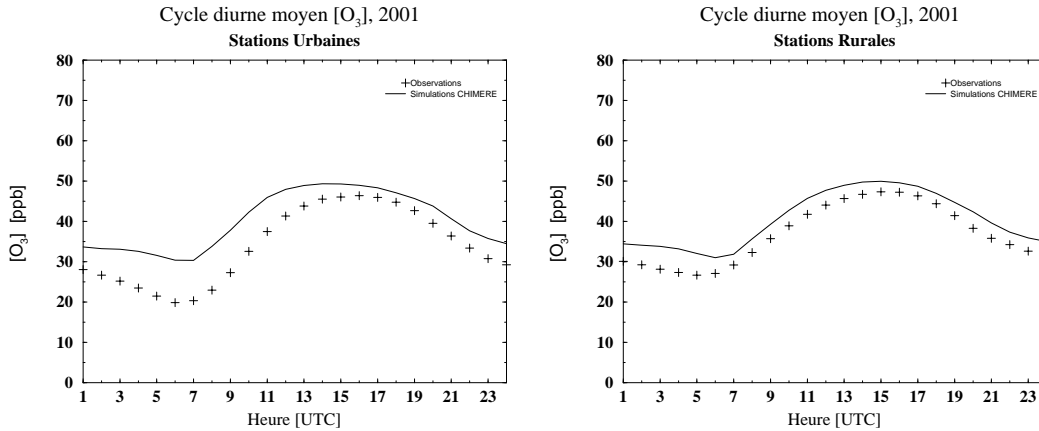


FIG. 3.8 – Profils diurnes moyens des concentrations d’ozone observées et simulées par le modèle CHIMERE continental sur les stations urbaines et sur les stations rurales. Unités des concentrations en ppb. Heure en UTC.

stations est représentée sur la figure 3.2.

L’ozone

Les figures 3.9 et 3.10 présentent respectivement les profils temporels observés et simulés par CHIMERE régional sur deux stations urbaines et sur les cinq stations rurales entourant l’agglomération de Paris. Le seuil d’information et de recommandation (90 ppb) a été dépassé plusieurs fois en 2001.

- ▷ L’épisode de pollution du 22 au 26 Juin est caractérisé par des concentrations fortes en zone urbaine (PARIS06, PARIS07). Le 23 et le 26 Juin, le panache urbain s’est développé au Nord-Ouest de l’agglomération parisienne (FREMAINVILLE, PRUNAY). Au coeur du panache, les concentrations atteignent 100 ppb. Des concentrations assez fortes sont observées sur l’ensemble des stations du réseau AIRPARIF.
- ▷ Les épisodes du 25-30 Juillet et du 22-26 août sont encore plus frappants. Des concentrations assez fortes, de l’ordre de 80 ppb, sont observées non seulement sur l’ensemble des stations parisiennes, mais aussi, comme nous l’avons remarqué dans le paragraphe 3.4, dans l’Ouest de la France et au Pays Bas.

Ces situations sont caractéristiques d’une *pollution* de type *continental*.

- ▷ Quelques épisodes courts sont à remarquer le 25 Mai et le 3 Juillet. Ces épisodes sont caractérisés par des niveaux de concentrations relativement faibles de l’ordre 60 ppb à l’entrée de l’agglomération en amont du panache (SAINTS, FONTAINEBLEAU). Tandis que des concentrations plus fortes sont constatées au Nord-Ouest de Paris, sous le vent, dans le panache de la ville (station FREMAINVILLE). Ces épisodes sont caractéristiques de *pollutions*, dites *locales*, dues aux émissions locales de la ville et de sa banlieue.

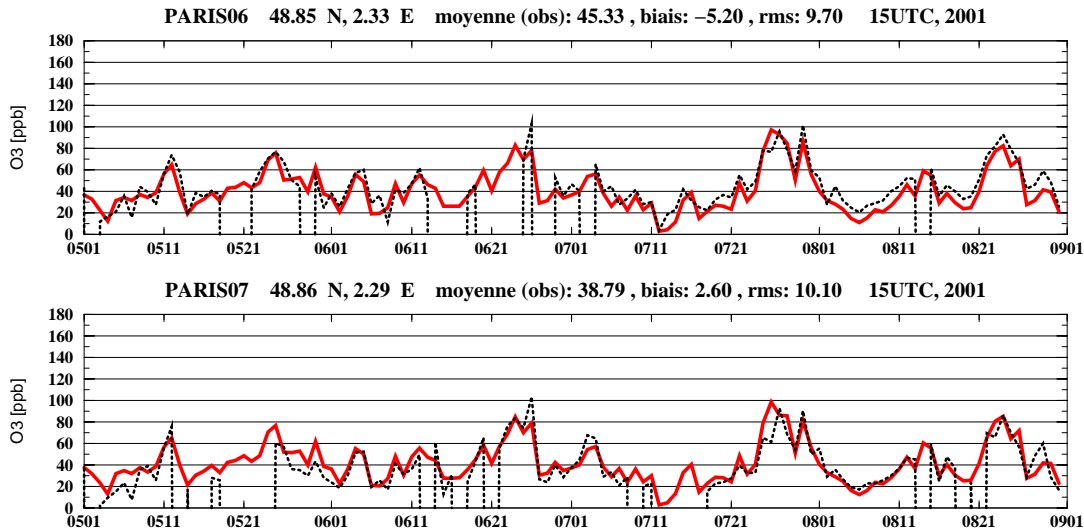


FIG. 3.9 – Séries temporelles des concentrations d’ozone observées (courbe noire pointillée) et simulées (courbe grise) par CHIMERE régional à 15 h UTC du 1^{er} Mai au 31 août 2001 sur deux stations urbaines proches (dans la même maille du modèle). L’unité des concentrations est en ppb.

Ces séries temporelles montrent que le modèle régional est capable de simuler les concentrations d’ozone aussi bien sur des stations urbaines, plus soumises aux émissions, que sur les stations rurales, plus éloignées. La comparaison avec les séries temporelles obtenues à l’échelle continentale sur les stations parisiennes (SAINTS et FREMAINVILLE) montre de faibles différences, prouvant ainsi la bonne représentativité de ces stations à 15 h UTC. Cependant, le modèle régional résout mieux les pics du 23 Juin, du 3 et 30 Juillet.

Le tableau 3.5 donne le détail de statistiques faites par la comparaison simulation/observation à 15 h UTC, station par station, et en moyenne sur l’ensemble des stations d’Île-de-France pour la période estivale 2001. Globalement le modèle a tendance à sous-estimer les concentrations d’ozone à 15 h UTC dans la zone urbaine et à les surestimer dans les zones rurales.

Ce tableau détaille également le pourcentage d’erreurs ”raisonnables” (E15) en filtrant les jours où soit l’observation soit la prévision du maximum est supérieure à un certain seuil (E15/seuil avec un seuil égal à 50, 75 ou 90 ppb). Nous notons en particulier que la performance du modèle diminue lorsque des valeurs élevées sont à prévoir ou sont prévues sur les stations rurales, suite à la plus forte sensibilité du modèle aux paramètres météorologiques lors des pointes [Vautard et al., 2000].

La comparaison de ces scores avec ceux obtenus avec le modèle continental sur les stations rurales parisiennes montre que les biais et la RMS sont généralement réduits. Par contre la corrélation n’est pas améliorée. Ceci n’est pas étonnant car les données météorologiques sont simplement interpolées sur le domaine régional, ce qui n’améliore pas leur résolution. Seules les émissions, assez stationnaires d’un jour à l’autre, sont plus résolues. Le modèle régional devient alors plus sensible aux paramètres météorologiques.

3.5 Évaluation statistique de la qualité de CHIMERE régional à l'aide des observations de surface

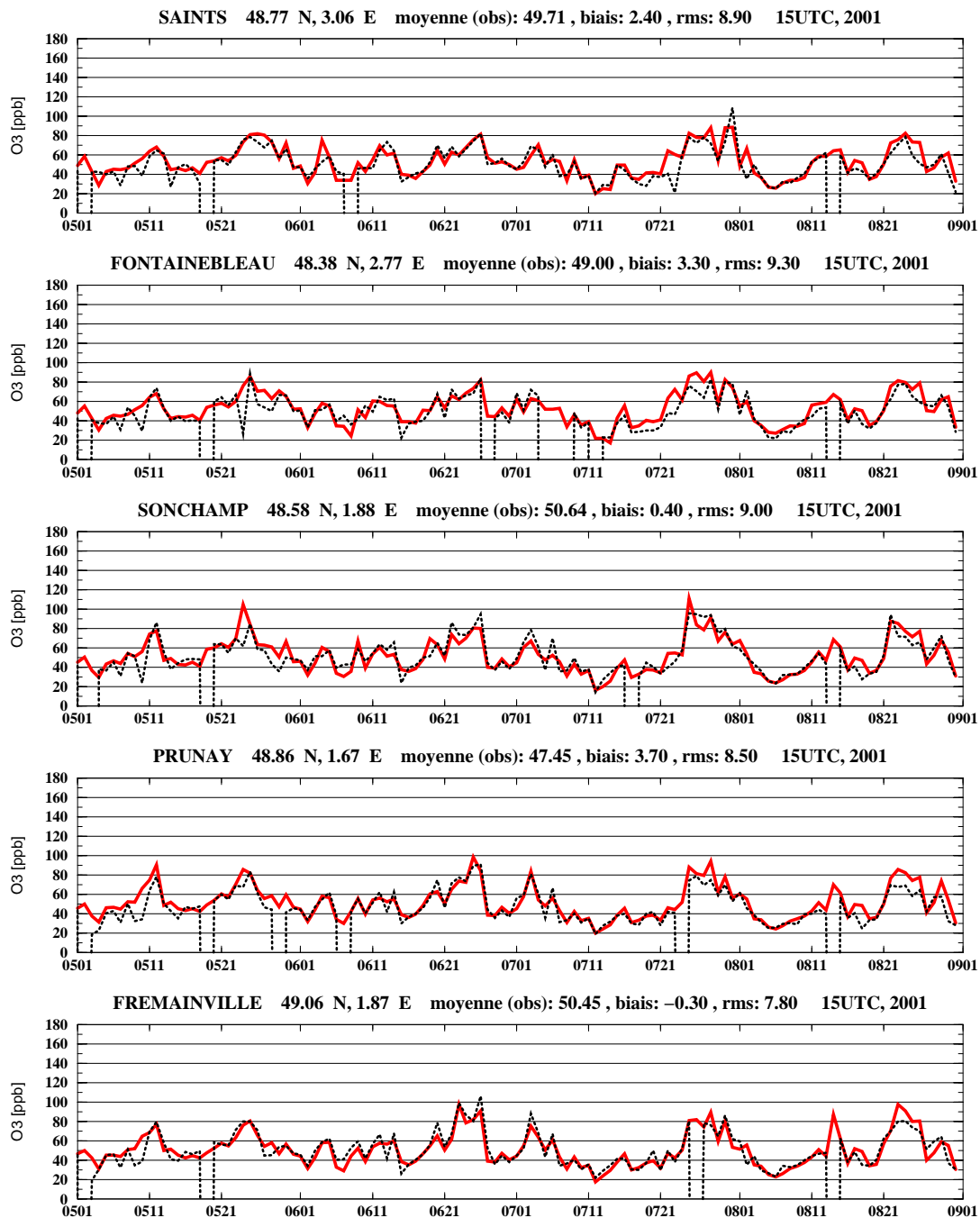


FIG. 3.10 – Séries temporelles des concentrations d’ozone observées (courbe noire pointillée) et simulées (courbe grise) par CHIMERE régional à 15 h UTC du 1^{er} Mai au 31 août 2001 sur les cinq stations rurales du réseau AIRPARIF en service à l’époque. L’unité des concentrations est en ppb.

TAB. 3.5 – Statistiques d'erreurs des comparaisons entre les concentrations d'ozone simulées par CHIMERE régional et les valeurs observées à 15 h UTC, station par station et en moyenne sur l'été 2001. Nous rappelons les scores de CHIMERE continental sur les stations aussi utilisées à l'échelle européenne. L'unité des concentrations est en ppb.

Station	Niv.	N	BIAIS	RMS	COR	E15	E15/50	E15/75	E15/90
PARIS06	1	105	-5.2	9.7	0.91	88.6	77.8	60.0	40.0
PARIS07	1	100	2.6	10.1	0.88	84.0	67.6	58.3	50.0
PARIS13	1	114	-6.2	10.7	0.91	81.6	67.4	61.5	60.0
PARIS18	1	111	-3.0	9.0	0.91	90.1	85.7	75.0	40.0
NEUILLY	1	119	-1.4	8.7	0.90	92.4	90.0	80.0	33.3
GENNEVILLIERS	1	119	-1.0	8.2	0.91	93.3	90.0	86.7	75.0
GARCHES	1	119	-2.7	9.5	0.90	89.9	85.5	64.7	50.0
AUBERVILLIERS	1	117	-2.8	8.2	0.92	94.9	93.5	90.0	80.0
VITRY	1	118	0.3	10.4	0.86	85.6	76.1	81.8	80.0
MONTGERON	1	120	5.3	11.0	0.88	83.3	76.2	50.0	44.4
CHARTRES-FULBERT	1	97	-1.8	9.3	0.85	89.7	84.0	91.7	100.0
CRETEIL	1	42	4.3	10.4	0.81	90.5	84.2	50.0	—
TREMBLAY	1	116	-2.2	8.7	0.90	94.0	93.0	92.9	100.0
MANTES	1	114	2.4	8.0	0.89	92.1	89.6	86.7	66.7
LES-ULIS	1	118	1.7	9.3	0.88	90.7	88.3	83.3	57.1
CERGY	1	120	0.8	8.6	0.90	92.5	85.5	84.2	60.0
MONTGE-EN-GOELE	1	118	2.1/1.2	8.1/8.5	0.90/0.88	92.6/90.7	86.9	70.6	33.3
SAINTS	1	119	2.8/2.4	9.0/8.9	0.85/0.85	90.2/89.9	85.7	80.0	0.0
FONTAINEBLEAU	1	114	4.0/3.3	9.9/9.3	0.84/0.84	89.0/92.1	88.7	69.2	100.0
SONCHAMP	1	118	0.6/0.4	9.2/9.0	0.87/0.87	92.6/93.2	88.5	88.2	71.4
PRUNAY	1	117	3.5/3.7	9.2/8.5	0.89/0.89	87.5/90.6	84.2	81.2	75.0
FREMAINVILLE	1	119	-0.2/-0.3	8.8/7.8	0.89/0.90	86.8/94.1	90.8	88.2	50.0
NOISY-SUR-OISE	1	92	1.2/-0.1	8.1/7.9	0.91/0.91	93.5/93.5	89.1	91.7	83.3
MOYENNE			0.1	9.1	0.88	90.3	84	76	61

La figure 3.11 montre la variation diurne des concentrations d'ozone observées et simulées. Le comportement de l'ozone est bien respecté mais une surestimation moyenne est constatée de 5 ppb la nuit sur les stations rurales. Cette surestimation a déjà été discutée lors de l'évaluation du modèle continental. Il faut noter cependant que, cette fois-ci, les problèmes de représentativité ne sont pas visibles sur les stations urbaines pourtant plus proches des sources d'émissions. L'explication de ce résultat est simple : la zone d'émission des oxydes d'azote est concentrée sur l'agglomération parisienne (d'extension spatiale d'environ 80 km²) et plusieurs mailles du modèle sont totalement consacrées à ces émissions, limitant ainsi leur dilution. Dans les zone rurales, les centres d'émission sont de petite extension et donc plus dilués dans des mailles de 6x6 km².

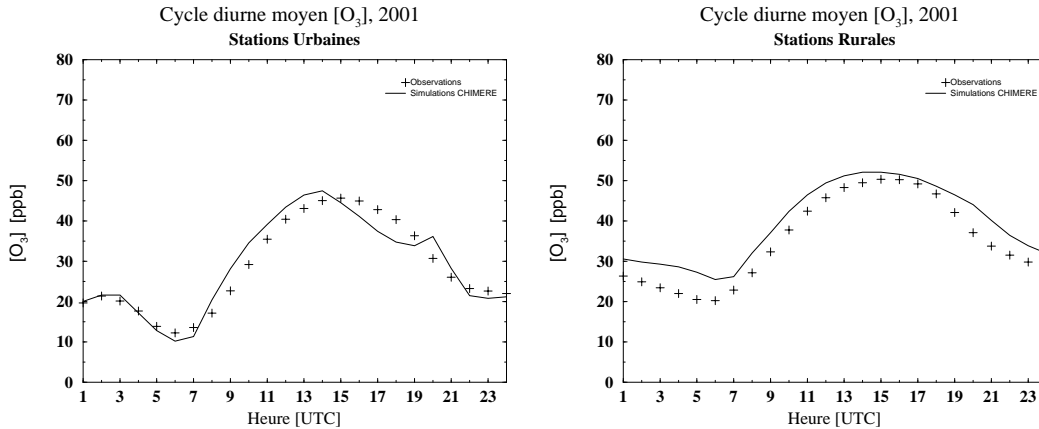


FIG. 3.11 – Profils diurnes des concentrations d'ozone observées (signes plus) et simulées (courbe) par le modèle CHIMERE régional sur les stations urbaines et rurales. L'unité des concentrations est en ppb, celle de l'heure en UTC.

L'anomalie constatée entre 15 et 21 h UTC sur l'ensemble des stations urbaines correspond à une mauvaise modulation temporelle des émissions. En effet, le cycle journalier des émissions de monoxyde d'azote présente normalement deux maximums : l'un le matin à l'heure de pointe de trafic, responsable de la forte titration de l'ozone vers 6-7 UTC ; l'autre le soir, plus amorti, correspondant au déplacement plus étalé dans le temps de la population active. Ce cycle est assez bien reproduit dans le modèle le matin, mais le pic est trop abrupte le soir, causant une sous-estimation de l'ozone (surestimation du monoxyde d'azote) entre 15 et 19 h UTC, suivi d'une légère surestimation du même composé (sous-estimation du monoxyde d'azote) à 20 h UTC. Une modification du profil d'émission a été effectuée en se basant sur les mesures des stations de trafic. Cette nouvelle version a été implémentée dans les versions ultérieures du modèle.

Le dioxyde d'azote

Le dioxyde d'azote, qui est plus problématique en hiver se manifeste aussi en été sous forme de pics matinaux dans l'agglomération. Il a connu deux journées de dépassement du seuil d'information et de recommandation en 2001 (seuil de $200 \mu\text{g}/\text{m}^3$, équivalent à environ 100 ppb), le 27 Juillet et le 24 août. La figure 3.12 montre les variations des concentrations de dioxyde d'azote observées et simulées à 9 h UTC sur deux stations localisées dans l'agglomération de Paris et sur une station située à l'extérieur de Paris. Les concentrations de dioxyde d'azote ne dépassent généralement pas 20 ppb en dehors de l'agglomération. Le modèle montre des difficultés à reproduire convenablement les variations du polluant sur la zone urbaine. Il sous-estime notamment largement les dépassements du seuil d'information. Notons que ces dépassements de seuil sont très localisés. Par exemple la station PARIS07, qui est très proche de la station NEUILLY, mesure des concentrations 20 ppb plus faibles le 24 août. En mai, le modèle surestime fortement les concentrations de dioxyde d'azote sur l'ensemble des stations parisiennes à cause de la présence de nuages. En effet, en condition nuageuse, de fortes erreurs sont souvent constatées sur ce polluant primaire, car le mélange vertical est très simplifié dans le modèle. Le problème est d'une extrême complexité car la variabilité des nuages de la couche limite est grande et généralement mal représentée dans les modèles météorologiques.

La figure 3.13 montre les cycles diurnes moyens calculés sur les stations urbaines ou rurales. Nous remarquons que la variation diurne du dioxyde d'azote est assez mal reproduite par le modèle sur la zone urbaine. En effet les pics matinaux, qui apparaissent généralement vers 7 h UTC, sont fortement surestimés. Cette surestimation peut avoir plusieurs origines : une surestimation des émissions et/ou un mauvais diagnostic de la hauteur de mélange. Un trop faible mélange vertical favorise une accumulation du dioxyde d'azote dans la première couche du modèle. De plus, la faible résolution verticale du modèle ne permet pas de prendre en compte correctement le gradient vertical relativement fort dans les premières couches de l'atmosphère. Ceci est probablement responsable des sous-estimations constatées lors de la croissance de la couche limite entre 8 et 11 h UTC.

Les cycles diurnes simulés du dioxyde d'azote présentent, comme les cycles diurnes d'ozone simulés, une anomalie en soirée. En effet, les émissions de monoxyde d'azote ne sont pas assez étalées dans le temps dans le modèle. Une surestimation des émissions de monoxyde d'azote conduit à une titration plus importante de l'ozone et ainsi qu'à une production trop forte de dioxyde d'azote. Ceci explique probablement la légère surestimation de ses concentrations entre 15 et 18 h UTC. La chute trop rapide des émissions de monoxyde d'azote est responsable quant à elle des sous-estimations

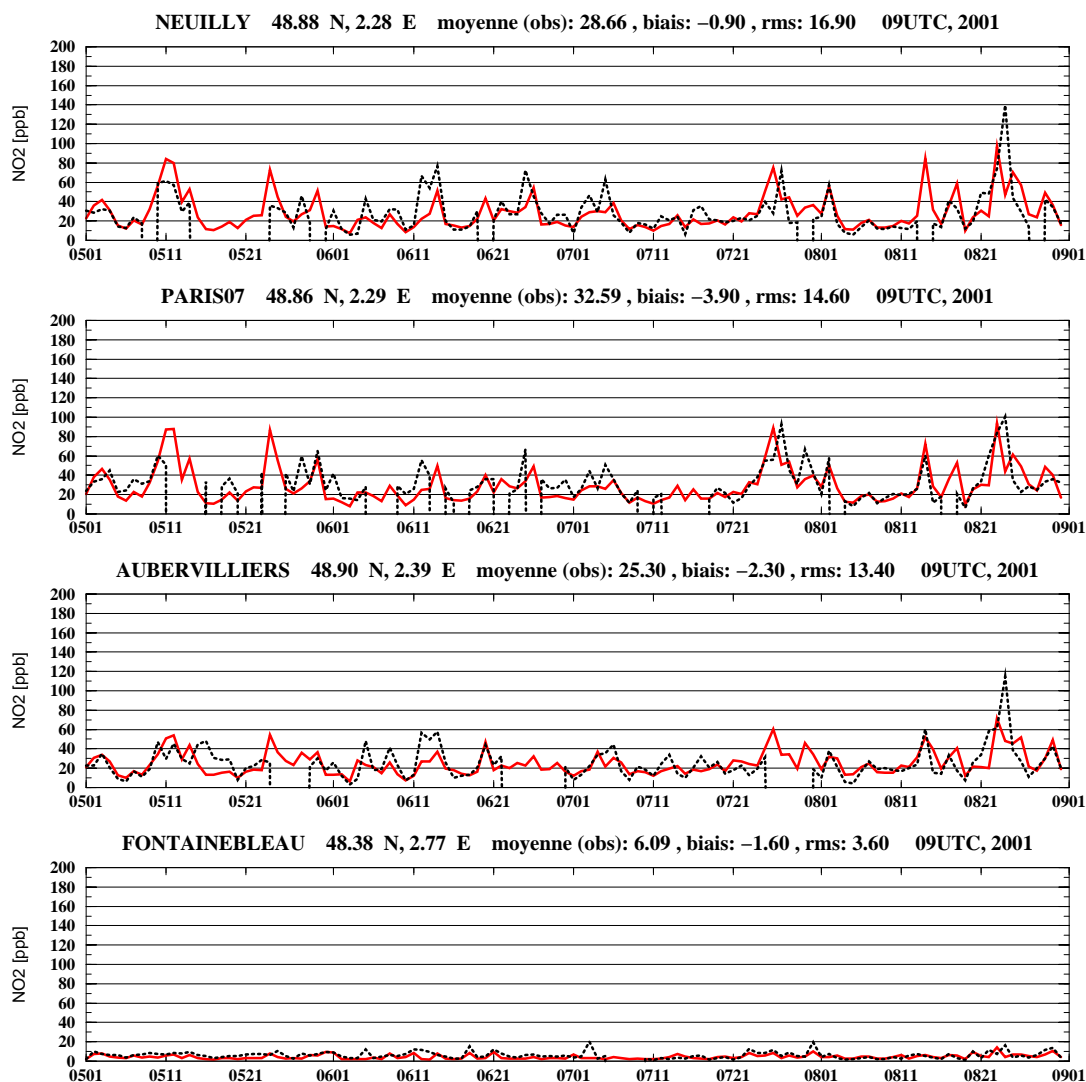


FIG. 3.12 – Séries temporelles des concentrations de dioxyde d'azote observées et simulées par CHIMERE régional à 9 h UTC du 1^{er} Mai au 31 août 2001 sur trois stations urbaines (NEUILLY, PARIS07, AUBERVILLIERS), et une station rurale (FONTAINEBLEAU) du réseau AIRPARIF. L'unité des concentrations est en ppb.

constatées à partir de 19 h UTC.

Le tableau 3.6 donne le détail des statistiques obtenues lors de la comparaison des concentrations de dioxyde d'azote simulées et les valeurs observées à 7 h UTC (qui correspond à peu près au pic matinal des concentrations de ce polluant), station par station, et en moyenne sur l'ensemble des stations sur la période estivale 2001. Nous remarquons que sur l'ensemble des stations intra-muros, éloignées de quelques kilomètres, les biais sont très variables d'une station à une autre, passant

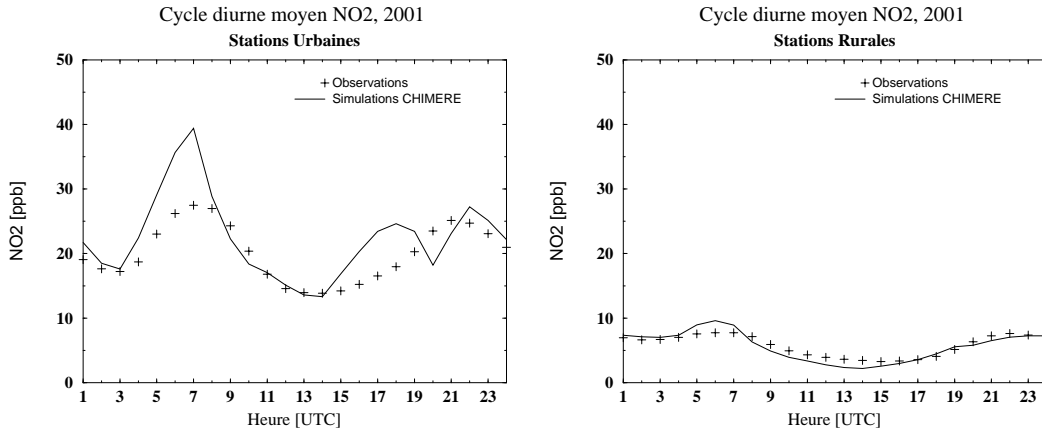


FIG. 3.13 – Profils diurnes des concentrations de dioxyde d'azote observées (signes plus) et simulées (courbe) par le modèle CHIMERE régional. L'unité des concentrations en ppb, l'heure étant en UTC

parfois du simple au double (ex : GENNEVILLIERS, NEUILLY). Ceci montre le caractère très localisé des concentrations de dioxyde d'azote. Les corrélations, de l'ordre de 0.7, indiquent que le modèle reproduit assez bien la variabilité des concentrations du dioxyde d'azote à 7 h UTC. Le modèle commet cependant de "grosses erreurs" (Cf. E15), et particulièrement lorsque de fortes concentrations sont observées ou simulées (Cf. E15/50, E15/75, E15/90). En effet, environ 30% des erreurs sont supérieures à 15 ppb, et environ 80% des erreurs sont supérieures au même seuil quand les concentrations observées ou simulées dépassent 50 ppb. Les dépassements de 90 ppb sont pratiquement toujours soit simulés et non observés, soit observés mais non reproduits par le modèle.

TAB. 3.6 – Statistiques d'erreurs de comparaison entre les concentrations de dioxyde d'azote simulées par CHIMERE régional et les valeurs observées à 7 h UTC, station par station et en moyenne sur l'été 2001. L'unité des concentrations est en ppb.

Station	BIAIS	RMS	COR	E15	E15/50	E15/75	E15/90
GARCHES	16.2	20.6	0.75	53	8	0	0
ARGENTEUIL	-0.1	10.5	0.77	84	50	50	—
ISSY	14.7	19.6	0.79	57	13	0	0
NEUILLY	17.1	21.2	0.83	54	8	6	0
PARIS07	15.6	20.0	0.82	53	9	7	0
GENNEVILLIERS	8.9	13.2	0.84	80	30	25	0
PARIS06	19.7	23.0	0.84	42	0	0	0
PARIS18	13.4	17.6	0.80	62	27	20	33
PARIS13	20.2	23.7	0.75	37	5	9	0
ST-DENIS	12.4	16.1	0.80	68	0	0	0
VITRY	13.9	19.0	0.72	57	11	0	0
AUBERVILLIERS	13.2	16.6	0.78	65	22	66	—
PARIS12	18.0	21.9	0.74	51	9	0	0
IVRY	16.1	19.9	0.78	57	9	10	0
EVRY	1.2	7.8	0.77	93	—	—	—
BOBIGNY	9.6	14.2	0.70	77	25	0	0
MONTGERON	6.5	10.6	0.73	87	0	—	—
TREMBLAY	4.5	10.4	0.61	89	0	—	—
CERGY	0.8	7.2	0.72	94	—	—	—
CACHAN	17.5	21.4	0.78	49	5	0	0
PARIS01	20.5	23.4	0.81	39	0	0	0
LA-DEFENCE	9.9	15.1	0.84	72	25	20	0
VERSAILLES	6.9	12.4	0.80	84	50	0	—
MELUN	-2.6	6.5	0.64	96	—	—	—
FONTAINEBLEAU	0.6	4.3	0.46	100	—	—	—
MANTES	-0.8	5.6	0.72	95	—	—	—
RAMBOUILLET	2.0	4.8	0.70	99	—	—	—
MOYENNE	10.2	15.1	0.75	70	15	11	2

3.5.2 Un terrain complexe : Bassins de Marseille et de Berre

La confrontation des simulations du modèle CHIMERE régional centré sur l'Île-de-France, avec les observations de surface a montré que le modèle reproduit assez correctement les concentrations d'ozone et moins bien les concentrations de dioxyde d'azote. Nous proposons dans ce paragraphe d'étudier les performances du même modèle sur la région des bassins de Berre et de Marseille. Cette région, au relief variable, au bord de la mer méditerranéenne est le siège de processus dynamiques complexes. Si l'utilisation de données météorologiques de grande échelle est justifiée sur la région de l'Île-de-France, loin des côtes maritimes et sans variation significative de la topographie, elle peut être cause d'erreurs sur la région méditerranéenne.

L'ozone

Les pics de concentrations d'ozone sont souvent observés entre 11 et 17 h UTC. Par un effet de brise, en fin d'après-midi, les masses d'air sont ramenées vers l'intérieur des terres. Nous avons choisi de présenter les séries temporelles des concentrations observées et simulées à 15 h UTC, sur une station au bord de la mer près de l'étang de Berre (PORT-DE-BOUC) et deux autres stations plus éloignées de la côte.

La figure 3.14 montrent donc les variations journalières des concentrations d'ozone observées et simulées à 15 h UTC sur ces trois stations pendant la période de Mai à Août 2001. Les épisodes de forte pollution, constatés à l'échelle européenne, sont également observés sur cette région. Nous remarquons que le modèle reproduit relativement bien la variabilité du polluant. Cependant, le modèle fait de très grosses erreurs lors d'épisodes de fortes concentrations. Le 25 Juin, par exemple, le panache, assez localisé, est simulé un peu trop au sud-ouest de la station AURIBEAU, le 26 un peu trop à l'ouest. Le 23 août, un dépassement de seuil est simulé sur la station AIX-PLATANE, alors qu'il est observé sur les stations PORT-DE-BOUC et AURIBEAU. En fait, ce jour-là, de fortes concentrations, de plus de 90 ppb, sont observées sur l'étang de Berre, alors que le modèle simule des concentrations 30 ppb plus faibles. Le modèle décale le panache vers l'Est. De manière générale, le modèle montre beaucoup de difficultés à positionner les panaches dans cette région. La circulation des masses d'air est difficile à modéliser, elle est déterminée par la combinaison de phénomènes de brise, de vents de pente, et de canalisation par les vallées. Le modèle sous-estime par exemple souvent les concentrations sur les stations de la ville d'Aix, car celle-ci se situe régulièrement en bordure de panache. Le modèle continental, qui moyenne sur 50x50 km² les concentrations, masque ces erreurs de positionnement. C'est pourquoi, il simule mieux les concentrations observées à la fin du mois d'août, que le modèle régional (comparaison faite avec la figure 3.7).

Le tableau 3.7 donne les statistiques de comparaison simulations/observations des concentrations d'ozone à 15 h UTC. Les scores sont, en moyenne, moins bons que ceux obtenus par CHIMERE régional sur l'Île-de-France. Les biais sont négatifs sur les stations autour de la zone industrielle de l'étang de Berre. Ils sont positifs en dehors de cette zone, et particulièrement sur les stations urbaines (Marseilles, Toulon, Avignon, Arles, Nîmes). Les corrélations inférieures à 0.7, et les pourcentages d'erreurs raisonnables inférieurs à 80% correspondent aux stations les plus proches de la mer, soumises à des effets de brise mal modélisés par le modèle.

Une comparaison des scores obtenus avec CHIMERE continental (Cf. tableau 3.3) sur quelques stations méditerranéennes montre que le modèle régional n'est pas beaucoup plus performant. Il détériore même la RMS sur la station AIX-PLATANES (modèle continental : 12.9 ppb ; modèle

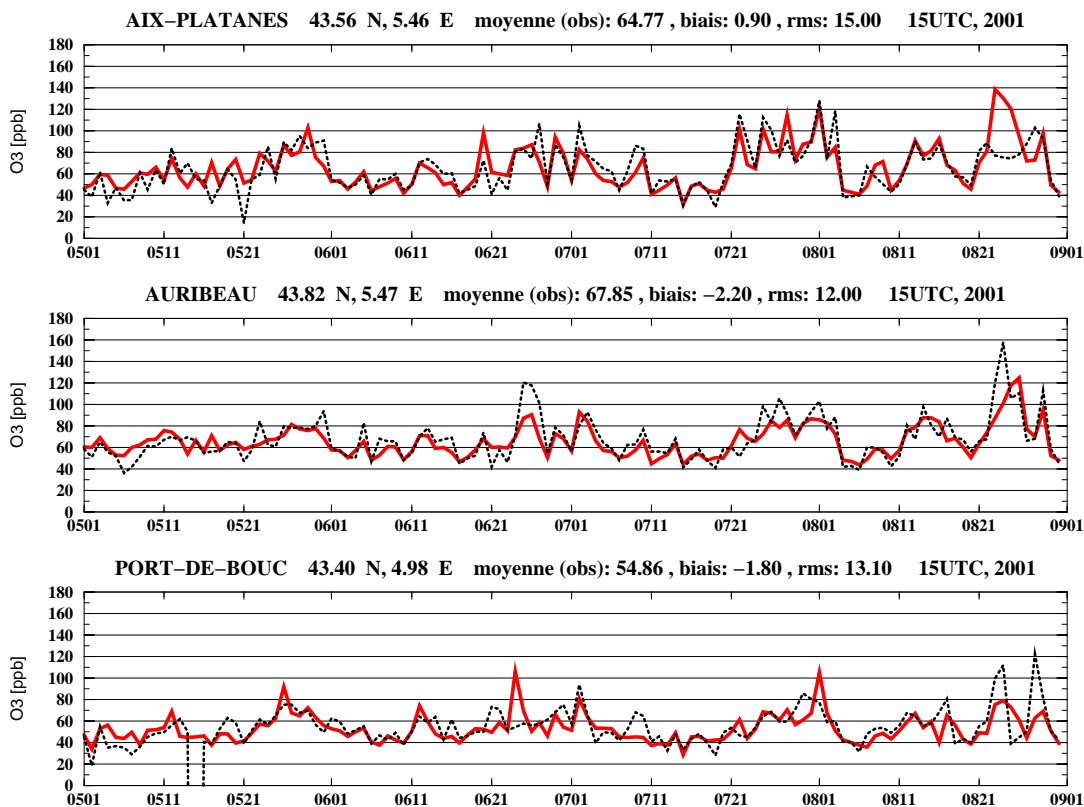


FIG. 3.14 – Séries temporelles des concentrations d’ozone observées (courbe noire pointillée) et simulées (courbe grise) par CHIMERE régional à 15 h UTC du 1^{er} Mai au 31 août 2001 sur trois stations des réseaux AIRMARAIX ou AIRFOBEP. Unités des concentrations en ppb.

régional : 15 ppb). Il faut cependant noter que la RMS est très sensible aux grandes erreurs singulières. Les erreurs commises lors de l’épisode de la fin Août augmente fortement cette RMS. Sans prendre en compte cet épisode, la RMS des erreurs sur AIX-PLATANE passe de 12.9 à 12.6 ppb avec CHIMERE continental et de 15 à 12.3 ppb avec CHIMERE régional, montrant tout de même une légère amélioration de ce score. Nous pouvons également remarquer (par comparaison avec la figure 3.7) que, sur la même station, les pics de l’épisode de fin Juillet sont mieux résolus par le modèle régional que par le modèle continental.

La figure 3.15 donne les profils diurnes moyens des concentrations d’ozone observées et simulées par CHIMERE régional, calculés sur plusieurs stations localisées dans ou autour des principales agglomérations (Avignon, Aix, Marseille, Toulon) et autour de l’Étang de Berre. Nous constatons que le profil diurne d’ozone est globalement surestimé sur les stations urbaines. Nous retrouvons ici une des caractéristiques du modèle CHIMERE, qui a déjà été discutée lors de l’évaluation du modèle continental et régional centré sur l’Île-de-France. Cependant, contrairement à l’Île-de-France, les centres d’émission sont d’extension plus faibles (de quelques à quelques dizaines de kilomètres carré) et éparpillés sur le domaine. Ces émissions ont donc tendance à être diluées dans

TAB. 3.7 – Statistiques d'erreurs des comparaisons entre les concentrations d'ozone simulées par CHIMERE régional et les observations à 15 h UTC, station par station et en moyenne sur l'été 2001. L'unité des concentrations est en ppb.

Station	Niv.	BIAIS	RMS	COR	E15	E15/50	E15/75	E15/90
AUBAGNE-PENITENTS	1	0.5	16.4	0.61	76	63	44	30
AIX-CENTRE-ART	1	0.9	15.5	0.72	73	56	48	23
AIX-OUEST	1	-0.9	13.5	0.77	80	69	60	33
MARSEILLE-5AVENUES	1	8.1	15.3	0.66	69	45	35	28
MARSEILLE-MARGUERITE	1	2.4	12.2	0.69	83	71	59	33
LA-CIOTAT	1	3.3	14.3	0.66	77	60	58	43
TOULON-ARSENAL	1	5.9	13.6	0.69	77	61	50	16
LA-SEYNE-SUR-MERS	1	7.8	15.2	0.67	64	43	42	25
AVIGNON-MAIRIE	1	1.8	9.9	0.78	88	81	76	66
AVIGNON-ARROUSAIRE	1	4.4	10.4	0.81	87	83	73	25
LE-PONTET-MARGUERITE	1	4.1	11.9	0.76	85	82	82	60
NIMES-GARDIES	1	-2.4	8.4	0.85	91	82	77	83
BERRE	1	-2.9	15.1	0.55	74	48	34	0
PAGODE	1	-8.6	16.2	0.42	68	19	11	0
PORT-DE-BOUC	1	-1.8	13.1	0.61	79	60	30	14
ISTRES	1	-6.8	16.2	0.50	74	25	14	0
FOS-Carabins	1	-12.1	19.1	0.53	64	26	11	0
MIRAMAS	1	-7.0	17.3	0.59	78	50	36	20
VITROLLES	1	-6.8	17.0	0.55	75	51	41	27
MARIGNANE-Ville	1	-1.2	17.1	0.43	80	54	33	22
ST-MARTIN-de-CRAU	1	-3.4	9.5	0.80	89	77	76	100
ARLES	1	6.0	10.0	0.83	86	75	84	100
SALON	1	0.5	11.9	0.83	82	72	62	25
ROGNAC	1	-7.9	16.3	0.66	74	48	39	26
SAUSSET	1	5.2	17.8	0.55	75	58	43	22
BOUC-BEL-AIR	1	1.5	15.1	0.58	73	50	41	30
ROUSSET	1	1.6	14.2	0.77	78	64	59	46
SAZE	1	-1.3	8.7	0.81	91	86	81	50
VALLABREGUES	1	-2.3	9.1	0.82	91	82	78	75
COMTAT-VENAISSIN	1	8.2	12.2	0.80	75	61	58	100
AIX-PLATANES	1	0.9	15.0	0.74	78	63	61	39
TOULON-OLIVE	1	3.1	12.2	0.68	79	64	48	22
PENNES-MIRABEAU	1	-1.4	15.5	0.57	71	52	34	21
CUQUES-ALLAUCH	1	-1.6	13.9	0.67	75	59	41	33
VALLEE-HUVEAUNE	1	1.3	15.1	0.61	74	53	44	20
LA-VALETTE-LA-GARDE	1	4.0	11.3	0.74	82	69	52	16
MONTPELLIER-NORD	1	-0.1	8.6	0.85	91	80	73	66
MONTPELLIER-SUD	1	1.9	10.3	0.81	87	81	72	60
ST-ESTEVE	2	6.0	10.2	0.78	85	77	66	25
AURIBEAU	3	-2.2	12.0	0.81	83	75	62	41
SAINT-REMY	1	1.0	8.8	0.84	88	82	75	80
ST-MARIE-de-la-MER	1	10.0	14.4	0.67	66	38	30	50
AUPS-STE-BAUME	2	4.7	15.1	0.70	75	64	60	48
CADARACHE	1	2.6	10.5	0.83	87	83	81	81
MOYENNE		0.6	13.3	0.70	79	62	53	39

le modèle, qui sous-estime alors les concentrations d'ozone la nuit, et durant les heures de trafic.

Les concentrations d'ozone sont sous-estimées sur la région de Berre dans l'après midi et en soirée. Comme nous allons le voir dans le paragraphe suivant, ces sous-estimations sont accompagnées d'une surestimation des concentrations de dioxyde d'azote.

Le dioxyde d'azote

La figure 3.16 montre les profils diurnes moyens des concentrations de dioxyde d'azote observées et simulées par CHIMERE régional. Ces profils sont calculés sur les mêmes sites que la figure 3.15. Nous constatons que le profil diurne du dioxyde d'azote est globalement sous-estimé sur les stations urbaines, tandis que les concentrations de dioxyde d'azote sont surestimées autour de l'Étang de Berre.

Ces écarts peuvent avoir plusieurs origines : l'incertitude dans les émissions, un mauvais diagnostic de la hauteur de mélange et/ou un problème de représentativité des stations.

Ces profils sont tout à fait cohérents avec ceux obtenus pour l'ozone. Il semble que les concentrations du monoxyde d'azote sont surestimées sur la région de Berre et sous-estimées sur le reste

3.5 Évaluation statistique de la qualité de CHIMERE régional à l'aide des observations de surface

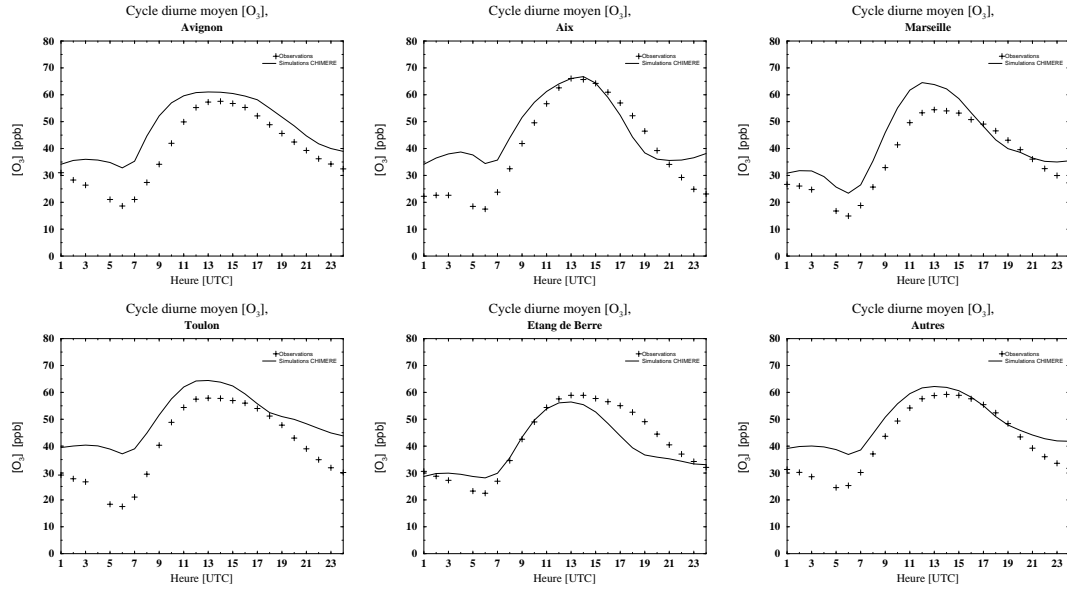


FIG. 3.15 – Profils diurnes moyens des concentrations d’ozone observées et simulées par le modèle CHIMERE régional sur quatre agglomérations, sur l’Étang de Berre et sur le reste des stations. Ils correspondent à des moyennes de profils calculés sur plusieurs stations les plus proches des sites. L’unité des concentrations est en ppb. L’heure est en UTC.

du domaine. La sous-estimation du monoxyde d’azote est probablement due à des problèmes de résolution du modèle. Les surestimations du dioxyde d’azote constatées sur la région de l’Étang de Berre proviennent sûrement de la combinaison de deux phénomènes : une couche de mélange simulée trop basse (due au contraste terre-étang) et une surestimation des émissions.

Finalement, le tableau 3.8 présente les statistiques d’erreur pour la comparaison simulations-observations des concentrations de dioxyde d’azote à 7 h UTC, station par station et en moyenne sur l’ensemble des stations. Nous retrouvons les principaux résultats :

- ▷ les scores sont moins bons que dans le cas de l’ozone.
- ▷ plus les concentrations simulées ou observées sont fortes, plus les capacités du modèle diminuent.
- ▷ le modèle sous-estime en général fortement les concentrations du dioxyde d’azote sur les stations dans ou près des agglomérations. Notons que ceci n’est pas vrai pour deux stations : MARSEILLE-5-AVENUES et MARSEILLE-MARGUERITE, montrant l’hétérogénéité des concentrations observées sur des stations relativement proches.
- ▷ le modèle surestime les concentrations du polluant sur la région de Berre

En outre, nous notons que les corrélations sont faibles, de l’ordre de 0.5, montrant la difficulté du modèle à représenter les variations journalières des concentrations du dioxyde d’azote. Le ca-

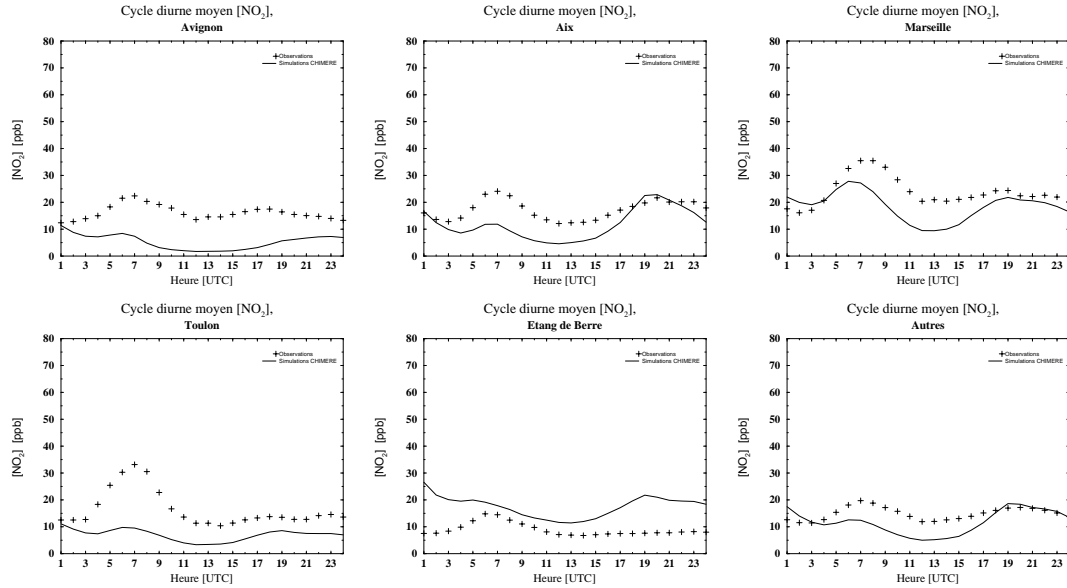


FIG. 3.16 – Profils diurnes des concentrations de dioxyde d’azote observées et simulées par le modèle CHIMERE régional sur quatre agglomérations, sur l’Étang de Berre et sur le reste des stations. Ils correspondent à des profils moyens calculés sur plusieurs stations les plus proches des sites. L’unité des concentrations est en ppb. L’heure est en UTC.

dastre d’émission utilisé est un cadastre préliminaire, fournit par la communauté de la campagne ESCOMPTE. Ce cadastre doit être amélioré dans l’avenir.

3.6 Évaluation statistique de la qualité de CHIMERE à l’aide des données aéroportées de la campagne ESQUIF

La comparaison des simulations de CHIMERE avec des *observations surfaciques* a montré la capacité du modèle à simuler les épisodes estivaux de pollution photochimique de 2001 à l’échelle continentale comme à l’échelle régionale.

Pour estimer la qualité du modèle à des niveaux supérieurs, moins soumis à l’influence directe des émissions, des mesures aéroportées sont indispensables. Ces mesures ne sont en général disponibles que pour des épisodes courts pendant des campagnes de mesure. La figure 3.17 montre les résultats d’une comparaison des concentrations d’ozone simulées et observées pendant le temps de vols des avions DIMONA ou ARAT de la campagne ESQUIF en 1999.

Globalement, le modèle continental, comme le modèle régional, repèrent les panaches urbains. Cependant, le premier ne reproduit pas systématiquement les structures fines des panaches urbains et sous-estime largement les concentrations de ces structures à cause de sa faible résolution horizontale. C’est le cas du 17 Juillet 1999 et du 18 Juillet 1999, où il détecte très difficilement le panache.

3.6 Évaluation statistique de la qualité de CHIMERE à l'aide des données aéroportées de la campagne ESQUIF

TAB. 3.8 – Statistiques d'erreurs des comparaisons entre les concentrations de dioxyde d'azote simulées par CHIMERE régional et observations à 7 h UTC, station par station et en moyenne sur l'été 2001. Unités des concentrations en ppb.

Station	Niv.	BIAIS	RMS	COR	E15	E15/50	E15/75	E15/90
MARSEILLE-PARADIS	1	-6.4	19.0	0.59	60	30	0	---
MARSEILLE-TIMONE	1	-2.7	18.8	0.33	56	0	---	---
MARSEILLE-PLOMBIERES	1	-28.3	34.6	0.38	19	4	0	0
MARSEILLE-RABATAU	1	-11.2	21.7	0.52	47	20	0	0
AUBAGNE-FOCH	1	-23.7	27.2	0.11	30	0	---	---
AIX-ROY-RENE	1	-15.2	18.8	0.50	47	0	---	---
AVIGNON-ROCADEOUIIS	1	-22.3	25.6	0.38	33	0	---	---
BERRE	1	-2.1	10.2	0.62	86	---	---	---
PAGODE	1	12.0	16.3	0.73	73	---	---	---
L-ILE	1	9.2	14.6	0.66	76	---	---	---
ISTRES	1	1.5	9.1	0.50	92	---	---	---
FOS-Carabins	1	7.0	12.6	0.66	81	0	---	---
VITROLLES	1	1.8	13.0	0.45	73	---	---	---
MARIGNANE-Ville	1	-2.1	9.4	0.77	86	---	---	---
PORT-DE-BOUC-EDF	1	-1.6	13.6	0.59	78	50	---	---
ST-MARTIN-de-CRAU	1	-2.2	8.2	0.43	91	---	---	---
ARLES	1	-8.1	11.4	0.51	82	---	---	---
SALON	1	-10.6	14.5	0.51	72	---	---	---
MARSEILLE-ST-LOUIS	1	-15.6	20.1	0.75	51	0	0	---
AUB-EST-PENITENTS	1	-5.1	10.5	0.51	85	---	---	---
AIX-CENTRE-ECOLE-ART	1	-9.1	13.5	0.45	76	---	---	---
AIX-OUEST-BOUFFANRES	1	-11.0	15.1	0.63	68	---	---	---
MARSEILLE-PRADOUS	1	-3.8	16.9	0.59	62	14	0	---
MARSEILLE-5-AVENUES	1	2.8	17.2	0.45	70	0	---	---
MARSEILLE-MARGUERITE	1	2.9	15.5	0.58	68	0	---	---
MARSEILLE-THIERS-NOA	1	0.4	17.9	0.38	60	0	---	---
TOULON-CHALUCET	1	-24.5	27.5	0.56	26	0	---	---
TOULON-ARSENAL	1	-23.5	26.6	0.59	29	0	---	---
LA-SEYNE-SUR-MERS	1	-19.6	23.6	0.56	36	---	---	---
AVIGNON-MAIRIENRES	1	-7.5	10.4	0.54	90	---	---	---
AVIGNON-ARROUSAIRE	1	-13.0	15.6	0.55	63	---	---	---
LE-PONTET-5-AVENUES	1	-16.5	19.2	0.60	48	---	---	---
MIRAMAS	1	1.3	7.4	0.47	96	---	---	---
VALLEE-HUVEAUNEAU	1	-5.7	12.3	0.56	81	---	---	---
ROGNAC	1	2.7	8.8	0.58	91	---	---	---
SAUSSET	1	7.7	14.6	0.58	80	0	0	---
BOUC-BEL-AIR	1	-2.3	9.1	0.73	90	---	---	---
BIVER	1	2.2	7.5	0.65	95	---	---	---
AUPS-STE-BAUME	2	-0.4	1.9	0.04	100	---	---	---
MOYENNES		-6.2	15.6	0.53	67	6	0	0

Le modèle régional, en revanche, reproduit assez correctement les concentrations d'ozone en altitude (niveau 2 ou 3 du modèle, Cf. ableau 3.2). Il est à noter cependant qu'il sous-estime largement les niveaux d'ozone le 30 Juillet 1999 sur tout le domaine de simulation. Il les sous-estime aussi le 17 et 18 Juillet 1999, mais seulement dans le panache. Ces sous-estimations sont probablement dues à une sous-estimation des précurseurs de l'ozone. En revanche, le modèle surestime les niveaux de fond le 2 Juillet 1999, due à un problème dans les conditions aux limites, provenant du modèle continental. Nous remarquons aussi un léger décalage dans le positionnement du panache du 18 Juillet 1999. Ce décalage est dû au fait que, ce jour-là, le vent horizontal était faible (inférieur à 3 m.s^{-1}) et que le modèle est particulièrement sensible au données météorologiques pendant les épisodes stagnants [Vautard et al., 2000].

La figure 3.18 montre les résultats d'une comparaison des concentrations du dioxyde d'azote simulées et observées pendant les vols du DIMONA. Le modèle continental, comme le modèle régional, sous-estime systématiquement le dioxyde d'azote. Le modèle régional suit cependant assez bien les variations spatiales du polluant.

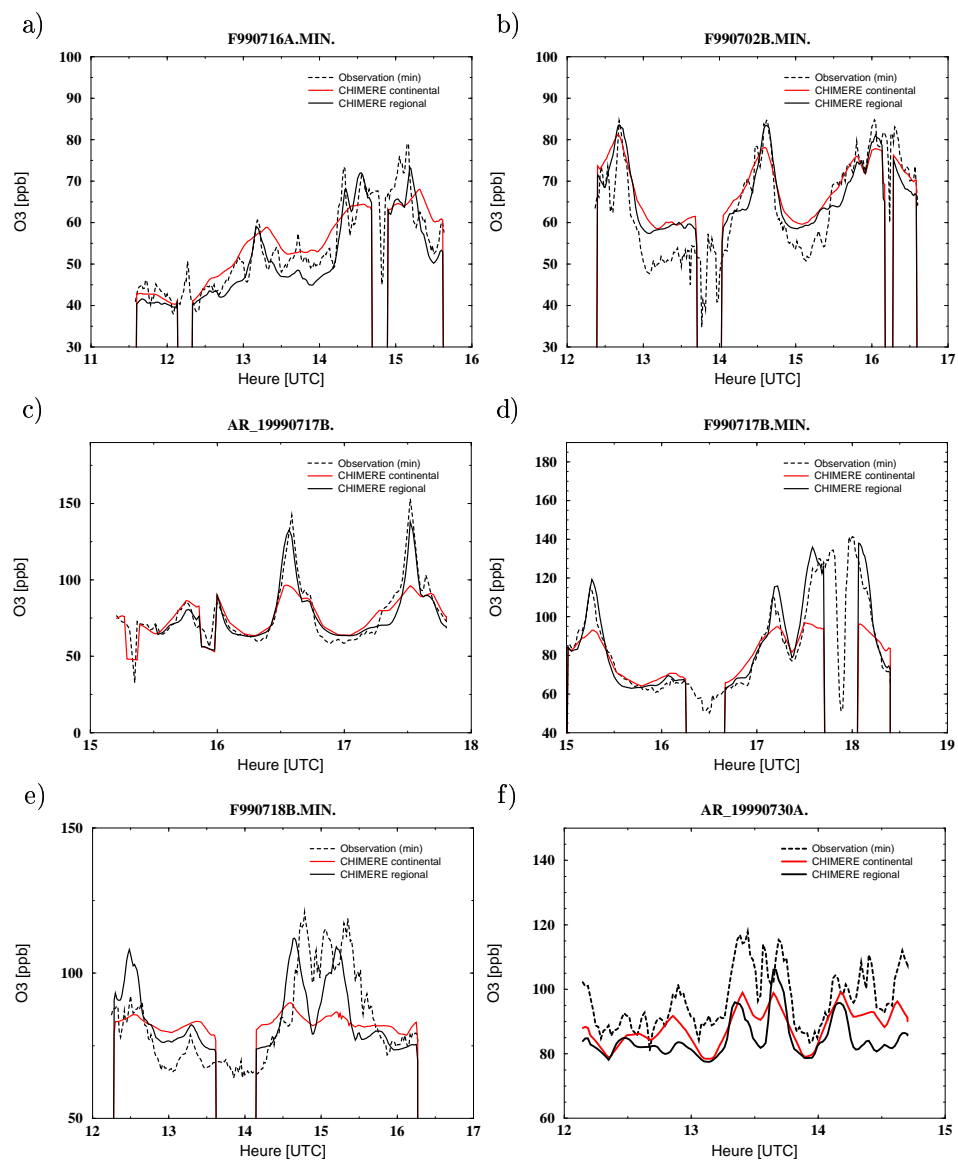


FIG. 3.17 – Séries temporelles des concentrations d’ozone mesurées (courbe pointillée) toutes les minutes par les avions DIMONA ou ARAT pendant trois POI, et simulées par les modèles CHIMERE continental (courbe pleine grise) et régional (courbe pleine noire). a) vol DIMONA du 16/07/1999 (POI 6), b) vol DIMONA du 02/07/1999 (POI 5), c) vol ARAT du 17/07/1999 (POI 6), d) vol DIMONA du 17/07/1999, e) vol DIMONA du 18/07/1999, f) vol ARAT du 30/07/1999 (POI 8). L’unité des concentrations est en ppb, l’heure étant en UTC.

3.6 Évaluation statistique de la qualité de CHIMERE à l'aide des données aéroportées de la campagne ESQUIF

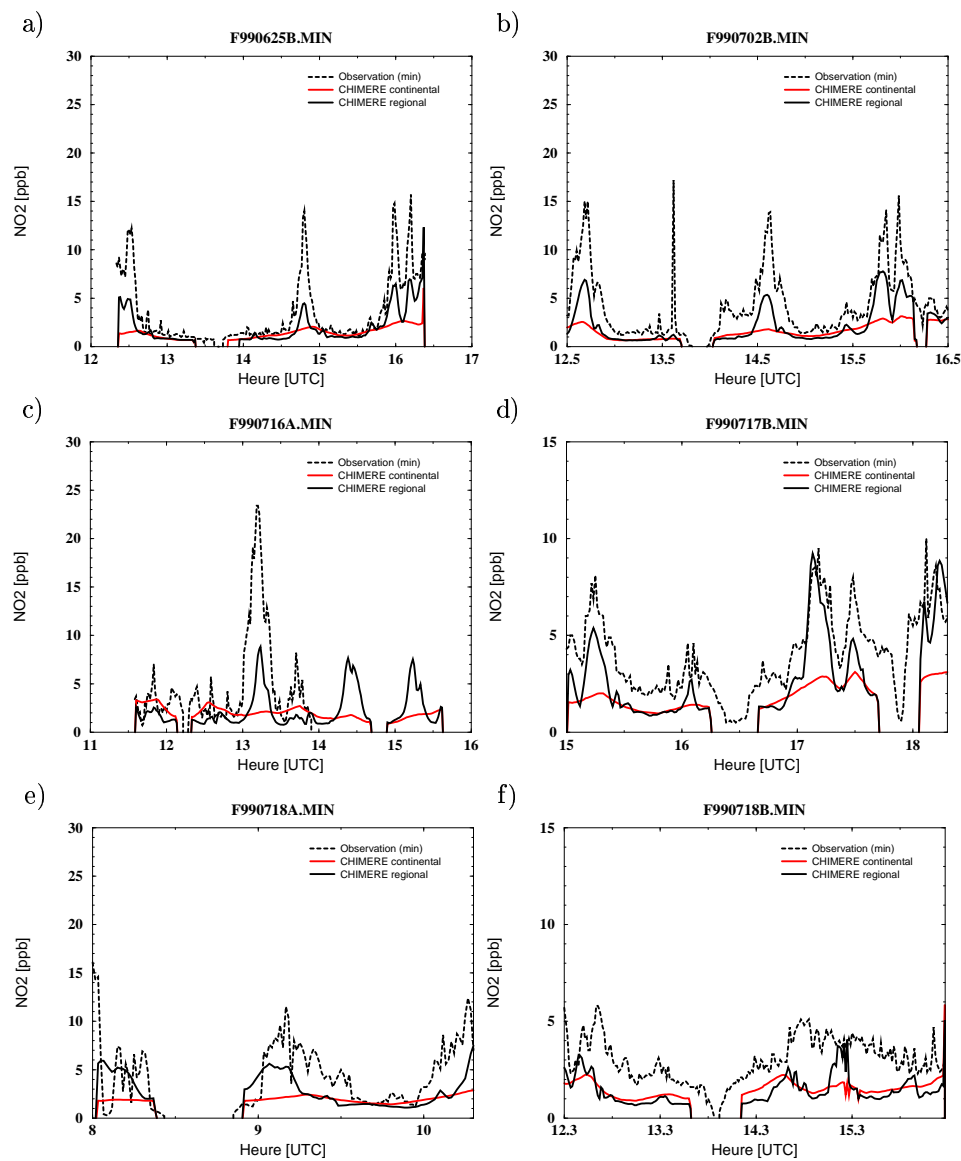


FIG. 3.18 – Séries temporelles des concentrations de dioxyde d'azote mesurées (courbe pointillée) toutes les minutes par les avions DIMONA pendant trois POI, et simulées par les modèles CHIMERE continental (courbe pleine grise) et régional (courbe pleine noire). a) 25/06/1999 b) 02/07/1999 c) 16/07/1999 d) 17/07/1999 e) 18/07/1999 le matin f) le 18/07/1999 l'après-midi. L'unité des concentrations en ppb, l'heure étant en ppb.

3.7 Conclusions

L'objectif initial de ce chapitre était de quantifier la qualité des simulations de ce modèle. Nous avons montré que le modèle continental et le modèle régional centré sur l'Île-de-France reproduisent généralement les concentrations d'ozone à 15 h UTC (environ l'heure du pic journalier), à moins de 15 ppb près. Le nombre d'erreurs supérieures à 15 ppb est de l'ordre de 10%.

Le modèle continental simule assez correctement les pics de concentrations d'ozone non seulement sur des sites proches des agglomérations mais aussi sur des zones éloignées des grandes sources d'émission, avec une tendance à la surestimation imputable aux résolutions horizontales et verticale du modèle. Il a, en outre, plus de difficulté à simuler les concentrations d'ozone dans les régions à relief variable (Alpes), le long des côtes (côte méditerranéenne), dans les vallées (Vallée du Rhône, de la Saône), et dans les régions de fortes émissions (Italie du Nord). De manière générale, il ne peut pas résoudre les structures fines des polluants. Les observations aéroportées de la campagne ESQUIF ont notamment montré que, contrairement au modèle régional, le modèle continental ne reproduit pas l'intensité des panaches urbains en altitude.

La qualité des simulations du modèle régional est moins bonne sur la région de Berre et de Marseille que sur la région d'Île-de-France. Les erreurs commises sur les pics de concentrations d'ozone sont cependant généralement inférieures à 15 ppb, 20% d'entre elles sont supérieures à 15 ppb. En revanche, environ 60% des dépassements du seuil d'information et de recommandation (90 ppb) ne sont pas bien simulés (erreur de plus de 15 ppb), le modèle faisant pratiquement systématiquement des erreurs dans le positionnement des panaches. La circulation des masses d'air complexe (à cause de la présence d'effet de côte et de relief) n'est pas suffisamment bien représentée par les données météorologiques de grande échelle (qui sont interpolées sur le domaine régional).

Nous avons également montré que le modèle régional simule plus difficilement les concentrations de son précurseur : le dioxyde d'azote. Les raisons de cette difficulté supplémentaire sont multiples. D'abord, le dioxyde d'azote est un polluant directement lié aux émissions, fortement incertaines. Ensuite, les pics de dioxyde d'azote apparaissent lors de la montée de la couche de mélange, quand le modèle est le plus sensible à tous les paramètres (émissions, rayonnement, vent ...). Sur la région d'Île-de-France, comme sur la région de Berre et de Marseille, les champs de concentrations de dioxyde d'azote de 7 h UTC (heure du pic) sont, en général, représentés à moins de 15 ppb près (30% d'erreur sont supérieures à 15 ppb). En revanche, des erreurs de plus de 15 ppb sont constatées sur pratiquement tous les dépassements du seuil 50 ppb (prévus ou observés).

La comparaison des cycles diurnes observés et simulés a montré que ce modèle commet des *erreurs systématiques* (qui se répètent), notamment la nuit et durant les heures de fortes émissions. Ces erreurs sont en partie liées à la résolution du modèle. Les séries temporelles ont en outre mis en évidence des *erreurs aléatoires*. L'objectif des prochains chapitres est précisément de réduire l'*erreur globale* commise sur ces champs de polluants afin à terme de pouvoir offrir des *cartes de polluants plus réalistes*.

Chapitre 4

Analyse des champs de concentrations d'ozone sur la région de l'Île-de-France

Le chapitre précédent a montré que le modèle CHIMERE représente assez correctement les variations des concentrations d'ozone sur la région de l'Île-de-France. Cependant, nous avons mis en évidence certaines erreurs du modèle. Ces erreurs sont maintenant comptabilisées dans une *erreur globale* que nous appellerons, l'*erreur du modèle* ou l'*erreur de modélisation*. De la même manière, les erreurs d'observation (incluant les erreurs instrumentales et les erreurs de représentativité) sont regroupées dans l'*erreur d'observation*.

Dans ce chapitre, nous proposons dans un premier temps de détailler les étapes de la mise en place de quatre algorithmes permettant de produire des analyses de *champs de concentration d'ozone sur la région d'Île-de-France*. Nous utilisons les observations d'ozone de surface, provenant de l'association agréée pour la surveillance de la qualité de l'air, AIRPARIF, et les sorties du modèle CHIMERE-régional, présentées au Chapitre 3.

Deux Interpolations Statistiques sont développées sur la base des données de simulation et d'observation de surface, constituées pendant l'été 1999. Deux types de modèles de covariance sont construits : un modèle *classique*, homogène et istrope, couramment utilisé en météorologie, et un autre modèle *plus original* hétérogène et anisotrope. En outre, un Krigeage est appliqué directement, soit sur les observations de surface, soit sur les différences constatées entre les simulations et les observations que l'on appelle couramment *les innovations*. Ces méthodes, dont les formulations mathématiques ont été présentées dans le chapitre 2, sont validées de façon objective et intercomparées sur la base de données (simulations et observations) des étés 1999, 2000, et 2001.

4.1 Mise en place des Interpolations Statistiques

La première hypothèse, qui a permis, à partir de l'équation générale de l'Interpolation Statistique 2.33, d'aboutir au système 2.35 donnant l'expression des poids optimaux, est que les observations, comme les simulations, ne sont pas biaisées. Si cette hypothèse est vérifiée, le champ

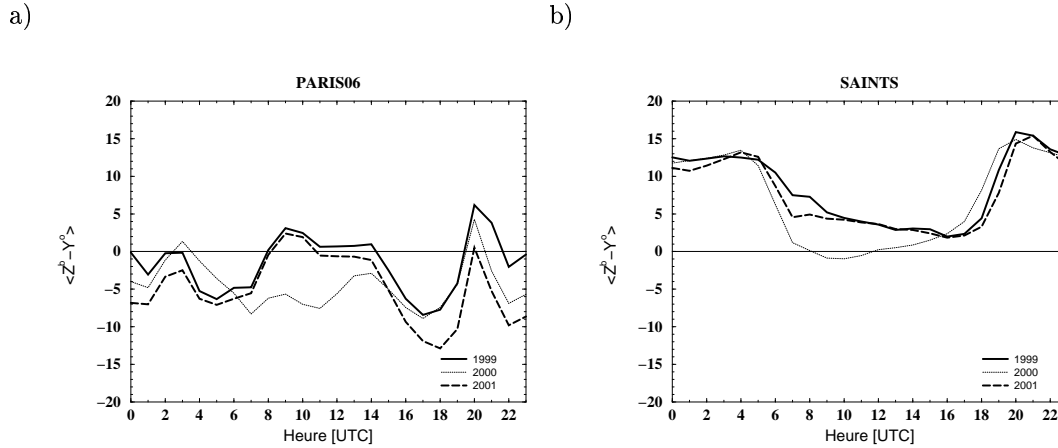


FIG. 4.1 – Évolution du biais global (écart moyen entre simulations Z^b et observations Y^o) en fonction de l'heure de la journée, sur deux stations a) PARIS06 b) SAINTS. Ce biais est calculé sur trois étés : 1999, 2000, 2001. L'unité des concentrations est en ppb. L'heure est en UTC.

analysé est alors non biaisé. Or nous avons précisément montré, dans le chapitre précédent, que les observations ou le modèle contiennent un biais. Nous montrons dans le paragraphe 4.1.1 comment ce problème est traité, tandis que le paragraphe 4.1.2 concerne la modélisation de la fonction de covariance.

4.1.1 Traitement des biais

La comparaison des cycles diurnes des concentrations d'ozone mesurées et simulées a mis en évidence un biais, qui résulte de la combinaison du biais d'observation (principalement du à des erreurs systématiques de représentativité) et du biais de modélisation (lié aux incertitudes sur les émissions, à des problèmes de mélange, de dépôt ...). La figure 4.1 montre l'évolution de ce biais, dit *global*, constaté sur deux stations, PARIS06 (au centre de Paris) et SAINTS (à l'extérieur, au nord-est de Paris), pendant les étés 1999, 2000, et 2001 en fonction de l'heure de la journée.

Puisque nous utilisons le même modèle (sans modifier le cadastre d'émission), la variation du biais est due à un changement dans les conditions météorologiques ou à une modification des émissions réelles. Nous remarquons que ce biais est variable d'une année sur l'autre sur la station urbaine, alors qu'il est assez reproductible sur la station rurale. L'été 2000 se différencie clairement des deux autres étés. Contrairement aux étés 1999 et 2001, les conditions météorologiques de l'été 2000 ont été particulièrement défavorables à la formation d'épisodes de pollution (situations nuageuses).

Sur la station SAINTS, un biais de plus de 10 ppb est constaté la nuit (19-6 h UTC). Ce biais est dû au manque de représentativité de la station. Sa reproductibilité suggère que les sources d'émissions de monoxyde d'azote, qui entourent la station, n'ont pas beaucoup évoluées de 1999 à 2001.

La situation en zone urbaine est très différente. Des changements de biais plus importants sont

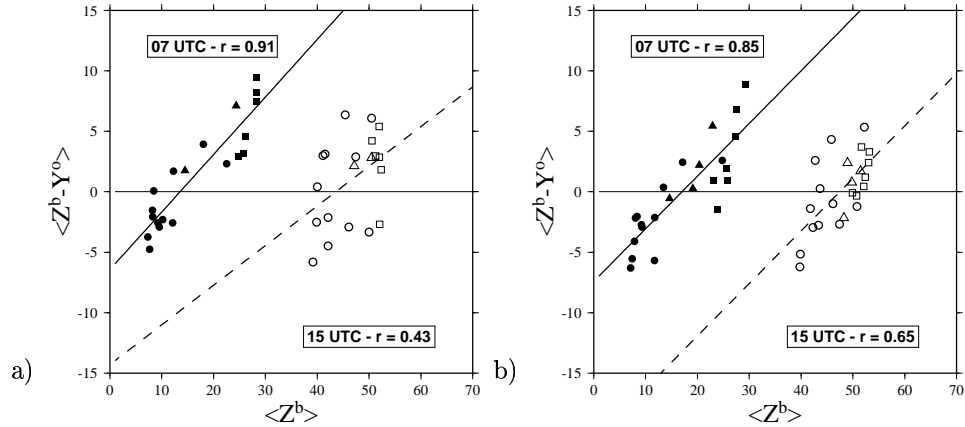


FIG. 4.2 – Écarts moyens entre simulations Z^b prises aux points de mesure et observations Y^o en fonction de la concentration moyenne simulée par le modèle aux mêmes points. Les symboles pleins correspondent aux valeurs de 7 h UTC, les symboles vides à celles de 15 h UTC. La courbe pleine est le modèle de biais utilisé à 7 h, la courbe pointillée celui de 15 h. Les cercles, les triangles et les carrés représentent respectivement les stations urbaines, périurbaines, et rurales. Les écarts sont présentés pour a) l'été 1999, b) l'été 2001. Les unités de l'abscisse et de l'ordonnée sont en ppb. r est le coefficient de corrélation des droites de régression.

constatés entre les années 1999 et 2001. Les émissions urbaines de monoxyde d'azote ont été modifiées d'une année à l'autre, ce qui n'est pas surprenant puisque ces émissions proviennent principalement du transport routier.

La difficulté pour corriger le biais d'observation et le biais du modèle est de faire la part, dans ce biais *global*, de ce qui est dû à une erreur de modélisation, de ce qui est dû aux observations. Considérons dans un premier temps que les observations ne sont pas biaisées. Dans ce cas, le biais constaté sur les stations est entièrement accordé au modèle. C'est ce que nous avons implicitement supposé dans le Chapitre 3. Nous avons notamment montré dans ce précédent chapitre que le modèle a tendance à légèrement sous-estimer les concentrations d'ozone sur la zone urbaine, la zone d'émission, et à les surestimer fortement en dehors pendant les heures de pointes de trafic. Par ailleurs, la zone urbaine est en général caractérisée par des concentrations plus faibles que la région environnante du fait des émissions de monoxyde d'azote. Tout se passe donc comme si "le modèle" surestimait les concentrations les plus élevées et sous-estimait les concentrations les plus faibles.

C'est pourquoi, nous utilisons *la concentration moyenne simulée par le modèle* comme un critère définissant la zone urbaine et nous nous limitons à supposer que le biais en dépend linéairement. La figure 4.2 montre la dépendance de ce biais en fonction de la concentration moyenne simulée pour l'année 1999 et l'année 2001. Si à 7 h UTC, cette dépendance est claire pour les deux années, à 15 h UTC, elle l'est un peu moins surtout pendant l'été 1999. Nous remarquons la grande variabilité des biais sur la zone urbaine. Il y a approximativement autant de biais négatifs que de biais positifs. Ceci montre que le cycle diurne simulé et moyenné sur l'ensemble des stations urbaines (Cf. figure 3.11) qui ne semble pas biaisé par rapport au cycle diurne observé, cache en fait des problèmes de représentativité.

Nous modélisons, de cette manière, un biais légèrement négatif (-4 et 0 ppb) sur Paris et sa

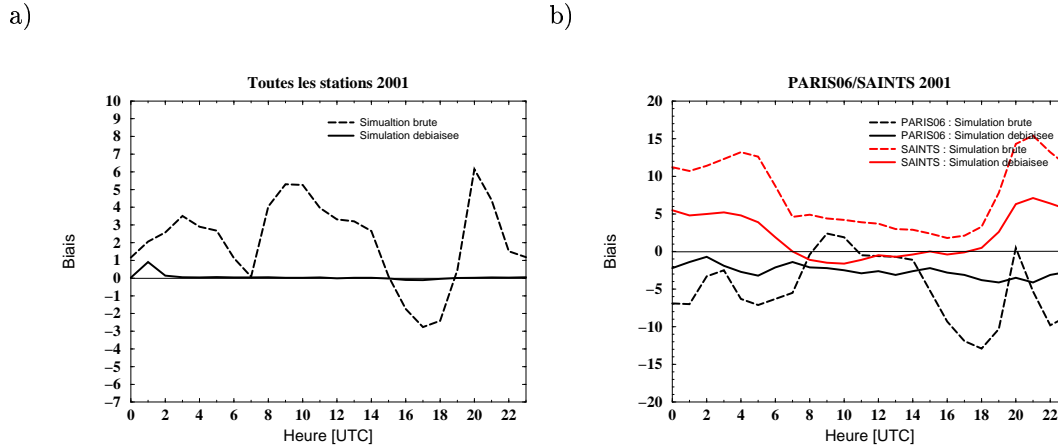


FIG. 4.3 – Comparaison des simulations brutes ou des simulations corrigées de leur biais avec les observations. Écart moyen (en ppb) en fonction de l'heure de la journée (en UTC). Le biais est modélisé à partir des données de l'été 1999 et est testé sur les simulations de l'été 2001. a) Résultats moyennés sur l'ensemble des stations du réseau, b) exemples sur deux stations.

banlieue à 7 h UTC et un biais positif sur le reste de la région, dépassant parfois 8 ppb. A 15 h, le biais du modèle est moins variable (de -2 à 2 ppb).

La figure 4.3 présente le biais des simulations brutes et le biais des simulations corrigées. La figure 4.3.a. montre la moyenne de ce biais, calculée sur l'ensemble des stations d'Île-de-France. Nous remarquons que notre paramétrisation du biais du modèle permet en moyenne de réduire considérablement les écarts entre observations et simulations à n'importe quelle heure de la journée. Le biais résiduel correspond à l'opposé du biais d'observation. Il est nul en moyenne sur l'ensemble des stations, ce qui ne veut pas dire qu'il l'est systématiquement sur chacune des stations (Cf. figure 4.3.b., qui détaille les biais sur deux stations du réseau).

Avant d'appliquer une Interpolation Statistique, le biais du modèle et les biais d'observation, ainsi modélisés, sont désormais soustraits respectivement des simulations et des observations. Le biais du modèle est supposé uniforme dans la direction verticale dans la couche de mélange. Au dessus de cette couche, il est supposé nul. Cette question est abordée dans le paragraphe 4.5.

4.1.2 Modélisation de la fonction de covariance

Fonction de covariance homogène et isotrope

Afin de modéliser la fonction de covariance, nous utilisons dans un premier temps la *méthode des observations*, décrite dans le paragraphe 2.5.2. Nous supposons donc que la fonction de covariance de l'erreur du modèle est stationnaire dans le temps, qu'elle est homogène et isotrope, et que les erreurs d'observation sont décorréées entre elles.

Pour respecter les variations diurnes des concentrations d'ozone, un covariogramme est construit à chaque heure de la journée. Pour cela, nous utilisons la base de données (simulations et observations) de l'été 1999. Les corrélations et les variances des innovations sont calculées sur les 153

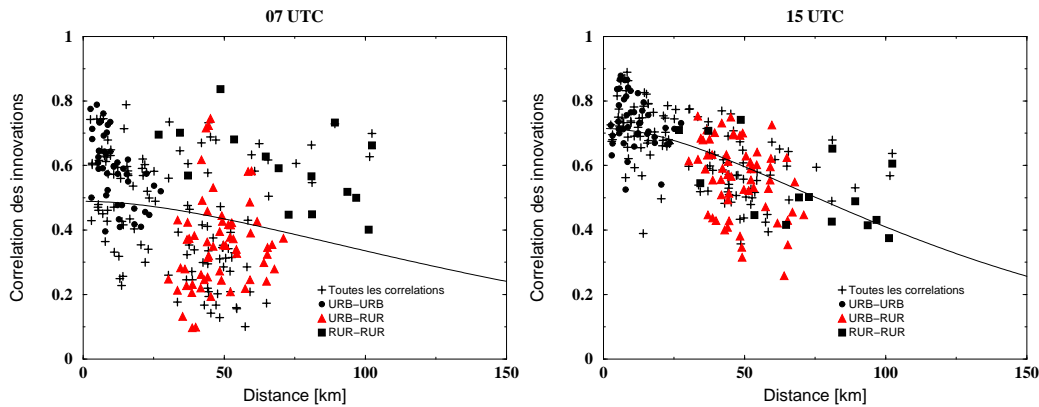


FIG. 4.4 – *Corrélogrammes homogènes et isotropes de 7 h UTC et de 15 h UTC* : ces graphes représentent les corrélations entre les innovations constatées en deux points de mesure en fonction de la distance qui sépare ces deux points. Les corrélations entre couples de stations rurales, entre couples de stations urbaines et les corrélations croisées entre stations urbaines et rurales, sont respectivement représentées par des ronds noirs, des carrés noirs, et des triangles gris. Les croix représentent les corrélations avec et entre les autres types stations. La courbe noire est le modèle de corrélation f_B de [Balgovind et al., 1983].

valeurs des concentrations d’ozone observées et simulées du 1^{er} Mai au 31 août, correspondant à une donnée par jour. La figure 4.4 présente les corrélogrammes de 7h UTC et de 15 h UTC. Nous distinguons sur cette figure les corrélations des innovations entre stations rurales, entre stations urbaines et les corrélations croisées entre stations urbaines et stations rurales.

A 7 h UTC, les corrélations des innovations varient de 0.4 à 0.8 pour des distances inférieures à 10 kilomètres. Ce sont essentiellement des corrélations correspondant à des couples de stations urbaines. La gamme de variation de ces corrélations augmente lorsque la distance est supérieure à 50 km. A cette distance, les corrélations entre couples du type [urbaine ; rurale] sont en général assez faibles, de l’ordre de 0.2-0.4. En revanche, à grande distance, des corrélations assez fortes, de l’ordre de 0.7-0.8, sont constatées. Ces corrélations de longue distance correspondent à des couples de stations rurales. L’hypothèse d’isotropie de la fonction de covariance est clairement mise en défaut. Pendant les heures de pointe de trafic, dans les basses couches mélangées, l’ozone est titré par le monoxyde d’azote. Sa concentration, qui dépend de l’intensité des sources d’émission, est variable et assez faible. Ceci est particulièrement valable dans l’agglomération parisienne, la zone principale d’émission. En revanche, des concentrations assez homogènes, provenant d’autres régions, sont advectées en Île-de-France et observées sur les stations rurales. Vu sa résolution, le modèle simule des concentrations plus homogènes que les concentrations observées. C’est pourquoi les innovations, qui reflètent surtout la variabilité des observations, sont plus corrélées entre deux points de la zone rurale qu’entre un point de la zone urbaine et un autre de la zone rurale.

A 15 h UTC, l’hétérogénéité spatiale des concentrations est beaucoup moins marquée car l’air est suffisamment brassé dans une couche de mélange bien développée. Cependant, dans la journée, d’autres violations de l’hypothèse d’isotropie sont constatées. Celles-ci sont en particulier dues au développement du panache urbain.

Un modèle de corrélation f_B est ajusté par la fonction de Balgovind (Cf. équation 2.43) quelque soit l'heure de la journée. Le modèle de covariance est obtenu en multipliant le modèle de corrélation par la moyenne des innovations. Les variances des erreurs d'observation et de modélisation, supposées homogènes, sont alors déduites de la précédente moyenne et de la valeur que prend le modèle de corrélation à la distance zéro. A 7 h UTC ce rapport est égal à 0.5, correspondant à des valeurs de la variance de l'erreur d'observation et de l'erreur du modèle égales à 50% de la variance de l'innovation. Les écarts types de ces erreurs sont d'environ 5 ppb, pour des concentrations moyennes de l'ordre de 10-20 ppb dans la zone urbaine, et de 40-50 ppb sur la zone rurale. Les erreurs sont donc, comme attendues, plus fortes sur la zone urbaine que sur la zone rurale.

Fonction de covariance hétérogène et anisotrope

Pour pallier les problèmes d'anisotropie spatiale des corrélations, évoqués précédemment, nous avons développé *un nouveau modèle de corrélation*. La méthode est basée sur le fait que les erreurs du modèle sont plus corrélées dans les régions soumises au même type de pollution (zone urbaine ou zone rurale) qu'entre régions soumises à des pollutions de type différent (entre zone urbaine et rurale).

La méthode de construction des nouveaux corrélogrammes est peu différente de la *méthode des observations* excepté qu'au lieu de tracer les corrélations des innovations en fonction de la distance qui les sépare, nous les traçons en fonction des *corrélations des concentrations d'ozone observées*. Un corrélogramme est construit à chaque heure de la journée sur la même base de données d'apprentissage (été 1999). La figure 4.5 montre les corrélogrammes de 7h UTC et de 15 h UTC. Nous distinguons également sur cette figure les corrélations des innovations entre stations rurales, entre stations urbaines et les corrélations croisées entre stations urbaines et rurales.

Nous remarquons que les corrélations sont beaucoup moins dispersées que sur la figure 4.4, prouvant déjà la pertinence de l'hypothèse avancée. Pour modéliser les covariances des erreurs du modèle, nous remplaçons simplement le critère de *distance* par le critère de *corrélation des concentrations d'ozone observées*. Or nous ne disposons pas d'observation en tout point de l'espace. C'est pourquoi, les termes de corrélations entre points de mesure et n'importe quel point, *a priori* non observé (i.e. les termes $B_{a,k}$ de l'équation 2.35), sont obtenus en utilisant les *corrélations des concentrations d'ozone simulées*. Pour éviter les discontinuités, les corrélations entre points de mesure (i.e. les termes $B_{k,l}$ de l'équation 2.35) sont finalement également déterminés à partir des corrélations données par le modèle CHIMERE.

Un modèle de corrélation f_C est ajusté par la fonction suivante :

$$f_C(c) = \left(1 + \frac{|1-c|}{L}\right) \exp\left(-\frac{|1-c|}{L}\right), \text{ avec } -1 \leq c \leq 1 \text{ et } L \geq 0 \quad (4.1)$$

La figure 4.5 montre que si les corrélations des innovations entre couples de stations du type [urbaine ; urbaine] ou [urbaine ; rurale] sont bien modélisées par f_C , les corrélations des innovations calculées entre les stations rurales sont systématiquement sous-estimées. Ces corrélations montrent néanmoins une forte dépendance avec les corrélations des concentrations. Ces résultats mettent en évidence la nécessité de construire un modèle qui dépende en plus de la typologie du site, comme

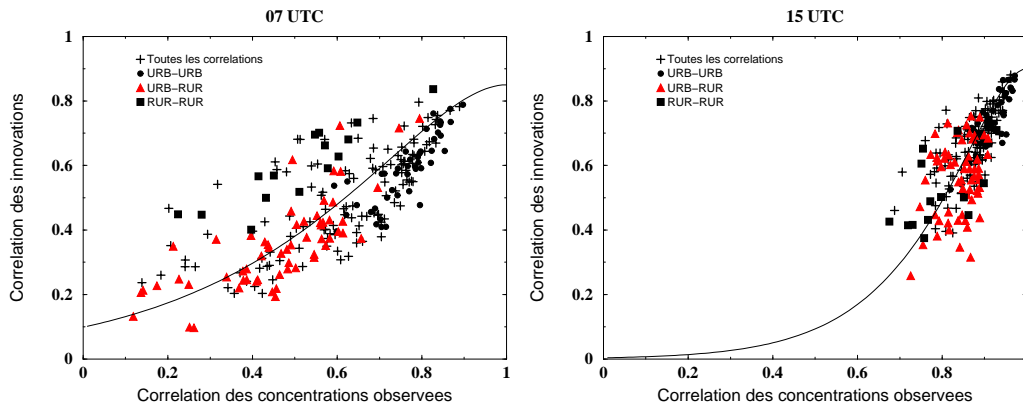


FIG. 4.5 – *Corrélogrammes hétérogènes et anisotropes de 7 h UTC et de 15 h UTC* : ces graphes représentent les corrélations entre les innovations constatées en deux points de mesure en fonction de la corrélation des concentrations observées. Les corrélations entre couples de stations rurales, entre couples de stations urbaines et les corrélations croisées entre stations urbaines et rurales, sont respectivement représentées par des ronds noirs, des carrés noirs, et des triangles gris. Les croix représentent le reste des corrélations. La courbe noire est le modèle de corrélation f_C .

l'ont récemment proposé *Hoelzemann et al.* [2001] (Cf. paragraphe 2.5.2).

Le modèle de covariance, B , est obtenu, comme pour la méthode précédente, en multipliant le modèle de corrélation, f_C , par la moyenne de la variance des innovations. La variance de l'erreur du modèle est obtenue à partir de la variance des innovations et de la valeur du modèle de corrélation quand la corrélation des concentrations simulées est égale à 1. Cette valeur est de 0.75 à 7 h UTC. Nous obtenons ainsi une deuxième estimation des erreurs de modélisation et d'observation, respectivement égales à environ 6 ppb et 4 ppb.

Le modèle de covariance de l'erreur, B , ainsi construit, est, *a priori*, hétérogène et anisotrope puisqu'il dépend de la structure de covariance du champ de concentrations d'ozone. Une observation, représentative d'une certaine région, aura une influence significative sur la correction des concentrations sur la région en question. En revanche, si une observation est très influencée par des sources locales d'émission, son influence sera limitée et dépendra de la position de ces sources.

La méthode que nous proposons n'est bien sûr valable que si le modèle reproduit correctement les corrélations des concentrations *observées*. La figure 4.6 montre précisément les corrélations entre les concentrations *observées* d'ozone en fonction des corrélations des concentrations d'ozone *simulées* par le modèle CHIMERE à 7 h, 9 h et 15 h UTC. Nous remarquons qu'à 15 h UTC, les corrélations des concentrations sont assez homogènes (entre 0.8 et 1) et légèrement surestimées par le modèle. En revanche, à 7 h et 9 h UTC, elle varient sur une gamme plus large, de 0.2 à 0.8, et sont largement surestimées. Pour corriger ces corrélations, nous utilisons une régression linéaire (coefficient de corrélation de 0.70 à 7 h, et de 0.60 à 15 h UTC). Les corrélations les plus mal représentées par le modèle, correspondent aux corrélations les plus faibles (inférieures à 0.4). Ces erreurs n'ont qu'un impact faible sur les analyses.

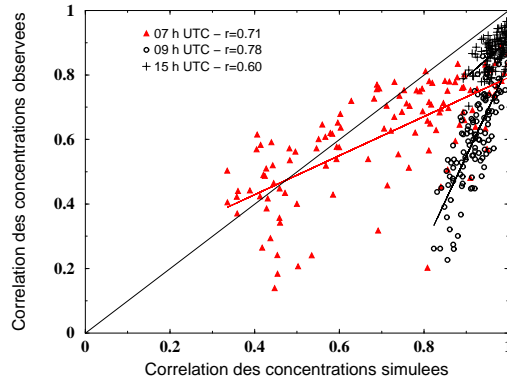


FIG. 4.6 – Corrélations entre concentrations d'ozone observées en fonction des corrélations entre concentrations d'ozone simulées par le modèle CHIMERE à 7 h, 9 h et 15 h UTC. Les droites correspondent aux modèles de correction des corrélations des concentrations d'ozone simulées. r est le coefficient de corrélation de la régression.

En outre, dans l'état actuel, les fondements théoriques de cette approche n'ont pas été investigués. Elle revêt pour l'instant un caractère purement heuristique. Le débat est encore ouvert quant à savoir si la *fonction de covariance* utilisée est bien effectivement une fonction de covariance, c'est à dire qu'elle est définie positive. En effet, le modèle f_C a été choisi par analogie avec la fonction de Balgovind. De simples tests, consistant à vérifier que les valeurs propres des matrices de covariances produites, ont été en permanence effectués.

4.1.3 Analyses 2D ou 3D

A chaque heure de la journée, deux modèles de covariance de l'erreur du modèle sont donc utilisés.

Le premier est homogène et isotrope. Basé sur un critère distance, ce modèle n'est utilisable que dans la dimension horizontale. Dans la journée, comme le montre la figure 4.7, qui présente les profils verticaux des concentrations d'ozone observées par l'avion DIMONA de la campagne ESQUIF le 17/07/1999 vers 15 h UTC, l'atmosphère est mélangée et les concentrations sont assez homogènes dans la couche de mélange. Les erreurs commises par le modèle ont donc une forte probabilité d'être corrélées dans cette couche et décorréliées des erreurs commises dans la couche supérieure non soumise aux processus de surface (émissions, dépôt ...). La nuit, l'atmosphère est stratifiée verticalement, créant de forts gradients verticaux de polluants dans les premiers mètres. Il existe alors d'intenses anisotropies entre la direction verticale et les directions horizontales. C'est pourquoi, l'interpolation statistique utilisant un modèle homogène et isotrope, qui sera maintenant appelée *méthode ISI* (pour Interpolation Statistique Isotrope), ne sera appliquée dans la suite que pour produire des analyses bidimensionnelles en surface.

Le deuxième modèle est hétérogène et anisotrope, basé sur une modélisation des covariances des erreurs du modèle en fonction de la covariance des concentrations. Ce modèle de covariance hori-

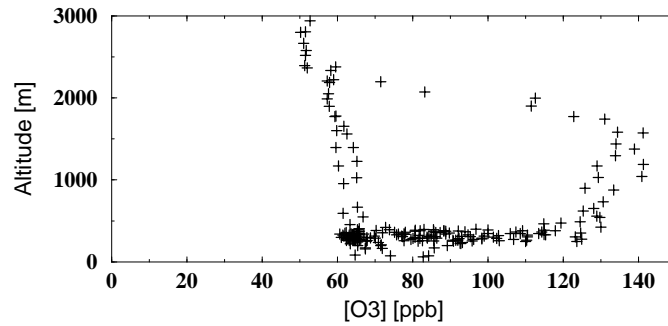


FIG. 4.7 – Profils verticaux des concentrations observées par l’avion DIMONA le 07/17/1999 dans l’après-midi. L’avion circule autour de l’agglomération parisienne et effectue deux profils, l’un dans le panache urbain, l’autre en dehors. L’unité des concentrations est en ppb. L’altitude est en mètre.

zontal peut *a priori* être étendu dans la direction verticale. En effet, la structure de corrélation d’une *champ de concentration* est différente selon les directions horizontales et verticale. Dans la couche de mélange, les concentrations d’ozone, pratiquement homogènes, sont très corrélées. En revanche, la nuit, la titration de l’ozone dans les basses couche de l’atmosphère décorrèle les concentrations de surface des concentrations d’altitude. Le modèle de covariance anisotropique que nous proposons devrait simuler des corrélations d’erreur suivant le même comportement et, ainsi, respecter l’anisotropie qui existe entre les directions horizontales et verticale. Or, le manque d’observation en altitude ne nous permet cependant pas de valider cette hypothèse *a priori*. C’est pourquoi, les observations aéroportées de la campagne ESQUIF sont utilisées pour la justifier *a posteriori*, tout au moins dans la couche limite : les observations sont comparées aux analyses tridimensionnelles. L’Interpolation Statistique utilisant un tel modèle hétérogène et anisotrope sera maintenant appelée : Interpolation Statistique Anisotrope, ou *méthode ASI*.

A partir des modèles de covariance d’erreur que nous venons de construire, nous pouvons calculer les poids d’interpolation. Ces poids dépendent de l’heure de la journée à laquelle est effectuée l’analyse et du nombre de stations disponibles. Tant que le nombre d’observations n’est pas modifié, les poids restent les mêmes et sont utilisés quelque soit la situation de pollution au temps de l’analyse. Si des observations manquent, les poids sont recalculés sur le nouvel ensemble d’observations. L’inconvénient majeur de ces méthodes est donc que les poids sont stationnaires dans le temps (i.e. ils sont les mêmes d’un jour à l’autre mais dépendant de l’heure de la journée). Pour pallier à ce problème, des méthodes de Krigeage ont été testées.

4.2 Mise en place de méthodes de Krigeage

L’intérêt des méthodes de Krigeage est de ne pas nécessiter obligatoirement d’information sur le passé du système étudié. Les statistiques des erreurs (biais, covariances) ne sont pas supposées *stationnaires dans le temps*, mais approchées par des moyennes spatiales, les ensembles statistiques étant construits à partir des concentrations observées disponibles au temps de l’analyse. Ces méthodes peuvent ainsi traiter une situation de pollution de l’air, à *un instant donné*, sans tenir compte du passé de l’atmosphère.

Le Krigeage Intrinsèque, décrit dans le paragraphe 2.7, est ici utilisé pour analyser les champs

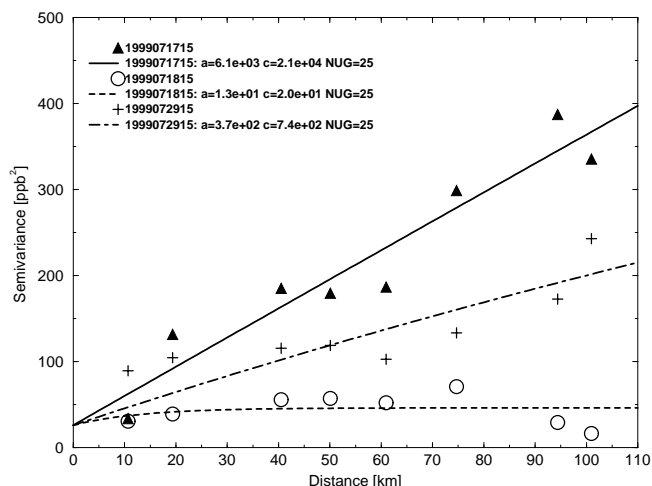


FIG. 4.8 – *Krigeage sur les observations. Variogrammes expérimentaux et modèles de variogramme pour trois jours : le 17/07/1999 (triangles, courbe pleine), le 18/07/1999 (cercles, courbe de tirets long), le 29/07/1999 (croix, courbe de tirets alternés long-court) à 15 h UTC. Distances en kilomètres, semivariances en ppb².*

de concentrations d'ozone. S'il est appliqué directement sur les observations, nous l'appellons OBK (pour Krigeage sur les OBservations). S'il est appliqué sur les innovations, nous appelons INK (pour Krigeage sur les INnovations).

4.2.1 Krigeage sur les observations, méthode OBK

Supposons temporairement que le modèle de chimie-transport n'est pas disponible. L'analyse ne peut alors être construite qu'à partir des observations, disponibles au temps d'analyse. Nous utilisons pour cela un Krigeage Intrinsèque. La méthodologie pour construire le variogramme est décrite en détail dans le paragraphe 2.7. Nous définissons des *classes de distance*, de 15 km de résolution, en nous accordant une tolérance de 10 km. Sur chacune des 8 classes de distance ainsi définies, la semivariance des différences de concentrations observées est calculée. Cette semivariance est ensuite tracée en fonction de la distance moyenne par classe. La figure 4.8 montre trois variogrammes expérimentaux construits à trois temps d'analyse différents : les 17, 18 et 29 Juillet 1999 à 15 h UTC. Les variogrammes sont ajustés par une fonction exponentielle donnée par l'équation 2.61, avec σ_o^2 égal à la variance d'erreur homogène d'observation déterminée lors de la mise en place de la méthode ASI (25 ppb à 15 h UTC). Nous remarquons qu'ils sont différents d'un jour à un autre.

Les variogrammes du 17 et 29 Juillet 1999 à 15 h UTC augmentent quasiment linéairement quand la distance augmente. Plusieurs hypothèses peuvent être émises pour expliquer cette croissance linéaire des variogrammes : soit il existe une structure de grande échelle qui n'est pas totalement détectée par les observations ; soit les observations détectent entièrement la structure de corrélation mais les concentrations d'ozone présentent un biais ou une anisotropie spatiale ; soit, enfin, la variance de ces différences est "infinie" (i.e. elle n'existe pas, l'hypothèse de stationnarité intrinsèque est alors mise en défaut). Le cas du 17/07/1999 est étudié de façon plus détaillée dans le paragraphe 4.4. Il est montré dans le paragraphe 4.2.2 que l'utilisation du modèle de chimie-

transport limite ces problèmes.

Le variogramme du 18 Juillet 1999 à 15 h UTC est plat à partir d'une vingtaine de kilomètres, indiquant une structure de corrélation spatiale de cette taille. Au-delà de cette distance, les concentrations sont décorrélatées. Ce jour-là, de fortes concentrations assez homogènes sont observées sur l'ensemble de la région d'Île-de-France.

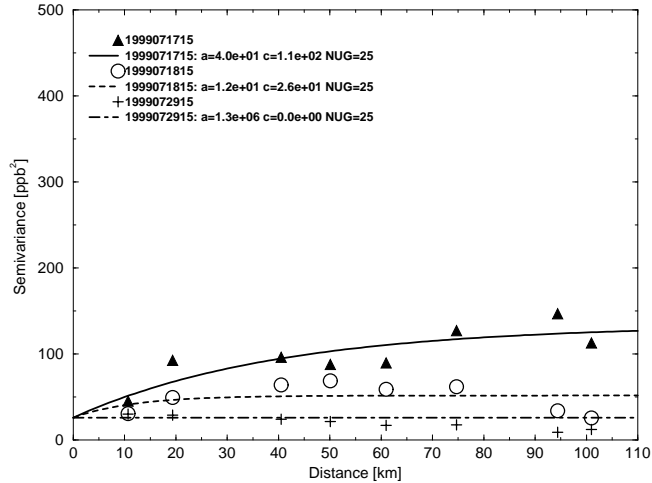


FIG. 4.9 – Krigeage sur les innovations. Variogrammes expérimentaux et modèles de variogramme pour trois jours : le 17/07/1999 (triangles, courbe pleine), le 18/07/1999 (cercles, courbe de pointillés long), le 29/07/1999 (croix, courbe de pointillés alternés long-court). Distance en kilomètres, semivariances en ppb^2 .

4.2.2 Krigeage sur les innovations, méthode INK

Le Krigeage Intrinsèque est maintenant appliqué sur les différences constatées entre les observations et les simulations, prises aux points de mesure. Nous avons montré dans le Chapitre 3 que le modèle reproduit à peu près convenablement les variations spatiales des concentrations d'ozone observées. On s'attend ainsi à ce que les différences simulations-observations soient plus homogènes que les observations. Les variogrammes construits sur ces différences devraient être plus souvent plats aux grandes distances. La figure 4.9 montre trois variogrammes expérimentaux construits pour les 17, 18, et 29 Juillet à 15 h UTC.

Tous ces variogrammes sont plats aux grandes distances, limitant ainsi les erreurs de Krigeage à grande distance. Le champ de différences simulations-observations du 29/07/1999 correspond à un *bruit blanc*, i.e. il n'y a pas de corrélation spatiale dans le champ des innovations et le modèle reproduit très bien les structures repérées par les observations (même s'il sous-estime de façon approximativement uniforme les concentrations ce jour-là, Cf. figure 5.7.a., paragraphe 5.3, Chapitre 5). En revanche, les variogrammes du 17 et du 18 Juillet mettent en évidence l'existence d'une structure spatiale dans l'erreur du modèle.

Comme la méthode ISI, les méthodes OBK et INK, basées sur un critère de distance, ne sont

appliquées que pour produire des champs de concentration horizontaux en surface.

4.3 Validation et comparaison des méthodes d'analyse de données, représentation des champs d'ozone de surface

Les modèles de covariance, construits sur la base de données de l'été 1999 sont maintenant utilisés pour produire les analyses des champs de concentrations d'ozone en surface des étés 1999 et 2001. Quant aux modèles de semivariances, ils sont construits directement à chaque temps d'analyse sur les données d'observation, ou sur les différences simulations-observations. Nous n'utilisons, pour le moment, que les données des stations existantes en 1999 pour analyser les champs de concentrations d'ozone de l'été 2001.

4.3.1 Méthode objective de validation : le *Leave-one-out*

Afin de contrôler la qualité des analyses obtenues, la méthode du *leave-one-out*, dite aussi *méthode de validation croisée* est utilisée. Cette méthode consiste à omettre certaines observations dans l'algorithme d'analyse, et à comparer ces observations aux analyses produites indépendamment. Nous avons procédé de la façon suivante : (1) nous mettons de coté les observations effectuées, pendant tout un été, par une station au point de mesure s_i (2) nous analysons les champs de concentrations d'ozone à partir des observations restantes (3) nous comparons les analyses résultantes, prises au point de mesure s_i , aux observations non utilisées et calculons des statistiques d'erreur (biais, RMS, corrélation ...). Finalement, nous appliquons le processus successivement à l'ensemble des stations du réseau d'Île-de-France. Ainsi, dans les paragraphes suivants, toutes les analyses sont comparées à des observations, non utilisées dans l'algorithme d'analyse de données.

TAB. 4.1 – Statistiques d'erreurs des comparaisons entre les concentrations de d'ozone simulées/analysées et les valeurs observées à 15 h UTC. Ces statistiques sont données station par station et en moyenne sur l'été 2001. L'unité des concentrations est en ppb.

Station	BIAS	RMS	COR	E15	E15/50	E15/75	E15/90
PARIS06	-5.2/-4.2	9.7/5.8	0.91/0.98	88/100	70/100	60/100	40/100
PARIS07	2.6/3.1	10.1/7.3	0.88/0.95	85/98	72/94	66/90	50/100
PARIS13	-6.2/-5.0	10.7/6.8	0.91/0.98	81/100	63/100	61/100	60/100
PARIS18	-3.0/-1.9	9.0/5.2	0.91/0.97	90/100	78/100	75/100	40/100
NEUILLY	-1.4/-0.6	8.7/4.2	0.90/0.98	92/99	84/94	80/88	33/50
GENNEVILLIERS	-1.0/-1.8	8.2/4.8	0.91/0.98	93/99	85/95	86/92	75/83
MARCHES	-2.7/-4.0	9.5/6.4	0.90/0.97	89/96	72/91	64/86	50/66
AUBERVILLIERS	-2.8/-2.6	8.2/4.3	0.92/0.99	94/100	95/100	90/100	80/100
VITRY	0.3/0.4	10.4/6.1	0.86/0.95	84/97	70/100	81/100	80/100
MONTGERON	5.3/2.7	11.0/6.1	0.88/0.96	83/95	65/84	52/66	44/60
CRETEIL	4.3/0.2	10.4/4.1	0.81/0.96	90/100	75/100	50/100	****/****
TREMBLAY	-2.2/-3.9	8.7/6.5	0.90/0.96	93/97	93/89	92/84	100/83
MANTES	2.4/0.3	8.0/3.7	0.89/0.98	91/100	85/100	80/100	66/100
MONTGE-EN-GOELE	1.2/-1.6	8.5/5.5	0.88/0.95	90/98	86/93	70/84	33/0
SAINTS	2.4/-0.6	8.9/7.0	0.85/0.90	89/93	83/77	80/72	0/0
FONTAINEBLEAU	3.3/0.3	9.3/7.0	0.84/0.90	92/96	80/88	69/90	100/****
SONCHAMP	0.4/-2.1	9.0/6.3	0.87/0.94	93/95	83/86	88/92	71/80
PRUNAY	3.7/1.0	8.5/4.2	0.89/0.97	90/99	76/95	81/91	75/100
FREMAINVILLE	-0.3/-2.7	7.8/4.5	0.90/0.98	94/100	92/100	88/100	50/100
MOYENNES	0.1/-1.2	9.2/5.6	0.88/0.96	89/98	79/94	74/90	58/78

Le tableau 4.1 donne les résultats d'une comparaison entre les analyses, produites à l'aide de la méthode ASI, et les observations d'ozone de surface. Nous reportons également dans ce tableau les statistiques obtenues avec les simulations brutes. Il apparait que la RMS des différences analyses-observations est plus faible que celle entre les simulations et les observations. Nous améliorons cette RMS d'environ 40%. La variabilité des concentrations est également nettement mieux représentée

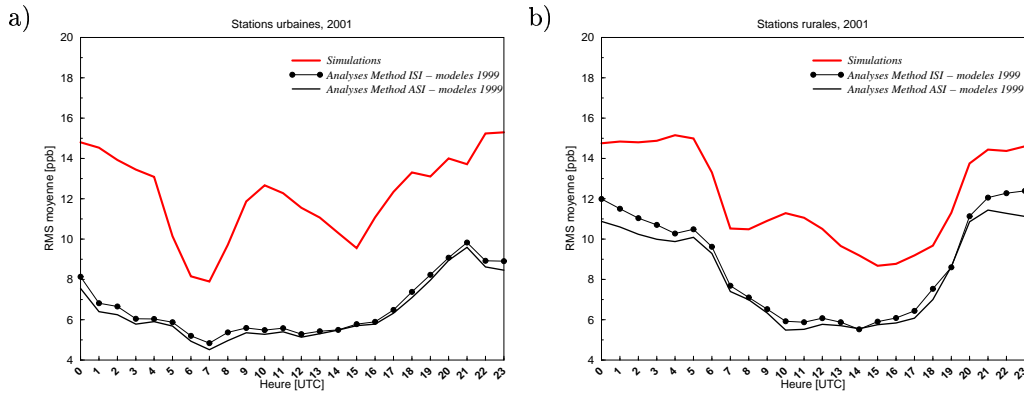


FIG. 4.10 – Comparaison, en terme de RMS, des simulations brutes et des observations (courbe grise), ou des analyses et des observations non utilisées dans ces analyses. La méthode du *leave-one-out* est appliquée à chaque station d'observation. Les analyses sont produites soit par la méthode ISI (courbe de points), soit par la méthode ASI (courbe noire). Les statistiques sont calculées et moyennées en fonction de l'heure (unité UTC) sur l'ensemble des stations a) urbaines, b) rurales. L'unité de la RMS est en ppb.

par les analyses car les corrélations sont meilleures, passant d'une moyenne de 0.88 à 0.96. Le nombre d'erreurs supérieures à 15 ppb est quant à lui aussi diminué passant de 10% (simulations) à 2% (analyses). En outre, les dépassements de seuil (50, 75, 90 ppb) sont bien simulés. Nombreux sont les cas, où par exemple, 100% des dépassements de seuil de 90 ppb sont simulés à moins de 15 ppb près.

4.3.2 Confrontation des méthodes ISI et ASI

La figure 4.10 montre le résultat d'une comparaison des simulations brutes ou des analyses avec les observations. Dans le cadre du *leave-one-out*, cette figure est le résultat de 19×24 analyses des champs de concentrations de l'été 2001, correspondant à une série estivale d'analyse par station et par heure. Les comparaisons sont présentées en terme de RMS des écarts. Les analyses sont produites soit par la méthode ISI, soit par la méthode ASI.

Premièrement, il est à noter que, quelque soit la méthode d'assimilation de données utilisée, la RMS des différences simulations-observations est réduite de 40 à 50% grâce à l'introduction des observations. Cette réduction est plus marquée sur les stations urbaines que sur les stations rurales (par exemple à 10 h UTC, une réduction de 7 ppb est constatée sur les stations urbaines, contre 5 ppb sur les stations rurales) car les stations urbaines sont en fait plus nombreuses et concentrées à l'intérieur de l'agglomération. Si nous retirons les observations d'une station urbaine, l'analyse effectuée à partir des autres observations urbaines proches reconstruit l'information manquante. En revanche, si nous retirons une station rurale, souvent beaucoup plus éloignée des autres points de mesure, l'analyse se rapproche du modèle numérique, qui est finalement la seule information disponible.

Deuxièmement, les RMS des différences (simulations ou analyses)-observations présente un cycle diurne assez marqué, avec des valeurs plus fortes la nuit que le jour. Ceci s'explique par la difficulté du modèle à simuler les concentrations la nuit (problème de représentativité, sensibilité du modèle

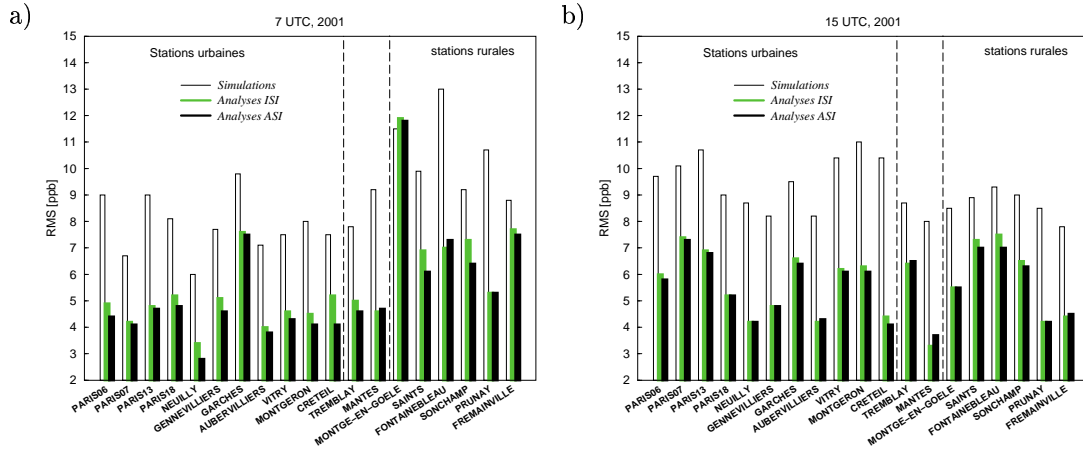


FIG. 4.11 – Comparaison, en terme de RMS, des simulations brutes (barres vides) ou des analyses avec les observations. La méthode du leave-one-out est appliquée à chaque station de mesure. Les analyses sont produites par la méthode ISI (barres grises), soit par la méthode ASI (barres pleines noires). la comparaison est effectuée a) à 7 h UTC, b) à 15 h UTC. L'unité de la RMS est en ppb.

quand la couche mélangée est mince, mélange trop fort).

Enfin, la comparaison des résultats obtenus par les méthodes ISI et ASI montre des différences de 1 à 1.5 ppb sur les stations rurales. Ces écarts apparaissent particulièrement la nuit, de 20 h UTC à 5 h UTC, et sont moins marqués sur les stations urbaines. Dans la journée, entre 5 h et 20 h, les méthodes sont pratiquement identiques. Dans tous les cas, la méthode ASI est légèrement plus performante (RMS plus faible) que la méthode ISI. Ces résultats sont la conséquence directe d'une anisotropie des corrélations des erreurs la nuit et pendant les heures de pointes de trafic, qui n'est pas respectée par la méthode ISI. En effet, nous avons montré dans le paragraphe 4.1.2 qu'à 7 h UTC, les innovations prises en deux stations rurales éloignées sont souvent plus corrélées que les innovations prises entre une station urbaine et une station rurale, alors que, pendant l'après midi, lorsque l'air est suffisamment mélangé, les concentrations comme les erreurs sont plus homogènes et isotropes.

La figure 4.11 présente la RMS des différences (simulations ou analyses)-observations, station par station à 7 h UTC et à 15 h UTC. Cette RMS montre de grandes variations d'une station à une autre. A 7 h UTC, la RMS moyenne des différences analyses-observations est d'environ 5 ppb sur les stations urbaines, tandis qu'elle est plus forte sur les stations rurales (environ de 7 ppb), car les concentrations sont souvent faibles dans la zone urbaine. Cette différence entre stations urbaines et stations rurales n'est pas constatée à 15 h UTC car les concentrations sont plus souvent homogènes. Nous notons également que si les méthodes ASI et ISI donnent des RMS pratiquement identiques à 15 h UTC, à 7 h UTC la méthode ASI permet de réduire la RMS des différences analyses-observations d'environ 5 à 20% de plus.

4.3 Validation et comparaison des méthodes d'analyse de données, représentation des champs d'ozone de surface

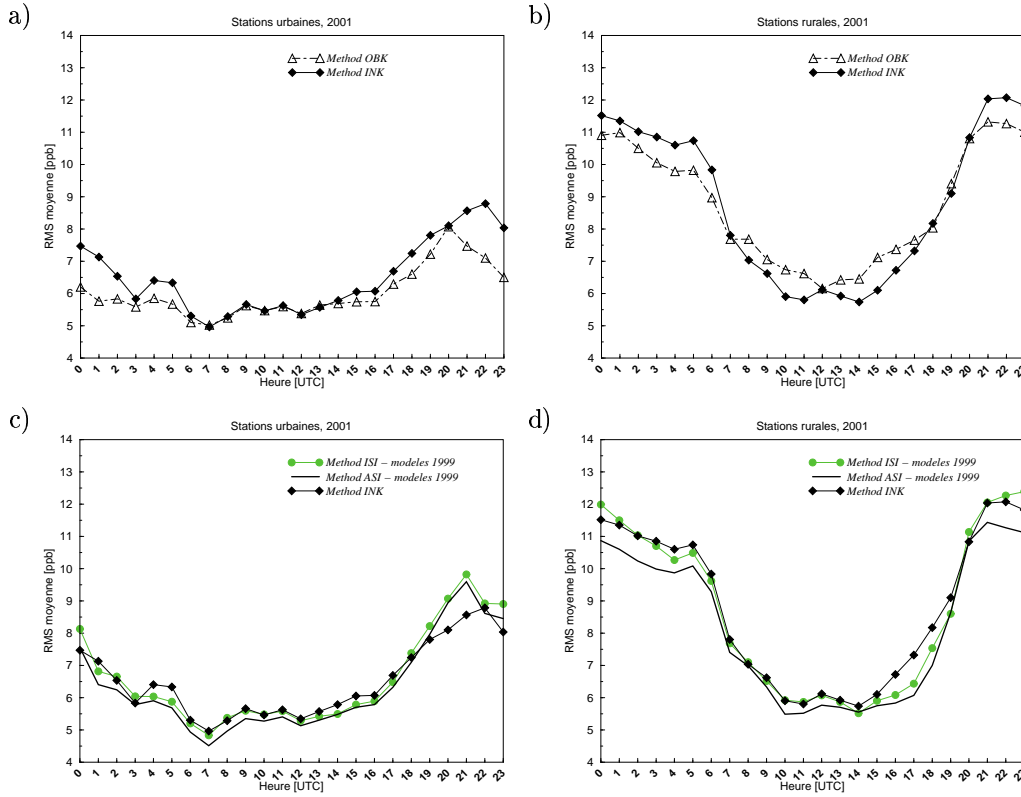


FIG. 4.12 – Comparaison, en terme de RMS, des analyses et des observations en fonction de l'heure (unité en UTC). Les analyses sont produites par les méthodes ASI (courbe noire), OBK (courbe pointillée et triangles vides), INK (ligne noire et carrés). Comparaison a) OBK/INK sur les stations urbaines b) OBK/INK sur les stations rurales c) ISI/ASI/INK sur les stations urbaines d) ISI/ASI/INK sur les stations rurales. L'unité de la RMS est en ppb, l'heure est en UTC.

4.3.3 Confrontation des méthodes ISI, ASI, OBK, et INK

Nous comparons maintenant les méthodes ISI et ASI et les méthodes de Krigeage. La comparaison OBK/INK doit nous indiquer si le modèle apporte ou non de l'information sur l'état de pollution de l'atmosphère. La comparaison INK/ISI permet de tester la pertinence de l'utilisation d'une climatologie pour estimer le modèle de covariance, et les biais (la méthode INK suppose un biais de modélisation uniforme, alors que dans le cas de la méthode ISI, le biais du modèle est estimé à partir de la climatologie). La figure 4.12 présente la RMS des différences analyses-observations sur l'ensemble des stations urbaines ou rurales.

La comparaison entre les méthodes INK et OBK, figure 4.12 a. et b., montre que la méthode OBK, qui n'utilise pas le modèle, est plus performante (RMS plus faible) la nuit (de 20 à 6 h UTC) sur l'ensemble des stations, et en fin d'après-midi (14-20 h UTC) sur les stations urbaines. Ces résultats ne sont pas surprenants car le modèle est moins performant la nuit que le jour. En fin d'après midi, les écarts entre les méthodes, sont constatés aux heures pendant lesquelles le modèle sous-estime les concentrations sur la zone urbaine dû à la mauvaise modulation des émissions.

CHIMERE ne peut que dégrader les analyses car sur la zone urbaine le nombre de stations est suffisamment grand pour que le problème soit surdéterminé.

Au contraire, sur les stations rurales, les analyses OBK sont clairement de moins bonne qualité que les analyses INK (RMS plus forte) entre 10 et 17 h UTC. Cette période de temps correspond aux heures pendant lesquelles la présence d'un panache urbain sur les zones rurales est possible. L'information apportée par le modèle est alors essentielle. En effet, nous rappelons que les expériences sont ici effectuées dans le cadre d'un *leave-one-out*. Cela signifie que, si autour du point analysé, aucune observation ne détecte de panache, la méthode OBK ne permettra pas de le représenter. En revanche, s'appuyant sur les structures modélisées par le modèle, les méthodes ISI, ASI et INK le repèront plus facilement (si bien sûr les simulations sont correctes).

Les figures 4.12 c. et 4.12 d. montrent que la méthode ASI n'est pas plus performante que la méthode INK sur les stations urbaines. Dans la journée, les méthodes donnent des résultats sensiblement équivalents. En revanche, sur les stations rurales, les analyses ASI sont de meilleure qualité, grâce à la combinaison d'une meilleure modélisation des biais et du caractère anisotrope du modèle de covariance. Les analyses obtenues par les méthodes isotropes INK et ISI, sont pratiquement équivalentes, sauf entre 14 et 19 h UTC. Ces heures, correspondent aux heures pendant lesquelles le modèle sous-estime les concentrations sur la zone urbaine. La méthode INK suppose que le biais est uniforme, or, comme les stations sont plus nombreuses, elles imposent un biais négatif sur tout le domaine, ce qui n'est pas correct.

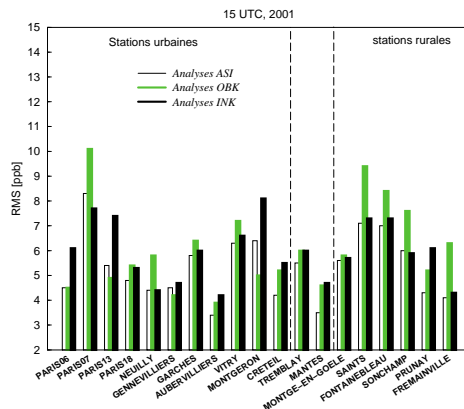


FIG. 4.13 – Comparaison, en terme de RMS, des analyses et des observations à 15 h UTC. Les analyses sont produites par les méthodes ASI (barres vides), soit par la méthode OBK (barres grises), soit par la méthode INK (barres noires). L'unité de la RMS est en ppb.

La figure 4.13 donne le détail de la RMS des différences analyses-observations station par station à 15 h UTC. Des différences de 2 ppb sont constatées sur la majeure partie des stations rurales.

La figure 4.14 présente l'écart moyen entre les analyses ISI, ASI, OBK, INK, et les observations. Il apparait clairement que, quelle que soit l'heure de la journée, les analyses OBK sont plus biaisées que les INK. L'utilisation du modèle permet en fait de réduire le biais spatial des champs de concentration. Dans la journée, la présence d'un panache induit un biais spatial des concentrations entre la zone urbaine (concentrations faibles) et la zone rurale (concentrations plus fortes). La

Enfin, les concentrations d'ozone sont plus faibles sur Paris et sa banlieue que sur la région environnante. Comme le modèle reproduit assez correctement ces structures spatiales, retirer les valeurs simulées aux valeurs observées permet ensuite de traiter un champ d'innovations plus homogène que le champ de concentration initial.

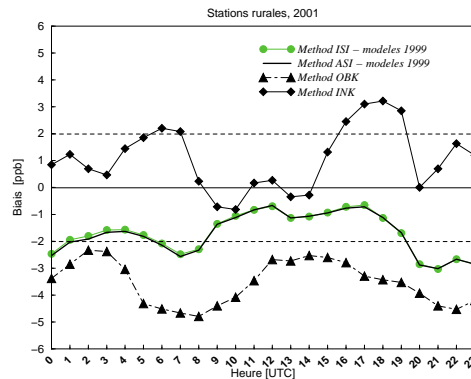


FIG. 4.14 – Écart moyen entre analyses et observations. Les analyses sont produites par les méthodes ISI (courbe grise et ronds), ASI (courbe noire), OBK (courbe pointillée et triangles), INK (courbe pointillée et carrés). L'unité du biais est en ppb, celle de l'heure en UTC.

Finalement, les différences constatées entre les méthodes ASI et INK apparaissent maintenant nettement entre 14 h et 19 h UTC. Les différences entre les analyses ASI et les observations sont biaisées à cause du biais d'observation, qui n'est pas corrigé ici.

4.4 Exemple d'analyses de champs de concentration d'ozone (étés 1999 et 2001)

Nous proposons dans ce paragraphe de regarder quelques cas particulièrement intéressants de pollution par l'ozone, constatés pendant les étés 1999 et 2001. Nous présentons les simulations du premier niveau du modèle CHIMERE régional centré sur l'Île-de-France, et les analyses produites pour le 17 Juillet 1999 et le 24 Mai 2001 à 15 h UTC. Le 17 Juillet 1999 est un jour où le modèle simule bien les concentrations du polluant, et l'analyse ajoute une faible correction. En revanche, le 24 mai 2001, les structures simulées s'écartent fortement de celles observées. Toutes les observations disponibles en 1999 et en 2001 sont désormais utilisées (i.e. nous n'utilisons plus de *leave-one-out*). Les observations supplémentaires, disponibles pendant l'été 2001, sont supposées ne pas avoir de biais d'observation.

4.4.1 Cartes de concentrations

Le 17 Juillet 1999

La figure 4.15 montre la simulation des concentrations d'ozone sur la région d'Île-de-France pour le 17 Juillet 1999 à 15 h UTC, et trois analyses produites à l'aide des méthodes ASI, OBK, et INK. Quelques observations sont inscrites directement sur les cartes.

Nous pouvons remarquer que le modèle simule un panache très concentré au nord-ouest de Paris. Si les concentrations d'ozone sont surestimées au centre du panache, elle sont en assez bon accord avec les observations sur le reste du domaine. Le modèle sous-estime visiblement légèrement l'étendue du panache.

Les analyses ASI et INK sont très peu différentes de la simulation initiale, montrant ainsi qu'une très faible correction a été effectuée. En revanche, l'analyse OBK donne au panache une structure tout à fait irréaliste, même si elle reproduit parfaitement les observations. Aucune source

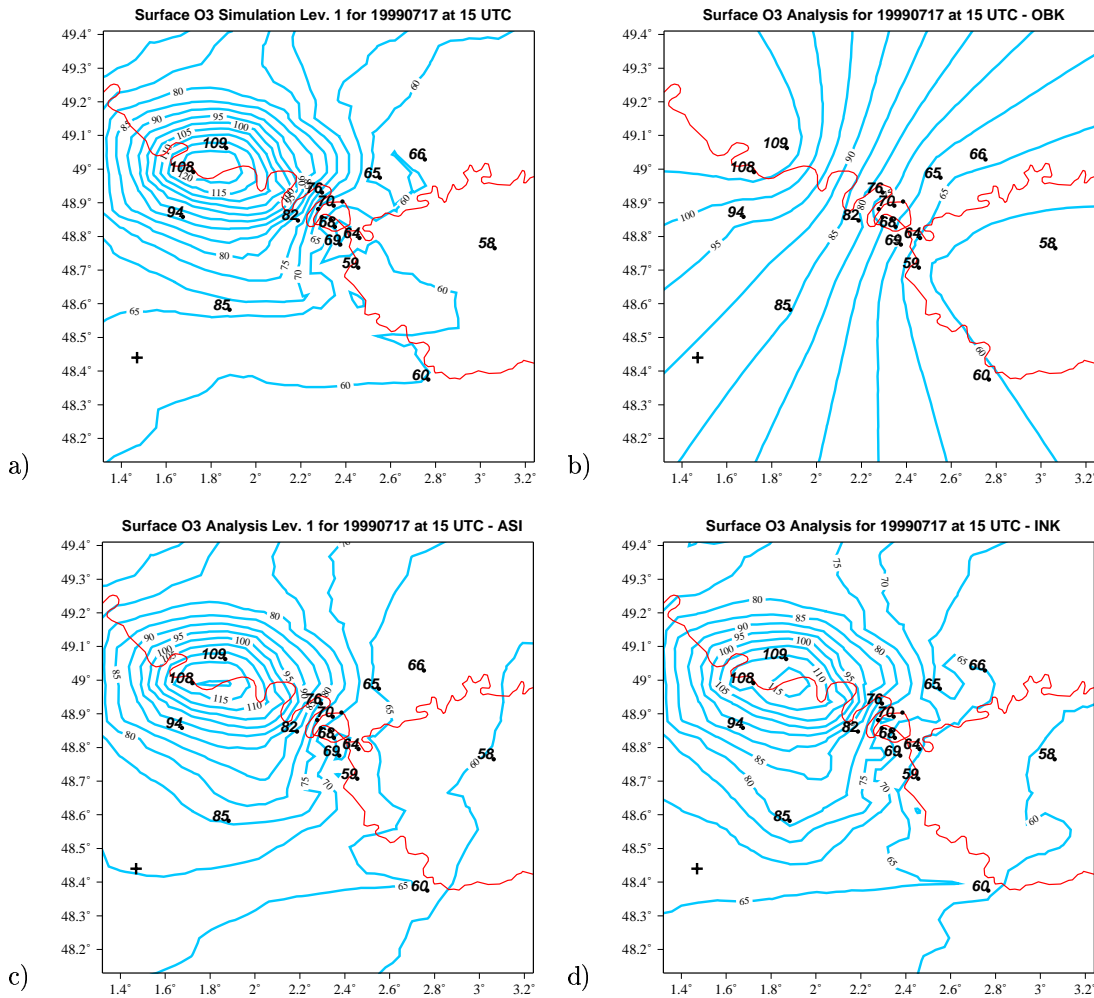


FIG. 4.15 – Simulations brutes et analyses (contours gris) des concentrations d'ozone par le modèle CHIMERE dans sa couche de surface pour le 17 Juillet 1999 à 15 h UTC. Comparaison avec quelques observations de surface (nombres noirs). Les points noirs représentent l'ensemble des stations utilisées dans les analyses. Paris est identifié par une ligne gris foncé, située au centre de la carte. Les cours de la Seine et de la Marne sont représentés. a) Simulation. Les analyses sont produites par b) la méthode OBK, c) la méthode ASI, d) la méthode INK. La croix localise la station CHARTRES-FULBERT. L'unité des concentrations est en ppb.

de précurseurs ne peut expliquer les fortes concentrations au nord et au sud-ouest du domaine. Pour juger de la qualité de cette analyse, la station CHARTRES-FULBERT gérée par l'association LIGAIR, notée d'une croix sur les cartes, est de très grande utilité. Celle-ci n'a pas été utilisée pour construire les analyses, elle donne une concentration de 77 ppb à 15 h UTC. Cette valeur prouve que l'analyse OBK n'est pas correcte. Elle est en outre plus proche de l'analyse ASI que de l'analyse INK pour la simple raison que les concentrations simulées sur les stations CHARTRES-FULBERT et SONCHAMP (station la plus proche de la précédente) sont très souvent corrélées (coefficient de corrélation : 0.99), comme les concentrations observées (0.96). Ainsi en augmentant les concentrations d'ozone sur la station SONCHAMP, la méthode ASI augmente aussi les concentrations d'ozone sur la station CHARTRES-FULBERT.

Le 24/05/2001

La figure 4.16 montre la simulation des concentrations d'ozone du 24 Mai 2001 à 15 h UTC sur l'Île-de-France, et trois analyses produites à l'aide des méthodes ASI, OBK, et INK. Les observations disponibles et utilisées sont également inscrites directement sur les cartes.

Les conditions météorologiques du 24 Mai 2001 sont caractérisées par un vent faible de nord-est et un ciel dégagé en début de journée. Dans l'après-midi, des nuages se développent au Sud du domaine. Le vent est bien représenté par le modèle du CEPMMT, en revanche, les nuages sont fortement sous-estimés. Le modèle simule alors un panache au sud-ouest de Paris, avec des dépassements du seuil d'information et de recommandation (90 ppb) sur deux stations de mesures. Cependant, la présence des nuages limite la photolyse et ces dépassements de seuil ne sont en réalité pas observés. Les observations sont même largement en dessous de ce seuil (48 et 62 ppb). De même, le modèle simule plus de 75 ppb au sud-est de Paris, à FONTAINEBLEAU, alors qu'il n'est observé que 26 ppb. Les simulations ne reproduisent pas les observations, qui semblent plutôt identifier un panache au nord du domaine.

Dans ce cas particulier, les analyses ASI et INK montrent de grandes différences. En effet, la méthode ASI montre plus de difficulté à garder une valeur de 75 ppb sur la station SAINTS, à l'est du domaine, et à diminuer les concentration dans le sud. Ceci est du au fait que les concentrations simulées sur la station SAINTS sont très souvent corrélées à celle simulées sur la station FONTAINEBLEAU. L'analyse ne reproduit pas les forts gradients de concentrations, inhabituels, dans la direction Nord-Sud. L'intensité du panache est néanmoins diminué de 25 ppb. La méthode INK, en revanche, reproduit très bien les gradients et particulièrement ceux qui sont constatés dans l'agglomération. L'analyse INK est alors plus proche de l'analyse OBK. Les grandes différences constatées, dans ce cas, entre les analyses ASI et INK constituent une situation extrême rare et très instructive. La méthode INK est plus performante que la méthode ASI pour la simple raison que les poids de l'interpolation ne sont pas stationnaires dans le temps, mais sont recalculés à chaque analyse.

4.4.2 Séries temporelles

La figure 4.17 présente les profils temporels des concentrations d'ozone observées, simulées par le modèle CHIMERE régional, et analysées par la méthode ASI, sur deux stations urbaines

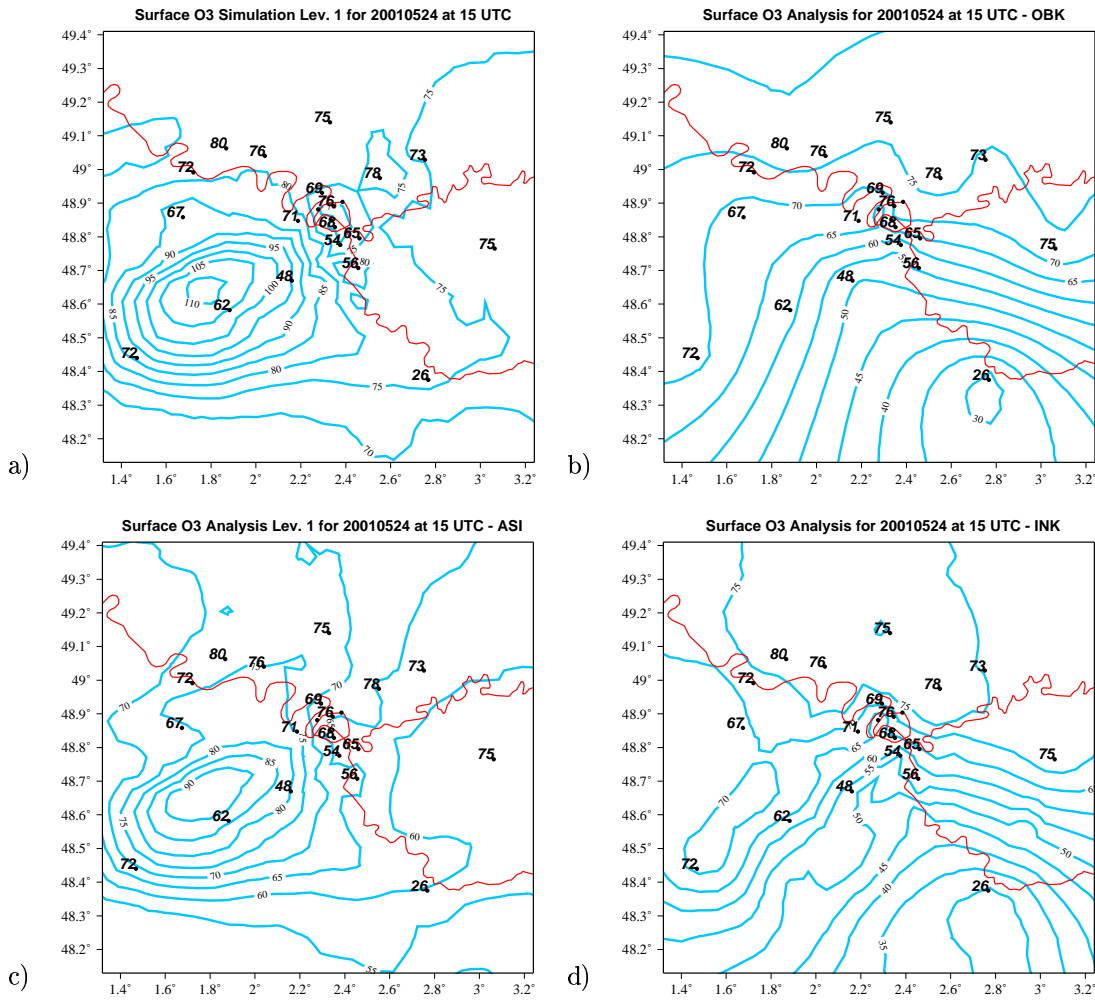


FIG. 4.16 – Simulations brutes et analyses (contours gris) des concentrations d'ozone par le modèle CHIMERE dans sa couche de surface pour le 24 Mai 2001 à 15 h UTC. Comparaison avec quelques observations de surface (nombres noirs). Les points noirs représentent l'ensemble des stations utilisées dans les analyses. Paris est identifié par une ligne gris foncé, située au centre de la carte. Les cours de la Seine et de la Marne sont représentés. a) Simulation. Les analyses sont produites par b) la méthode OBK, c) la méthode ASI, d) la méthode INK. L'unité des concentrations est en ppb.

(PARIS06 et PARIS07). Nous remarquons que globalement les analyses sont plus proches des observations que les simulations initiales. Le biais et la RMS, respectivement de -5.2 ppb et de 9.70 ppb, constatés sur la station PARIS06 sont réduits à -1.9 ppb et 4.7 ppb. Sur la station PARIS07, la RMS est plus faiblement diminuée à cause d'une augmentation du biais. Dans le centre urbain, bien couvert par la mesure, l'analyse a tendance à se rapprocher de la moyenne des observations et à lisser les concentrations. Plus une observation s'écarte de cette moyenne, plus elle marque un comportement non représentatif des autres mesures. Les erreurs de représentativité sont alors mises en évidence.

4.4 Exemple d'analyses de champs de concentration d'ozone (étés 1999 et 2001)

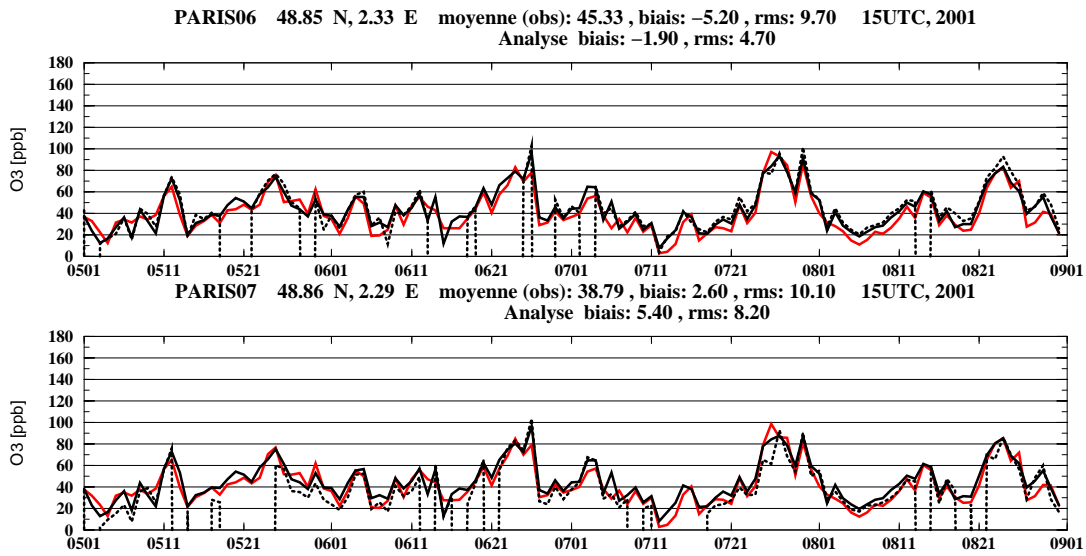


FIG. 4.17 – Séries temporelles des concentrations d'ozone observées, simulées par CHIMERE régional, et analysées par la méthode ASI à 15 h UTC du 1^{er} Mai au 31 août 2001 sur deux stations urbaines proches (dans la même maille du modèle). La courbe pointillée correspond aux observations, la courbe grise aux simulations brutes, et enfin la courbe noire aux analyses. L'unité des concentrations est en ppb.

La figure 4.18 montre les cycles diurnes moyens des concentrations d'ozone observées et analysées. Nous remarquons le parfait accord de ces cycles. Ces profils sont à comparer à ceux de la figure 3.11. L'anomalie de fin d'après-midi constatée sur les stations urbaines et les surestimations nocturnes, mises en évidence sur les stations rurales ont été corrigées.

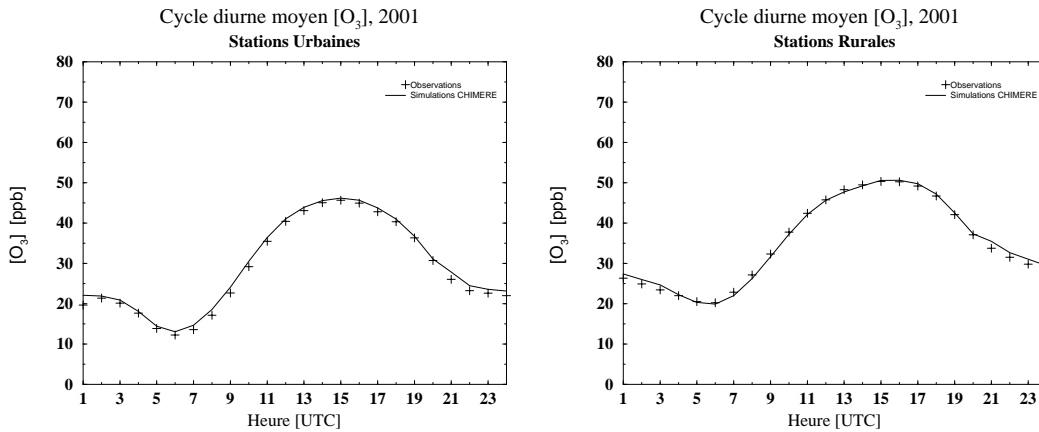


FIG. 4.18 – Profils diurnes des concentrations d'ozone observées et analysées sur les stations urbaines et rurales. L'unité des concentration est en ppb. L'heure est en UTC.

4.4.3 Cartes d'erreurs d'analyse

L'Interpolation Statistique et le Krigeage ont l'avantage, par rapport à un 4D-VAR, d'être des méthodes qui permettent, de manière très simple, d'estimer l'*erreur d'analyse*. Nous appelons par *erreur d'analyse*, la racine carré de sa variance, c'est à dire son écart type. Nous proposons maintenant d'étudier sa structure spatiale.

Dans le cas de la méthode d'analyse ASI, pour une distribution de station fixe, le champ d'erreur d'analyse est supposé constant dans le temps et est estimé à partir de la base de données de l'été 1999. Dans le cas des méthodes de Krigeage, un champ d'erreur de Krigeage est calculé à *chaque instant d'analyse*. Ainsi pour comparer les structures d'erreur d'analyse, nous moyennons

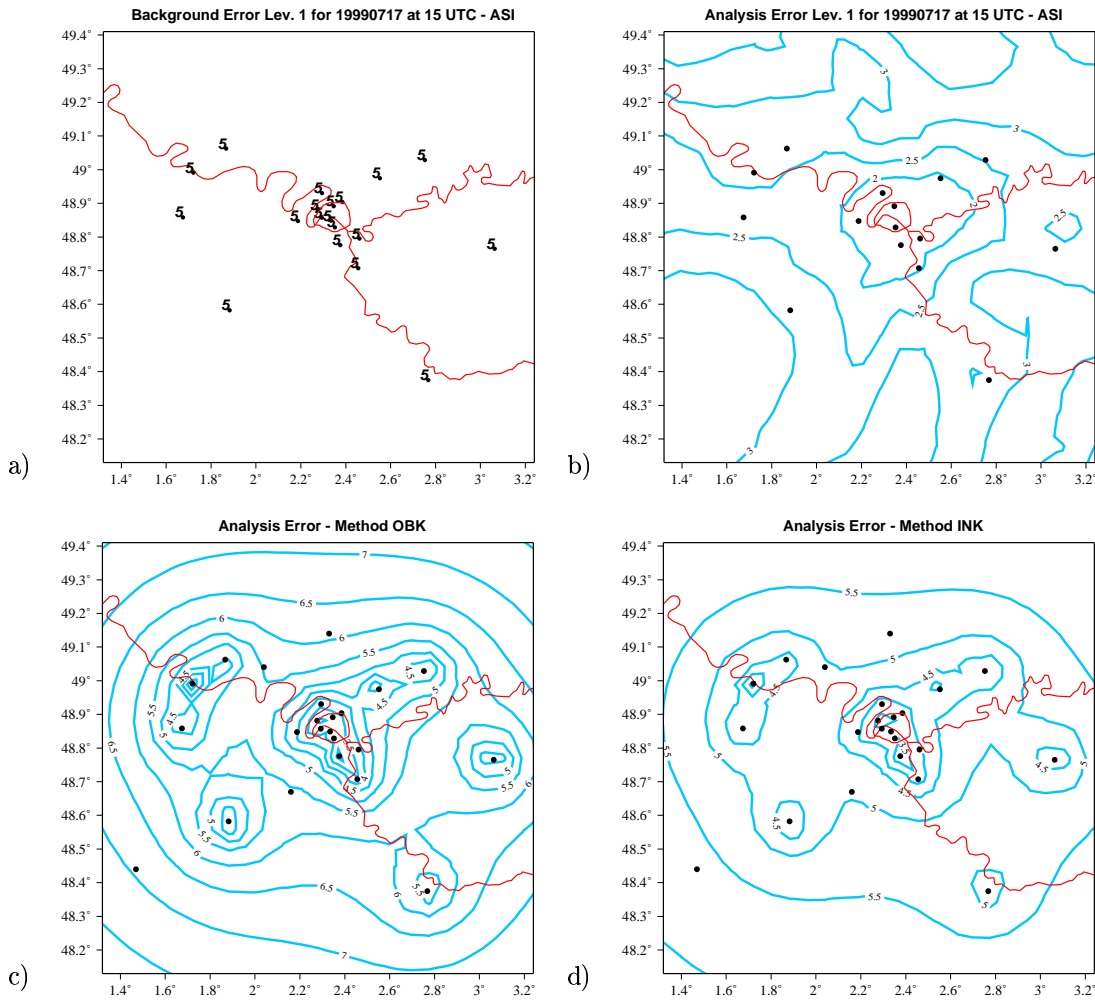


FIG. 4.19 – Cartes d'erreur d'observation et d'analyse. a) erreurs d'observation, l'erreur de modélisation étant homogène de 7 ppb, b) erreur d'analyse obtenue par la méthode ASI, c) erreur moyenne de Krigeage obtenue par la méthode OBK, d) même figure qu'en c) pour la méthode INK. L'unité des erreurs est en ppb.

les erreurs de Krigeage sur l'été 1999.

La figure 4.19 a. présente l'erreur homogène d'observation, estimée par la méthode ASI pour un réseau d'observation tel que celui du 17 Juillet 1999. L'erreur d'observation est uniforme de 5 ppb, l'erreur de modélisation de 7 ppb. La figure 4.19 b. montre l'erreur d'analyse ASI. Cette erreur d'analyse est assez homogène et systématiquement inférieure à l'erreur d'observation. Ce résultat est cohérent avec le test de validation *a posteriori* des méthodes d'assimilation de données, proposé par Talagrand et Bouttier [2000]. Nous notons que cette erreur d'analyse est minimale là où le réseau d'observation est le plus dense, i.e. près du centre urbain et, à partir de ce centre, dans les directions des stations de mesures rurales. Elle est maximale dans les régions non observées, et notamment au nord et au sud du domaine. Depuis 1999, l'association AIRPARIF a installé des stations supplémentaires dans ces régions.

Les figures 4.19 c. et 4.19 d. présentent respectivement les champs de l'erreur moyenne de krigeage obtenus par les méthodes OBK et INK. Contrairement à l'erreur d'analyse ASI, l'erreur de Krigeage est minimale près des stations de mesure. L'erreur d'analyse INK est en moyenne plus faible que l'erreur d'analyse OBK montrant que le modèle apporte de l'information, et particulièrement loin des observations.

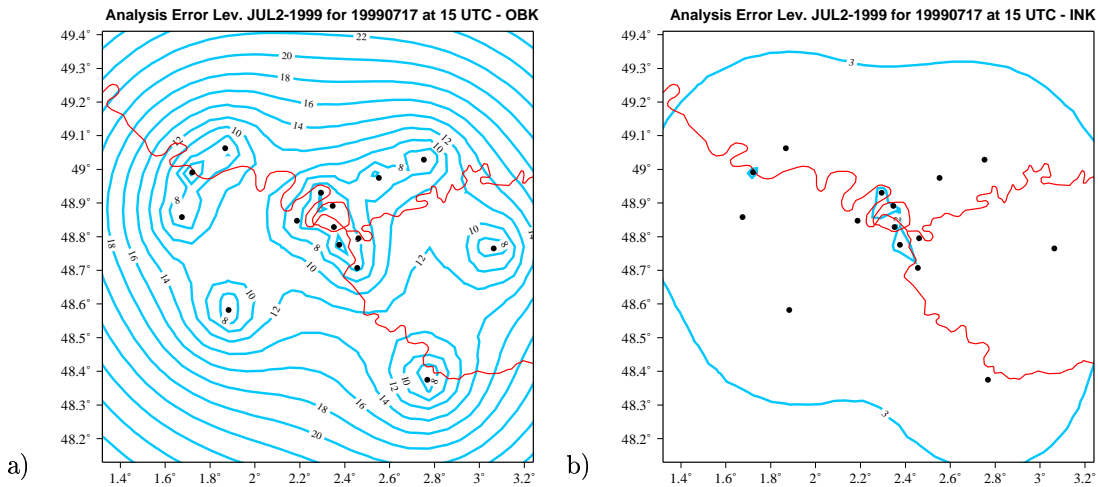


FIG. 4.20 – Cartes d'erreur d'analyse obtenues pour le 17 Juillet 1999 à 15 h UTC par a) la méthode OBK b) la méthode INK. L'unité des erreurs est en ppb.

Finalement, la figure 4.20 donne un exemple de cartes d'erreur de Krigeage pour le 17 Juillet 1999 à 15 h UTC. Nous remarquons que l'erreur d'analyse OBK est particulièrement élevée, indiquant le peu de confiance que nous devons apporter à l'analyse OBK obtenue sur ce jour-là (Cf. figure 4.15.b.). Ce résultat est en accord avec ce qu'avait suggéré le semivariogramme rectiligne (Cf. figure 4.8). La raison de la forte croissance du variogramme est due à la présence du panache au nord-ouest du domaine. En effet, la figure 4.15 montre que trois stations sont dans le panache et mesurent des concentrations fortes, tandis les autres stations mesurent des concentrations beaucoup plus faibles. Les accroissements $Y^o(\mathbf{s}) - Y^o(\mathbf{s} + \mathbf{h})$ sont donc d'autant plus grands que la distance h augmente. Comme les simulations de CHIMERE reproduisent très bien les variations spatiales

des concentrations ce jour-là, les différences simulations-observations sont plus homogènes que les concentrations observées. Le semivariogramme construit sur les innovations (Cf. figure 4.9) est alors aplati aux grandes distances et l'analyse de meilleure qualité (l'erreur d'analyse est inférieure à 3 ppb sur la majeure partie du domaine). L'information apportée par le modèle devient cruciale.

4.5 Validation en altitude des analyses obtenues par la méthode ASI

L'objet de ce paragraphe est de montrer que la méthode ASI peut être utilisée pour construire des analyses tridimensionnelles de polluants à partir seulement des observations de surface.

Le modèle de covariance horizontal est étendu dans la direction verticale et le biais du modèle, principalement dû aux émissions, est supposé verticalement uniforme dans la couche de mélange. Il n'est pas corrigé dans les couches supérieures non influencées par les processus de surface. Nous comparons les simulations et les analyses tridimensionnelles produites à l'aide de la méthode ASI aux observations d'ozone aéroportées effectuées durant la campagne ESQUIF par les vols ARAT et DIMONA présentés dans le chapitre 3. La figure 4.21 présente les résultats de cette comparaison en terme de RMS et d'écart type (STD) des écarts. La RMS est systématiquement réduite indiquant que l'assimilation des observations de surface a un impact positif non négligeable sur les analyses en altitude. La STD est aussi réduite, mais en moindre mesure, montrant que cet impact se traduit principalement par une réduction du biais du modèle.

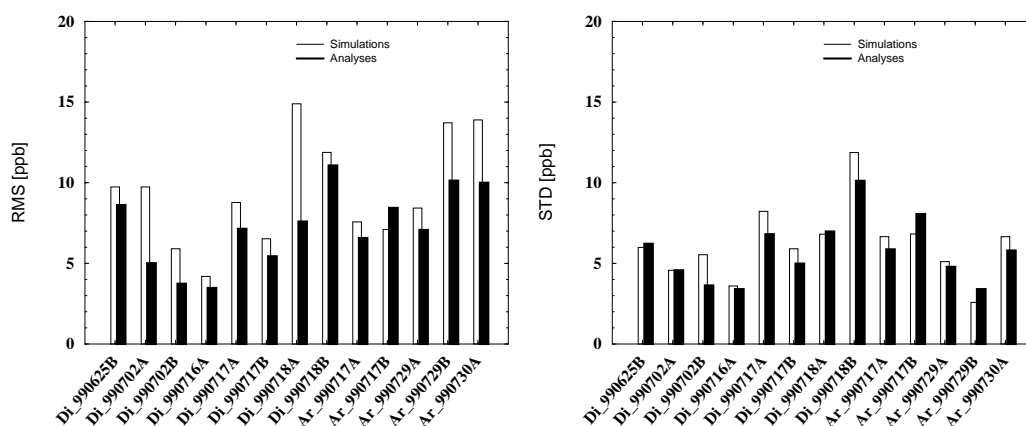


FIG. 4.21 – RMS des différences constatées entre les mesures d'ozone effectuées par les avions de la campagne ESQUIF, et les simulations de CHIMERE (niveau 2 ou 3) ou les analyses. Les barres vides correspondent aux simulations brutes. Les barres pleines aux analyses. L'unité de la RMS est en ppb. Di pour DIMONA. Ar pour ARAT. Le vol Di_990717A correspond au vol DIMONA effectué le 17 Juillet 1999 le matin. A pour matin, B pour après-midi.

Enfin, la figure 4.22 donne la série temporelle des concentrations d'ozone mesurées toutes les minutes par l'ARAT pendant la Période d'Observation Intensive (POI) du 30/07/1999 dans l'après-

midi. Même si les structures les plus fines du polluant ne sont pas plus résolues par l'analyse que par les simulations brutes, le champ de concentration est tout de même mieux représenté. Les observations de surfaces permettent de réduire le biais de 5 à 10 ppb. Le comportement du modèle est, de plus, amélioré dans le panache.

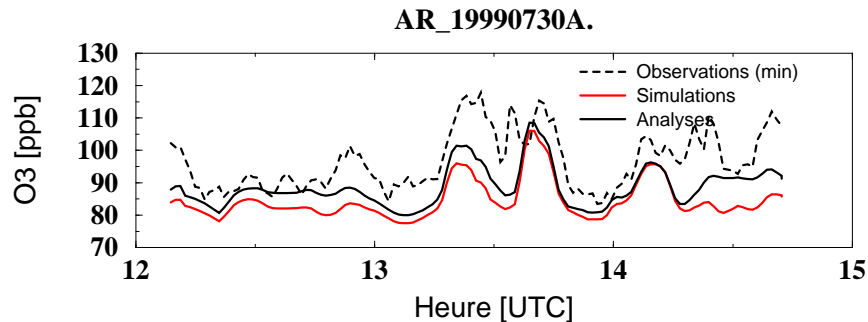


FIG. 4.22 – Série temporelle des concentrations d'ozone mesurées par l'avion ARAT pendant la POI du 30/07/1999 dans l'après-midi. Les données ont été moyennées toutes les minutes. L'unité des concentrations est en ppb, celle de l'heure en UTC.

4.6 Conclusions

Nous avons présenté dans ce chapitre quatre méthodes différentes pour analyser le mieux possible les champs 2D ou 3D de concentrations d'ozone. Ces méthodes sont appliquées sur la région d'Île-de-France. Elles sont comparées et discutées.

La méthode ISI est une méthode très utilisée en météorologie pour produire des analyses. Cette méthode a été utilisée comme référence initiale. Les covariances des erreurs du modèle sont estimées en fonction de la distance qui les sépare. Son utilisation a montré quelques inconvénients, notamment durant les heures de fortes émissions et la nuit, où les émissions rendent l'atmosphère inhomogène et anisotrope. De plus, vu le manque d'observation en altitude, la méthode est limitée à des analyses de champs bidimensionnels de polluants en surface.

C'est pourquoi nous avons développé une nouvelle méthode (méthode ASI) pour construire un modèle de covariance inhomogène et anisotrope. Les covariances des erreurs du modèle sont modélisées en fonction de la covariance des concentrations simulées. Cette méthode simple s'est montrée plus efficace quand la méthode ISI présentait des faiblesses. En outre, suffisamment générale, elle est utilisée pour corriger des champs tridimensionnels de polluants. Les mesures de la campagne ESQUIF ont permis une validation de la méthode dans la couche de mélange. Cet algorithme a été testé en temps quasi réel au Laboratoire de Météorologie (IPSL) dans le cadre du projet PIONNER. Il est installé à AIRPARIF depuis mai 2001. Chaque jour de l'été 2001, les données d'observation de la veille ont été introduites dans le modèle CHIMERE (autre version que celle qui est présentée dans ce document) juste après leur réception. Elles étaient, en général, disponibles en fin de matinée, sur le site <http://www.airparif.asso.fr>.

Les méthodes de Krigeage, empruntées aux sciences minières, consistent à interpoler spatialement des données. Ces méthodes sont utilisées pour éviter l'hypothèse de stationnarité du modèle de covariance supposée dans l'utilisation des méthodes précédentes. Les poids de l'interpolation sont ici estimés à partir des données disponibles à l'instant de l'analyse, sans utiliser leur dimension temporelle. Les poids sont alors variables dans le temps. Deux sortes de données ont été interpolées : les observations et les différences simulations-observations.

La méthode OBK consiste à interpoler spatialement les observations. Cette approche est d'autant plus performante que le nombre d'observations bien réparties est suffisant. Ce nombre "minimum" dépend de la complexité de la structure du champ des espèces que l'on souhaite analyser. Cette méthode montre inévitablement des faiblesses à détecter des panaches hors des zones observées.

La méthode INK est basée sur le même principe que la méthode OBK, mais au lieu d'interpoler les observations, l'idée originale a été d'interpoler les différences simulations-observations. Cette méthode conduit à de bien meilleures performances par rapport à la méthode OBK car le modèle structure le champ estimé. C'est à dire que, plus le modèle est correct, plus les différences entre les observations et les simulations sont homogènes spatialement, et plus un Krigeage est efficace. Il est évident que la qualité de cette méthode, comme celle de la méthode ASI, dépend de la qualité du modèle utilisé. Les principales différences entre la méthode INK et la méthode ASI ont été observées dans des cas "atypiques", où les structures observées sur les champs d'ozone s'écartaient des structures "usuelles" observées.

Toutes ces méthodes ont permis d'analyser les champs d'ozone, simulés par modèle CHIMERE régional centré sur Paris pendant les étés 1999 et 2001 (du 1^{er} Mai au 31 Septembre). Nous avons montré que, quelque soit la méthode utilisée, la RMS des différences analyses-observations est 30% à 50% plus faible (en *leave-one-out*) que la RMS des différences simulations-observations. La combinaison des observations et des simulations améliore incontestablement la représentation des concentrations du polluant.

En ce qui concerne l'initialisation du modèle pour la prévision de la qualité de l'air, la méthode ASI semble plus appropriée dans la mesure où elle permet une correction en altitude. Corriger les concentrations dans la couche de mélange pendant les heures où elle est le plus développée, c'est assurer une correction partielle des concentrations de l'ozone contenues dans la couche résiduelle la nuit. Or, le lendemain, lors du lever de la couche limite, l'ozone des couches supérieures est mélangé aux polluants contenus dans la couche de mélange. La correction de la veille devrait avoir un impact sur les prévisions du jour. Le prochain chapitre est précisément consacré à l'application de la méthode ASI pour produire des champs tridimensionnels de concentrations d'ozone à l'échelle continentale, en vue d'utiliser ensuite les analyses produites pour initialiser le modèle de prévision.

Chapitre 5

Analyse des champs de concentrations d’ozone à l’échelle continentale et prévision de la pollution

Ce chapitre est destiné à l’étude de la prévisibilité de la pollution photo-oxydante à l’échelle européenne. Les travaux ont été menés dans le cadre du projet PIONEER (Prévisibilité et Incertitudes de l’Ozone à l’Echelle Européenne et Régionale). L’objectif de ce chapitre est en particulier de quantifier l’impact d’une initialisation du modèle CHIMERE avec des analyses tridimensionnelles sur les prévisions à court terme.

La méthode ASI (Interpolation Statistique Anisotrope) est utilisée pour produire des analyses de champs de concentration d’ozone à l’échelle européenne. Nous utilisons pour cela la version continentale du modèle CHIMERE. La mise en place de l’algorithme d’Interpolation Statistique à cette échelle n’a pas présenté de difficulté supplémentaire par rapport à l’échelle régionale (Chapitre 4). Cependant, il a fallu, dans un premier temps, collecter au niveau européen un ensemble de données d’observation. En effet, depuis le début de l’été 2001, une interface modèles/données, comprenant l’acquisition en temps quasi-réel de données d’ozone sur l’Europe, a été développée. Les données d’ozone, partiellement validées nous sont gratuitement mises à disposition par une trentaine d’associations de surveillance de la qualité de l’air, listées en Annexe C dans les tableaux C.1 et C.2.

Ce chapitre est organisé de manière analogue au chapitre précédent, avec premièrement, une description de la mise en place de l’algorithme d’assimilation de données (nous insisterons particulièrement sur les développements nouveaux par rapport à l’échelle régionale), puis une validation objective de l’algorithme (méthode du *leave-one-out*). Enfin, les expériences de prévision sont présentées.

5.1 Mise en place d'une Interpolation Statistique

5.1.1 Problème des biais d'observation et de modélisation

L'étude statistique des écarts simulations-observations a mis en évidence un biais *global*. Celui-ci englobe le biais de modélisation et le biais d'observation, qui est principalement un biais de représentativité. Comme les concentrations moyennes simulées par le modèle CHIMERE continental sont plus homogènes que celles simulées par le modèle régional, la paramétrisation du biais de modélisation en fonction du critère *concentration moyenne* n'est pas envisageable. La sélection des stations est finalement la meilleure solution pour limiter les biais d'observation.

5.1.2 Choix des stations

Toutes les observations disponibles en temps réel lors de l'expérience PIONEER 2001 n'ont pas été utilisées. En effet, nous avons sélectionné les stations, dont la représentativité est la plus proche de celle associée à la résolution du modèle, c'est à dire principalement les stations rurales. Pour certaines grandes régions, où il y a très peu de stations de mesure d'ozone hors des zones urbaines, nous avons opté pour des stations péri-urbaines, ou si ce n'était pas possible, les stations urbaines qui donnaient des valeurs moyennes observées les plus fortes possible. Au total les observations de 232 stations sont utilisées. Leur distribution est représentée sur la figure 3.1. Cette sélection de stations a été effectuée pour limiter les erreurs de représentativité. Nous avons cherché en particulier (1) à éviter les trop forts biais d'observation, et (2) à limiter les corrélations d'erreur d'observation.

5.1.3 Modélisation de la fonction de covariance

La méthodologie pour construire le modèle de covariance hétérogène et anisotrope suit celle qui a été exposée au chapitre 4, dans le paragraphe 4.1.2. C'est à dire que les corrélations des innovations sont modélisées en fonction de la corrélation des concentrations d'ozone observées. Cependant, par rapport à l'échelle régionale, la variance des innovations n'est plus supposée homogène. Cette variance est modélisée en fonction de la variance des concentrations d'ozone observées. Autrement dit, les variances des erreurs de modélisation et d'observation dépendent maintenant du lieu où elles sont estimées. Seul le rapport de la variance de l'erreur de modélisation par la variance de l'innovation est supposé homogène. Nous rappelons que ce rapport est donné par la valeur de la fonction de corrélation quand la corrélation des concentrations simulées est égale à 1. La fonction de corrélation, qui est ajustée au corrélogramme, est, comme à l'échelle régionale la fonction f_C (proposée par l'équation 4.1. Nous avons ainsi construit sur la base de données de l'été 1999, un modèle de covariance à chaque heure de la journée.

5.2 Qualité des analyses des champs de concentrations d'ozone

Ce paragraphe est consacré à l'étude de la qualité des analyses des champs de concentrations de l'été 2001 à l'échelle européenne.

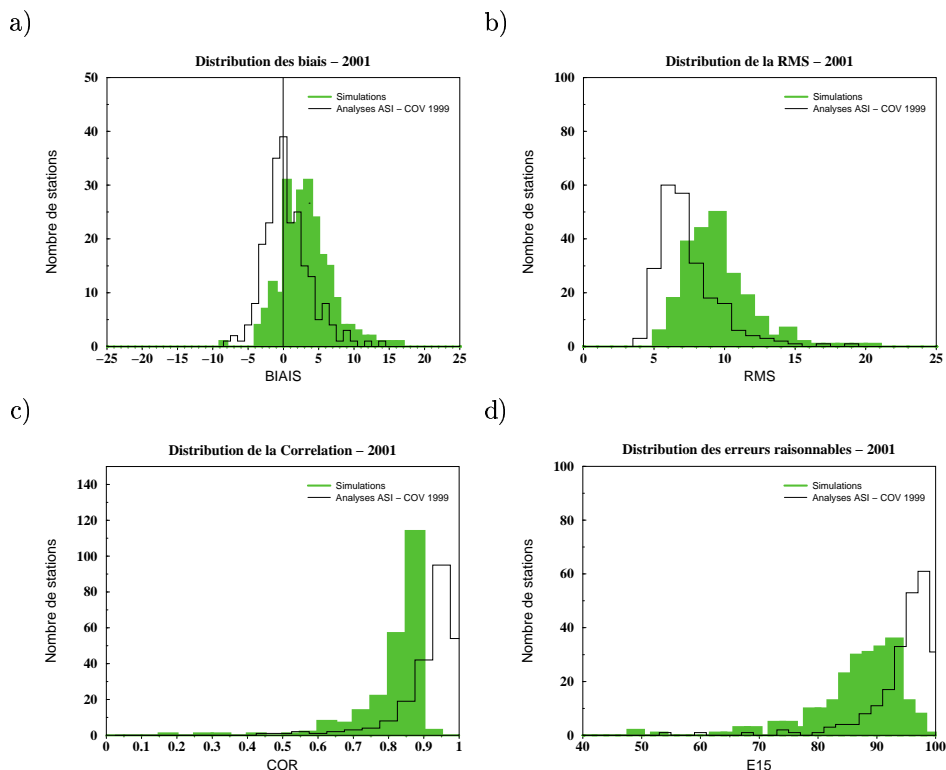


FIG. 5.1 – Statistiques effectuées sur les différences entre les concentrations d'ozone observées et les valeurs simulées par CHIMERE continental (barres pleines grises) ou analysées (barres noires vides) à 15 h UTC. Ces statistiques sont calculées sur 232 stations de mesure européennes. a) Biais b) RMS c) Corrélation d) Erreurs raisonnables. L'unité des concentrations est en ppb.

5.2.1 Utilisation des observations de surface

Validation Objective : le *leave-one-out*

Pour quantifier l'apport des observations dans la représentation des champs de concentrations d'ozone, nous utilisons la méthode du *leave-one-out*, exposée dans le paragraphe 4.3.1. La figure 5.1 présente les valeurs statistiques de vérification (biais, RMS, corrélation, erreur raisonnable) sous forme d'histogrammes. Ces valeurs sont calculées sur les différences entre les concentrations journalières d'ozone à 15 h UTC observées et les valeurs analysées, pour chacune des 232 stations. Nous avons également reporté sur cette figure les histogrammes, que nous avons obtenus avec les simulations et discutées dans le paragraphe 3.4. Le tableau 5.1 donne le détail de ces statistiques.

Le biais des analyses est centré sur zéro. L'assimilation des observations permet globalement de réduire le biais de 3.2 ppb (biais initial des simulations). En effet, les analyses se rapprochent de la moyenne des observations. Or, comme sur la plus grande partie des stations, les concentrations sont surestimées par le modèle, l'analyse réduit les biais positifs et augmente les biais négatifs. Environ 95% des biais d'analyse sont compris entre -5 et 5 ppb. Ces biais résiduels correspondent approximativement à l'opposé du biais d'observation.

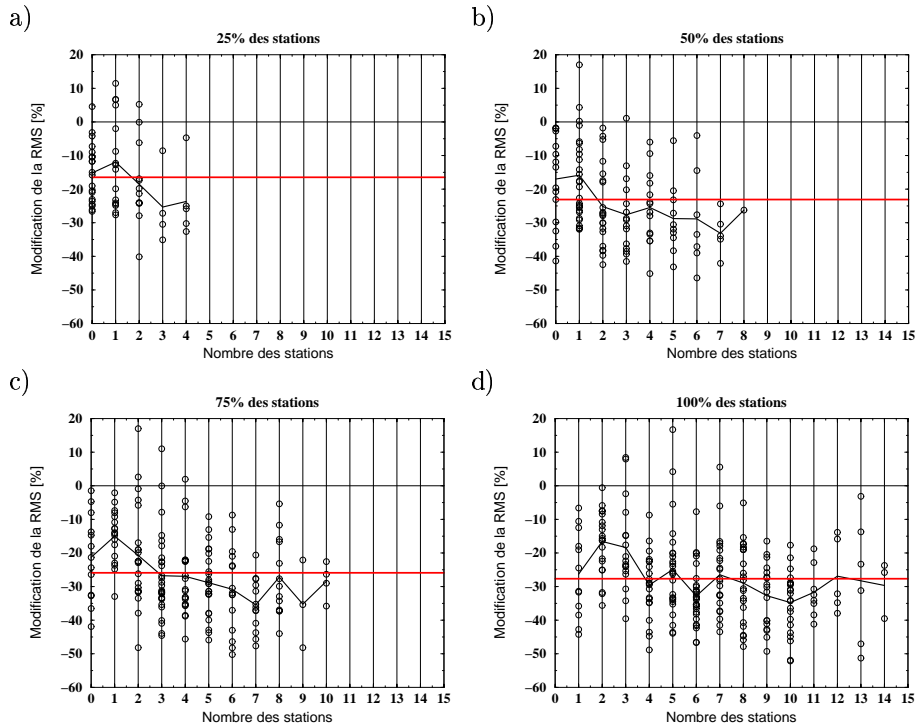


FIG. 5.2 – Modification de la RMS des différences entre simulations et observations après assimilation des observations (en pourcentage) en fonction du nombre de stations entourant le point analysé dans un rayon de 100 kilomètres. Le trait horizontal gris marque la moyenne sur l'ensemble des stations.

Les corrélations entre les concentrations analysées et celles observées sont pour la majeure partie comprises entre 0.9 et 1. Elles sont systématiquement plus fortes que celles calculées entre les simulations et les observations. Par rapport aux simulations, l'assimilation des observations permet une augmentation des plus fortes corrélations de l'ordre de 10%. Les corrélations les plus faibles, constatées sur les stations de bord de mer ou dans les régions à topographie complexe, sont quant à elles, augmentées de plus de 10%. La variabilité des concentrations est ainsi bien mieux représentée par les analyses sur le Massif Central, au relief relativement peu accidenté (les corrélations sont augmentées de 15% à 20%). En revanche, certaines corrélations, qui correspondent à des stations situées dans les régions plus montagneuses, au pied des Pyrénées, dans les Alpes, ou alors près des côtes, restent faibles.

Le pourcentage d'erreurs raisonnables se situe généralement au dessus de 95%. L'assimilation des observations permet ainsi de réduire le nombre de "grosses erreurs".

Enfin, la RMS est réduite en moyenne d'environ 30%, passant d'une moyenne de 9.3 à une moyenne de 6.9 ppb. Elle est réduite en général sur les régions les plus densément couvertes par la mesure. Cependant, ce n'est pas toujours le cas. Ainsi, pour étudier l'influence de la densité de stations sur la qualité de l'analyse, une expérience a été menée. Elle consiste à produire des

5.2 Qualité des analyses des champs de concentrations d'ozone

TAB. 5.1 – Noms, locations et statistiques d'erreurs de comparaison analyses-observations faites sur les concentrations d'ozone à 15 h UTC, station par station et en moyenne sur l'été 2001. Nous avons reporté les valeurs statistiques obtenues avec les simulations du tableau 3.3 pour le biais et la RMS : simulations/analyses. L'unité des concentrations est en ppb.

Station	BIAS	RMS	COR	E15	Station	BIAS	RMS	COR	E15
PLAN-D-AUPS	7.4/5.2	14.1/11.7	0.82	75	KOENIGSTEIN	3.1/-0.6	10.4/7.3	0.92	93
CADARACHE	2.6/-1.1	10.8/11.4	0.78	88	SPESSART	-1.6/-3.7	8.4/6.6	0.95	97
COMTAT-VENAISSIN	9.1/6.3	12.4/10.2	0.86	83	FRANKENBERG	-2.2/-4.2	8.3/6.0	0.96	98
AIX-PLATANES	0.1/-3.4	12.9/12.3	0.84	87	FURTH-ODENWALD	2.8/-0.4	9.2/6.0	0.94	97
TOULON-OLIVE	1.1/0.0	11.3/9.1	0.79	90	BURG-HERZBERG	-2.0/-3.7	7.6/5.4	0.96	100
NANTES-JARDIN	8.3/5.8	11.5/8.1	0.92	95	BAD-AROLSSEN	1.9/-0.0	8.0/5.3	0.94	98
ANGERS-ALLARD	1.9/-2.3	7.4/4.6	0.96	100	LOCKENITZ	0.6/-0.7	5.9/4.5	0.91	98
ST-NAZAIRE-BLUM	4.1/0.7	8.6/5.6	0.90	95	GOEHLEN	5.3/3.9	8.6/6.7	0.87	96
LE-MANS-BELAIR	5.8/1.6	9.7/5.4	0.93	96	SOLLING-DASSEL	4.0/2.2	9.0/6.2	0.92	97
CHOLET-ST-EXUPERY	2.2/-1.3	7.5/4.7	0.96	98	LINGEN	0.7/-3.2	10.7/8.2	0.95	93
LAVAL	5.4/0.7	9.9/5.3	0.94	98	WURMBERG-BRAUNLAGE	3.8/1.9	7.5/5.0	0.94	100
LA-ROCHE-SUR-YON	5.2/2.4	9.5/6.5	0.91	96	NORDERNEY	1.6/-0.9	8.9/8.4	0.72	92
ST-GOBAIN-SINCENY	12.2/8.2	14.7/9.9	0.92	89	CUXHAVEN	2.1/-0.1	7.6/6.3	0.86	97
CRECY-EN-PONTHIEU	-2.2/-7.5	9.7/10.5	0.91	85	JORK	2.9/0.6	6.3/5.0	0.91	97
CHATEAU-THIERRY	9.6/5.9	12.6/8.4	0.91	93	HERTEN	1.8/-1.4	9.9/7.0	0.93	95
ALBERT	3.6/-1.3	10.2/6.7	0.91	95	HUERTH	3.6/-0.5	10.1/6.9	0.94	95
FORET-DE-BROTONNE	3.6/-1.8	8.4/6.1	0.93	96	EGGEBIRGE	4.7/3.0	9.1/5.8	0.94	100
EUREUX-ST-MICHEL	1.5/-3.6	8.0/6.3	0.95	96	EPFEL	3.0/1.2	7.6/5.8	0.93	98
PHARE-D-AILLY	4.6/-0.4	12.0/8.7	0.80	91	ROTHARGEIRGE	3.8/1.1	8.2/5.0	0.95	100
ST-ROMAIN-DE-COLBOSC	3.9/-2.2	9.3/7.2	0.89	95	NETTETAL	2.8/-0.9	9.7/7.2	0.92	95
FREMAINVILLE	-0.2/-5.0	8.8/7.4	0.95	95	NIEDERZIER	1.8/-2.3	9.3/6.1	0.96	96
FRUNAY	3.5/-0.8	9.2/4.7	0.96	100	FINNENTROP	6.5/3.0	9.4/5.9	0.95	98
SAINTS	2.8/-1.0	9.0/6.4	0.91	97	SOLINGEN	0.7/-3.1	10.7/8.2	0.95	93
NOISY-SUR-OISE	1.2/-3.1	8.1/6.3	0.96	96	BORKEN-GEMEN	1.3/-1.8	8.9/6.2	0.95	96
SONCHAMP	0.6/-2.7	9.2/6.3	0.94	97	AACHEN-BURTSCHIED	0.6/-3.5	9.8/7.2	0.95	97
FONTAINEBLEAU	4.0/1.7	9.9/7.5	0.89	94	WESTEIFEL-WASCHEID	3.0/1.3	7.6/5.6	0.93	98
MONTGE-EN-GOELE	2.1/-2.3	8.1/5.6	0.96	99	WESTERWALD-HERDORF	3.5/0.3	9.0/5.4	0.95	99
MONTPELLIER-NORD	-2.6/-3.5	9.1/7.3	0.93	95	PFÄLZERWALD-H.	4.2/0.1	10.0/7.0	0.92	97
MONTPELLIER-SUD	-1.0/-0.6	11.3/8.6	0.87	92	NEUSTADT-STROHMARKT	15.9/13.1	18.2/14.8	0.92	61
ST-ESTEVE	1.4/-0.9	8.9/7.8	0.81	95	WESTERWALD-NEUHSL	3.5/-0.5	9.1/6.6	0.93	99
NIMES-GARDIES	-4.0/-5.3	8.6/7.1	0.95	96	ALTENDEICH	0.8/-1.9	5.8/4.9	0.93	99
SAZE	-1.3/-3.6	7.9/6.4	0.93	99	BRUNSBUETTTEL	-0.6/-3.2	6.4/5.3	0.93	99
VALLABREGUES	-2.2/-4.0	9.0/7.3	0.91	94	BEHMARN	-3.6/-3.3	10.8/9.0	0.53	87
NE-ALSACE	3.2/-0.9	9.7/6.3	0.94	97	ST.-PETER-ORDING	-1.1/-2.9	7.3/5.8	0.90	97
VOSGES-MOYENNES	2.7/-1.4	7.0/4.3	0.97	100	GLUECKSBURG	-1.7/-3.4	6.5/6.0	0.85	97
VOSGES-NORD	5.8/1.6	9.7/5.7	0.94	98	SAARBUECKEN	6.7/2.0	10.3/6.8	0.93	95
VOSGES-HAUTES	9.1/5.3	11.4/7.7	0.94	96	CARLSFELD	6.0/3.4	9.7/6.0	0.93	98
3FRONTIERES	5.1/0.5	11.7/6.3	0.95	97	HOVERSVERDA	2.5/-1.5	7.7/4.5	0.96	100
MULHOUSE-EST	2.1/-2.8	9.4/6.7	0.94	95	RADEBEUL-WAHNSDORF	3.9/0.5	9.0/4.3	0.96	100
COLMAR-SUD	0.2/-4.5	9.4/7.9	0.94	95	ZINNWALD	0.7/-4.0	7.9/6.3	0.94	97
STRASBOURG-UEST	1.4/-2.7	10.0/8.6	0.92	94	FICHTELBERG	1.5/-2.0	8.2/5.8	0.92	98
MONTFLUCON-HIPPODROME	0.4/-2.5	6.9/5.2	0.94	100	MITTELNDORF	6.2/1.9	10.7/5.8	0.93	98
AURILLAC-AERODROME	4.6/-0.1	8.2/5.6	0.86	98	COLLMBERG	3.1/0.4	8.2/4.8	0.95	99
PUY-DE-DOME	2.2/-1.2	7.1/6.0	0.99	96	WERNIGERODE	4.4/2.7	7.8/5.6	0.93	99
VALS-PRES-LE-PUY	4.5/0.8	9.2/6.1	0.87	98	POUCH	3.8/1.2	8.4/5.0	0.94	99
BRENNILIS	1.4/-2.0	5.7/4.7	0.92	98	HARZGERODE	0.7/-1.6	6.6/4.4	0.96	99
GUIPRY	7.0/3.1	10.4/6.8	0.89	98	NEUHAUS	3.9/2.4	7.8/4.4	0.96	100
BREST-JEAN-MACE	0.2/-0.6	7.6/7.0	0.74	93	POSSEN	0.9/-0.8	6.2/3.4	0.97	100
RENNES-COURTEL	2.5/-2.8	7.7/6.1	0.93	98	WESTERLAND	-0.3/-1.3	7.3/5.7	0.83	97
LORIENT-VARLIN	2.2/-1.4	6.8/5.3	0.91	98	DEUSELBACH	3.8/-0.8	7.2/5.2	0.94	99
GAUDONVILLE	3.0/-2.3	7.1/5.2	0.92	99	BROTJAACKLRIEGEL	5.6/0.8	9.3/5.3	0.91	97
PEYRUSSE-VEILLE	8.6/4.4	10.7/6.5	0.91	97	SCHAUNSLAND	4.3/0.0	10.0/6.0	0.94	97
ALBI-CANTEPAU	10.1/6.4	12.3/8.4	0.89	91	WALDHOF	-0.1/-3.0	7.2/6.6	0.92	94
BELESTA-EN-LAURAGAIS	4.3/-1.2	8.8/5.8	0.89	100	BASSUM	6.0/2.7	10.9/8.3	0.85	94
REVIN	2.9/-2.2	8.7/5.9	0.94	99	GITTRUP	4.5/1.9	9.7/5.8	0.94	97
ST-PARRES	3.3/-0.0	8.0/5.2	0.93	99	REGNITZLOSAU	0.9/-2.0	7.7/5.2	0.94	100
REIMS-VAL-DE-MURIGNY	6.9/2.4	10.4/6.1	0.93	98	UECKERMUENDE	0.4/0.2	6.3/5.0	0.88	98
HALLUIN	7.1/5.5	12.7/9.4	0.91	91	ZINGST	-2.4/-1.6	10.8/9.0	0.40	89
CHAMBORD	0.8/-1.6	9.2/6.2	0.91	96	SCHMUECKE	3.2/1.3	7.3/4.0	0.96	100
ORLEANS-LA-SOURCE	0.0/-2.5	7.8/5.2	0.95	100	NEUGLOBSOW	3.5/1.6	9.0/6.1	0.86	96
TOURS-DAMES	0.8/-2.0	9.0/6.8	0.91	96	LUECKENDORF	11.2/7.5	13.2/9.0	0.91	94
BOURGES-LEBLANC	-0.5/-2.6	7.3/5.4	0.94	98	LEINEFELDE	-0.2/-2.0	6.7/4.3	0.96	100
CHARTRES-FULBERT	-0.1/-4.1	8.1/6.9	0.95	94	MELPITZ	0.8/-2.0	9.4/6.4	0.91	97
CHATEAUROUX-SUD	0.2/-1.8	7.8/5.2	0.93	97	HELGOLAND	5.6/3.8	9.4/7.1	0.84	94
AIRVAULT	1.5/-0.9	6.6/3.9	0.97	100	LEHNMUEHLE	4.9/1.3	8.5/5.6	0.93	98
CHIZE	-1.4/-3.7	7.6/5.5	0.97	98	AUKRUG	2.2/0.3	5.8/4.4	0.92	100
LA-ROCHELLE-AYTRE	0.4/-1.8	9.0/6.7	0.90	95	FALKENBERG	0.9/-2.3	6.3/4.6	0.95	99
L-ISLE-D-ESPAGNAC	1.3/-1.5	7.4/5.2	0.93	98	SCHORFHEIDE	0.9/-1.5	7.0/4.7	0.94	100
POITIERS-CHASSENEUIL	0.8/-1.4	7.5/5.0	0.95	98	ASTON-HILL	2.3/0.8	7.8/5.3	0.88	97
NIORT-JEAN-ZAY	4.1/2.1	8.3/5.0	0.95	97	SOMERTON	-1.5/-3.8	8.7/7.8	0.84	94
COGNAC	3.5/1.3	8.0/4.2	0.96	100	LADYBOWER	0.1/-1.6	5.7/6.1	0.86	100
FLOIRAC	5.3/2.1	9.2/7.0	0.90	99	LOUGH-NAVAR	3.1/1.9	7.4/6.9	0.81	94
LE-TEMPLE	-0.1/-3.7	10.4/9.6	0.81	88	SIBTON	1.1/1.7	9.4/8.3	0.72	90
PAU-LE-HAMEAU	6.9/2.7	9.5/6.6	0.84	98	WICKEN-FEN	1.8/1.5	7.9/6.7	0.89	96
BIARRITZ	3.3/-0.4	9.6/8.2	0.76	93	NARBERTH	1.9/0.1	8.0/7.2	0.77	93
IRATY	7.7/3.6	10.8/8.2	0.63	91	LULLINGTON-HEATH	-2.7/-3.5	10.5/10.4	0.70	90
DIJON-DAIX	4.0/-0.1	11.9/9.3	0.84	97	HIGH-MUFFLES	-1.7/-2.4	8.3/6.2	0.86	98
ST-BRISSON	7.5/4.4	10.4/7.5	0.89	95	EKSDALEMUIR	5.7/3.8	7.9/6.2	0.87	99
CHALON-HILAIRE	7.5/2.6	11.5/7.7	0.90	92	HARWELL	2.5/1.5	11.1/9.0	0.81	93
MONTCEAU-LES-MINES	8.0/3.9	12.3/7.9	0.92	94	YARNER-WOOD	1.5/-0.6	6.2/5.4	0.90	97
MACON-CHAMPLEVERT	5.1/-0.5	10.2/5.9	0.94	98	GLAZEBURY	5.2/4.5	9.2/8.7	0.82	91
VALENCE-SUD	7.6/3.2	12.4/7.7	0.90	93	BUSH-ESTATE	1.4/-1.7	6.4/5.8	0.78	97

TAB. 5.2 – Suite du tableau 5.1.

Station	BIAS	RMS	COR	E15	Station	BIAS	RMS	COR	E15
CHARAVINES	6.4/1.2	11.1/6.5	0.91	95	BOTTESFORD	-3.8/-5.7	9.6/9.3	0.86	90
ST-ETIENNE-COUBERTIN	2.4/-2.1	13.5/10.9	0.86	85	ROCHESTER	-3.0/-3.4	11.9/11.0	0.62	84
DIEME	2.4/-2.4	11.5/9.1	0.88	92	REDCAR	0.7/0.7	8.8/6.6	0.80	96
ANNEMASSE-GAILLARD	3.2/-2.5	10.8/8.0	0.92	92	NL-934	3.1/0.8	8.1/6.3	0.84	96
ANNECY-NOVEL	6.7/1.7	10.5/5.8	0.93	99	NL-538	2.4/1.7	7.5/5.0	0.92	100
CHAMBERY-LE-HAUT	6.6/1.7	12.3/7.2	0.91	96	NL-107	4.4/0.4	10.1/4.9	0.97	99
ROCHES-DE-CONDRIEU	2.4/-0.7	9.4/7.7	0.90	93	NL-918	3.9/0.9	8.3/5.3	0.92	99
ALENCON	3.0/-2.4	8.0/5.6	0.93	98	NL-913	4.5/1.4	8.1/5.2	0.90	98
LISIEUX	12.2/6.7	14.8/9.3	0.92	90	NL-318	-1.2/-1.9	9.1/6.3	0.91	95
OUISTREHAM	16.6/11.3	19.5/13.4	0.80	76	NL-235	6.1/4.0	11.3/7.6	0.93	95
RANVILLE	14.8/9.1	17.6/11.4	0.83	81	NL-230	6.8/3.1	12.7/7.3	0.94	95
CHERBOURG	5.8/2.1	11.2/8.5	0.49	96	NL-227	5.3/1.2	11.0/5.3	0.96	98
FORST	4.5/0.8	8.2/4.6	0.94	100	NL-722	6.5/3.1	11.6/6.4	0.94	97
WITTENBERGE	2.1/-0.4	5.9/3.9	0.95	100	NL-724	4.3/1.2	9.2/5.5	0.95	98
PREMNITZ	0.4/-2.6	6.0/4.8	0.95	100	NL-818	4.3/0.8	9.9/5.6	0.91	97
PRENZLAU	0.8/-1.1	6.0/4.8	0.91	98	NL-232	-8.2/-8.7	14.0/12.0	0.90	82
NAUEN	6.3/4.1	11.1/9.0	0.80	95	NL-633	3.5/-0.5	10.0/5.5	0.96	99
BERLIN-MARIENFELDE	2.0/-0.9	6.9/4.0	0.96	99	NL-633	0.3/-0.8	9.7/6.6	0.89	98
BERLIN-FR HAFEN	4.7/2.4	8.6/5.0	0.94	97	PAYERNE	2.0/-1.4	7.4/5.3	0.94	98
WIESLOCH	8.3/5.3	11.5/7.3	0.96	95	CHAUMONT	4.4/1.5	9.1/5.5	0.94	97
WEIL-AM-RHEIN	3.7/-0.9	13.2/8.9	0.89	90	JUNGFRAUJOCH	-0.0/-4.2	6.8/7.9	0.68	96
WELZHEIMER-WALD	1.9/-0.5	9.6/6.2	0.95	98	LAEGGEREN	6.2/3.0	9.3/5.9	0.96	98
SCHWARZWAL-SUED	4.3/-0.1	10.3/5.5	0.96	98	RIGI	3.6/-0.0	9.8/7.2	0.88	93
FREUDENSTADT	2.6/-0.8	8.0/6.1	0.94	98	TAENIKON	3.5/-0.2	9.2/4.8	0.96	100
KONSTANZ	7.3/3.4	11.7/6.6	0.94	96	MAGADINO	9.6/4.5	16.1/13.2	0.69	75
BADEN-BADEN	-1.9/-6.2	9.5/9.2	0.95	89	DAVOS	8.0/2.7	10.8/6.7	0.77	97
SCHWAEBISCHE-ALB	-1.3/-5.4	8.0/7.5	0.95	97	PARMA-CITADELLA	10.6/5.1	15.6/11.0	0.77	87
BAD-WALDSEE	4.6/0.7	8.6/4.4	0.95	100	MODENA-20-SETTEMBRE	2.5/-4.9	12.6/9.8	0.91	87
NEUSTADT-A.D.DONAU	0.7/-4.0	8.1/6.6	0.94	98	BOLOGNA-CUCCOLINO	13.4/8.3	20.1/18.6	0.50	54
BAD-REICHENHALL	11.5/6.7	14.2/9.1	0.89	89	GHERARDI	-1.6/-7.7	15.4/16.0	0.58	68
GARM.-PART.-WANK	5.4/-0.3	8.9/5.0	0.90	100	COTIGNOLA	2.5/0.0	14.1/11.7	0.79	85
BREMERHAVEN	7.8/5.4	11.6/9.5	0.83	88	FORLI-GIARDINI	4.7/-1.8	10.1/7.5	0.88	92
GREBENAU	0.5/-1.0	7.3/4.1	0.96	100	RIMINI-RICCIONE	10.7/8.4	14.8/12.6	0.68	82
WITZENHAUSEN	-3.1/-4.8	8.1/7.0	0.96	100	CASTEL-SAN-GIOVANNI	6.8/-0.1	13.8/8.9	0.85	89
MOYENNES	3.2/0.0	9.3/6.9	0.90	95					

analyses en n'utilisant que 75%, 50% ou 25% des séries estivales d'observations disponibles (les observations retenues sont sélectionnées au hasard parmi les 232 séries). Les analyses sont ensuite comparées aux mesures non utilisées (en leave-one-out). Nous distinguons parmi ces mesures, celles qui constituent les seules observations dans un rayon de 100 km, celles qui sont effectuées à moins de 100 km d'une, deux, puis trois, etc, autres mesures.

La figure 5.2 montre le pourcentage de modification de la RMS, engendrée par l'assimilation des observations, en fonction du nombre de stations entourant le point analysé dans un rayon de 100 km. Nous avons calculé à chaque expérience, quand 25%, 50%, 75% ou 100% des stations sont utilisées, le pourcentage moyen de modification de la RMS, celui-ci prend respectivement les valeurs, -17%, -23%, -26%, -28%. Ceci suggère que globalement, plus le nombre de stations est grand, plus la RMS est réduite. Dans chaque cas, nous notons que, même en des points éloignés, où les stations les plus proches sont à plus de 100 kilomètres (cas nombre de stations nul), la RMS peut être réduite de 10 à 20%. Nous remarquons également que les scores sont très variables d'une station à une autre. En général, plus le nombre de stations entourant le point analysé est grand, plus la RMS est réduite. Cependant ce n'est pas toujours le cas. Certaines stations détériorent même les analyses, mais ceci reste assez rare. Ces cas apparaissent lorsque le nombre de stations est faible (1, 2, 3 stations à moins de 100 km). Le problème provient du biais. En effet, celui-ci est localement corrigé en fonction de l'écart moyen des différences simulations-observations constatées sur les stations environnantes. Plus le nombre de stations est faible, plus une mesure très perturbée par des effets locaux, au biais d'observation fort, peut avoir de l'influence dans l'interpolation aux points analysés les plus proches, et ainsi détériorer les analyses. La RMS est généralement réduite de 10 à 50%, les plus fortes réductions de la RMS étant constatées sur les points souvent les plus entourés de stations de mesure.

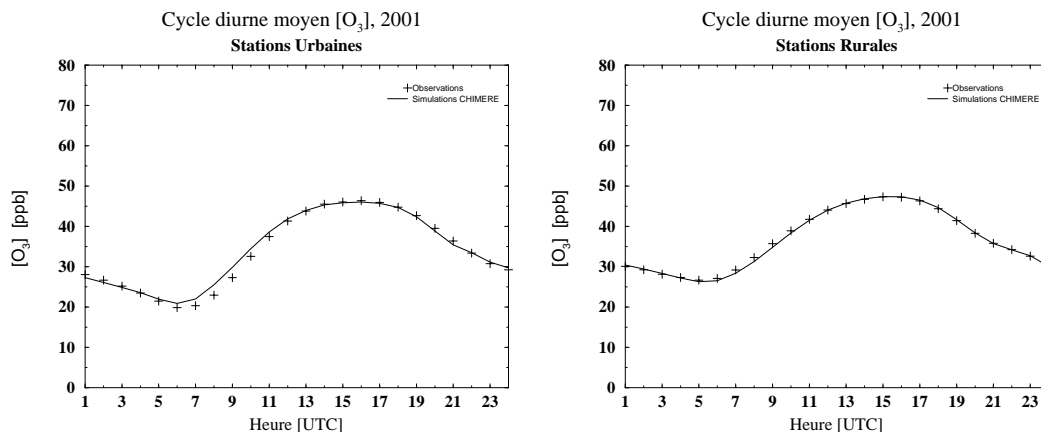


FIG. 5.3 – Profils diurnes moyens des concentrations d'ozone observées et analysées sur les stations urbaines et sur les stations rurales. Les analyses sont produites à partir des simulations du modèle CHIMERE continental et des observations de surface à l'aide de la méthode ASI. L'unité des concentrations est en ppb, celle de l'heure en UTC.

Les cycles diurnes

Les analyses sont produites à chaque heure de la journée. Pour donner une idée des performances de la méthode ASI sur un cycle diurne, nous présentons sur la figure 5.3 les résultats d'une comparaison entre le cycle diurne moyen observé et le cycle diurne moyen analysé sur les stations urbaines ou les stations rurales. Cette figure est à comparer avec la figure 3.8. Alors que les simulations surestiment le cycle diurne, les analyses le représentent très correctement.

Les profils temporels

Pour illustrer quelques analyses des concentrations, nous présentons sur la figure 5.4 des séries temporelles des concentrations d'ozone observées et analysées à 15 h UTC du 1^{er} Mai au 31 août 2001. Toutes les observations sont utilisées pour produire les analyses (pas de *leave-one-out*). Nous reportons également les valeurs simulées par le modèle continental que nous avons déjà commentées au chapitre 3. Globalement, les analyses représentent mieux les variations des concentrations d'ozone que les simulations. Les surestimations du mois de mai sont très bien corrigées. Les problèmes de représentativité, constatés sur la station NANTES-JARDIN, sont un peu réduits (le biais passe de 8.3 ppb à de 5.0 ppb). Les épisodes de fin Juillet, et de fin août sont plus correctement reproduits par les analyses que par les simulations brutes sur l'ensemble des stations. En revanche, sur la station SAINTS, le pic dépassant 100 ppb n'est pas corrigé, nous verrons dans la suite, que ceci est dû au caractère trop localisé du panache urbain.

5.2.2 Utilisation des observations d'altitude de la campagne ESQUIF

De la même manière qu'au Chapitre 4, nous utilisons les données aéroportées de la campagne ESQUIF pour contrôler la qualité des analyses en altitude. La figure 5.5 présente les résultats de la comparaison des concentrations observées, simulées et analysées en terme de RMS et d'écart type (STD). La RMS n'est pas systématiquement réduite, cependant ces résultats indiquent que

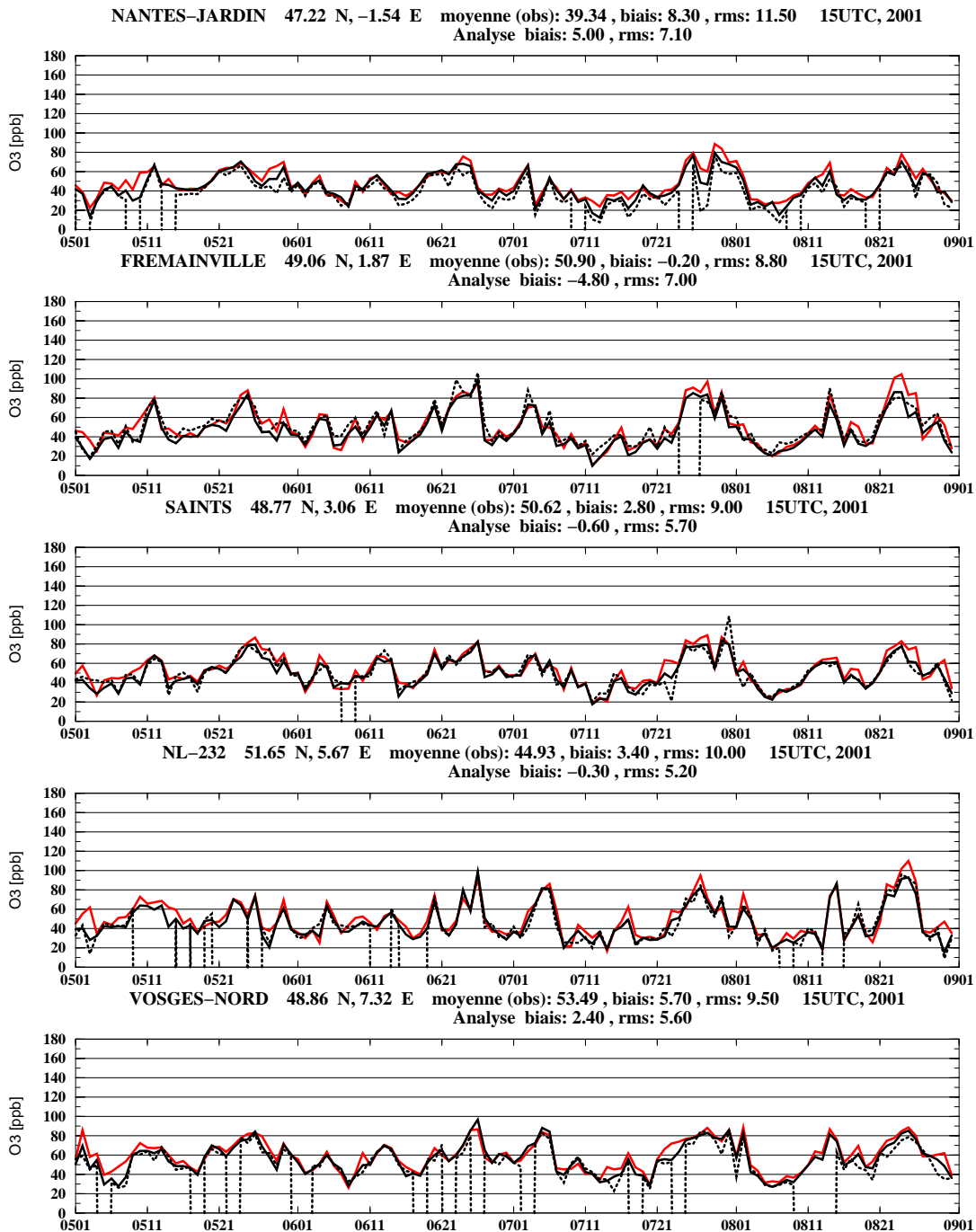


FIG. 5.4 – Séries temporelles des concentrations d'ozone observées (courbe pointillées), simulées par le modèle continental (courbe grise) et analysées (courbe noire) à 15 h UTC, du 1^{er} Mai au 31 août 2001. L'unité des concentrations est en ppb

l'assimilation des observations de surface a tout de même un impact positif non négligeable sur les analyses en altitude. Les cas où la RMS est réduite correspondent à une correction d'erreurs de grande échelle, comme des erreurs sur les concentrations de fond. En revanche, la STD n'est pas réduite, montrant que l'impact de l'introduction des observations de surface se traduit uniquement par une réduction du biais du modèle.

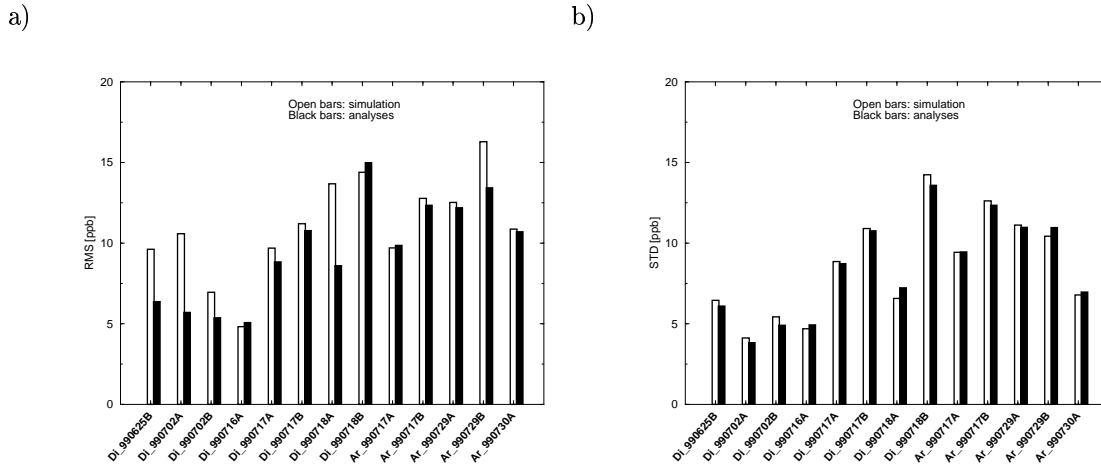


FIG. 5.5 – RMS des différences constatées entre les mesures d’ozone effectuées par les avions de la campagne ESQUIF, et les simulations de CHIMERE continental (niveau 2 ou 3) ou les analyses. Les barres vides correspondent aux simulations brutes. Les barres pleines aux analyses. L’unité de la RMS est en ppb. Di pour DIMONA. Ar pour ARAT. Le vol Di_990711A correspond au vol DIMONA effectué le 17 Juillet 1999 le matin. A pour matin, B pour après-midi.

Enfin, la figure 5.6 donne la série temporelle des concentrations d’ozone mesurées toutes les minutes par l’ARAT pendant la POI du 29 Juillet 1999 dans l’après-midi. Même si les structures les plus fines du polluant ne sont pas plus résolues par l’analyse que par les simulations brutes, le champ de concentration est en moyenne mieux représenté. Les observations de surface permettent de réduire le biais de 10 ppb dans le panache de l’agglomération.

5.3 Application à la prévision

5.3.1 Une série d’expériences en temps réel

L’expérience POLLUX

Depuis 1996, l’IPSL (LMD, SA) est activement engagé dans la prévision de la pollution atmosphérique. La première expérience de prévision en temps réel, POLLUX, a été menée en collaboration avec AIRPARIF, soutenu par l’Agence de l’Environnement et de Maitrise de l’Energie (ADEME) dans le cadre du programme de recherche PRIMEQUAL (1996) du Ministère de l’Aménagement du Territoire et de l’environnement. Durant l’été 1998, un serveur internet, destiné à un usage scientifique, et non à un usage opérationnel ou à une lecture grand public, proposait la visualisation des prévisions du modèle CHIMERE à trois jours des concentrations d’ozone et de dioxyde d’azote. A l’époque, le modèle CHIMERE simulait les concentrations moyennes des

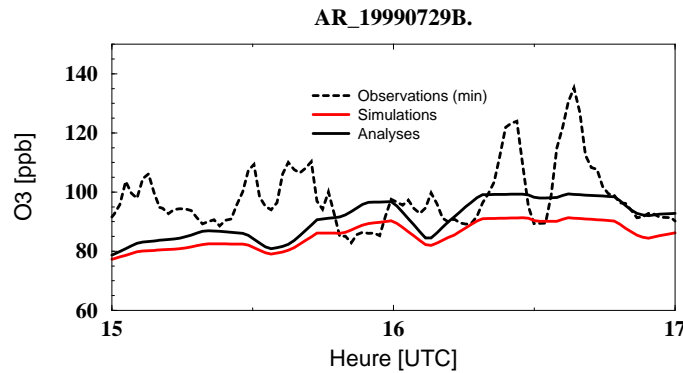


FIG. 5.6 – Série temporelle des concentrations d'ozone mesurées par l'ARAT pendant la POI du 29 Juillet 1999 dans l'après-midi. Les observations ont été moyennées toutes les dix minutes. L'unité des concentrations est en ppb, celle de l'heure en UTC.

polluants dans 10 boîtes sur la région d'Île-de-France. Les flux de polluants arrivant en amont de la région étaient calculés à l'aide d'un modèle de rétrotrajectoires et d'un modèle statistique, opérationnel sur l'Europe de l'Ouest [Vautard *et al.*, 2001]. Le système était forcé par les données du CEPMMT. Ces premières expériences en temps réel ont mis en évidence deux problèmes majeurs : la mauvaise prise en compte des conditions aux limites et le transport de l'erreur de prévision.

Le premier problème a été résolu par une approche de modèles grille emboîtés à différentes résolutions : le système CHIMERE multi-échelles est né. Le deuxième a motivé le travail de thèse exposé dans ce document.

Les cas du 29 et du 30 Juillet 1999 sont le résultat d'un problème dû aux conditions aux limites, et d'une accumulation d'erreur de prévision. La figure 5.7 présente des simulations de la version régionale 200102 du modèle CHIMERE centré sur Paris (version antérieure à celle présentée dans ce document), forcée aux limites par la version 200102 du modèle continental. Le modèle est forcé par les prévisions météorologiques du CEPMMT à 6 heures (qui correspondent aux *first guesses* ou premières ébauches). Les simulations de CHIMERE, présentées sur cette figure, sont consécutives, ce qui signifie que les simulations du 29 Juillet servent de conditions initiales aux prévisions du 30 Juillet (à minuit). La figure 5.7.a. présente les concentrations d'ozone pour le 29 Juillet 1999 à 15 h UTC. Ce jour-là, un temps sec et ensoleillé, associé à une absence de vent (environ 2 m.s^{-1}) et à des températures élevées (plus de 30°C), persiste en région d'Île-de-France. Le modèle CHIMERE simule un panache au sud-ouest de Paris. Il sous-estime de plus de 10 ppb les concentrations d'ozone sur tout le domaine. Cette sous-estimation est due à une sous-estimation des conditions aux limites du domaine régional. L'erreur est issue du modèle continental. Sans aucune correction des champs de concentrations d'ozone à l'échelle continentale, le modèle sous-estime encore plus les concentrations d'ozone le lendemain, et particulièrement au nord-ouest de Paris. La figure 5.7.b. montre la simulation pour le 30 Juillet 1999 à 15 h UTC. La situation météorologique n'a pas évolué par rapport au jour précédent. Comme les masses d'air stagnent dans le Nord de la France, les erreurs de la veille s'accumulent et s'ajoutent aux erreurs du jour même, et le modèle sous-estime cette fois-ci les concentrations d'ozone de plus de 25 ppb. Ainsi, si l'erreur commise par le modèle

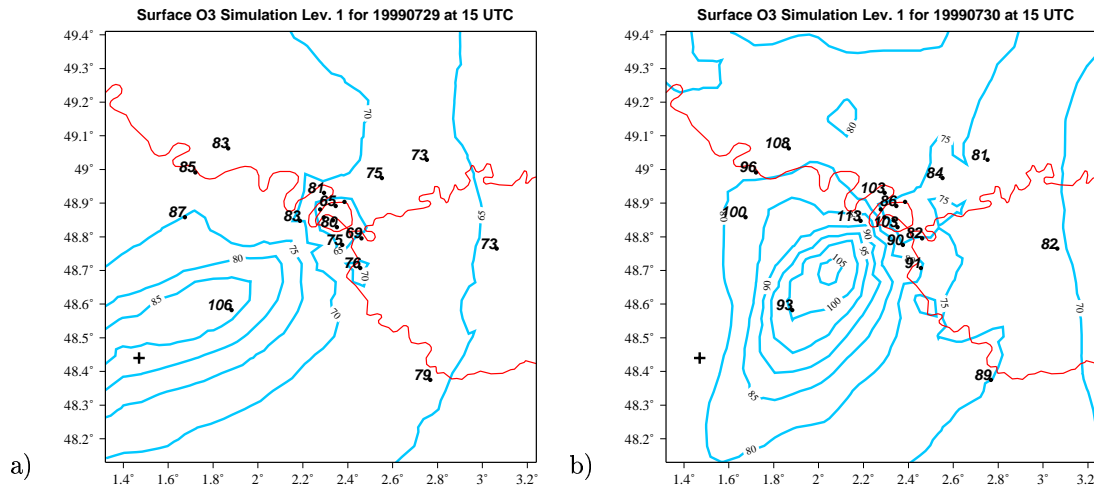


FIG. 5.7 – Simulations (contours gris) des concentrations d’ozone par le modèle CHIMERE régional centré sur l’Île-de-France dans sa couche de surface pour a) le 29 Juillet 1999, et b) le 30 Juillet 1999, à 15 h UTC. Les nombres noirs indiquent la valeur de quelques observations de surface. Les points noirs représentent l’ensemble des stations utilisées dans les analyses. Paris est identifié par une ligne gris foncé, située au centre de la carte. Les cours de la Seine et de la Marne sont également représentés. L’unité des concentrations est en ppb.

n’est pas corrigée à temps, elle se propage et perturbe la prévision du lendemain.

De manière générale, il a été noté que les modèles de chimie-transport européens ont tendance à sous-estimer les concentrations d’ozone lors des épisodes de fortes chaleurs (dépassant 30°C). Ces erreurs sont, encore aujourd’hui, très mal comprises par l’ensemble de la communauté scientifique. Il est soupçonné que des émissions (anthropiques ou biogéniques), liées à une quelconque évaporation, ne sont pas prises en compte.

L’expérience EUROPOLLUX

En 2000, le système CHIMERE multi-échelles, décrit dans le paragraphe 3.1, est testé en temps réel via le projet EUROPOLLUX, soutenu par l’ADEME et Elf Aquitaine. L’expérience est menée à l’IPSL en collaboration avec Météo-France, AIRPARIF et l’ASPA (AASQA en Alsace). Elle est destinée à étudier la prévisibilité de la pollution photo-oxydante à différentes échelles spatiales. Le modèle CHIMERE continental force des modèles régionaux de plus grande résolution, centrés sur l’Île-de-France et sur l’Alsace. Le modèle est, cette fois-ci, forcé soit par les prévisions du CEPMMT (et pas par les *first guesses*), soit par celles du modèle ARPEGE (Météo-France). Un serveur internet propose : les prévisions de CHIMERE à trois jours des concentrations d’ozone et de dioxyde d’azote à l’échelle européenne, avec la possibilité de zoomer sur les deux régions précédemment citées, et les comparaisons entre valeurs prévues et observées. (Quelques résultats sont accessibles actuellement sur le site internet <http://euler.lmd.polytechnique.fr/europollux/>.)

Pour la première fois, une Interpolation Statistique est testée pour produire des analyses de

champs de concentration d'ozone sur la région d'Île-de-France et à l'échelle continentale. Les analyses de la veille sont mises à disposition chaque jour en fin de matinée. Nous ne disposions à l'époque que d'une cinquantaine d'observations par jour. Ces observations nous étaient gratuitement mises à disposition par les associations pour la surveillance de la qualité de l'air suivantes : AIRMARAIX (PROVENCE), AIRPARIF (Île-de-France), ATMOAUVERGNE (Auvergne), AIR NORMAND (Haute Normandie), AIR PAYS DE LA LOIRE (Pays de La Loire), ASQAP (Picardie), AIR LANGUEDOC-ROUSSILLON (Languedoc Roussillon), ASPA (Alsace-Vosges), NABEL/SAEFL (Suisse), LML (Pays Bas), DETR (Royaume Uni).

Ces expériences en temps réel nous ont permis, jour après jour, d'observer la pertinence des approches et des modèles utilisés et de les améliorer en fonction des résultats. Elles ont été déterminantes pour définir les axes de développements du système CHIMERE.

Le projet PIONEER (Prévisibilité et Incertitudes de l'Ozone à l'Echelle Européenne et Régionale)

Le projet PIONEER, soutenu par l'ADEME, le PNCA (Programme National de recherche sur la Chimie Atmosphérique), et TOTALFINAELF, est destiné à évaluer l'incertitude des prévisions d'ozone à différentes échelles spatiales. Ainsi, depuis 2001, les modèles CHIMERE et MOCAGE [Peuch *et al.*, 1999] sont utilisés simultanément afin de tester les prévisions d'ozone, évaluer l'incertitude sur ces prévisions, valider indépendamment les modèles, et initier une comparaison entre les modèles pour analyser et éventuellement expliquer les différences. Ce projet inclut aussi la mise en place d'une interface modèles/observations comprenant l'acquisition en temps quasi-réel de données d'ozone sur l'Europe et leur intégration dans un processus d'assimilation de données produisant des champs initiaux et des analyses d'ozone. Depuis l'été 2001, une plate-forme multi-modèle de prévision de l'ozone sur l'Europe en temps réel est développée à l'IPSL, en collaboration avec Météo-France et le Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA). Un serveur internet, dont l'adresse est <http://euler.lmd.polytechnique.fr/pioneer/>, est opérationnel tous les jours. Il fournit des prévisions de maxima et des valeurs horaires sur l'Europe de l'Ouest et sur trois régions françaises (Ile-de-France, l'Alsace, et la région de Fos Berre-Marseille) à l'aide des modèles CHIMERE et MOCAGE ; des prévisions d'incertitude sur ces mêmes prévisions, avec notamment des probabilités de dépassement du seuil d'information et de recommandation concernant l'ozone (90 ppb) ; une cartographie à l'échelle européenne de l'ozone effectuée chaque jour pour la veille à 15h UTC à partir des sorties de CHIMERE continental et des données que nous recevons de plusieurs associations françaises et étrangères (UK, Allemagne, Suisse, Pays-Bas, Italie) ; une cartographie horaire sur l'Île-de-France en collaboration avec AIRPARIF ; des graphes de vérification des prévisions sur l'ensemble des stations.

Ce travail de thèse a principalement porté sur la mise en place de l'interface modèles/données et de l'algorithme d'assimilation de données. Durant l'été 2001, une expérience d'assimilation en continu a été menée. Cette expérience a consisté à initialiser le modèle continental de prévision chaque jour à partir des analyses de la veille. L'objectif était de quantifier l'apport d'une assimilation des observations en continu à la qualité des prévisions. Nous présentons dans la suite quelques résultats concernant ces derniers points.

5.3.2 Cartographie en quasi temps réel

Depuis l'été 2001, nous recevons chaque jour les données horaires de la veille concernant les observations des concentrations d'ozone, de monoxyde d'azote et de dioxyde d'azote de plusieurs réseaux français et étrangers (UK, Allemagne, Suisse, Pays-Bas, Italie) par mail, par *ftp*, ou via internet. Les données d'ozone, reçues avant 10 LT (heure locale), sont utilisées pour produire à l'aide des sorties de CHIMERE, des analyses des concentrations d'ozone de la veille à 15 h UTC, à l'échelle régionale, et continentale. Elles sont généralement disponibles en fin de matinée, vers 11 h LT. Quelques cas intéressants sont maintenant décrits, les analyses météorologiques associées aux cas présentés sont données en Annexe D. Ils correspondent aux épisodes pour lesquels le modèle CHIMERE prévoit des fortes concentrations d'ozone. Nous comparons les prévisions d'ozone et les analyses aux situations effectivement observées, les analyses étant produites à l'aide de la méthode ASI.

Cas du 30 Juillet au 2 août 2001

L'épisode de pollution photochimique de fin juillet 2001 est un épisode long qui s'étend du 25 Juillet au 2 août sur la France, l'ozone étant ensuite évacué à l'Ouest vers l'Allemagne. Les figures 5.8, 5.9 et 5.10, présentent respectivement les simulations de CHIMERE continental des concentrations d'ozone en surface du 30 Juillet au 2 Août 2001, les valeurs observées et analysées correspondantes à 15 h UTC.

Le 30 Juillet 2001, des conditions anticycloniques sont établies sur l'ensemble du territoire français avec la présence d'une dorsale d'altitude et le prolongement de l'anticyclone des Açores en surface [*Météo.Hebdo*, 2001]. Des températures maximales très élevées sont observées sur la quasi-totalité du pays (de l'ordre de 30-35°C). Un bas géopotentiel au sud du Portugal engendre des remontées de cirrus parfois denses et d'altocumulus sur l'Aquitaine. Les structures de concentration d'ozone simulées par le modèle CHIMERE sont en très bon accord avec celles observées, avec des concentrations élevées partout en France, mais moins fortes sur la région Aquitaine jusqu'au centre du pays à cause de la présence de nuages. Des concentrations beaucoup plus faibles sont simulées et observées au Nord de la France à cause de la présence d'un front froid qui longe la côte Nord. Toutefois, une surestimation significative de la simulation est à noter vers l'estuaire de la Loire et au Nord-Ouest de l'Italie, ainsi qu'une sous-estimation dans la région de Cologne. En dehors de ces régions où les erreurs sont comprises entre 5 ppb et 15 ppb, les concentrations sont, sur la majeure partie des stations, correctes à moins de 5 ppb près. Les analyses respectent les forts gradients au Nord du pays et montrent des concentrations plus faibles sur la partie Centre et Sud-Ouest du pays. Les sous-estimations sont bien corrigées dans la région de Cologne. En revanche, les plus fortes concentrations, dépassant 90 ppb ne sont pas représentées par les analyses, ce qui n'est pas étonnant vu leur caractère très localisé (comme par exemple sur 2 stations au Nord-Ouest de Paris).

La situation météorologique est similaire le 31 Juillet 2001, cependant des nuages bas advectés par le courant maritime alimentent en journée le littoral, principalement sur le nord du Finistère et le Pas-de-Calais. Une forte canicule affecte les régions méridionales, et il n'est pas rare de voir la température dépasser les 37°C. Les concentrations d'ozone et leurs forts gradients sont très bien représentées par le modèle. Une faible correction, inférieure à 5 ppb, est effectuée sur la majeure partie du domaine européen. Seule une sous-estimation des concentrations est corrigée de plus de 5 ppb au centre de l'Allemagne. Les plus fortes concentrations sont mieux reproduites, excepté sur

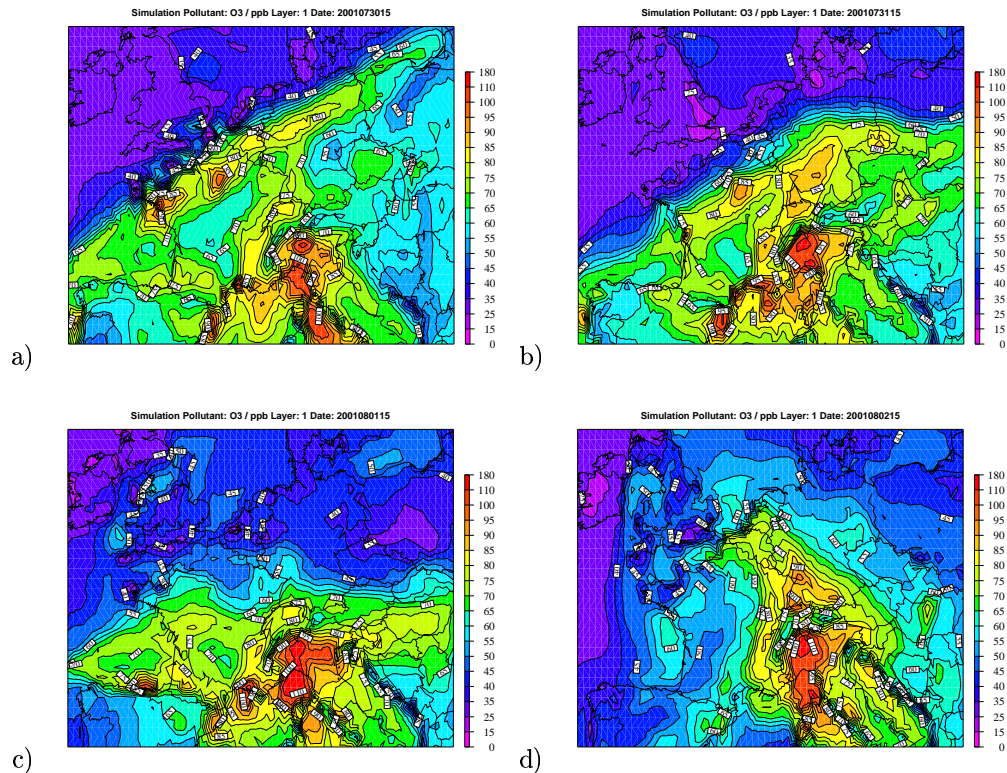


FIG. 5.8 – Simulations de CHIMERE continental des concentrations d'ozone en surface du 30 Juillet au 2 Août 2001 à 15 h UTC. a) 30/07/2001 b) 31/07/2001 c) 01/08/2001 d) 02/08/2001. L'unité des concentrations est en ppb.

la station SAINTS près de Paris car le pic mesuré par cette station est très localisé et non détecté par d'autres stations.

Le 1^{er} août, les conditions restent anticycloniques en surface, cependant la goutte froide présente sur le Portugal remonte vers le golfe de Gascogne apportant une dégradation orageuse dans la nuit sur les régions proches de l'Atlantique. Les nuages bas se dissipent sous l'effet conjugué de l'évolution diurne et du flux orienté à l'est. L'impact de ces nuages bas, mal prévus par le modèle météorologique, est aussi mal représenté par le modèle CHIMERE. Dans ce cas, la photochimie n'est pas assez freinée sur l'Ouest de la France. Le modèle CHIMERE surestime alors de plus de 5 ppb les concentrations d'ozone sur cette partie, ce que l'Interpolation Statistique permet de corriger efficacement.

Le 2 août, un talweg atlantique d'altitude pénètre la façade ouest de la France. Il s'accompagne en surface d'un minimum dépressionnaire qui circule sur le quart nord-ouest. La frontogenèse engendrée provoque de nombreux foyers orageux qui épargnent toutefois les régions méridionales. Le modèle CHIMERE accentue trop les gradients de concentrations du Centre de la France vers l'est, et surestime fortement (de plus de 25 ppb) les concentrations d'ozone du Nord de la France

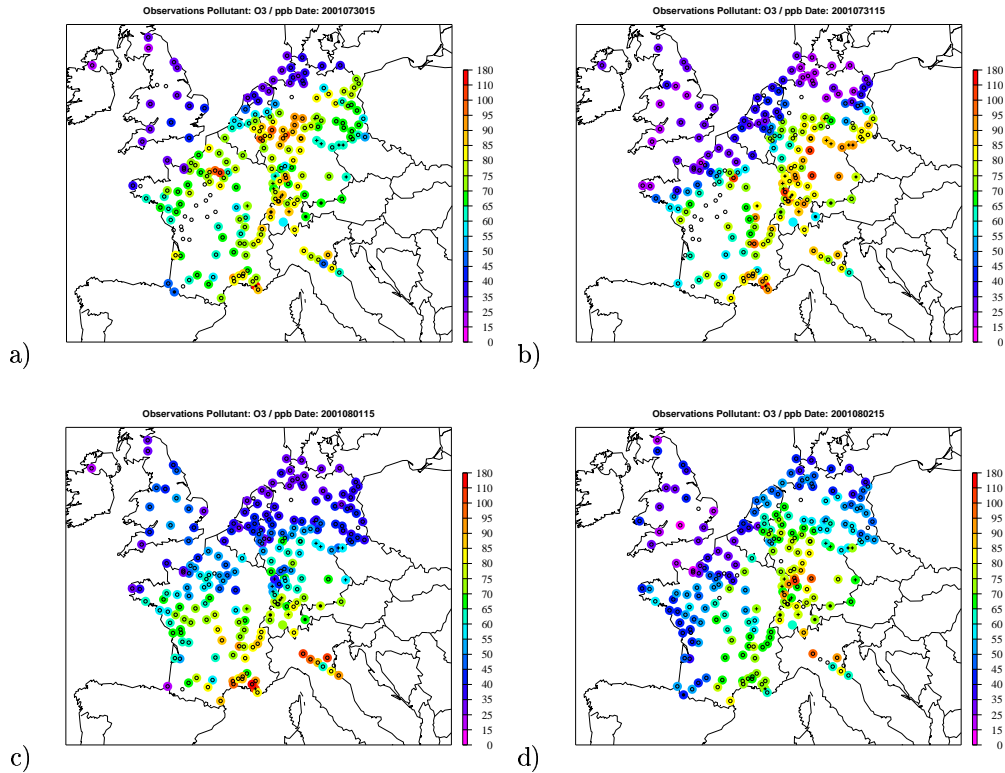


FIG. 5.9 – Observations des concentrations d’ozone en surface du 30 Juillet au 2 Août 2001 à 15 h UTC. a) 30/07/2001 b) 31/07/2001 c) 01/08/2001 d) 02/08/2001. L’unité des concentrations est en ppb.

jusqu’aux Pays Bas. Les analyses améliorent visiblement la représentation des concentrations du polluant. Durant les jours qui suivent, la ”bulle” de pollution est transportée vers l’est, les concentrations sont atténuées (de l’ordre de 70 ppb), laissant la France sous de faibles concentrations (25 ppb), dues à des précipitations ou à la présence de nuages sur la majeure partie du pays.

Cas du 28 Juillet au 2 août 2002

L’épisode de fin Juillet, début août 2002 est un épisode très intéressant, car contrairement au cas présenté ci-dessus, les simulations et les analyses montrent de fortes différences, qui se répètent de jour en jour. Les figures 5.11, 5.12 et 5.13, présentent respectivement les simulations de CHIMERE continental des concentrations d’ozone en surface du 28 Juillet au 2 Août 2002, les valeurs observées et analysées correspondantes.

Le 28 Juillet, la France se situe dans une situation de col en altitude, entre des flux cycloniques orientés au sud-ouest sur la Bretagne et au nord-est sur le Sud-Est du pays [Météo.Hebdo, 2002]. La circulation de talwegs engendre une instabilité orageuse sur les Cévennes et les Alpes. Le déclenchement d’orages est favorisé par un marais barométrique en surface. Le courant océanique

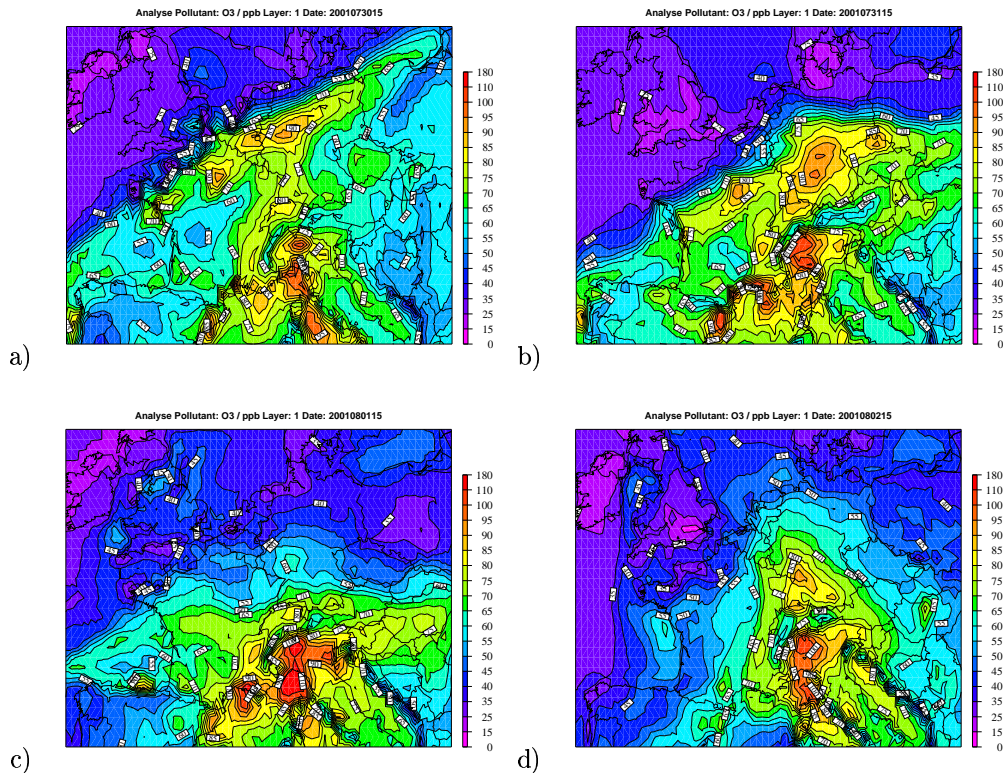


FIG. 5.10 – Analyses des concentrations d'ozone en surface du 30 Juillet au 2 Août 2001 à 15 h UTC, obtenues (par la méthode ASI) à partir des simulations de CHIMERE continental et des observations de la figure 5.9. a) 30/07/2001 b) 31/07/2001 c) 01/08/2001 d) 02/08/2001. L'unité des concentrations est en ppb.

advecte beaucoup d'humidité des basses couches du Pays Basque au Tarn sous forme de stratus. Surmontés de nuages d'altitude, des nuages bas s'étendent également le matin du Finistère au Calvados. Quelques altocumulus instables circulent sur la façade ouest du pays ainsi qu'à l'est du Rhône. Le ciel est ainsi très nuageux sur le Comté niçois. Hormis sur le Finistère où les entrées maritimes perdurent toute la journée, les nuages disparaissent au fil de la matinée alors qu'un temps très chaud et ensoleillé se généralise. L'après-midi, la canicule envahit le pays, à l'exception du littoral de la Manche et de l'Atlantique. Pendant que des entrées maritimes affectent les côtes du Languedoc-Roussillon, des orages se développent sur les Alpes du Sud puis sur les Cévennes, apportant localement de fortes précipitations. En Allemagne et en Suisse, des cumulus sont observés dans la journée. Ces situations nuageuses et orageuses affectent le cycle diurne de l'ozone, en réduisant les taux de photolyse. A 15 h UTC, de fortes surestimations, de plus de 15 ppb sont constatées sur la Normandie, l'Alsace et à l'ouest de l'Allemagne. En effet, le modèle simule des concentrations comprises entre 65 et 75 ppb sur l'ouest de l'Allemagne, alors que des concentrations de l'ordre de 55 à 60 ppb sont observées. Globalement, le modèle CHIMERE surestime les concentrations entre 5 et 15 ppb sur l'ensemble de l'Europe. Les simulations et les analyses montrent de grandes différences.

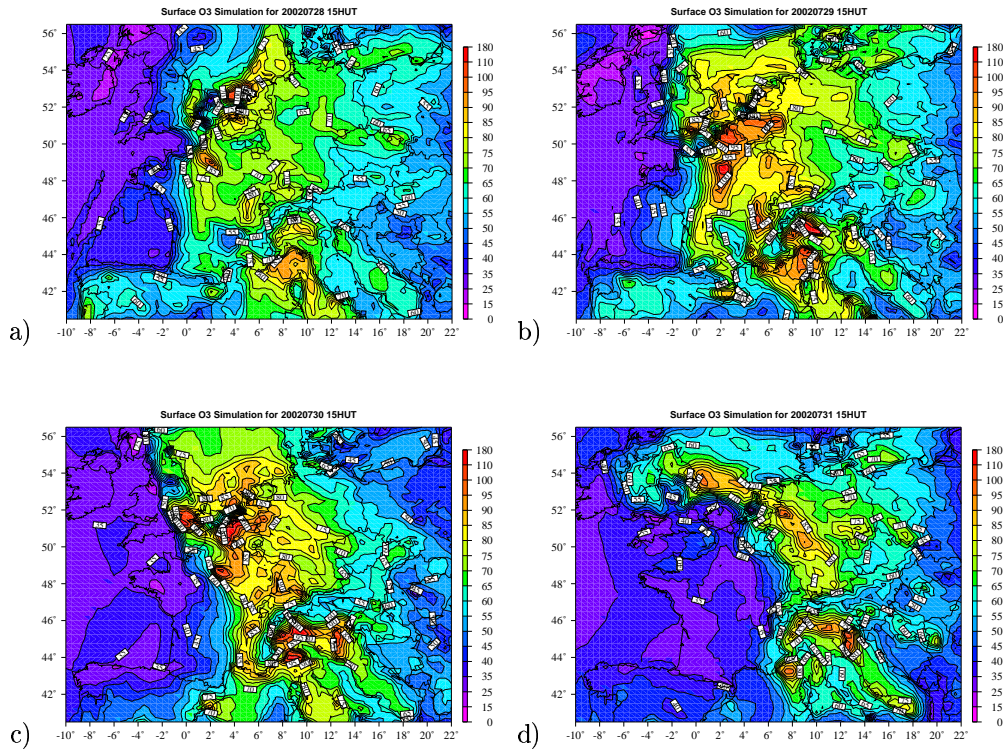


FIG. 5.11 – Simulations de CHIMERE continental des concentrations d’ozone en surface du 28 au 31 Juillet 2002 à 15 h UTC. a) 28/07/2002 b) 29/07/2002 c) 30/07/2002 d) 31/07/2002. L’unité des concentrations est en ppb.

Le 29 Juillet, la France reste sous l’influence du col d’altitude, associé à l’arrivée d’une pseudo-limite froide en surface (i.e. un front froid peu marqué) sur la côte atlantique. Le passage de talwegs déclenche encore une activité orageuse sur la façade atlantique, le Massif central et le Sud-Est, tandis que la canicule touche presque la globalité du pays. A l’ouest de l’Allemagne, et sur l’Alsace des cumulus sont observés. Les concentrations d’ozone sont alors encore surestimées sur ces régions (CHIMERE simule environ 85 ppb au lieu de 70 ppb). L’Interpolation Statistique permet de corriger considérablement ces erreurs, cependant les concentrations sont un peu trop diminuées à l’est de la Belgique, et dans le Nord de la France.

Le 30 Juillet, un talweg d’altitude aborde la Bretagne l’après-midi, à l’avant d’une goutte froide centrée à l’ouest de l’Irlande. Affectant d’abord l’Ouest du pays, de forts orages s’étendent aux régions de l’Est (de la baie de la Somme au Nivernais, des foyers se multiplient des frontières du Nord-Est à la Provence et au Midi-Pyrénées). Une forte couverture nuageuse se maintient sur la moitié ouest de la France.

Le 31 Juillet, la pseudo-limite froide continue de se décaler vers l’est en cours de journée. Le ciel est très nuageux sur l’ensemble du pays et les averses sont nombreuses. Des averses orageuses

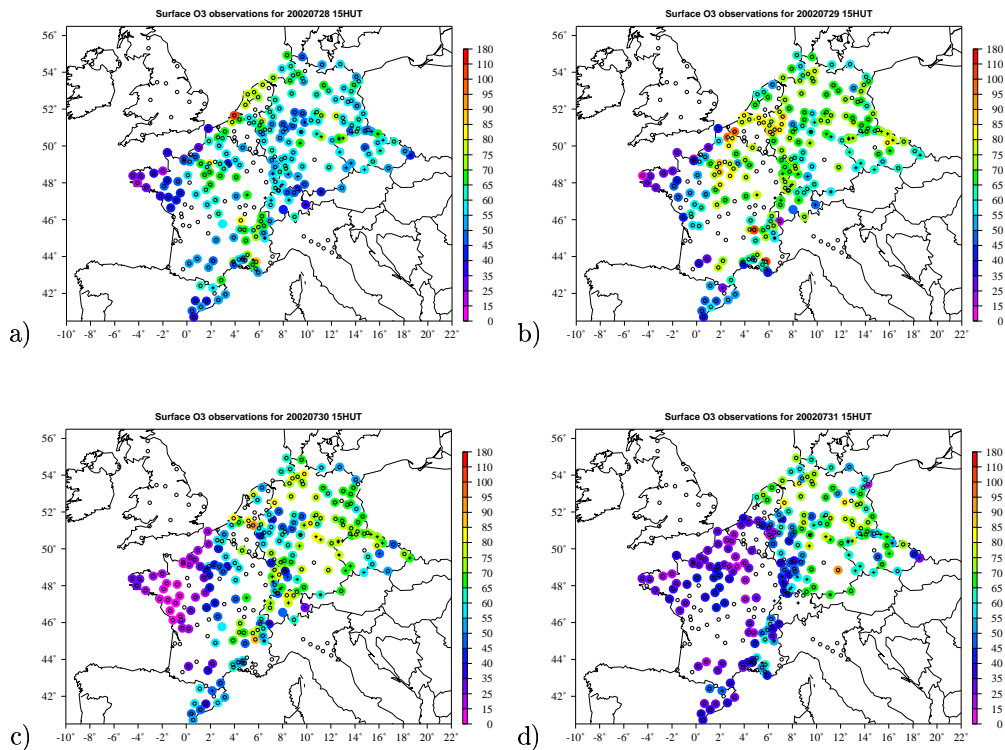


FIG. 5.12 – Observations des concentrations d'ozone en surface du 28 au 31 Juillet 2002 à 15 h UTC. a) 28/07/2002 b) 29/07/2002 c) 30/07/2002 d) 31/07/2002. L'unité des concentrations est en ppb.

se développent notamment en milieu de journée des Monts du Lyonnais jusqu'à la Côte d'Azur. La goutte froide se stabilise sur le Nord-Ouest de la France pendant les deux jours qui suivent et la situation météorologique évolue très lentement pendant les jours qui suivent, le 1^{er} et le 2 août. Tandis que le modèle n'a aucune difficulté à simuler les très faibles concentrations à l'arrière du front, les erreurs se repètent à l'avant de ce front.

Ce genre d'erreur a été largement mis en évidence durant les étés 2000, 2001 et 2002. En effet, il a été noté que le modèle est très sensible à la nébulosité. En général, de fortes erreurs ont été constatées durant les épisodes orageux à l'avant des fronts car la nébulosité est chaotique et difficile à prévoir par les modèles météorologiques. De plus, l'impact de cette nébulosité sur la photolyse est incertain. La représentation de ce processus essentiel pour la photochimie est une problématique qui reste à explorer.

La figure 5.14 montre les cartes de différences entre les simulations et les analyses à 15 h UTC du 28 Juillet au 2 août 2002, résumant ainsi les erreurs commises par le modèle sur cet épisode. Nous remarquons un déplacement progressif de ces erreurs vers le nord-est de l'Europe, suivant l'évolution du front froid. Ces cartes nous ont amené à réfléchir sur ces questions : est-ce que les

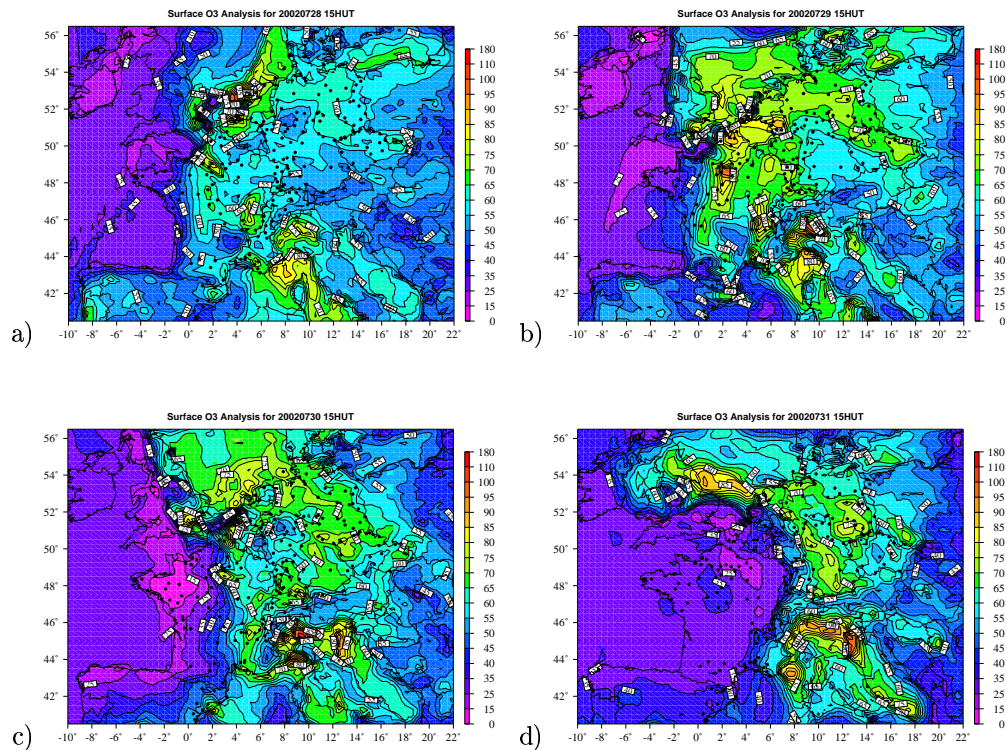


FIG. 5.13 – Analyses des concentrations d’ozone en surface du 28 au 31 Juillet 2002 à 15 h UTC. Ces analyses résultent de la combinaison des simulations de la figure 5.11 et des observations de la figure 5.12. a) 28/07/2002 b) 29/07/2002 c) 30/07/2002 d) 31/07/2002. L’unité des concentrations est en ppb.

erreurs commises le 28 Juillet s’accumulent et s’ajoutent aux erreurs du 30 Juillet et se propagent vers l’est? Ou est-ce que ces erreurs sont décorréélées des erreurs de la veille? Pour répondre ces questions, nous avons étudié l’impact d’une initialisation avec des analyses sur les prévisions à court terme. C’est précisément le sujet du paragraphe suivant.

5.3.3 Impact des analyses sur la prévision

Les erreurs commises par le modèle sur la prévision quotidienne des concentrations d’ozone sont multiples. Elles sont la combinaison d’une erreur sur les valeurs météorologiques analysées et prévues; d’une erreur sur les émissions de polluants primaires; d’une erreur sur les paramétrisations du modèle, et notamment sur l’impact de la nébulosité sur la photolyse; d’une erreur due aux résolutions horizontales et verticale du modèle; et enfin d’une erreur sur les concentrations initiales de la prévision. Les erreurs météorologiques sont dans le cas du modèle CHIMERE des erreurs extérieures qui ne peuvent pas être corrigées. Quant à l’erreur commise sur les émissions est liée à l’incertitude que nous avons sur leur intensité, les processus d’émissions (notamment quand il fait très chaud), et sur leur distribution.

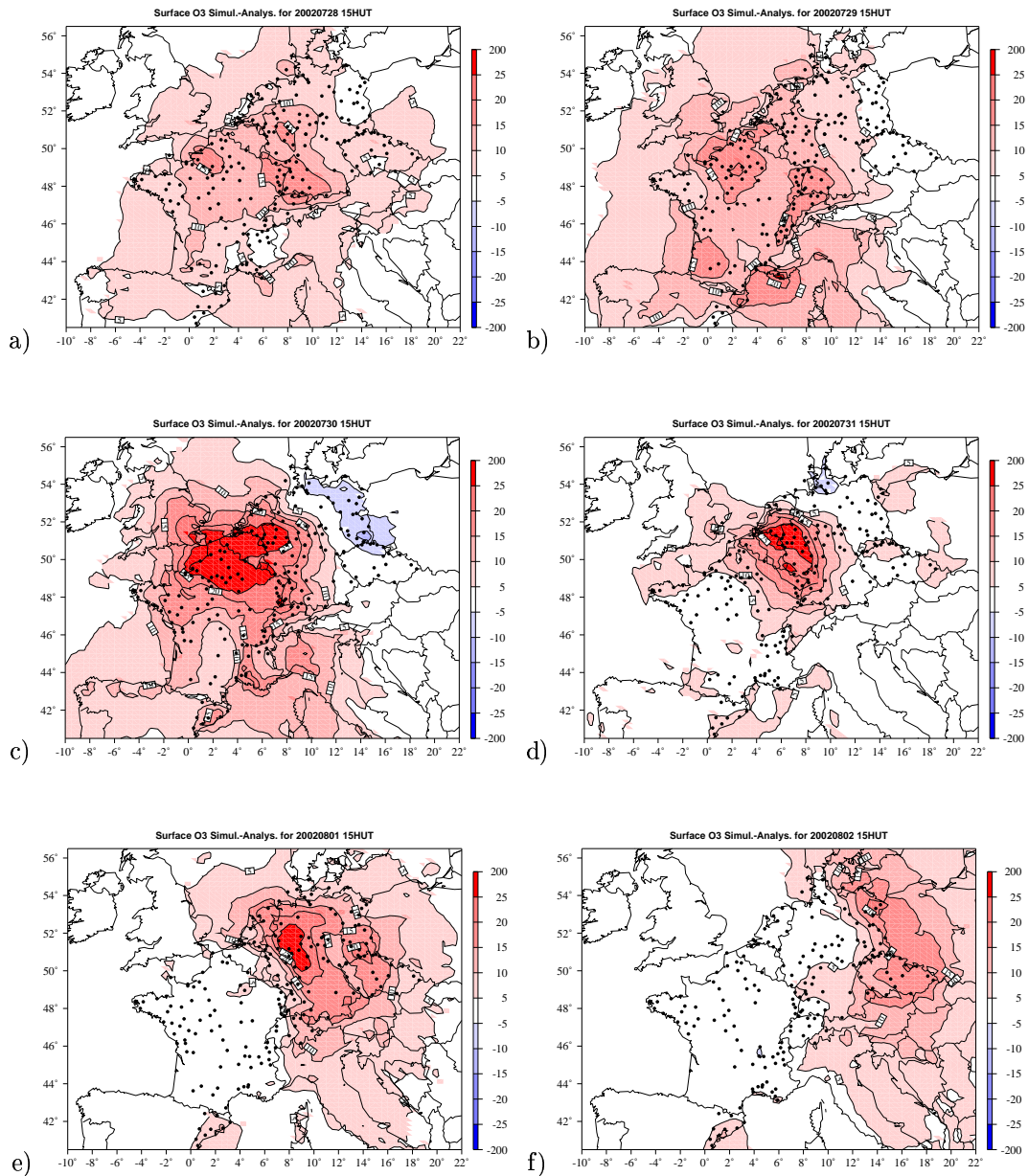


FIG. 5.14 – Cartes de différences entre les simulations et les analyses des concentrations d'ozone en surface du 28 Juillet au 2 Août 2002 à 15 h UTC. L'unité de ces différences est en ppb.

Les erreurs de prévision, ainsi citées, peuvent éventuellement se propager d'un jour à un autre, d'une région vers une autre région. Nous avons cherché à quantifier l'apport de l'assimilation des

observations sur les prévisions à court terme.

Sensibilité du système photochimique aux conditions initiales

Depuis très peu de temps, les "prévisionnistes du temps chimique" s'attachent à définir correctement l'état initial d'un système atmosphérique photochimique. Ce manque d'intérêt, manifesté jusqu'à aujourd'hui pour les conditions initiales, est lié au fait que la chimie atmosphérique ne semble pas présenter de sensibilité similaire à la sensibilité de l'écoulement d'air à ces conditions initiales. Le système photochimique est bien plus sensible aux paramètres météorologiques, aux émissions et au dépôt des polluants. Ceci constitue une différence majeure par rapport au cas météorologique, qui facilite plutôt le problème.

Une autre différence par rapport au cas météorologique est qu'il faut prendre en compte explicitement la cinétique des réactions chimiques. Nous n'avons pas du tout abordé ce sujet, mais ce problème semble maintenant bien maîtrisé. Des outils d'assimilation de données (adjoints de modèles de chimie atmosphérique), permettant de prendre en compte cette cinétique ont été développés, et utilisés avec succès [Schmidt, 1999; Elbern et al., 1997, 2000; Elbern et Schmidt, 1999, 2001].

La figure 5.15 présente les résultats de prévision à 42 h pour deux jours, le 5 et le 7 août 1997, obtenus par Elbern et Schmidt [2001]. Leur travaux montrent que (1) le système chimique a une mémoire d'environ deux jours (2) l'assimilation des données d'observation, par un 4D-VAR, sur la fenêtre d'assimilation 6-12 h UTC, permet d'améliorer les prévisions entre 6 à 12 heures après la fin de la fenêtre d'assimilation. Aux échéances plus longues, les analyses ont encore un impact mais celui-ci est moins significatif.

Deux expériences similaires ont été menées avec le modèle CHIMERE en mode de prévision mais sur une période plus longue, i.e. un été. La première expérience a consisté à ré-initialiser le modèle continental tous les jours avec des analyses produites à l'aide de la méthode ASI, construites à partir des simulations brutes et des observations disponibles en surface. La deuxième expérience consiste à ré-initialiser le modèle continental à partir d'analyses construites à l'aide des prévisions de la veille, issues elles-mêmes des analyses de l'avant veille. Il s'agit alors de tester l'apport d'une assimilation en continu des observations.

Initialisation du modèle CHIMERE avec des analyses de concentrations d'ozone, cas de l'été 1999

L'heure d'initialisation du modèle est choisie à 15 h UTC parce que cette heure correspond à peu près au pic journalier des concentrations d'ozone et à la hauteur maximale de la couche limite. Ainsi, nous espérons, à partir des analyses tridimensionnelles de 15 h UTC, assurer une correction des concentrations d'ozone dans la couche résiduelle, isolée de la surface la nuit et donc des sources d'émissions surfaciques.

La première expérience a consisté à simuler une situation de prévision. Ainsi, les premières ébauches (i.e. les *first guesses*) météorologiques du CEPMMT, correspondant aux prévisions à six heures, sont dans un premier temps utilisées pour prévoir la pollution à trois jours pendant l'été 1999. Chaque jour J, le modèle est initialisé à partir des analyses effectuées à 15 h UTC la veille (J-1). A partir de ces analyses, le modèle est intégré pour produire les pseudo-prévisions (sans

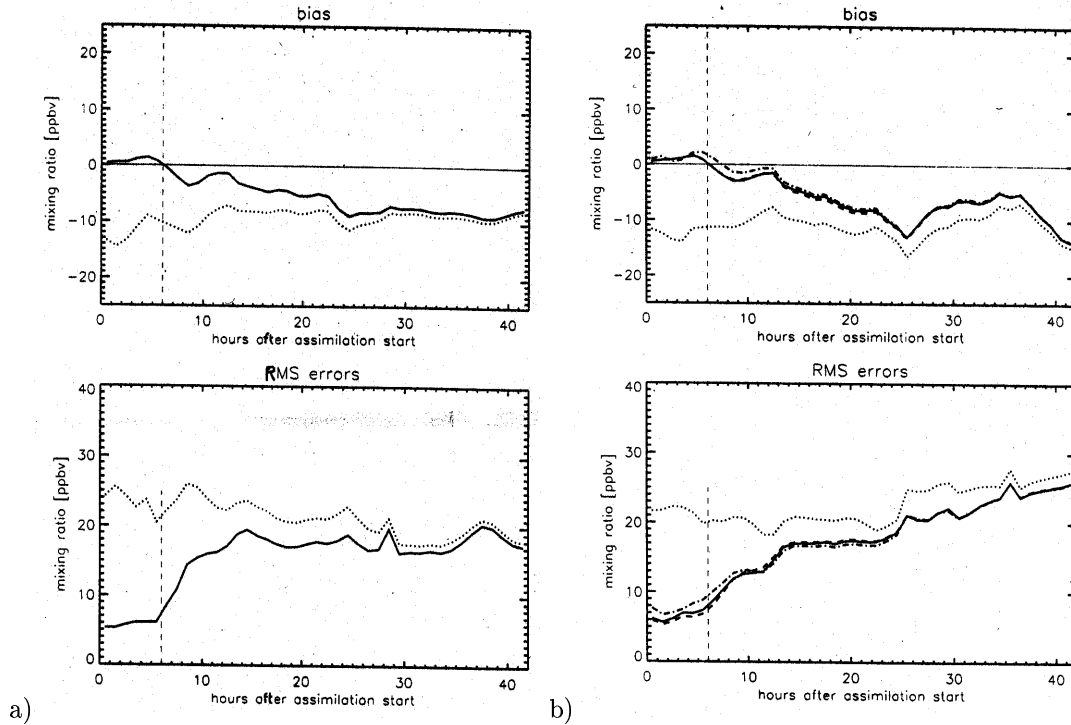


FIG. 5.15 – Biais et RMS moyens (en ppb) calculés sur 400 stations (principalement localisées en Allemagne, et au Royaume Uni) en fonction de l'heure de la journée (en UTC). a) Biais et RMS des différences observations-simulations avec initialisation ou non des simulations le 5 août 1997 b) même chose que a) concernant le 7 août 1997. Le modèle utilisé est le modèle EURAD. Les observations sont assimilées sur la fenêtre 6-12 h UTC, et les simulations sont initialisées à 6 h UTC. La courbe pointillée verticale marque la fin de la fenêtre d'assimilation à 12 h UTC. La courbe pointillée correspond aux simulations brutes sans initialisation avec une analyse, la courbe pleine, aux simulations initialisées avec un état analysé à 6 h UTC. Source : [Elbern et Schmidt, 2001]. Le 7 août, ils étudient l'impact du rayon d'influence (courbes tirets-pointillées).

prévision météorologique en utilisant seulement les premières ébauches) à trois jours et demi (J-1 à partir de 15 h UTC, J, J+1, J+2). L'utilisation des prévisions météorologiques en temps réel est discutée dans le paragraphe suivant.

Le tableau 5.3 montre les valeurs statistiques résultant d'une comparaison des concentrations d'ozone prévues en surface par la version 200102 de CHIMERE continental (prévisions effectuées pour les échéances J-1, J, J+1) avec les valeurs observées en surface. Les prévisions sont, tous les jours de l'été 1999 (du 1^{er} mai au 30 septembre), soit initialisées à partir des concentrations prévues de la veille à 15 h UTC, soit à partir des analyses tridimensionnelles des champs de concentrations d'ozone. Ce tableau donne aussi les valeurs statistiques de *persistance*, qui sont obtenues par une prévision qui consisterait à prendre les valeurs simulées du jour J pour prévoir les concentrations du lendemain.

Au temps d'initialisation (échéance J-1), les scores des analyses sont meilleurs que ceux des prévisions de référence. Ils sont aussi meilleurs que les scores moyens obtenus en *leave-one-out*

TAB. 5.3 – Statistiques d’erreurs de comparaison entre les concentrations d’ozone prévues et les valeurs observées à 15 h UTC pendant l’été 1999. E15 en pourcentage. E15/S correspond au pourcentage d’erreurs inférieures à 15 ppb quand les concentrations observées ou les valeurs simulées dépassent un certain seuil S. S=75 ppb ou S=90 ppb. Les scores P concernent les prévisions brutes. Les scores P+I concernent les prévisions initialisées avec les analyses. L’unité des concentrations est en ppb.

Echéance	BIAIS		RMS		COR		E15		E15/75		E15/90	
	P	P+I	P	P+I	P	P+I	P	P+I	P	P+I	P	P+I
J-1	1.0	-0.3	9.8	5.9	0.75	0.91	88	97	71	87	50	74
J	1.0	0.9	9.8	9.4	0.75	0.77	88	89	71	73	50	55
Persistence	1.0		13.1		0.56		77		39		21	
J+1	1.0	1.0	9.8	9.8	0.75	0.75	88	88	70	71	50	53

dans le tableau 5.1, ce qui n’est pas étonnant puisque, dans cette expérience, nous utilisons *toutes* les observations disponibles. La persistance est systématiquement beaucoup moins bonne. Il est bien sûr important de remarquer que la RMS des différences prévisions-observations est faiblement réduite, que la corrélation est faiblement améliorée à l’échéance J. Nous devons également noter que les pourcentages d’erreurs raisonnables E15/75, et E15/90 sont augmentés respectivement de 2% et 5%. Cela signifie que l’utilisation des analyses de la veille permet d’améliorer surtout les prévisions des fortes valeurs des concentrations.

La figure 5.16 illustre un cas particulier, en montrant les séries temporelles des concentrations d’ozone observées, prévues par le modèle CHIMERE non initialisé par des analyses, et celles prévues par le modèle CHIMERE initialisé le 29 Juillet 1999 à 15 h UTC par des analyses. Les prévisions concernent le 29 Juillet à partir de 15 h UTC jusqu’au 31 Juillet à 15 h UTC. Nous remarquons

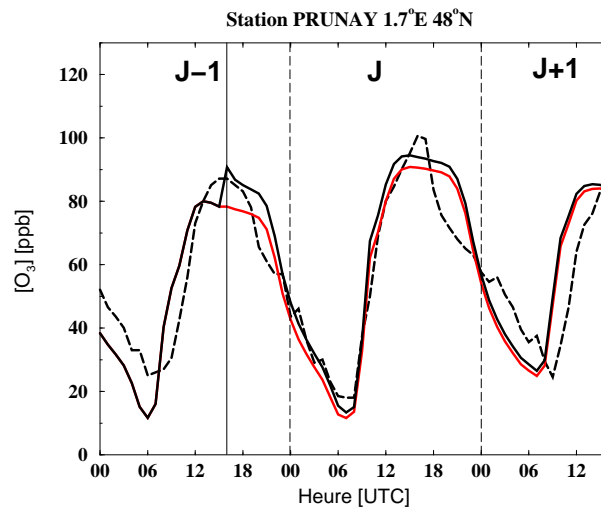


FIG. 5.16 – Séries temporelles des concentrations d’ozone observées (courbe pointillée), prévues par le modèle CHIMERE non initialisé par des analyses (en gris) ou initialisé à 15 h UTC par les analyses du 29 Juillet 1999 (en noir) sur la station PRUNAY, située au nord-ouest de Paris. La ligne verticale pleine marque le temps d’analyse, et les lignes pointillées délimitent les échéances J-1, J, J+1. L’unité des concentrations est en ppb, celle de l’heure en UTC.

qu'une correction d'environ 10 ppb est effectuée le 29 Juillet 1999, et que cette correction à un impact de 3 ppb sur les prévisions du pic d'ozone le lendemain. En revanche, au delà du temps d'analyse, la trajectoire du modèle n'est pas modifiée et les surestimations constatées en soirée, vers 21 h UTC, sont même accentuées.

La même expérience de prévision a été effectuée avec la nouvelle version du modèle (version 200108) pour prévoir les concentrations d'ozone de l'été 2001. Les résultats de la comparaison des

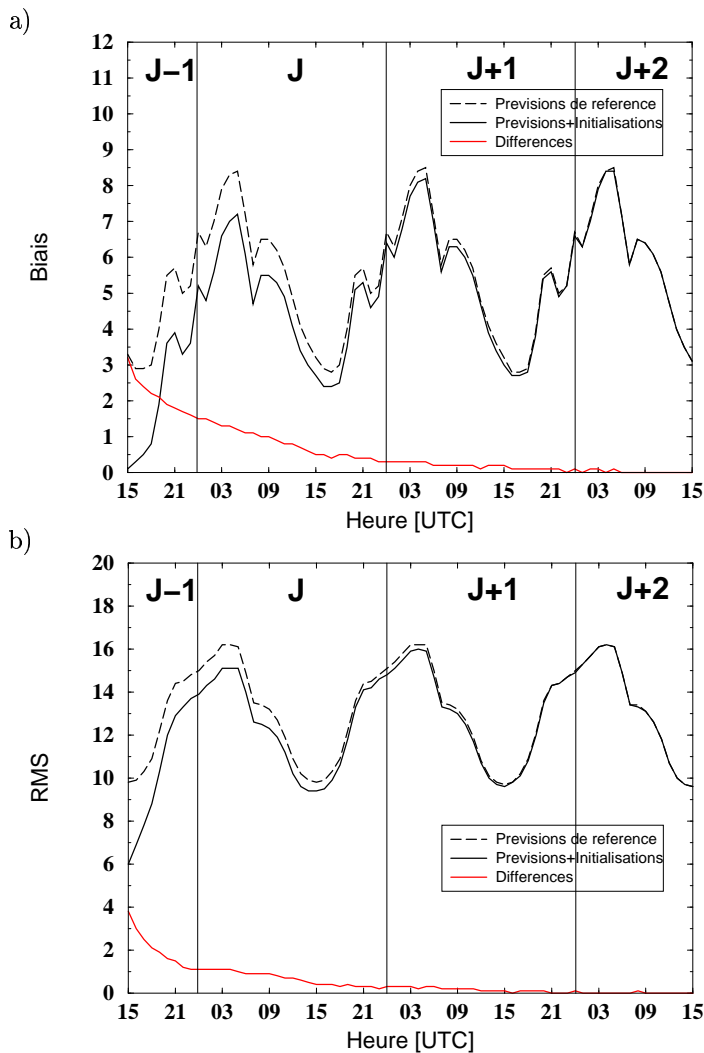


FIG. 5.17 – a) Moyenne des différences entre les concentrations d'ozone observées, et les valeurs prévues par le modèle CHIMERE non initialisé par des analyses (courbe pointillée) ou initialisé chaque jour de l'été 2001 à 15 h UTC par des analyses (courbe noire pleine). b) Même figure que a) pour la moyenne de la RMS des différences. Ces moyennes sont calculées sur l'ensemble des 232 stations. La courbe grise représente l'évolution des écarts entre les deux types de prévisions. Les unités du biais et de la RMS sont en ppb, celle de l'heure est en UTC.

prévisions à trois jours et demi, effectuées à partir ou non des analyses de la veille, sont présentés sur la figure 5.17.

La figure 5.17.a. montre l'évolution de l'écart moyen entre les concentrations d'ozone observées et les valeurs prévues, en fonction de l'échéance de la prévision. Nous notons qu'à partir de l'instant de l'analyse (J-1 à 15 h UTC), le biais augmente lentement durant les trois premières heures, jusqu'à 18 h UTC, l'heure de descente de la couche de mélange. Ensuite, ce biais augmente très rapidement jusqu'à 21 h UTC. Ceci résulte en grande partie de la sous-estimation de la titration de l'ozone par le monoxyde d'azote, dont les concentrations sont sous-estimées vu les résolutions horizontale et verticale du modèle (il y a aussi des erreurs dans le mélange, Cf. paragraphe 3.4). Le biais atteint sa valeur maximale vers 5 h UTC. Les prévisions initialisées par les analyses sont en moyenne moins biaisées que les prévisions de référence de J-1 à J+1, au delà les prévisions sont identiques. De même, la figure 5.17.b. montre que la RMS des différences prévisions-observations est plus faible dans le cas des prévisions initialisées par des analyses. Ces résultats montrent que l'initialisation a un impact positif sur les prévisions. Cependant, cet impact est particulièrement visible pendant les trois premières heures avant la descente de la couche de mélange. Le biais est alors réduit d'environ 3 ppb. Quant à la RMS, elle est réduite de 4 ppb au temps de l'analyse. Les écarts de RMS entre les deux prévisions diminuent ensuite très rapidement de 15 à 18 h UTC. A 18 h UTC, ces écarts sont de 2 ppb. Ainsi, 50% de la correction est "perdue" dans les trois heures car une partie de l'ozone est titré par le monoxyde d'azote dans la couche surfacique. Vingt quatre heures après l'initialisation, les écarts entre les deux types de prévisions sont de 0.5 ppb (biais et RMS), correspondant à une réduction de la RMS de 5% et du biais de 15%.

Expérience en temps réel, assimilation en continu, PIONEER 2001

La deuxième expérience a été menée en temps réel avec la version 200102 du modèle. Les prévisions de CHIMERE, sans initialisation par les analyses, suivent le processus suivant. Premièrement, les prévisions météorologiques sont extraites du CEPMMT à 0 h UTC, et sont ensuite transférées au Laboratoire de Météorologie Dynamique, où les prévisions de la pollution photochimique sont effectuées. Les prévisions météorologiques concernent les échéances J-2, J-1, J, J+1, J+2, J+3, les quatre dernières sont issues des analyses de la veille (J-1) effectuée à 12 h UTC. De J-1 à J (0 h UTC), seuls les *first guesses* sont utilisés. Le modèle CHIMERE est initialisé à la première heure (0 h UTC) de J-1, les concentrations des polluants prévues la veille sur l'ensemble du domaine de simulation servant de conditions initiales aux concentrations du jour. Il produit finalement les prévisions des concentrations des polluants aux échéances J-1, J, J+1, et J+2. Les simulations de J-1, permettent, par comparaison avec les prévisions de la veille, de quantifier la part d'erreur due à la chimie par rapport à l'erreur météorologique.

Les observations collectés à J-1, reçues avant 10 h LT à J sont combinées avec les *first guesses* de J-1 de 15 h UTC en utilisant la méthode ASI. Les analyses tridimensionnelles des concentrations d'ozone, ainsi produites, sont ensuite injectées, sans modifier les concentrations des autres polluants que l'ozone dans le système de prévision. Les nouvelles prévisions à trois jours et demi (J-1 à partir de 15 h UTC, J, J+1, J+2) issues de la ré-initialisation, servent le lendemain, comme first-guess pour construire les analyses. La figure 5.18 montre l'évolution du biais (moyenne des différences entre les concentrations d'ozone observées et les valeurs prévues) en fonction de l'échéance de la prévision. Les résultats sont très peu différents des résultats de la première expérience, montrant que l'introduction des observations n'a qu'un impact de 24 h.

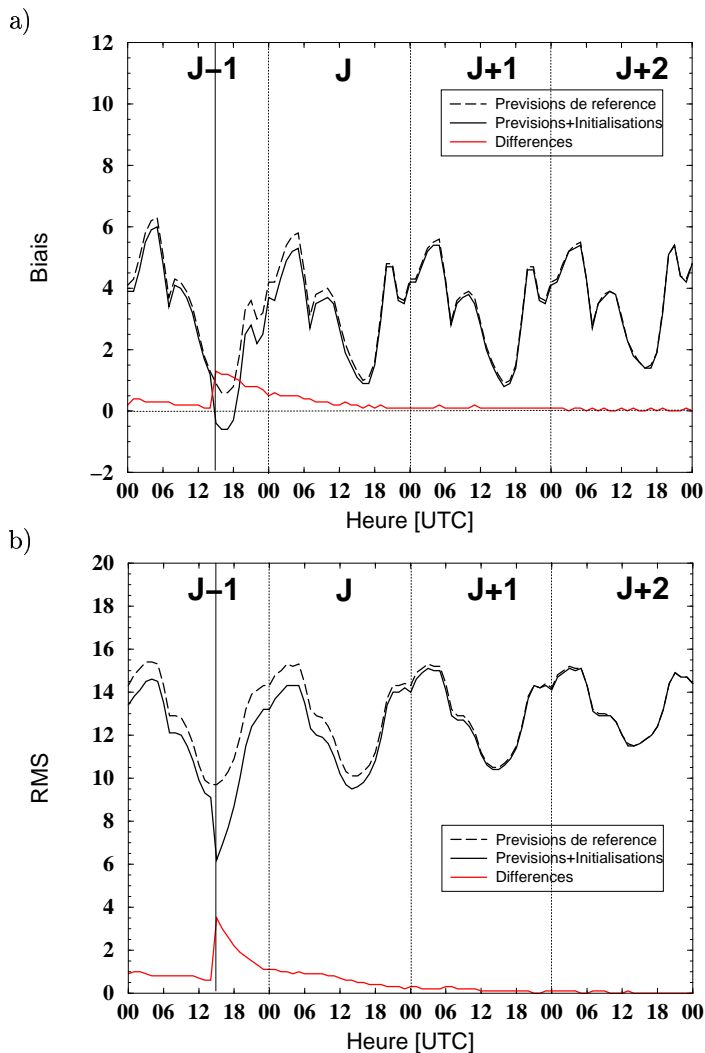


FIG. 5.18 – a) Moyenne des différences entre les concentrations d'ozone observées, et les valeurs prévues par le modèle CHIMERE non initialisé par des analyses (courbe pointillée) ou celles prévues par le modèle CHIMERE initialisé chaque jour de l'été 2001 à 15 h UTC par des analyses (courbe noire pleine). b) Même figure que a) pour la moyenne de la RMS des différences. Ces moyennes sont calculées sur l'ensemble des 232 stations. La courbe grise représente l'évolution des écarts entre les deux types de prévisions. Les unités du biais et de la RMS sont en ppb, celle de l'heure est en UTC.

Le tableau 5.4 détaille les pourcentages d'erreurs raisonnables aux différentes échéances. À part à J-1, les scores sont pratiquement identiques. Seulement, 2% des erreurs supérieures à 15 ppb sont corrigées à J quand les concentrations d'ozone observées ou prévues sont supérieures aux seuils de 50, 65, 75 et 90 ppb.

TAB. 5.4 – Statistiques d’erreurs de comparaison entre les concentrations d’ozone prévues et les valeurs observées à 15 h UTC pendant l’été 2001. L’unité des concentrations est en ppb. E15 en pourcentage. E15/S correspond au pourcentage d’erreurs inférieures à 15 ppb quand les concentrations observées ou les valeurs simulées dépassent un certain seuil S. S prend pour valeur 50, 65, 75, ou 90 ppb. Les scores P concernent les prévisions brutes. Les scores P+I concernent les prévisions initialisées avec les analyses.

Echéance	COR		E15		E15/50		E15/65		E15/75		E15/90	
	P	P+I	P	P+I	P	P+I	P	P+I	P	P+I	P	P+I
J-1	0.82	0.93	89	96	83	94	78	100	71	85	49	67
J	0.81	0.82	88	89	82	84	76	78	70	72	48	50
J+1	0.79	0.79	86	87	80	80	72	73	64	65	44	44
J+2	0.75	0.75	83	83	75	75	69	69	61	61	39	39

Correction simultanée des précurseurs de l’ozone : le dioxyde d’azote, le PAN

Dans les sections précédentes, seuls les champs de concentrations d’ozone, simulés par le modèle CHIMERE, sont corrigés à l’aide des observations du polluant. L’initialisation du modèle est effectuée de façon brusque, en introduisant les concentrations d’ozone analysées sans modifier les concentrations d’autres polluants. L’état de pollution de l’atmosphère n’est alors probablement pas auto-cohérent. Pour essayer d’améliorer la cohérence entre les différents polluants, nous avons testé une correction simultanée de deux des précurseurs de l’ozone : le dioxyde d’azote (NO_2), et le peroxy-acétyl-nitrate (PAN). En effet, nous avons corrigé ces précurseurs en tenant compte de la corrélation entre leurs concentrations et celles de l’ozone. Nous avons supposé que les polluants sont liés linéairement entre eux. Lorsque le coefficient de corrélation de la régression linéaire dépassait 0.6, nous avons modifié les concentrations des précurseurs d’une correction proportionnelle à celle appliquée à l’ozone. Ces études ont été vaines (0.1 ppb de plus est alors gagné sur la RMS des différences prévisions-observations à l’échéance J).

5.4 Conclusions

L’objectif de ce chapitre était de savoir si avec une méthode d’assimilation de données, il est possible d’améliorer les prévisions d’un modèle de chimie-transport. Nous avons montré que l’assimilation de données ne résout en fait qu’une petite partie du problème.

Des analyses de champs de concentration d’ozone ont été produites à partir des sorties du modèle CHIMERE continental et des observations de surface en utilisant la méthode ASI. Une analyse diagnostique des résultats de comparaison des analyses avec des observations indépendantes (utilisation d’un *leave-one-out*) a permis de mettre l’accent sur la qualité de ces analyses. Nous avons notamment montré que les pics de concentration d’ozone (vers 15 h UTC) sont en moyenne reproduits à 7 ppb près, le nombre d’erreurs supérieures à 15 ppb étant de l’ordre de 5%. Les cycles diurnes sont aussi très bien reproduits par les analyses (bien mieux que par les simulations brutes).

Les expériences de cartographie en temps réel ont également montré que le modèle CHIMERE commet de fortes erreurs lors de la présence de nuages, dont l’impact sur la photochimie est très mal représenté. Ceci suggère que l’impact de la nébulosité est un processus, qui devra faire l’objet de recherches plus approfondies dans le futur.

Les analyses des champs de concentration d’ozone ont ensuite été utilisées pour initialiser le

modèle de prévision à 15 h UTC. L'impact des analyses est visible jusqu'à 24 h après l'initialisation, incluant l'heure du pic du lendemain. Cependant, malgré la qualité des analyses utilisées, nous avons montré que cet impact reste assez faible. En effet, les prévisions sont particulièrement améliorées dans les premières heures qui suivent l'initialisation, puis l'impact des analyses diminue très rapidement (0.5 ppb est gagné sur la RMS des différences prévisions-observations, 24 heures après l'initialisation). Plusieurs raisons peuvent expliquer ces résultats. La première raison est que la qualité des prévisions du temps chimiques ne dépend pas uniquement de celle des conditions initiales. En effet, ces prévisions sont aussi sensibles à d'autres paramètres : émissions, aux paramètres météorologiques ... Les précurseurs devraient être corrigés, à l'heure des fortes émissions, c'est à dire dans la matinée. La deuxième raison tient peut-être au manque d'information en altitude. La correction en surface de l'ozone n'est pas "conservée" dans la couche résiduelle, mais "diluée" pendant la nuit dans la couche résiduelle. Une question reste donc ouverte : ce serait-il possible d'améliorer les prévisions à court terme si nous disposions d'observations en altitude, telles que des données satellitales ?

Chapitre 6

Autres applications des méthodes d'analyse de données ASI et INK

Dans les chapitres précédents, nous avons montré que l'Interpolation Statistique Anisotrope (méthode ASI) et le Krigeage des innovations (méthode INK) permettent de corriger très correctement les champs de concentrations d'ozone simulés par le modèle CHIMERE à l'échelle régionale, sur l'Île-de-France, comme à l'échelle européenne. Nous cherchons, dans ce dernier chapitre, à montrer le degré de généralité de ces méthodes en les appliquant à un autre polluant : le dioxyde d'azote. Cette première expérience est tout d'abord effectuée sur la région d'Île-de-France. Comme dans le cas de l'ozone, les sorties du modèle CHIMERE régional sont combinées avec des observations de dioxyde d'azote afin d'obtenir des cartes plus réalistes. La deuxième expérience consiste à tester également la méthode ASI sur la région de Berre et de Marseille pour produire des analyses de champs de concentration d'ozone et de dioxyde d'azote.

Ces deux expériences sont regroupées ici car, comme nous l'avons montré au chapitre 3, le modèle CHIMERE régional présente une certaine difficulté à simuler correctement les concentrations d'ozone sur la région méditerranéenne, ainsi que celles de dioxyde d'azote sur les deux régions. La question est alors de savoir s'il est possible d'améliorer la représentation des champs de concentration des polluants, même lorsque le modèle commet initialement de fortes erreurs.

6.1 Les limites du modèle CHIMERE

6.1.1 Le dioxyde d'azote

Comme le dioxyde d'azote est non seulement un polluant direct, mais aussi un précurseur de l'ozone, sa représentation spatiale et temporelle est une étape importante dans la compréhension des épisodes photochimiques. Ce polluant est le produit de trois voies de formation : l'émission directe, l'oxydation du monoxyde d'azote par l'ozone, et l'oxydation du monoxyde d'azote par les radicaux (formation photochimique).

La première voie de formation est prépondérante sous faible rayonnement, particulièrement lors de conditions de faible dispersion (vent faible, stabilité de l'atmosphère). La seconde dépend

des concentrations d'ozone, et, est primordiale lors des jours assez ventés, quand la photochimie n'a pas le temps de se mettre en place. Enfin, la dernière voie de formation est importante par fort ensoleillement, le matin, et par faible dispersion. Ces conditions sont assez rares en région parisienne, et la production photochimique du dioxyde d'azote reste une voie de formation assez marginale par rapport aux deux premières.

Nous avons montré dans le chapitre 3 que le modèle CHIMERE simule plus difficilement le dioxyde d'azote que l'ozone. Plusieurs raisons peuvent expliquer ceci. Premièrement, le dioxyde d'azote est un polluant dont la concentration est plus directement reliée aux émissions que l'ozone. Or la plus grande incertitude plane sur ces émissions. Dans le cas de la formation de l'ozone, le nombre de processus et de paramètres mis en jeu est tellement important que des erreurs peuvent se compenser. Le modèle peut alors donner de bonnes concentrations d'ozone sans pour autant les donner pour les bonnes raisons. (Lors de l'évaluation d'un modèle, il est important donc de vérifier par exemple que le modèle reproduit correctement les sensibilités des concentrations d'ozone à des paramètres tels que les émissions [*Sillman*, 1999 ; *Hodzic*, 2001]).

Deuxièmement, les pics de dioxyde d'azote apparaissent le matin, lorsque la couche limite n'est pas encore totalement développée. Or plus la couche limite est mince, plus le modèle est sensible à tous les paramètres (émissions, vent, nébulosité, dépôt ...), et ses erreurs sont accentuées. Enfin, la durée de vie du monoxyde et du dioxyde d'azote est de quelques secondes à quelques heures, de sorte que ces polluants sont très localisés autour de leur source. Les problèmes de représentativité sont alors encore plus présents que pour l'ozone.

6.1.2 La région de Berre et de Marseille

La modélisation de la pollution photochimique sur la région de Berre et de Marseille est difficile car les mécanismes dynamiques qui siègent sur cette région sont complexes. Le climat méditerranéen chaud et ensoleillé est propice à l'établissement d'épisodes de pollution (régime anticyclonique avec absence de vent). La transition terre-mer, sur une côte découpée, donne lieu à des cycles complexes de brise. En outre, les circulations de mésoéchelle sont canalisées par les reliefs (Alpes, vallée du Rhône et de la Durance). Il faut aussi savoir que cette région connaît de fortes émissions, notamment sur la zone industrielle de l'étang de Berre. Les types de surface sont variées (zone urbaine de Marseille, zone industrielle, zone montagneuse) et les couverts végétaux sont multiples. Ceux-ci imposent une variabilité spatiale du bilan d'énergie de surface, et ainsi de la couche limite.

C'est la région, qui a été choisie pour effectuer la campagne ESCOMPTE. Cette campagne a été mis en place avec pour objectif majeur de constituer une base de données très détaillée en vue de contribuer au développement et à la validation des modèles de chimie de transport sur un terrain complexe. Les résultats de la comparaison des données aéroportées provenant de cette campagne, trop préliminaires, ne sont pas exposés dans ce document. La comparaison entre les simulations de CHIMERE régional et les observations de surface (ozone et dioxyde d'azote) a montré, dans le chapitre 3, que le modèle simule plus difficilement les concentrations d'ozone et de dioxyde d'azote sur cette région que sur la région d'Île-de-France. En particulier, lors des épisodes de forte pollution, les panaches urbains concentrés en ozone sont alors très souvent mal positionnés. De plus la corrélation moyenne entre les simulations du dioxyde d'azote et les valeurs observées correspondantes, calculée sur l'ensemble des stations mesurant le polluant à 9 h UTC, est faible,

de 0.53.

6.2 Représentation des champs de concentrations du dioxyde d'azote sur l'Île-de-France

Dans le cas du dioxyde d'azote, en plus du fait que le modèle ne simule pas très bien les concentrations du polluant (notamment sur les stations les plus proches des sources c'est à dire sur les stations de trafic), nous sommes confrontés à plusieurs problèmes. Premièrement, les stations de mesure sont principalement situées dans les agglomérations (ici Paris et sa banlieue). De larges zones du domaine de simulation ne sont donc pas observées, comme l'indique la figure 3.2. Deuxièmement, les observations, effectuées autour de l'agglomération, sont très probablement surestimées à cause de la non spécificité de la méthode de mesure (Cf. paragraphe 3.2.2).

6.2.1 Méthodes d'analyse de données

Les méthodes ASI, INK, et OBK, exposées au chapitre 4, sont testées pour produire des analyses de champs de concentration de dioxyde d'azote. Les méthodologies sont quasiment identiques, excepté que dans le cas de la méthode ASI, nous modélisons séparément les covariances et la variance des innovations. En effet, cette variance n'est plus supposée homogène mais est estimée en fonction de la variance des concentrations du dioxyde d'azote. Les modèles statistiques (modèles de biais, de covariance et modèles de variance) sont déterminés sur la base des observations et des simulations du dioxyde d'azote de l'été 1999.

Pour éviter les plus forts biais de représentativité, nous n'utilisons pas les stations de type trafic. Sur ces stations, le biais *global* (différences moyennes entre les simulations et les observations) atteint -10 à -15 ppb à toute heure de la journée. Alors que sur les autres stations, il montre une forte dépendance linéaire avec les concentrations moyennes simulées. Le coefficient de corrélation de la régression linéaire est par exemple de 0.93 à 7 h UTC (nous constatons un biais globalement positif sur la zone urbaine et un biais plus faible sur le reste du domaine). Dans le cas de la méthode ASI, le biais de modélisation est donc estimé de la même manière que dans le paragraphe 4.1.1, c'est-à-dire en fonction de la concentration moyenne du polluant simulée par le modèle. Le biais résiduel est supposé être le biais d'observation.

En ce qui concerne les méthodes de Krigeage, le faible nombre de couples de stations éloignées de plus de 20 km, nous conduit à réduire l'intervalle de distance L à 5 km et la tolérance à 4 km.

6.2.2 Validation objective des méthodes d'analyse

La figure 6.1 présente les résultats de la comparaison des analyses produites pendant l'été 2001 par les trois méthodes d'analyse (ASI, OBK, INK), avec les observations non utilisées (la méthode du *leave-one-out* étant de nouveau appliquée à chacune des stations de mesure). Les figures 6.1.a. et 6.1.c. montrent respectivement ces résultats en terme de RMS et de STD des différences, ces deux quantités étant moyennées sur l'ensemble des stations urbaines et présentées en fonction de l'heure de la journée. Nous reportons également les valeurs moyennes des RMS et STD des différences simulations-observations. Notons que, quelle que soit la méthode d'interpolation utilisée, la RMS

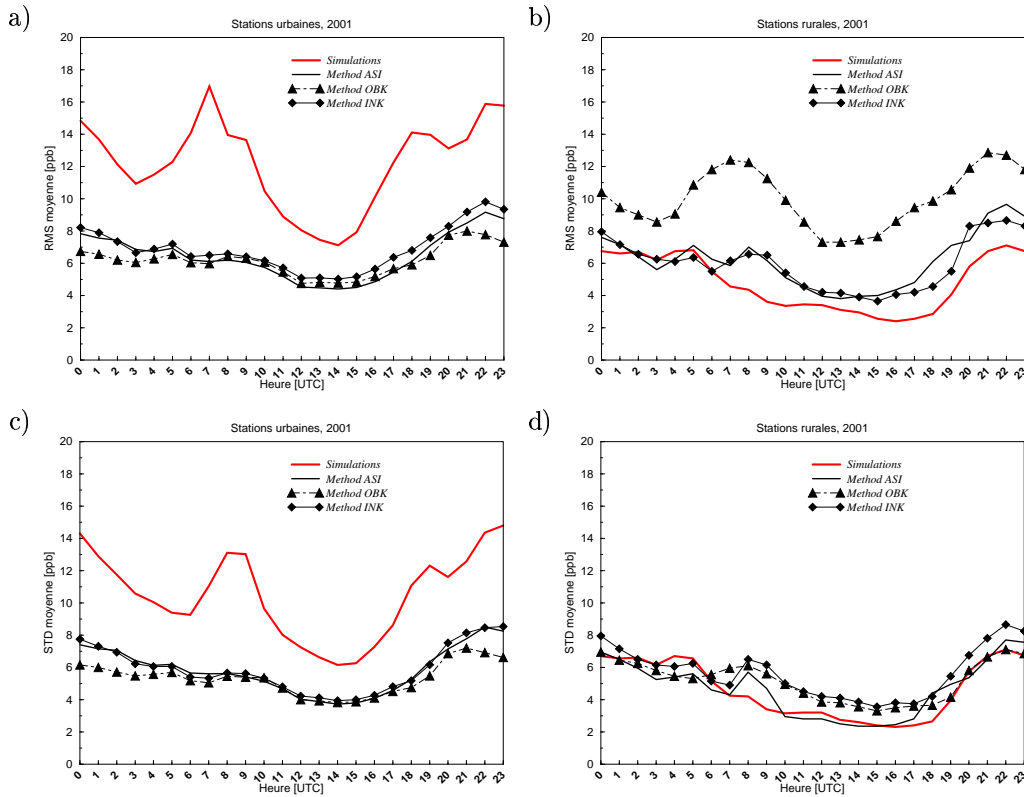


FIG. 6.1 – Comparaison des simulations (courbe grise) ou des analyses avec les observations de surface en fonction de l'heure (unité UTC) en terme de RMS et de STD (en ppb). Les analyses sont produites par les méthodes ASI (courbe noire), OBK (courbe pointillée et triangles), INK (courbe et carrés). a) RMS sur les stations urbaines b) RMS sur les stations rurales c) STD sur les stations urbaines d) STD sur les stations rurales.

et la STD des différences analyses-observations sont en moyenne beaucoup plus faibles que celles calculées sur les différences simulations-observations. Cette réduction est particulièrement remarquable aux heures d'émissions les plus fortes, c'est à dire vers 8-9 h et 18-19 h UTC. La RMS est alors réduite de plus de 50%, et la STD est réduite d'environ 50%. Ceci signifie que, non seulement le biais des différences simulations-observations est réduit, mais aussi que la variabilité des concentrations est mieux reproduite par les analyses. Ces figures montrent également que les méthodes sont sensiblement équivalentes pour représenter les champs de concentration du polluant durant la journée. En revanche, la méthode OBK est plus performante la nuit, indiquant que le modèle a alors tendance à détériorer les analyses.

Les figures 6.1.b. et 6.1.d. présentent les résultats de la même comparaison, moyennés sur les stations rurales. De manière assez surprenante, les simulations brutes (i.e. non modifiées par l'introduction des observations) sont plus proches des observations que les analyses. La méthode OBK apparaît clairement comme la moins performante. En effet, cette méthode suppose que les observations ont une moyenne uniforme spatialement. Or, ce n'est pas correct car les moyennes des concentrations observées sont plus élevées sur la zone urbaine (25-30 ppb) que sur la zone

rurale (10-15 ppb). D'après les scores obtenus par la méthode INK, le modèle permet de réduire considérablement ces problèmes de biais. Par ailleurs, la STD nous indique que les observations, principalement urbaines, n'apportent pratiquement pas d'information sur la variabilité des concentrations sur la zone rurale. Dans le cas de la méthode ASI, par exemple, les analyses et les simulations sont très proches, excepté à 8-9 h UTC et à 18 h UTC. A ces heures, la méthode ASI corrèle plus fortement les erreurs commises sur les stations rurales à celles commises sur les stations urbaines.

Le tableau 6.1 donne le détail des statistiques d'une comparaison des simulations et des analyses (méthode ASI) avec les observations, effectuée à 7 h UTC, station par station, et en moyenne sur l'ensemble des stations. Le biais et la RMS sont fortement réduits, sauf sur les stations rurales. Les corrélations sont également améliorées, passant d'une moyenne de 0.75 à 0.86. La variabilité des concentrations est donc bien mieux représentée par les analyses. En ce qui concerne les erreurs supérieures à 15 ppb, elles sont généralement évitées grâce à l'introduction des observations. En effet, les erreurs commises sur les dépassements des seuils de 75 et 90 ppb (souvent simulés mais non observés) sont pour la plus grande partie réduites. A quelques exceptions près (sur les stations GENNEVILLIERS, ST-DENIS, et VITRY), les dépassements des seuils 50 et 75 ppb sont systématiquement reproduits à moins de 15 ppb près par les analyses.

TAB. 6.1 – Statistiques d'erreurs de comparaison des concentrations de dioxyde d'azote simulées/analysées avec les valeurs observées à 7 h UTC. Ces statistiques sont données station par station et en moyenne sur l'été 2001. L'unité des concentrations est en ppb.

Station	BIAS	RMS	COR	E15	E15/50	E15/75	E15/90
GARCHES	16.2/5.0	20.6/8.4	0.75/0.89	53/94	8/100	0/---	0/---
ARGENTEUIL	-0.1/-7.2	10.5/9.8	0.77/0.87	84/89	50/100	50/---	---/---
ISSY	14.7/-0.1	19.6/5.1	0.79/0.94	57/99	13/100	0/---	0/---
NEUILLY	17.1/-0.1	21.2/5.3	0.83/0.95	54/97	8/100	6/100	0/---
PARIS07	15.6/-1.4	20.0/6.2	0.82/0.93	53/99	9/100	7/---	0/---
GENNEVILLIERS	8.9/-1.8	13.2/6.6	0.84/0.91	80/96	30/66	25/100	0/---
PARIS06	19.7/2.2	23.0/5.1	0.84/0.96	42/100	0/100	0/---	0/---
PARIS18	13.4/-2.7	17.6/6.6	0.80/0.92	62/95	27/100	20/100	33/---
PARIS13	20.2/2.7	23.7/6.5	0.75/0.91	37/95	5/100	9/---	0/---
ST-DENIS	12.4/0.3	16.1/5.7	0.80/0.91	68/96	0/33	0/---	0/---
VITRY	13.9/-1.0	19.0/6.4	0.72/0.90	57/99	11/0	0/0	0/---
AUBERVILLIERS	13.2/-0.2	16.6/4.7	0.78/0.93	65/100	22/100	66/100	---/---
PARIS12	18.0/1.4	21.9/6.0	0.74/0.91	51/98	9/100	0/---	0/---
IVRY	16.1/0.0	19.9/5.1	0.78/0.94	57/99	9/100	10/---	0/---
EVRY	1.2/-3.5	7.8/6.5	0.77/0.87	93/95	---/---	---/---	---/---
BOBIGNY	9.6/-1.8	14.2/6.8	0.70/0.85	77/96	25/100	0/---	0/---
MONTGERON	6.5/1.2	10.6/4.6	0.73/0.89	87/97	0/---	---/---	---/---
TREMBLAY	4.5/-1.5	10.4/6.7	0.61/0.75	89/96	0/---	---/---	---/---
CERGY	0.8/-0.2	7.2/5.8	0.72/0.76	94/97	---/---	---/---	---/---
CACHAN	17.5/2.5	21.4/5.6	0.78/0.94	49/98	5/---	0/---	0/---
PARIS01	20.5/2.5	23.4/4.7	0.81/0.96	39/100	0/100	0/---	0/---
LA-DEFENCE	9.9/-3.4	15.1/6.0	0.84/0.94	72/100	25/100	20/---	0/---
VERSAILLES	6.9/-0.9	12.4/5.5	0.80/0.91	84/98	50/---	0/---	---/---
MELUN	-2.6/-1.4	6.5/5.0	0.64/0.76	96/98	---/---	---/---	---/---
FONTAINEBLEAU	0.6/3.6	4.3/5.3	0.46/0.44	100/99	---/---	---/---	---/---
MANTES	-0.8/0.3	5.6/4.9	0.72/0.74	95/99	---/---	---/---	---/---
RAMBOUILLET	2.0/4.4	4.8/6.4	0.70/0.56	99/99	---/---	---/---	---/---
MOYENNE	10.2/-0.0	15.1/6.0	0.75/0.86	70/97	15/88	11/80	2/---

6.2.3 Profils temporels et cycles diurnes

Pour illustrer les analyses produites à partir de toutes les observations, nous présentons sur la figure 6.2, les séries temporelles des concentrations du dioxyde d'azote, observées, simulées par le modèle CHIMERE, et analysées à l'aide de la méthode ASI à 9 h UTC. Les variations du polluant sont très bien reproduites par les analyses, en particulier sur la station AUBERVILLIERS. En effet, en ce point de mesure, la RMS est réduite de plus d'environ 70% (elle passe de 13.4 à 4.3 ppb). Le 24 août, les analyses donnent des concentrations de dioxyde d'azote d'environ 100 ppb sur les

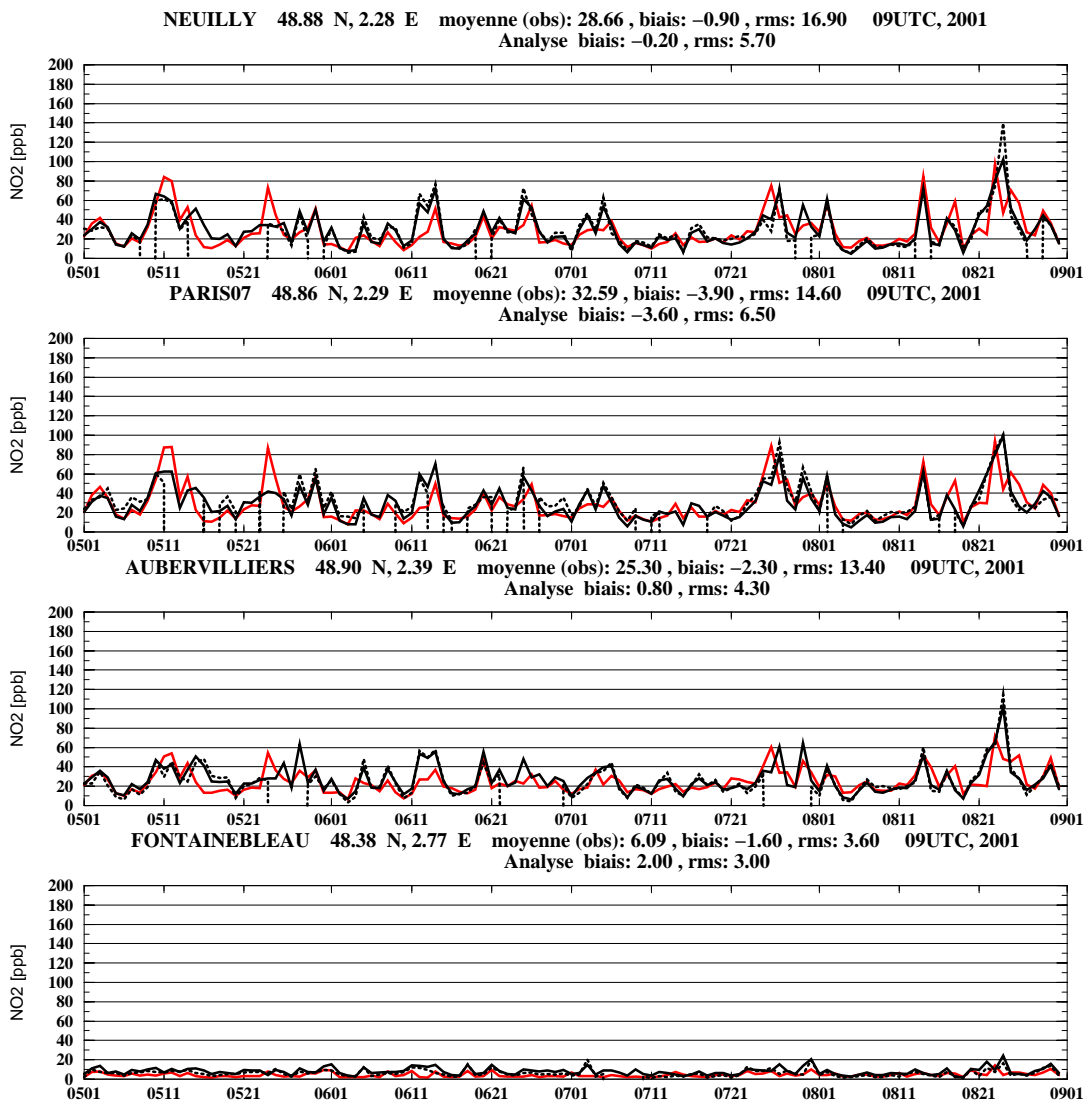


FIG. 6.2 – Séries temporelles des concentrations de dioxyde d'azote observées (courbe noire pointillée), simulées par le modèle CHIMERE (courbe pleine grise) et analysées (courbe pleine noire) à 9 h UTC du 1^{er} Mai au 31 août 2001 sur trois stations urbaines (NEUILLY, PARIS07, AUBERVILLIERS), et une station rurale (FONTAINEBLEAU) du réseau AIRPARIF. L'unité des concentrations est en ppb.

trois stations urbaines, indiquant trois dépassements du seuil d'information, effectivement observés mais initialement non simulés. Cependant, il est à noter que le dépassement de seuil constaté sur la station NEUILLY est toujours sous-estimé (d'environ 40 ppb) dû au caractère trop localisé de ce pic de pollution.

De même, pour donner une idée des performances de la méthode en fonction de l'heure de la journée, nous montrons sur la figure 6.3 les cycles diurnes des concentrations du dioxyde d'azote,

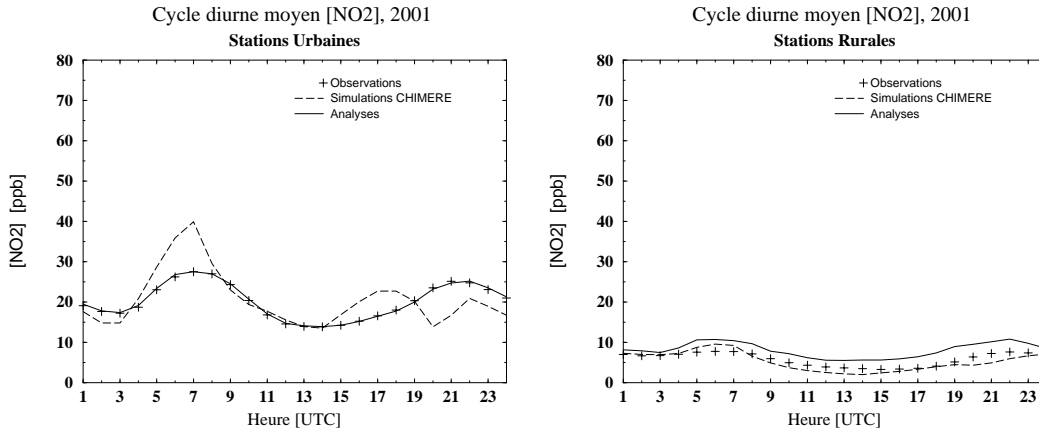


FIG. 6.3 – Profils diurnes des concentrations de dioxyde d’azote, observées (croix), simulées (courbe pointillée) et analysées à l’aide de la méthode ASI (courbe pleine). L’unité des concentrations est en ppb, celle de l’heure en UTC.

observées, simulées par le modèle CHIMERE, et analysées à l’aide de la méthode ASI. Le cycle diurne est parfaitement reproduit par les analyses sur les stations urbaines, ce qui n’est pas le cas sur les stations rurales. Il semble que les observations en zone urbaine est trop de poids dans la correction des concentrations simulées en zone rurale. Il serait bien, pour limiter cette influence, de coupler un modèle de covariance basé sur un critère de corrélation avec un modèle basé sur un critère de distance. Les différences constatées sur les stations rurales devraient en plus être amplifiées par les erreurs de mesure car le dioxyde d’azote est potentiellement surestimé sur les zones rurales (en effet, la mesure du dioxyde d’azote est *a priori* non spécifique, Cf. paragraphe 3.2.2, chapitre 3). Cependant, ces erreurs ne sont pas réellement préoccupantes, vu le faible niveau de concentration du polluant hors de la zone urbaine (environ 10 ppb).

6.3 Étude de la pollution sur les bassins de Berre et de Marseille

Nous présentons, dans cette section, les analyses des champs de concentration d’ozone et de dioxyde d’azote sur les bassins de Berre et de Marseille, produites pendant l’été 2001. La méthodologie suivie pour construire les modèles de biais, de corrélation et de variance n’est pas modifiée par rapport au paragraphe précédent.

6.3.1 Cas de l’ozone

Le tableau 6.2 donne le détail des statistiques de comparaison des simulations ou des analyses avec les observations des concentrations d’ozone de surface à 15 h UTC. Ces valeurs sont calculées sur les données de l’été 2001, station par station et en moyenne sur toutes les stations, dans le cadre d’un *leave-one-out*.

Globalement, les scores concernant les analyses sont meilleurs que ceux qui sont calculés pour les

simulations. La prise en compte des observations permet, par exemple, de réduire les biais les plus forts, principalement constatés sur le bassin de Berre. Nous notons, en revanche, une dégradation de ces biais sur les stations proches de la ville d'Aix. De manière générale, nous avons remarqué que les analyses reproduisent parfaitement les valeurs observées durant les jours très peu pollués car les concentrations d'ozone sont alors relativement homogènes. Pendant les épisodes de forte pollution photochimique, les observations mettent en évidence un panache, souvent très localisé, que le modèle CHIMERE a beaucoup de difficulté à positionner. Dans ces situations, l'algorithme d'assimilation de données a du mal à corriger des surestimations et des sous-estimations voisines. Ceci se traduit systématiquement par une sous-estimation du pic journalier. Comme la ville d'Aix se trouve en bordure de panache, lors des épisodes de pollution photochimique, les analyses sous-estiment souvent les fortes concentrations mesurées par les stations situées au bord de cette ville. Notons également que les scores sont très peu modifiés sur la station CADARACHE, située au nord-est du domaine de simulation, loin des autres stations de surface.

TAB. 6.2 – Statistiques d'erreurs de comparaison des concentrations d'ozone simulées/analysées avec les valeurs observées à 15 h UTC, station par station et en moyenne sur l'été 2001. L'unité des concentrations est en ppb.

Station	Niv.	BIAS	RMS	COR	E15	E15/50	E15/75	E15/90
AUBAGNE-PENITENTS	1	0.5/-0.7	16.4/9.3	0.61/0.91	76/90	63/76	44/65	30/41
AIX-CENTRE-ART	1	0.9/-3.6	15.5/10.2	0.72/0.90	73/84	56/66	48/56	23/38
AIX-OUEST	1	-0.9/-3.5	13.5/8.7	0.77/0.92	80/91	69/83	60/75	33/58
MARSEILLE-5AVENUES	1	8.1/7.9	15.3/11.2	0.66/0.88	69/83	45/68	35/64	28/25
MARSEILLE-MARGUERITE	1	2.4/1.6	12.2/7.2	0.69/0.90	83/94	71/84	59/78	33/60
LA-CIOTAT	1	3.3/-2.3	14.3/8.9	0.66/0.87	77/91	60/85	58/73	43/40
TOULON-ARSENAL	1	5.9/2.7	13.6/6.8	0.69/0.92	77/95	61/92	50/94	16/83
LA-SEYNE-SUR-MERS	1	7.8/3.7	15.2/8.8	0.67/0.88	64/88	43/80	42/80	25/77
AVIGNON-MAIRIE	1	1.8/-0.1	9.9/5.2	0.78/0.94	88/98	81/96	76/100	66/100
AVIGNON-ARROUSAIRE	1	4.4/2.5	10.4/5.9	0.81/0.94	87/97	83/93	73/100	25/100
LE-PONTET-MARGUERITE	1	4.1/2.6	11.9/7.8	0.76/0.91	85/92	82/89	82/85	60/50
NIMES-GARDIES	1	-2.4/-2.5	8.4/6.2	0.85/0.92	91/97	82/93	77/88	83/80
BERRE	1	-2.9/-1.4	15.1/6.9	0.55/0.92	74/94	48/88	34/80	0/50
PAGODE	1	-8.6/0.5	16.2/8.6	0.42/0.80	68/94	19/88	11/80	0/33
PORT-DE-BOUC	1	-1.8/1.6	13.1/8.0	0.61/0.87	79/90	60/73	30/55	14/40
ISTRES	1	-6.8/0.2	16.2/7.3	0.50/0.92	74/96	25/86	14/73	0/50
FOS-Carabins	1	-12.1/-4.7	19.1/9.2	0.53/0.92	64/92	26/83	11/58	0/20
MIRAMAS	1	-7.0/-3.3	17.3/11.3	0.59/0.85	78/99	50/74	36/62	20/44
VITROLLES	1	-6.8/-4.7	17.0/8.2	0.55/0.92	75/92	51/80	41/75	27/70
MARIGNANE-Ville	1	-1.2/1.4	17.1/9.1	0.43/0.84	80/92	54/72	33/57	22/44
ST-MARTIN-de-CRAU	1	-3.4/-1.1	9.5/6.7	0.80/0.90	89/92	77/80	76/61	100/25
ARLES	1	6.0/7.2	10.0/9.0	0.83/0.93	86/94	75/83	84/76	100/100
SALON	1	0.5/0.0	11.9/10.0	0.83/0.88	82/89	72/72	62/64	25/25
ROGNAC	1	-7.9/-6.1	16.3/9.1	0.66/0.94	74/90	48/76	39/71	26/57
SAUSSET	1	5.2/3.4	17.8/12.1	0.55/0.79	75/90	58/73	43/61	22/40
BOUC-BEL-AIR	1	1.5/-0.1	15.1/9.1	0.58/0.86	73/88	50/77	41/66	30/55
ROUSSET	1	1.6/-5.8	14.2/12.1	0.77/0.87	78/87	64/73	59/69	46/47
SAZE	1	-1.3/-2.5	8.7/5.8	0.81/0.93	91/98	86/96	81/96	50/100
VALLABREGUES	1	-2.3/-2.8	9.1/6.5	0.82/0.92	91/96	82/92	78/88	75/66
COMTAT-VENAISSIN	1	8.2/5.8	12.2/9.5	0.80/0.87	75/88	61/71	58/80	100/100
AIX-PLATANES	1	0.9/-5.0	15.0/11.4	0.74/0.88	78/87	63/76	61/67	39/50
TOULON-OLIVE	1	3.1/0.2	12.2/5.5	0.68/0.93	79/97	64/92	48/82	22/57
PENNES-MIRABEAU	1	-1.4/-1.6	15.5/8.8	0.57/0.87	71/91	52/83	34/70	21/50
CUQUES-ALLAUCH	1	-1.6/-2.2	13.9/7.2	0.67/0.93	75/92	59/81	41/77	33/77
VALLEE-HUVEAUNE	1	1.3/0.8	15.1/8.3	0.61/0.90	74/91	53/80	44/61	20/33
LA-VALETTE-LA-GARDE	1	4.0/1.2	11.3/6.0	0.74/0.92	82/98	69/95	52/90	16/83
MONTPELLIER-NORD	1	-0.1/0.7	8.6/7.3	0.85/0.89	91/93	80/85	73/76	66/66
MONTPELLIER-SUD	1	1.9/2.4	10.3/8.4	0.81/0.89	87/94	81/92	72/87	60/60
SAINT-REMY	1	1.0/-1.2	8.8/5.7	0.84/0.94	88/99	82/98	75/96	80/100
ST-MARIE-de-la-MER	1	10.0/8.4	14.4/13.0	0.67/0.76	66/75	38/40	30/14	50/0
CADARACHE	1	2.6/-2.0	10.5/10.3	0.83/0.84	87/88	83/72	81/63	81/33
MOYENNE		0.6/-0.3	13.3/8.7	0.70/0.89	79/91	62/80	53/72	39/54

Pour illustrer les performances de la méthode ASI sur tout l'été 2001, la figure 6.4 donne quelques exemples de séries temporelles de concentration d'ozone analysées à 15 h UTC. Nous comparons ces séries aux observations et aux simulations brutes. Il apparaît nettement que l'assimilation de données permet de mieux reproduire la variabilité du polluant, et principalement les concentrations plus faibles. En revanche, les pics sont systématiquement sous-estimés. Par exemple, les dépassements de seuil d'information des 25 et 26 Juin 2001 ne sont pas mieux représentés par les analyses sur la station AURIBEAU. Cette station est, comme la station AIX-PLATANES en

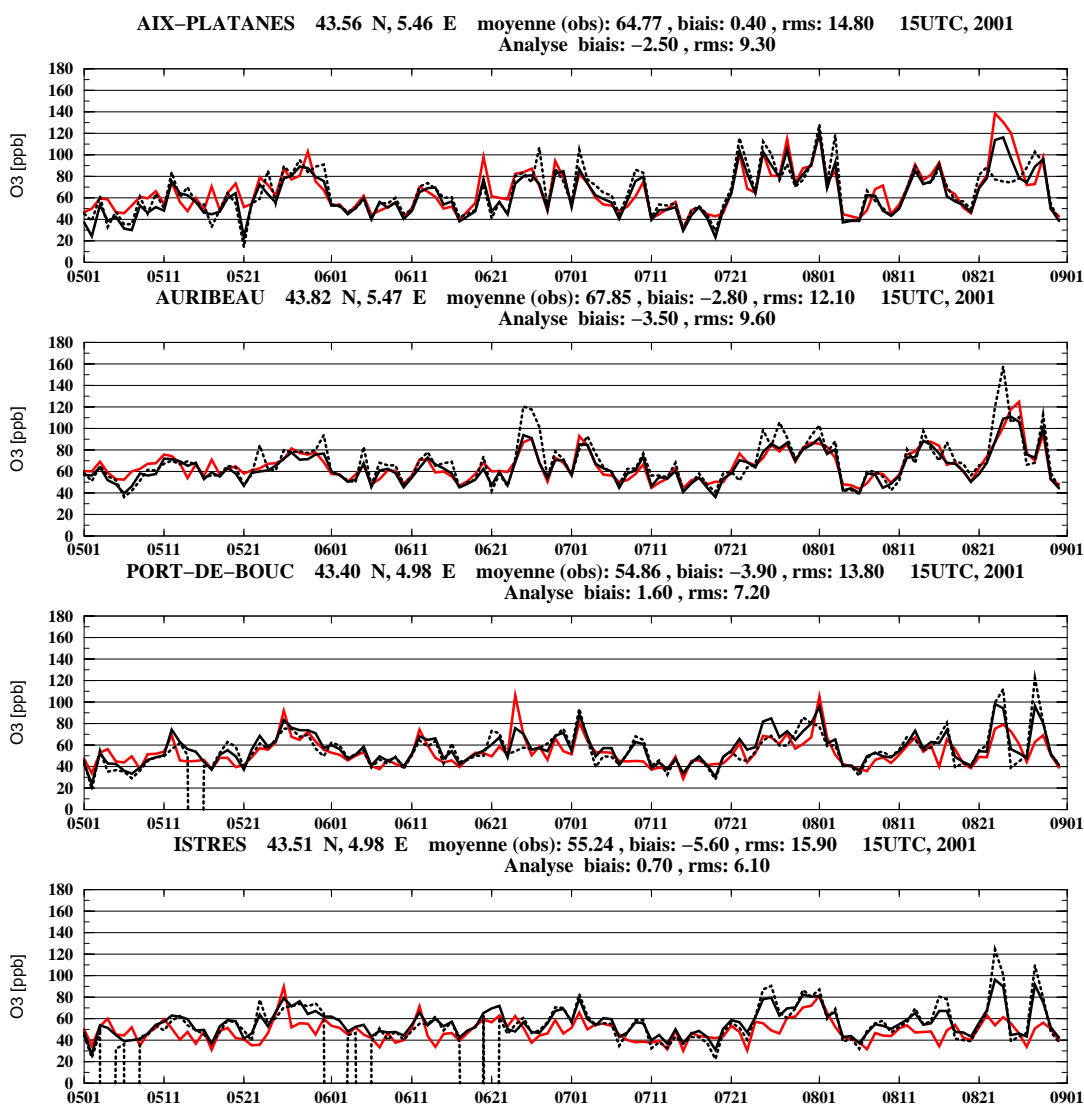


FIG. 6.4 – Séries temporelles des concentrations d’ozone observées (courbe pointillée), simulées par CHIMERE régional (courbe pleine grise) et analysées (courbe pleine noire) à 15 h UTC du 1^{er} Mai au 31 août 2001 sur quatre stations des réseaux AIRMARAIX ou AIRFOBEP. L’unité des concentrations est en ppb.

bordure de panache. Le manque d’observations dans cette région ne permet pas de repositionner le panache plus à l’est ces jours-là.

La situation du 23 août est complexe et les simulations sont difficiles à corriger. Le modèle simule de fortes concentrations d’environ 100 ppb entre Marseille et l’étang de Berre, alors qu’il devrait plutôt les simuler sur l’étang de Berre, un peu plus à l’ouest. En des endroits très rapprochés, le modèle surestime et sous-estime. L’analyse a alors tendance à se rapprocher de la moyenne.

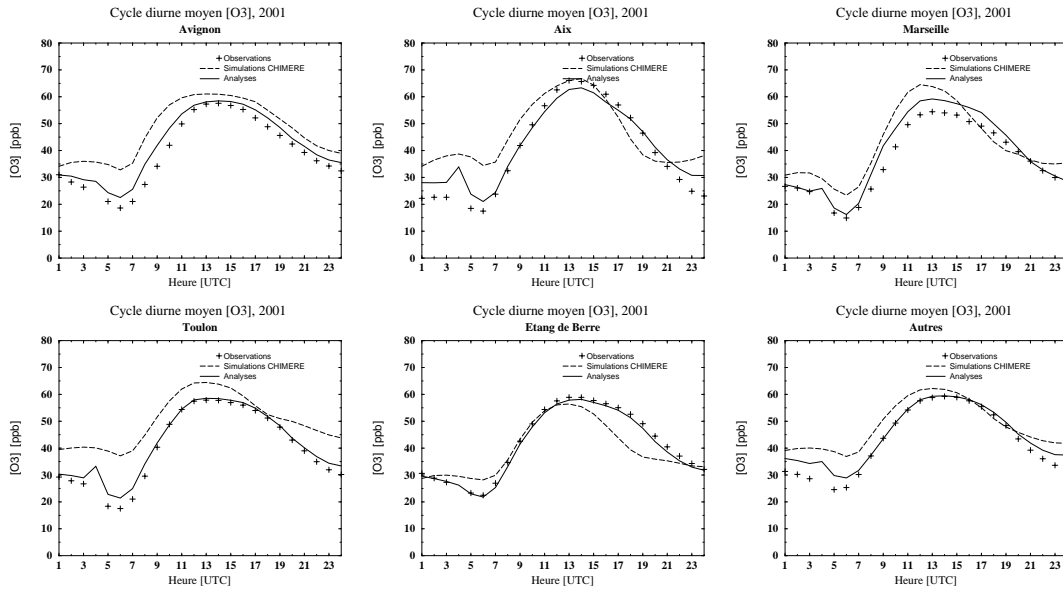


FIG. 6.5 – Profils diurnes moyens des concentrations d’ozone observées (croix) et simulées par le modèle CHIMERE régional (courbe pointillée) et analysées (courbe pleine) sur quatre agglomérations, sur l’Étang de Berre et sur le reste des stations. Ils correspondent à des moyennes de profils calculés sur plusieurs stations les plus proches des sites. L’unité des concentrations est en ppb.

Le 28 août, la situation est mieux analysée car le modèle simule des concentrations relativement homogènes, de 70 à 75 ppb à l’ouest de Marseille et autour de l’étang de Berre, et de 55-60 ppb sur l’étang. Les observations sont à peu près en accord avec les simulations, sauf sur les stations les plus proches de l’étang de Berre. Treize stations de mesure indiquent un panache très concentré (les concentrations d’ozone atteignent 95 à 110 ppb). Les analyses sont donc suffisamment contraintes par les observations pour s’approcher de leur valeur. Néanmoins, une des stations de l’étang de Berre, la station PAGODE, mesure 50 ppb, limitant alors les analyses sur la station voisine PORT-DE-BOUC à 100 ppb.

Enfin, nous présentons, sur la figure 6.5, les cycles diurnes moyens des concentrations d’ozone observées, simulées, et analysées. Ils sont ici calculés sur plusieurs stations localisées dans ou autour des principales agglomérations (Avignon, Aix, Marseille, Toulon) et autour de l’Étang de Berre. L’anomalie, qui apparaît à 4 h UTC, correspond à l’heure où l’association AIRMARAIX teste le fonctionnement de chacune des stations de son réseau de mesure. Sur les 44 observations usuellement disponibles, il en manque alors 22 et les analyses se rapprochent des simulations. Globalement, les analyses représentent mieux les cycles diurnes que les simulations initiales. Cependant, le cycle diurne moyen sur Marseille est surestimé. Celui-ci est en fait calculé à partir des données prises sur deux stations, MARSEILLE-5 AVENUES et MARSEILLE-MARGUERITE. Les concentrations moyennes observées à 15 h UTC sur ces stations sont respectivement de 50 et de 56 ppb. La première station est donc plus influencée par des sources d’émission que la deuxième. De manière générale, nous notons que les analyses sont moins performantes la nuit à cause des

problèmes de représentativité des stations et de la distribution assez dispersée des stations. En effet, les stations sont éloignées l'une des autres, n'apportant ainsi qu'une petite information dans l'analyse en d'autres points de mesure.

6.3.2 Cas du dioxyde d'azote

Nous présentons dans ce paragraphe les analyses des concentrations de dioxyde d'azote de l'été 2001, produites à partir de toutes les observations disponibles (observations de trafic, urbaines, périurbaines, rurales, et industrielles).

Généralement, la comparaison des séries temporelles de concentrations observées, simulées et analysées, montre qu'en *leave-one-out*, les corrélations sont largement améliorées après analyse, passant d'une moyenne de 0.53 à une moyenne de 0.76. En revanche, la RMS moyenne des différences simulations-observations est plus faiblement réduite. Elle passe de 15.6 à 11.8.

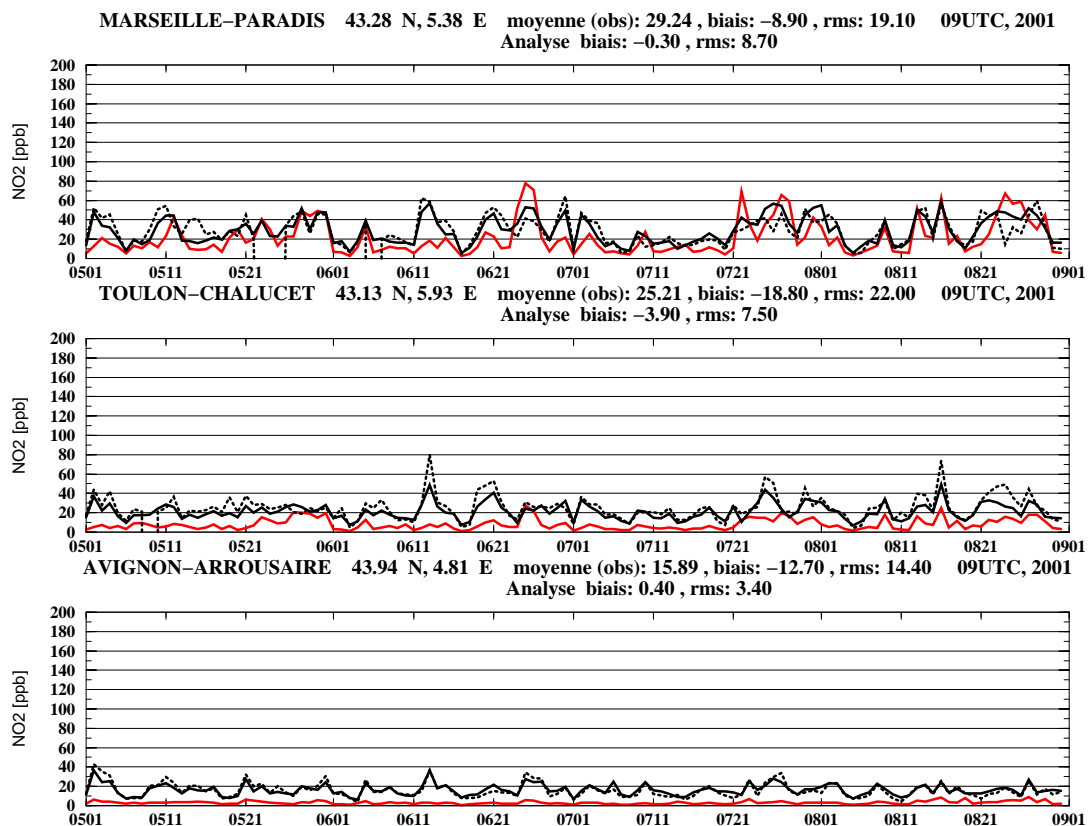


FIG. 6.6 – Séries temporelles des concentrations de dioxyde d'azote observées (courbe noire pointillée), simulées par le modèle CHIMERE régional (courbe pleine grise) et analysées (courbe pleine noire) à 9 h UTC du 1^{er} Mai au 31 août 2001 sur trois stations des réseaux AIRMARAIX ou AIRFOBEP. L'unité des concentrations est en ppb.

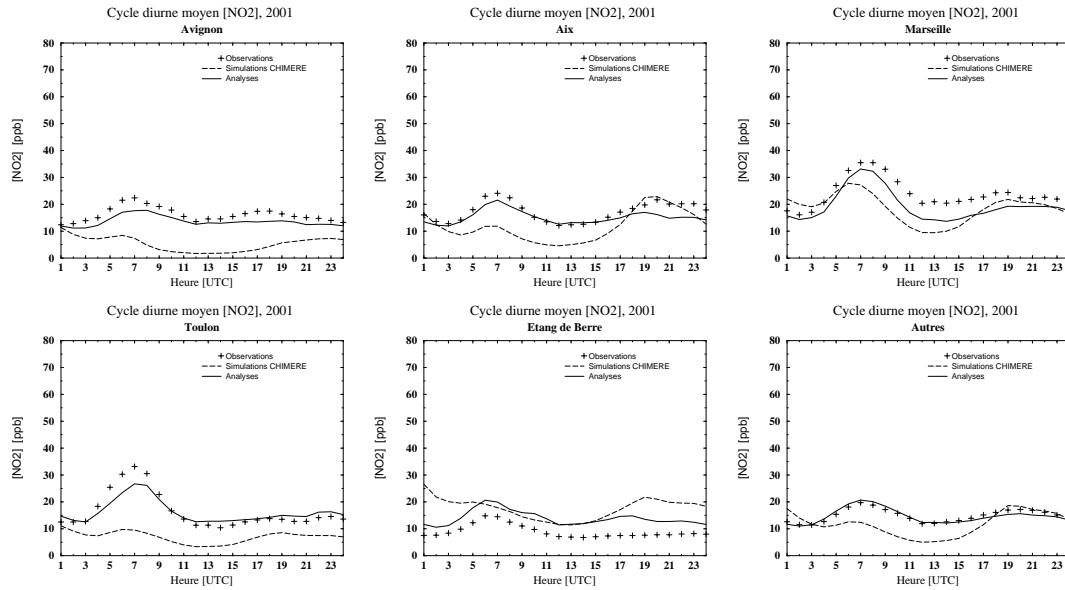


FIG. 6.7 – Profils diurnes des concentrations de dioxyde d'azote observées (croix), simulées par le modèle CHIMERE régional (courbe pointillée) et analysées (courbe pleine) sur quatre agglomérations, sur l'Étang de Berre et sur le reste des stations. Ils correspondent à des moyennes de profils calculés sur plusieurs stations les plus proches des sites. L'unité des concentrations est en ppb, l'heure est en UTC.

La figure 6.6 présente quelques exemples de séries temporelles à 9 h UTC. La variabilité du polluant est bien mieux reproduite par les analyses que par les simulations. Les résultats sont remarquables sur la station AVIGNON-ARROUSSAIRE, le biais passant de -12.7 à 0.4, et la RMS de 14.4 à 3.4. Ceci est dû à la présence de sources d'émission de faible intensité. Comme dans le cas de l'ozone, nous avons remarqué que nous améliorons plus particulièrement les valeurs faibles. Dès que les concentrations sont plus fortes, les analyses les sous-estime. C'est le cas sur la station TOULON-CHALUCET. En fait, le centre ville de Toulon relève des niveaux de dioxyde d'azote assez élevés, de 80 ppb, deux fois pendant l'été 2001. Ces plus fortes concentrations sont repérées par les analyses mais largement sous-estimées à cause de leur caractère très local.

Enfin, comme l'indique la figure 6.7, les cycles diurnes sont également mieux représentés par les analyses. La difficulté rencontrée sur Marseille est due à l'utilisation des stations de trafic très peu représentatives, comme les stations MARSEILLE-PLOMBIERES et AUBAGNE-FOCH.

6.4 Conclusions

A la question "Est-ce qu'il est possible d'améliorer la représentation des champs de concentration de l'ozone et du dioxyde d'azote quand le modèle a initialement beaucoup de mal à les simuler?", nous pouvons maintenant répondre et la réponse est oui.

Nous avons montré, en effet, que les pics de concentrations du dioxyde d'azote simulés à 15 ppb près sur la région d'Île-de-France, sont représentés à moins de 6 ppb près par les analyses (ASI

ou INK). La combinaison des observations et des simulations permet de réduire non seulement les biais, mais aussi d'améliorer la représentation de la variabilité des concentrations du polluant. Le nombre des erreurs supérieures à 15 ppb sont également réduites, passant de 30% (simulations) à moins 5% (analyses). Le cycle diurne est, en outre, parfaitement reproduit sur la zone urbaine. En revanche, une étude plus approfondie doit être menée pour mieux modéliser le biais et réduire l'influence des stations urbaines sur les zones rurales.

Sur la région de Berre et de Marseille, les résultats sont plus discutables, mais préliminaires. Nous avons utilisé toutes les observations disponibles pour produire les analyses des champs de concentration d'ozone et du dioxyde d'azote. Une sélection des stations, les plus représentatives, devra être effectuée dans le futur. Néanmoins, nous avons montré que la méthode ASI permet d'obtenir des champs de concentration d'ozone cohérents avec les observations à moins de 9 ppb à 15 h UTC (correspondant à peu de chose près à l'heure du pic journalier), le nombre d'erreurs supérieures à 15 ppb étant limité à 10%. Par rapport aux simulations, les corrélations entre observations et analyses sont également nettement meilleures. Elles passent de 0.7 en moyenne à 0.89, avec de plus fortes améliorations sur la région de l'étang Berre, où les corrélations peuvent passer d'environ 0.5 à 0.9.

Les analyses du dioxyde d'azote sont nettement moins biaisées que les simulations. Elles représentent, en outre, bien mieux la variabilité du polluant.

Chapitre 7

Conclusions et Perspectives

Ce chapitre de conclusion résume brièvement la contribution de cette thèse à la représentation de la pollution photochimique et sa prévision. Les perspectives offertes par ce travail y sont ensuite discutées.

Par la loi de la qualité de l'air, l'état s'est engagé à surveiller les polluants atmosphérique à l'échelle du territoire français. Depuis le vote de cette loi, le réseau d'observation s'est effectivement étendu des grandes vers les petites agglomérations, jusqu'aux régions rurales éloignées des sources d'émission. Cette extension du réseau, couplé au développement de modèle de chimie-transport a permis de considérables avancées dans le domaine de la qualité de l'air. Cependant, jusqu'à aujourd'hui les polluants étaient toujours représentés soit par des cartes d'observations interpolées (mesures de surface ou mesures aéroportées), soit par des simulations numériques. Or, nous savons que d'un côté, les mesures donnent une information plus ou moins localisées en fonction de leur localisation, et que l'autre les simulations offrent une visualisation plus ou moins réaliste des champs des polluants.

L'objectif majeur de cette thèse a été d'étudier précisément la représentation des polluants atmosphériques. Le travail de thèse s'est donc focalisé sur le développement et la validation d'outils statistiques permettant de mixer les informations apportées par les observations et les simulations, afin d'obtenir des cartes, i.e. des analyses, de champs de polluants plus réalistes que les simulations.

Les expériences de cartographie menées en temps réel sur l'Europe pendant ces trois années de travail ont mis en évidence une difficulté (qui existe dans une moindre mesure à l'échelle régionale) : le transport d'erreurs de prévision. Afin de savoir s'il est possible de limiter ce transport d'erreurs, les analyses tridimensionnelles, produites à l'échelle européenne, ont été utilisées pour initialiser un modèle de prévision. Des études ont donc été menées afin de quantifier l'effet de ces analyses sur les prévisions à court terme.

7.1 Synthèse des travaux entrepris

Les résultats de ce travail peuvent être synthétisés en trois grandes parties. La première partie a consisté à diagnostiquer les erreurs commises par le modèle CHIMERE multi-échelles, c'est-à-dire à examiner dans quelle mesure le modèle CHIMERE est capable de simuler les concentrations

des polluants aux échelle continentale et régionale. L'étude s'est principalement focalisée sur deux polluants : l'ozone et le dioxyde d'azote. La seconde partie est destinée à la mise en place de méthodes d'assimilation de données et de Krigeage. Enfin, la dernière partie est consacrée à l'étude de la prévisibilité de la pollution photochimique.

7.1.1 Diagnostic des erreurs de modélisation du modèle CHIMERE

Nous avons dans un premier temps cherché à quantifier l'*erreur globale* commise sur les sorties du modèle CHIMERE. Nous savons que celle-ci est la combinaison d'une erreur liée à l'état d'avancement du modèle (en effet, celui-ci n'est que la traduction de l'état des connaissances scientifiques du moment), d'une erreur sur ces entrées (émissions, météo, conditions aux limites ...), et sur les conditions initiales. S'il est possible d'estimer la sensibilité des sorties du modèle à tous les paramètres, il est difficile de s'exprimer sur la part d'erreur due à chacun de ces paramètres.

Les comparaisons des sorties du modèle CHIMERE, effectuées aussi bien avec des observations de surface qu'avec des mesures aéroportées, ont montré que le modèle simule de façon *assez* réaliste les concentrations d'ozone, à l'échelle continentale et à l'échelle régionale. Les concentrations du dioxyde d'azote sont, en revanche, moins bien reproduites.

Nous avons montré, sur la base des données de l'été 2001, que les concentrations des deux polluants sont généralement représentées à moins de 15 ppb près par le modèle continental et régional. Le nombre d'erreurs supérieures à 15 ppb commises par le modèle régional sur les concentrations d'ozone est de 10% en Île-de-France et de 20% sur la région de Berre et de Marseille. Nous avons montré, en outre, que 60% des dépassements du seuil d'information et de recommandation concernant l'ozone (90 ppb) sont mal simulés (erreurs supérieures à 15 ppb) sur les stations méditerranéennes, contre 40% sur les stations parisiennes. En effet, le modèle, dans sa version actuelle, ne permet pas de représenter de façon détaillée les écoulements de méso-échelle, les circulations complexes le long des côtes et dans les régions montagneuses.

Sur les deux régions étudiées, nous avons également montré que 30% des erreurs commises sur les concentrations du dioxyde d'azote sont supérieures à 15 ppb. Les dépassements du seuil de 50 ppb sont pratiquement systématiquement mal représentés (non observés ou non prévus). Ces erreurs sont associées aux faibles résolutions verticale et horizontales du modèle, à l'incertitude sur les émissions, et à la mauvaise représentation du mélange lors d'épisodes nuageux.

Enfin, les expériences en temps réel de cartographie des champs de concentrations à l'échelle européenne ont mis en évidence des erreurs systématiques, liées à la présence de nuages. De manière générale, l'utilisation du modèle CHIMERE est principalement limitée par ses résolutions horizontales et verticale, et par la simplification de processus essentiels tels que le mélange et l'atténuation des taux de photolyse en présence de nuages.

7.1.2 La cartographie

Nous avons ensuite cherché à produire des champs de concentration de polluants plus réalistes que les simulations. Le travail a alors consisté soit à corriger l'*erreur globale* commise sur les simulations des champs de concentration de polluants à l'aide des observations en surface, soit à interpoler des observations (sans utilisation du modèle numérique). Plusieurs techniques d'assimi-

lation de données ou de Krigeage ont été testées, validées objectivement et intercomparées. Nous détaillons maintenant les principaux résultats que nous avons obtenus.

Les Interpolations Statistiques

Ces méthodes, empruntées à la météorologie, consistent à interpoler linéairement les différences constatées entre une simulation et observations, et à ajouter le champ de différences obtenu à la simulation initiale. Le résultat de l'opération est ce qu'on appelle couramment une *analyse*. Les poids de l'interpolation optimaux (qui minimise la variance de l'erreur d'analyse) sont liés aux covariances des erreurs du modèle et d'observation. La détermination de ces covariances, non accessibles directement, repose sur une étude statistique des erreurs. Cette étude est généralement menée sur une longue période de temps : on utilise une *climatologie*, i.e. un ensemble de données passées. L'une des principales hypothèses de ces méthodes est de considérer que les covariances des erreurs sont stationnaires dans le temps. Les poids de l'interpolation ne dépendent alors que de la distribution des stations. Deux méthodes ont été appliquées.

- ▷ La première méthode, appelée méthode ISI (*Isotropic Statistical Interpolation*), est classique, régulièrement utilisée autrefois pour produire des analyses météorologiques. Elle a été utilisée comme référence initiale. Elle repose sur l'hypothèse que les covariances des erreurs du modèle entre deux points dépendent uniquement de la distance qui les séparent. Autrement dit, la fonction de covariance de l'erreur du modèle est supposée homogène et isotrope.

Nous avons adapté cette méthode au cas de la pollution atmosphérique en proposant une estimation des biais de modélisation et d'observation. Ces biais sont souvent "oubliés" dans la littérature, mais primordiaux pour représenter correctement des champs de concentration de polluants.

L'utilisation de cette méthode d'interpolation a montré quelques inconvénients, notamment durant les heures de fortes émissions et la nuit. Une étude de la covariance des erreurs du modèle pendant ces heures a montré que celle-ci est alors clairement anisotrope.

- ▷ La deuxième méthode, dite méthode ASI (*Anisotropic Statistical Interpolation*) est plus originale. Les covariances des erreurs du modèle sont modélisées en fonction de la covariance des concentrations simulées par le modèle numérique (le modèle de covariance est quant à lui estimé à partir des concentrations observées). Comme la fonction de covariance des concentrations n'est pas *a priori* isotrope, la méthode permet de modéliser une fonction anisotrope de covariance des erreurs du modèle. En outre, suffisamment générale, elle est utilisée pour corriger des champs tridimensionnels de polluants. L'utilisation des mesures de la campagne ESQUIF a permis une validation de la méthode en altitude. Cet algorithme est, depuis l'été 2001, utilisé en routine par l'association AIRPARIF pour produire des analyses d'ozone journalières.

Les méthodes de Krigeage

Ces méthodes sont empruntées aux sciences minières : elles consistent à interpoler spatialement des données. Elles ont été utilisées pour éviter l'hypothèse de stationnarité supposée dans l'utilisation des méthodes précédentes. Les poids de l'interpolation sont donc ici estimés à partir des données disponibles à *l'instant de l'analyse*, sans utiliser leur dimension temporelle. Les poids sont

alors variables dans le temps. Deux sortes de quantités ont été interpolées : les observations ou les différences simulation-observation.

- ▷ Un Krigeage intrinsèque a été appliqué aux observations. C'est la méthode OBK. Cette méthode est la plus couramment utilisée par les agences de surveillance de la qualité de l'air. Elle consiste à interpoler spatialement les observations. Les poids de l'interpolation sont issus d'un calcul de statistique spatiale. Ils sont recalculés à chaque analyse, et sont donc variables dans le temps. Cette approche est bien évidemment d'autant plus performante que le nombre d'observations, bien réparties, est suffisant. Ce nombre minimum est lui même dépendant de la complexité de la structure du champ de concentration des espèces que l'on souhaite analyser. Cette méthode a montré quelques faiblesses à détecter des panaches urbains hors des zones observées.
- ▷ La méthode INK a été développée pour améliorer la précédente. Elle se base sur le même principe que la méthode OBK, mais au lieu d'interpoler les observations, nous interpolons les différences simulations-observations. Cette méthode conduit à de bien meilleures performances par rapport à la méthode OBK car le modèle structure plus correctement le champ estimé. Si le modèle simule à peu près convenablement les concentrations du polluant, les différences simulations-observations sont plus homogènes que les observations elles-même et un Krigeage intrinsèque, appliqué sur ces différences, se montre plus efficace. La méthode INK est très proche de la méthode ASI. La différence est liée à la détermination des poids de l'interpolation. Avec la méthode INK, les poids de l'interpolation sont recalculés à chaque analyse, alors que dans le cas de la méthode ASI, ceux-ci sont quasi-stationnaires dans le temps. Les seules différences entre la méthode INK et la méthode ASI ont été observées dans des cas "atypiques", où les structures observées sur les champs d'ozone s'écartaient fortement des structures "usuelles" (statistiquement déterminées à l'aide des covariances des concentrations simulées). C'est pourquoi, si le réseau est suffisamment dense (i.e. si, à chaque heure, il y a assez de mesures pour construire les modèles de semivariances), nous conseillons plutôt aux personnes intéressées d'utiliser la méthode INK pour produire des analyses.

Ces méthodes ont été utilisées pour produire les analyses des concentrations d'ozone, et de dioxyde d'azote sur quatre étés (1999, 2000, 2001, et 2002). Elles ont été objectivement validées et testées à l'échelle régionale (Île-de-France et région de Berre et de Marseille), comme à l'échelle continentale. Les données d'observations en surface proviennent de différents réseaux de la qualité de l'air existant en Europe. La description, la validation, et la comparaison de ces méthodes ont été en particulier effectuées sur l'Île-de-France dans le cas de l'ozone, et sont détaillées dans un article soumis au *Journal of Geophysical Research* [Blond et al., soumis].

Qualité des analyses

L'utilisation combinée des sorties d'un modèle numérique et d'observations de surface permet d'améliorer la représentation des champs de concentration des polluants. Nous avons montré qu'il est alors possible de représenter les pics de concentrations d'ozone en surface à moins de 7 ppb près à l'échelle régionale, sur l'Île-de-France, le nombre d'erreurs supérieures étant alors limité à 2%.

Alors que le modèle régional simulait beaucoup moins correctement les concentrations du dioxyde d'azote sur l'Île-de-France, les analyses se sont montrées très performantes. En effet, nous

avons montré qu'il est également possible de représenter les pics de concentration du polluant à moins de 7 ppb. Le nombre d'erreurs supérieures commises par les analyses est alors inférieur à 5%. De la même manière, les résultats préliminaires, présentés sur la région de Berre et de Marseille, sont très encourageants. Dans tous les cas, la variabilité des polluants est bien représentée par les analyses.

L'application directe de ces méthodes d'analyse est la production journalière de cartes de concentration de polluant à la surface pour aider à la surveillance de la qualité de l'air. Puisque la méthode INK semble s'adapter plus facilement aux situations extrêmes que la méthode ASI, nous suggérons d'utiliser cette méthode pour ce genre d'application. Cependant il faut garder en mémoire que celle-ci est une méthode isotrope, et que des résultats différents pourraient être obtenus dans des régions présentant de fortes hétérogénéités de terrain (côtes, montagnes ...) ou d'émissions.

7.1.3 La prévision

Les analyses tridimensionnelles des champs de concentrations d'ozone, produites (par la méthode ASI) à l'échelle continentale ont également été utilisées pour initialiser le modèle de prévision à 15 h UTC. La question a été de savoir si nous pouvions améliorer la prévision des pointes de pollution au moyen des techniques d'assimilation de données.

Nous avons montré que l'assimilation de données ne résout qu'une petite partie du problème de prévision. En effet, l'impact des analyses est visible jusqu'à 24 h après l'initialisation, incluant l'heure du pic du lendemain. Cependant cet impact est assez faible. Plusieurs raisons peuvent expliquer ces résultats. La première raison est que la qualité des prévisions du temps chimiques ne dépend pas uniquement de celle des conditions initiales. En effet, ces prévisions sont aussi sensibles à d'autres paramètres : émissions, aux paramètres météorologiques ... Les précurseurs devraient être corrigés, à l'heure des fortes émissions, c'est à dire dans la matinée. La deuxième raison tient peut-être au manque d'information en altitude. La correction en surface de l'ozone n'est pas "conservée" dans la couche résiduelle, mais "diluée" pendant la nuit dans la couche résiduelle. Une question reste donc ouverte : ce serait-il possible d'améliorer les prévisions à court terme si nous disposions d'observations en altitude, telles que des données satellitales ?

7.2 Perspectives

Sachant que les analyses construites sont de meilleures représentations des champs de concentrations d'ozone que les simulations, les perspectives pratiques sont diverses et variées (diagnostique et étude de la pollution, modélisation inverse pour aider à la détermination des sources de polluants primaires, validation d'autres modèles, ...). Des études climatologiques de la pollution peuvent par exemple être menées. Nous avons quantifié la qualité de ces analyses, il sera donc toujours possible d'associer une barre d'erreur à ces estimations.

Vers l'étude de la pollution de proximité

Les prévisions d'un modèle de la qualité de l'air sont particulièrement sensibles aux émissions. Cependant, sans la possibilité de disposer de cadastres d'émissions plus détaillés, les modèles numériques utilisent en entrée des bases de données d'émissions passées, parfois même de plusieurs années. En outre, ces émissions sont souvent accessibles sous forme de totaux annuels et nécessitent donc d'être modulées en temps en utilisant des profils type. Ainsi, malgré la forte implication et la variabilité des émissions dues au transport, il n'y a aucun lien direct et permanent entre les émissions réelles du trafic et celles utilisées en entrée du modèle numérique. Nous avons montré que les concentrations du dioxyde d'azote sont difficiles à modéliser correctement, en particulier dû à leur lien direct avec ces émissions. La méthode ASI améliore la représentation de ce polluant, cependant il est primordial d'améliorer les estimations des émissions.

Le projet européen HEAVEN (<http://heaven.rec.org>) vise précisément à évaluer et à gérer les émissions liées au trafic dans plusieurs villes. Il est prévu de coupler une description du trafic en temps réel (comptage automatique des véhicules) avec un modèle numérique. Le produit final devrait fournir en temps *quasi* réel des cartes d'émissions des polluants primaires, et des cartes de champs de concentration des autres polluants. Ces cartes permettront probablement d'améliorer les simulations numériques et d'étudier de façon plus détaillée la pollution de proximité. Elles pourront servir, à termes, à mettre en place des stratégies de gestion du transport (restrictions de circulation, limitation de vitesse, amélioration du parc roulant).

Vers la détermination et la quantification des sources d'émission

Les expériences d'assimilation variationnelles menées par *Elbern et al.* [1997], *Elbern et Schmidt* [1999] montrent que si les observations sont en nombre suffisamment grand, il est possible de retrouver les sources d'émissions et d'optimiser leur intensité dans un modèle. Ces expériences montrent aussi que, lorsque le nombre d'observations est faible, la méthode 4D-VAR ne permet d'obtenir une bonne estimation des précurseurs que si l'ébauche est suffisamment proche de l'état réel. Ces résultats sont très encourageants pour l'avenir, si un jour des mesures satellitaires de données troposphériques sont disponibles. Il serait très intéressant d'examiner l'impact des analyses sur la détermination des sources et de leur intensité.

Vers la prévision d'ensemble

L'expérience de prévision avec initialisation par des analyses, que nous avons réalisée, repose sur l'hypothèse que les covariances d'erreur sont stationnaires dans le temps. Or nous savons aujourd'hui que ces covariances évoluent dans le temps. C'est pourquoi, il serait intéressant de développer un outil qui permette de modéliser l'évolution de ces covariances. Le filtre d'ensemble semble être un outil prometteur de prévision. Les méthodes proposées dans ce document peuvent être adaptées pour construire des cartes perturbées respectant les structures de corrélations d'erreur des champs simulés et ainsi des *ensembles statistiques* d'état initiaux.

Annexes

Annexe A

La pollution photochimique troposphérique

L'atmosphère terrestre peut être découpée en plusieurs couches verticales en fonction du gradient vertical de température observé. La troposphère est la partie de l'atmosphère la plus proche du sol, située entre la surface et environ 10 km d'altitude (son épaisseur dépend de la latitude). Elle est caractérisée par un gradient vertical négatif de température, c'est à dire que la température diminue avec l'altitude. La stratosphère, partie de l'atmosphère située au dessus de la troposphère, est caractérisée par un gradient de température positif. La troposphère se divise elle-même en la couche limite atmosphérique (CLA) et la troposphère libre [Stull, 1988], la CLA étant la partie de l'atmosphère directement sous l'influence des processus de surface.

Alors que la couche d'ozone (située en moyenne vers 25 km d'altitude) est bénéfique dans la stratosphère, la présence d'ozone dans l'air troposphérique est nocive. La chimie de production de l'ozone dans la CLA fait appel à un nombre élevé de réactions [Bey, 1997]. Ces réactions comprennent la photolyse du dioxyde d'azote (NO_2) et de l'ozone (O_3), la photo-oxydation du monoxyde de carbone (CO) et des Composés Organiques Volatils (COV) en présence d'oxydes d'azote (NO_x) et de radicaux hydroxyyles OH et hydroperoxyyles HO_2 , la titration de l'ozone par le monoxyde d'azote. Nous rappelons dans cette annexe, les caractéristiques de quelques espèces (ozone, NO_x , COV) et les principales étapes, qui interviennent dans la formation de l'ozone : l'équilibre photochimique (processus rapide) et le cycle des radicaux (processus plus lent).

A.1 Quelques espèces chimiques

A.1.1 L'ozone

La stratosphère contient environ 90% du contenu total de l'atmosphère en ozone. A cette altitude, l'ozone filtre les rayons ultraviolets les plus nocifs et protège ainsi notre peau. Dans la troposphère, qui contient les 10% restant, l'ozone est l'un des principaux polluants de la pollution dite "photo-oxydante". Le préfixe "photo" indique que cette pollution nécessite de la lumière. Elle est "oxydante" car elle augmente le pouvoir oxydant de l'atmosphère via la formation de composés réactifs oxydants tels que l'ozone et les espèces radicalaires. L'ozone est un gaz agressif pour les voies respiratoires et un irritant pulmonaire. Ces effets, très variables selon les individus,

sont surtout remarqués chez les personnes les plus sensibles (les enfants, les personnes âgées, les asthmatiques en général). Il a également un impact négatif sur la végétation, sa présence dans l'atmosphère entraînant notamment une diminution des rendements agricoles. À un moindre degré, il contribue à l'effet de serre.

L'ozone est un polluant secondaire, qui fait la une des journaux pendant l'été. Autrement dit, il n'existe pas d'émetteur direct d'ozone. Il résulte de la transformation chimique d'autres espèces présentes dans l'atmosphère (essentiellement des oxydes d'azote NO_x et COV). Les pics d'ozone sont fréquents par fort ensoleillement et forte chaleur et sont souvent localisés en dehors des zones urbaines. Ils apparaissent en début d'après midi. La durée de vie de l'ozone dépend de son altitude et varie entre quelques secondes et quelques jours. Plus l'ozone est en altitude, plus il peut être transporté loin. Parmi les épisodes de pollution par l'ozone, nous distinguons les épisodes de pollution locale des épisodes de pollution continentale. Les premiers sont caractérisés par l'arrivée de masses d'air "propre" sur une agglomération, et par la formation d'un panache en aval. Les seconds sont caractérisés par l'arrivée de masses d'air déjà chargées en ozone.

A.1.2 Les oxydes d'azote

Les oxydes d'azote sont principalement produits sous forme de monoxyde d'azote lors des processus de combustion sous haute température. Le monoxyde d'azote est ensuite très rapidement transformé en dioxyde d'azote par des oxydants (ozone, radicaux). Ils contribuent au phénomène des retombées acides via la formation et le dépôt d'acide nitrique. Ces composés sont à la base de la formation des *smog*, et sont, comme l'ozone, des irritants respiratoires.

Le dioxyde d'azote est un polluant plus problématique en hiver qu'en été. Cependant il se manifeste aussi en été sous forme de pics matinaux. C'est un polluant qui a une courte durée de vie (quelques secondes à quelques heures) et qui est ainsi très localisé autour des sources de monoxyde d'azote.

A.1.3 Les composés organiques volatils

Les Composés Organiques Volatils (COV) sont multiples. Il s'agit d'hydrocarbures émis par évaporation des bacs de stockages pétroliers ou remplissage des réservoirs automobiles. Ce sont aussi des composés organiques provenant de procédés industriels ou de combustions incomplètes. Beaucoup de solvants sont également émis lors de l'application des peintures, des encres, le nettoyage des surfaces métalliques et des vêtements. Ils sont aussi émis par l'agriculture et le milieu naturel. Les effets sont très divers selon les polluants : ils vont de la simple gêne olfactive à une irritation (aldéhydes) et à une diminution de la capacité respiratoire, jusqu'à des risques cancérogènes (benzène).

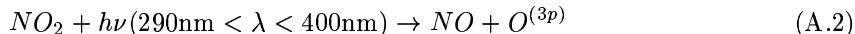
Les COV interviennent dans le processus de formation d'ozone dans la basse atmosphère.

A.2 Équilibre photochimique

La formation d'ozone résulte de la combinaison d'un atome d'oxygène dans l'état fondamental (O^(3p)) avec une molécule de dioxygène :



Dans la troposphère, la source notable d'atomes d'oxygène $O^{(3p)}$ est la photodissociation du dioxyde d'azote NO_2 aux longueurs d'ondes du visible et du proche ultraviolet (inférieures à 400 nm).



Le monoxyde ainsi formé réagit rapidement avec l'ozone (nous parlons alors de *titration* de l'ozone par le monoxyde d'azote) pour reformer du dioxyde d'azote. Un état stationnaire s'établit ainsi entre NO_2 , NO et O_3 .



En considérant des concentrations égales de NO et NO_2 , ce seul *équilibre photochimique* impliquerait des concentrations d'ozone de l'ordre de 20 ppb. Or, en dehors des zones urbaines, les concentrations d'ozone observées sont de l'ordre de 40 ppb les jours très nuageux, et certains jours d'été très ensoleillés ces concentrations peuvent très souvent dépasser les 100 ppb. Un tel équilibre ne suffit donc pas à expliquer la production d'ozone.

A.3 Production nette d'ozone

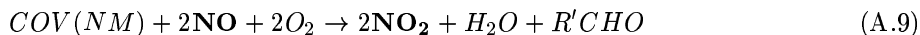
La production nette d'ozone dans la CLA, en région rurale ou en région urbaine provient essentiellement d'un cycle de radicaux à trois réactions dont voici une brève description. Une analyse plus complète des mécanismes peut-être trouvée dans [Toupance *et al.*, 1988].

A.3.1 Cycle des radicaux

En région rurale et dans la CLA, les COV ainsi que le monoxyde de carbone (CO) et le méthane (CH_4) sont oxydés par le radical hydroxyle OH en présence de monoxyde d'azote NO et d'ensoleillement. En milieu urbain, la même réaction a lieu, exception faite pour les COV méthaniques et le CO car le temps de réaction de ces composés est trop lent par rapport au temps de passage d'une masse d'air sur une agglomération. Ainsi, seuls les Composés Organiques Volatiles Non Méthaniques (COVNM) sont oxydés. Le cycle d'oxydation d'un COV(NM) s'effectue généralement selon les réactions suivantes :



où R est un radical, et RO_2 est un peroxy-radical très réactif. Le bilan de ces réactions s'écrit :



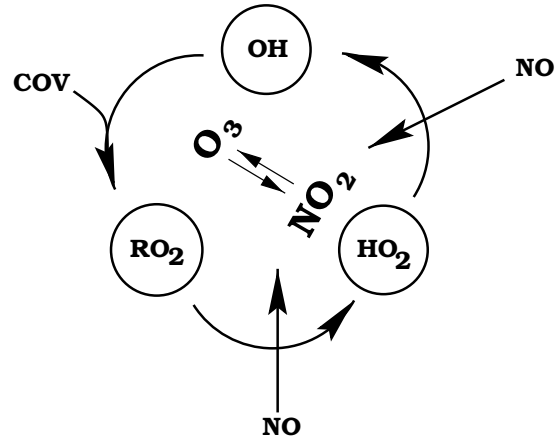


FIG. A.1 – Cycle d'oxydation des Composés Organiques Volatils

$R'CHO$ est un composé carbonylé qui peut à son tour être oxydé par OH . Par exemple, le formaldéhyde $HCHO$ se décompose en



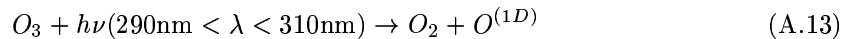
Enfin, le radical HO_2 réagit avec NO pour former du NO_2 .



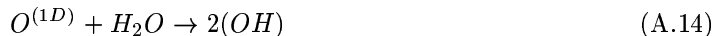
Nous soulignons que les trois radicaux (OH , RO_2 , HO_2) sont sans cesse recyclés au cours de ces réactions. C'est pourquoi nous parlons de *cycle d'oxydation des COV* ou *cycle des radicaux*. Ces radicaux ont pour effet de transformer NO en NO_2 , qui est ensuite photodissocié (suivant la réaction A.2). L'ozone est alors formé suivant la réaction A.1. Cette production est d'autant plus efficace que le composé organique volatil permet de cycles d'oxydation. Une molécule de méthane permet la création de 4 molécules d'ozone, une molécule de C_2H_6 de 6 molécules d'ozone, et une molécule comme C_5H_{12} de 12 molécules d'ozone. Les réactions sont résumées dans la figure A.1. La production d'ozone est donc subordonnée, par l'intermédiaire de NO_2 , à la présence de ces radicaux et à la quantité de NO disponible.

A.3.2 Puits et sources de radicaux

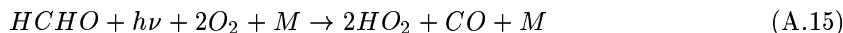
Malgré leur faible concentration (10^5 à 10^7 radicaux. cm^{-3}), l'ozone a des concentrations de l'ordre de 10^{12} molécules. cm^{-3}), les radicaux OH et HO_2 ont un rôle primordial dans la production de l'ozone troposphérique. Ces radicaux sont produits par différentes réactions dont la principale est la photolyse de l'ozone :



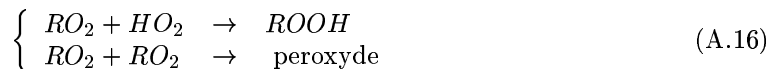
suivie de



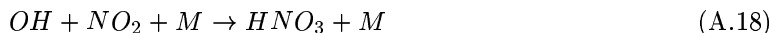
et la photolyse des composés carbonylés, comme par exemple le formaldéhyde



Enfin, les radicaux peuvent disparaître en suivant deux façons très distinctes, en réagissant entre eux,



ou avec un oxyde d'azote,



Ce sont ces deux voies de pertes qui distinguent essentiellement la photochimie d'un environnement pollué (riche en NO_x , souvent en site urbain) et celle d'un non pollué (pauvre en NO_x , souvent en site rural) [Kleinman, 1997].

A.3.3 Régimes chimiques

Il n'y a aucun doute que, pour la plupart des polluants, il existe à l'échelle d'une agglomération, comme Paris et sa banlieue, une relation directe entre émissions et concentrations. Néanmoins ce n'est pas le cas de l'ozone. Pour une situation météorologique estivale donnée, une augmentation (resp. une réduction) des concentrations de polluants primaires ne conduira pas nécessairement à une augmentation (resp. une réduction) des concentrations d'ozone.

En environnement non pollué, la perte de radicaux est dominée par les premières réactions A.16. Ainsi les réactions A.17 influent peu et une augmentation des émissions de NO va activer la formation d'ozone. La production d'ozone sera directement proportionnelle à la concentration de NO_x . En environnement pollué, comme c'est le cas dans une agglomération, les réactions A.17 contrôlent la perte de radicaux. Ainsi une augmentation des émissions de NO peut au contraire réduire la formation d'ozone car le cycle des radicaux est bridé par cette perte, sauf si les émissions de COVNM augmentent en proportion. Ainsi la production d'ozone est sensible à la composition des émissions et plus précisément au rapport des émissions COVNM/ NO_x [Honoré, 2000; Honoré et al. 2000]. C'est pourquoi, c'est la composition du mélange des émissions plus que leur intensité qui est cruciale pour la formation d'ozone. En fonction du milieu où l'on se place (urbain/rural), différents régimes chimiques sont observés. Ils peuvent s'exprimer efficacement en termes de comparaison des concentrations de COV et de NO_x [Sillman, 1999]. Une étude approfondie de la sensibilité de ces régimes peut être trouvée dans [Vautard et al., 1997] et [Menut et al., 2000].

A.4 Conditions favorables aux épisodes de pollution photochimique

Les mesures de polluants, tels que l'ozone et les oxydes d'azote, montrent une tendance assez stationnaire de leur concentration moyenne, mais présente une grande variabilité d'un jour à l'autre, voire d'une année à l'autre. Cette variabilité résulte de la variabilité naturelle des conditions météorologiques. En effet, les émissions anthropiques (dues à l'activité humaine) et biogéniques (dues aux processus naturels) de polluants ont un cycle diurne assez stable d'un jour à l'autre. Un jour très pollué se caractérise surtout par des conditions météorologiques assez particulières.

Les situations anticycloniques persistantes accompagnées d'une absence de vent au sol (inférieur à 2 m/s) et d'un ciel clair ou encore les situations de forte inversion de température se traduisent par une accumulation importante des polluants au-dessus d'une l'agglomération. En revanche, les situations, à caractère instable, caractérisées par du vent fort et des précipitations favorisent la dispersion le brassage et le lessivage des polluants.

Située en plaine, l'agglomération parisienne bénéficie généralement d'un climat océanique pluvieux ou venteux (vents d'ouest dominants). Cela explique la rareté des situations de forte pollution : moins de 10 jours par an.

A.5 Les niveaux d'alerte

L'Île-de-France a été la première région à se doter, le 25 avril 1994, d'une procédure d'information et d'alerte du public en cas d'épisode de pollution atmosphérique. Elle a été modifiée et renforcée par l'arrêté interpréfectoral du 24 juin 1999, prenant en compte les dispositions de la loi sur l'air du 30 décembre 1996 et de ses décrets d'application. Cette procédure a été révisée par l'arrêté interpréfectoral du 12 juillet 2002. Elle concerne trois polluants : le dioxyde de soufre (SO_2), l'ozone (O_3), le dioxyde d'azote (NO_2). Elle comporte deux niveaux de gravité croissante.

Les seuils des différents polluants intervenant dans la procédure d'information et d'alerte du public en cas d'épisode de pollution sont listés dans le tableau A.1.

A.5.1 Le niveau d'information et de recommandation

Ce niveau comprend des actions d'information de la population, des recommandations sanitaires aux catégories de la population particulièrement sensibles, ainsi que des recommandations et des mesures visant à réduire certaines des émissions polluantes. Ce niveau est déclenché lorsque le seuil d'information d'un des trois polluants est atteint. Chaque seuil d'information correspond à un niveau de concentration d'un polluant dans l'atmosphère au-delà duquel une exposition de courte durée a des effets limités et transitoires sur la santé des catégories de la population particulièrement sensibles (enfants, personnes âgées, asthmatiques et insuffisants respiratoires chroniques).

A.5.2 Le niveau d'alerte

Ce niveau comprend, outre les actions prévues au niveau d'information et de recommandation, des mesures de restriction ou de suspension des activités concourant à la pollution, y compris, le

cas échéant, de la circulation des véhicules. Il est déclenché lorsque le seuil d'alerte d'un des trois polluants est atteint ou risque de l'être.

Chaque Association Agréée pour la Surveillance de la Qualité de l'Air est chargée par délégation des Préfets des départements, de la détection des dépassements des différents seuils relatifs à cette procédure. Elles sont aussi chargées de la diffusion des communiqués de niveau d'information aux médias, ainsi qu'au grand public.

TAB. A.1 – *Seuils des différents polluants intervenant dans la procédure d'information et d'alerte du public en cas d'épisode de pollution en Île-de-France*

Polluant	Dioxyde d'azote NO ₂	Ozone O ₃	Dioxyde de soufre SO ₂
Niveau d'information et de recommandation	200 $\mu\text{g}/\text{m}^3$	180 $\mu\text{g}/\text{m}^3$	300 $\mu\text{g}/\text{m}^3$
Niveau d'alerte	400 $\mu\text{g}/\text{m}^3$ ou 200 $\mu\text{g}/\text{m}^3$ *	360 $\mu\text{g}/\text{m}^3$	500 $\mu\text{g}/\text{m}^3$ **

* Si la procédure d'information et de recommandation a été déclenchée la veille et le jour même et si les prévisions font craindre un nouveau risque de déclenchement pour le lendemain.

** dépassé pendant 3 heures consécutives.

Annexe B

Rappels de notions statistiques

Cette annexe est consacrée à la définition de quelques notions mathématiques de statistique utilisées abondamment tout au long de ce manuscrit.

B.1 L'espérance mathématique

L'état de pollution de l'atmosphère est défini à un instant t par les concentrations de toutes les espèces présentes dans l'air à cet instant. L'ensemble Ω de tous les états possibles de pollution de l'atmosphère étant supposé muni d'une certaine distribution de probabilité, chaque concentration Z est considérée comme une variable aléatoire. Une observation ou une valeur simulée z est *une réalisation* de la variable *concentration*.

Notons $E(Z)$ la moyenne statistique de la variable *concentration* Z , qui correspond à la moyenne d'un ensemble théoriquement infini de réalisations de Z . Cette moyenne est également appelée *l'espérance mathématique* et est définie par

$$E(Z) = \int_0^{\infty} zp(z)dz \quad (\text{B.1})$$

B.2 Approximation de l'espérance mathématique

Si la variable Z représente, par exemple, une concentration de polluant mesurée en un point et à un instant donné, nous ne disposons évidemment pas d'un ensemble statistique de mesures de Z (i.e ensemble d'observations effectuées au même moment et au même lieu), mais bien d'une seule et unique mesure, ou alors de plusieurs mesures effectuées au même instant en différents lieux. L'espérance mathématique, telle que définie ci-dessus, n'est donc pas calculable. Il faut donc définir un *ensemble statistique* sur lequel il sera possible d'approximer cette espérance par une moyenne algébrique. Deux solutions sont tour à tour proposées dans cette thèse pour approximer l'espérance mathématique d'une variable.

Une première possibilité consiste à calculer une moyenne dans le temps de Z . L'ensemble statistique choisi pour approximer $E(Z)$ est alors la succession des mesures effectuées en un point donné au cours d'un intervalle de temps fixé. Procéder ainsi, c'est supposer implicitement que

la grandeur Z est stationnaire, et que l'hypothèse d'ergodicité est valable. L'autre solution utilisée consiste à définir comme ensemble statistique, l'ensemble des mesures effectuées dans une région de l'espace, à un instant donné. Nous approximations alors l'espérance mathématique par une moyenne dans l'espace.

B.3 Variance et covariance

La variable $Z - E(Z)$, dont la moyenne statistique est nulle par définition, est dite centrée. Nous définissons la *variance* de la variable Z par $E((Z - E(Z))^2)$. Chaque variable est en plus supposée avoir une moyenne quadratique $E(Z^2)$ finie. On dit alors qu'elle appartient à l'ensemble $L^2(\Omega)$ des fonctions définies sur Ω de carré sommable. L'ensemble $L^2(\Omega)$ constitue un espace vectoriel. Étant données deux variables Z_i et Z_j appartenant à $L^2(\Omega)$, leur produit possède une moyenne finie $E(Z_i Z_j)$, et nous définissons leur covariance par $E((Z_i - E(Z_i))(Z_j - E(Z_j)))$. Le coefficient de corrélation de deux variables Z_i et Z_j et de variances non nulles est défini par

$$\frac{E((Z_i - E(Z_i))(Z_j - E(Z_j)))}{\sqrt{E((Z_i - E(Z_i))^2)E((Z_j - E(Z_j))^2)}} \quad (\text{B.2})$$

Un champ de concentration $Z(\mathbf{s})$, défini en tout point \mathbf{s} de l'espace, constitue une *fonction aléatoire*. Pour \mathbf{s} donné, $Z(\mathbf{s})$ est une variable aléatoire. La *fonction de covariance* de la fonction aléatoire Z est la fonction qui à tout couple de points $(\mathbf{s}_i, \mathbf{s}_j)$ de l'espace, fait correspondre la covariance statistique des variables aléatoires $Z(\mathbf{s}_i)$ et $Z(\mathbf{s}_j)$, soit $C(\mathbf{s}_i, \mathbf{s}_j) = E((Z(\mathbf{s}_i) - E(Z(\mathbf{s}_i)))(Z(\mathbf{s}_j) - E(Z(\mathbf{s}_j))))$. La fonction C est évidemment symétrique par rapport à ses deux arguments. Si $C(\mathbf{s}_i, \mathbf{s}_j)$ ne dépend que de la différence $\mathbf{s}_i - \mathbf{s}_j$, la fonction de covariance est dite *homogène*. Si $C(\mathbf{s}_i, \mathbf{s}_j)$ ne dépend que du module $\|\mathbf{s}_i - \mathbf{s}_j\|$, la fonction est dite *isotrope*.

Les fonctions de covariance possèdent de nombreuses propriétés mathématiques remarquables. Étant donnés n points de l'espace s_1, s_2, \dots, s_n et n coefficients scalaires $\lambda_1, \lambda_2, \dots, \lambda_n$, la variable aléatoire

$$y = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \quad (\text{B.3})$$

a pour moyenne quadratique

$$E(y^2) = E\left(\left(\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)\right)^2\right) = \sum_{i,j}^n \lambda_i \lambda_j E(Z(\mathbf{s}_i)Z(\mathbf{s}_j)) \quad (\text{B.4})$$

$$= \sum_{i,j}^n \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) \geq 0 \quad (\text{B.5})$$

La fonction C possède donc la propriété suivante : quel que soit le nombre n de points de l'espace $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$, la forme quadratique, $\sum_{i,j}^n \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j)$, définie par les $C(\mathbf{s}_i, \mathbf{s}_j)$ est semi-définie positive (i.e positive ou nulle). Les fonctions possédant cette propriété sont aussi semi-définies positives. Toutes les fonctions n'ayant pas forcément cette propriété, elles ne sont pas toutes acceptables pour modéliser des fonctions de covariances [Gaspari et Cohn, 1999].

Annexe C

Réseau européen de surveillance de la qualité de l'air

Depuis l'été 2000, une trentaine d'organismes européens se sont progressivement reliées au projet EUROPOLLUX, puis au projet PIONEER en 2001. Leur nom est listé dans les tableaux C.1 et C.2. Les données que nous utilisons dans ce document ont été gratuitement mises à disposition par ces organismes.

TAB. C.1 – Liste des associations pour la surveillance de la qualité de l'air en Europe

En France
AIRMARAIX (Provence)
AIRPARIF (Île-de-France)
AIR LANGUEDOC-ROUSSILLON (Languedoc Roussillon)
ATMO-POITOU-CHARENTES (Poitou-Charentes)
ATMOSFAIR (Bourgogne)
AIRFOBEP (Région Fos-Berre)
ATMOAUVERGNE (Auvergne) LIGAIR (Centre)
AIR PAYS DE LA LOIRE (Pays de La Loire)
ATMO-PICARDIE (Picardie)
ATMO-CHAMPAGNE-ARDENNES (Champagne-Ardenne)
COPARLY/AirDes2Savoies/ASCOPARG/AMPASEL/ASQUADRA/SUPAIRE
AIRBREIZH (Bretagne)
AIRAQ (Aquitaine)
AIR NORMAND (Haute Normandie)
AREMA, Lille (Nord)
OPALAIR (Nord)
ORAMIP (Midi-Pyrénées)
En Allemagne
UBA
En Suisse
NABEL/SAEFL
En Norvège
LML
En Italie
ARPA (Emilia-Romagna)
Au Royaume Unis
DETR
En Thèque République
Czech Hydrometeorological Institute
En Espagne
GENCAT

TAB. C.2 – Liste de 232 stations de mesure d'ozone localisées en Europe. À noter : N le niveau du modèle utilisé pour la comparaison, "typo" est la typologie (RUR=rurale, PUR=périurbaine, URB=urbaine de fond). Lorsque l'information est manquante, le symbole "****" est utilisé.

Nom	Organisme	Longi	Lati	Alti	N	Typo	Pays
PLAN-D-AUPS	AIRMARAIK	05.72	43.33	0690	2	RUR	FR
CADARACHE	AIRMARAIK	05.75	43.72	0300	1	RUR	FR
COMTAT-VENAISSIN	AIRMARAIK	05.03	44.10	****	1	PUR	FR
AIX-PLATANES	AIRMARAIK	05.46	43.56	****	1	PUR	FR
TOULON-OLIVE	AIRMARAIK	05.90	43.14	0180	1	PUR	FR
NANTES-JARDIN	AIR-PL	-1.54	47.22	0017	1	URB	FR
ANGERS-ALLARD	AIR-PL	-0.55	47.45	****	1	URB	FR
ST-NAZAIRE-BLUM	AIR-PL	-2.21	47.28	0007	1	URB	FR
LE-MANS-BELAIR	AIR-PL	00.21	48.01	****	1	URB	FR
CHOLET-ST-EXUPERY	AIR-PL	-0.88	47.06	****	1	URB	FR
LAVAL	AIR-PL	-0.75	48.08	****	1	URB	FR
LA-ROCHE-SUR-YON	AIR-PL	-1.35	46.65	****	1	URB	FR
ST-GOBAIN-SINCENY	ATMO-PICARDIE	03.28	49.58	****	1	RUR	FR
CRECY-EN-PONTHIEU	ATMO-PICARDIE	01.83	50.26	0037	1	RUR	FR
CHATEAU-THIERRY	ATMO-PICARDIE	03.40	49.04	****	1	***	FR
ALBERT	ATMO-PICARDIE	02.67	50.01	****	1	***	FR
FORET-DE-BROTONNE	AIRNORMAND	00.69	49.41	****	1	RUR	FR
EVREUX-ST-MICHEL	AIRNORMAND	01.15	49.04	****	1	URB	FR
PHARE-D-AILLY	AIRNORMAND	00.96	49.92	0010	1	RUR	FR
FREMAINVILLE	AIRPARIF	01.87	49.06	0122	1	RUR	FR
PRUNAY	AIRPARIF	01.67	48.86	0098	1	RUR	FR
SAINTS	AIRPARIF	03.06	48.77	0120	1	RUR	FR
NOISY-SUR-OISE	AIRPARIF	02.33	49.14	0080	1	RUR	FR
SONCHAMP	AIRPARIF	01.88	48.58	0148	1	RUR	FR
FONTAINEBLEAU	AIRPARIF	02.77	48.38	0127	1	RUR	FR
MONTGE-EN-GOELE	AIRPARIF	02.75	49.03	0162	1	RUR	FR
MONTPELLIER-NORD	AIR-LR	03.86	43.72	0080	1	PUR	FR
MONTPELLIER-SUD	AIR-LR	03.92	43.57	0030	1	PUR	FR
ST-ESTEVE	AIR-LR	02.84	42.72	0050	1	PUR	FR
NIMES-GARDIES	AIR-LR	04.31	43.84	0130	1	PUR	FR
SAZE	AIR-LR	04.68	43.94	0085	1	RUR	FR
VALLABREGUES	AIR-LR	04.64	43.83	0015	1	RUR	FR
NE-ALSACE	ASPA	08.16	48.92	0114	1	RUR	FR
VOSGES-MOYENNES	ASPA	07.13	48.49	0755	2	RUR	FR
VOSGES-NORD	ASPA	07.32	48.86	0330	2	RUR	FR
VOSGES-HAUTES	ASPA	07.20	48.22	1107	2	RUR	FR
3FRONTIERES	ASPA	07.57	47.60	0247	1	PUR	FR
MONTLUCON-HIPPODROME	ATMO-AUVERGNE	02.60	46.32	0205	1	PUR	FR
AURILLAC-AERODROME	ATMO-AUVERGNE	02.42	44.90	0640	1	PUR	FR
PUY-DE-DOME	ATMO-AUVERGNE	02.96	45.77	1460	3	RUR	FR
VALS-PRÉS-LE-PUY	ATMO-AUVERGNE	03.88	45.03	0600	1	PUR	FR
BRENNILIS	AIRBREIZH	-3.87	48.36	0160	1	RUR	FR
GUIPRY	AIRBREIZH	-1.88	47.83	0160	1	RUR	FR
BREST-JEAN-MACE	AIRBREIZH	-4.49	48.39	0075	1	URB	FR
RENNES-COURTEL	AIRBREIZH	-1.63	48.13	0160	1	URB	FR
LORIENT-VARLIN	AIRBREIZH	-3.39	47.74	0050	1	URB	FR
REVIN	ATMO-CHAMP-ARD	04.63	49.91	****	1	RUR	FR
ST-PARRES	ATMO-CHAMP-ARD	04.09	48.30	****	1	PUR	FR
REIMS-VAL-DE-MURIGNY	ATMO-CHAMP-ARD	04.02	49.23	0107	1	PUR	FR
HALLUIN	AREMALM	03.15	50.78	****	1	***	FR
CHAMBORD	LIGAIR	01.57	47.63	0071	1	RUR	FR
ORLEANS-LA-SOURCE	LIGAIR	01.93	47.90	0107	1	URB	FR
TOURS-DAMES	LIGAIR	00.76	47.38	0050	1	URB	FR
BOURGES-LEBLANC	LIGAIR	02.36	47.06	****	1	***	FR
CHARTRES-FULBERT	LIGAIR	01.47	48.44	0153	1	URB	FR
CHATEAUROUX-SUD	LIGAIR	01.68	46.81	****	1	***	FR
AIRVAULT	ATMO-POITOU-CH	-0.17	46.83	****	1	URB	FR
CHIZE	ATMO-POITOU-CH	-0.43	46.15	****	1	RUR	FR
LA-ROCHELLE-AYTRE	ATMO-POITOU-CH	-1.11	46.14	****	1	PUR	FR
L-ISLE-D-ESPAGNAC	ATMO-POITOU-CH	00.20	45.66	****	1	PUR	FR
POITIERS-CHASSENEUIL	ATMO-POITOU-CH	00.38	46.66	****	1	PUR	FR
NIORT-JEAN-ZAY	ATMO-POITOU-CH	-0.48	46.32	****	1	PUR	FR
COGNAC	ATMO-POITOU-CH	-0.32	45.70	****	1	URB	FR
FLOIRAC	AIRAQ	-0.53	44.84	0065	1	PUR	FR
LE-TEMPLE	AIRAQ	-0.94	44.87	0043	1	RUR	FR
PAU-LE-HAMEAU	AIRAQ	-0.32	43.31	0227	1	PUR	FR
BIARRITZ	AIRAQ	-1.55	43.45	0065	1	URB	FR
IRATY	AIRAQ	-1.04	43.03	1400	3	RUR	FR
DIJON-DAIX	ATMOSFAIR-N	05.00	47.35	****	2	PUR	FR
ST-BRISSON	ATMOSFAIR-N	04.09	47.27	****	1	RUR	FR
CHALON-HILAIRE	ATMOSFAIR-S	04.87	46.77	0182	1	PUR	FR
MONTCEAU-LES-MINES	ATMOSFAIR-S	04.37	46.68	0280	1	PUR	FR
MACON-CHAMPLEVERT	ATMOSFAIR-S	04.81	46.30	0213	1	PUR	FR
VALENCE-SUD	ASQUADRA	04.88	44.87	0110	1	PUR	FR
CHARAVINES	ASCOPARG	05.52	45.43	0491	1	RUR	FR
ST-ETIENNE-COUBERTIN	AMPASEL	04.39	45.47	0480	1	PUR	FR
DIEME	COPARLY	04.47	45.96	0540	2	RUR	FR
ANNEMASSE-GAILLARD	AIR2SAVOIES	06.22	46.19	0420	1	URB	FR
ANNECY-NOVEL	AIR2SAVOIES	06.14	45.92	0460	1	URB	FR
CHAMBERY-LE-HAUT	AIR2SAVOIES	05.92	45.60	0367	1	URB	FR
ROCHES-DE-CONDRIEU	SUPAIRE	04.77	45.45	0152	1	IND	FR

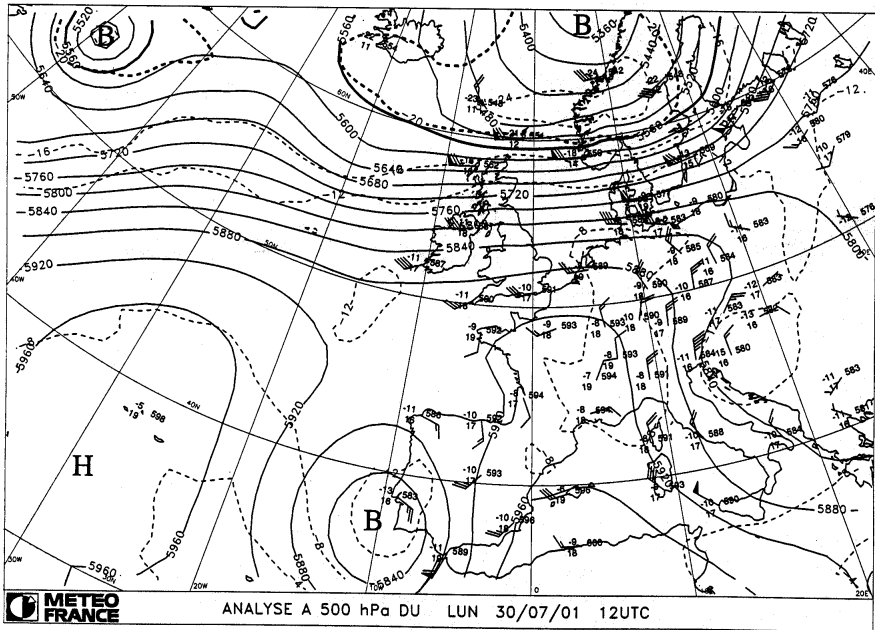
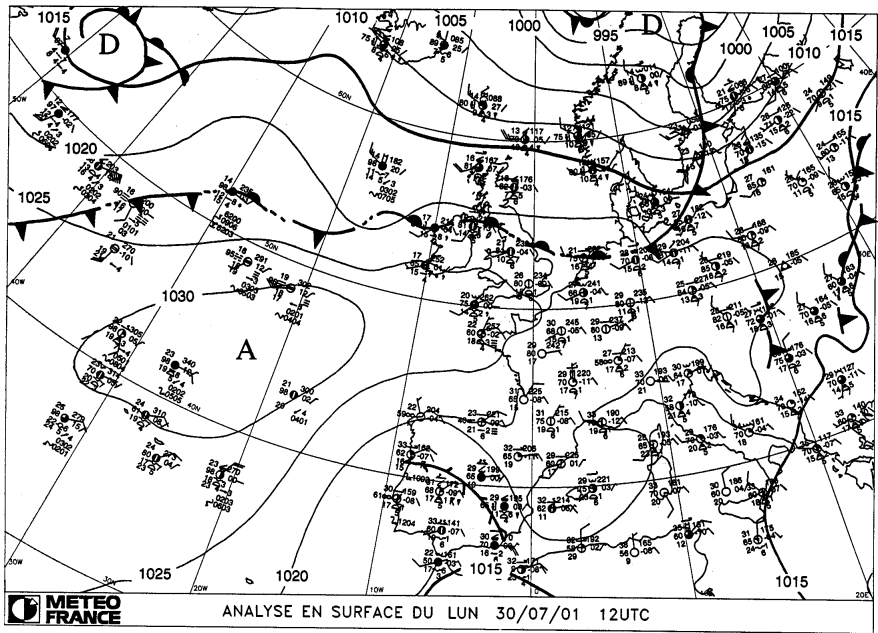
FORST	UBA	14.64	51.75	0075	1	PUR	DE
WITENBERGE	UBA	11.76	53.00	0022	1	PUR	DE
PREMNITZ	UBA	12.34	52.53	0030	1	PUR	DE
PRENZLAU	UBA	13.87	53.32	0030	1	PUR	DE
NAUEN	UBA	12.88	52.61	0032	1	PUR	DE
BERLIN-MARIENFELDE	UBA	13.37	52.40	0040	1	PUR	DE
BERLIN-FR. HAFEN	UBA	13.65	52.45	0035	1	PUR	DE
WIESLOCH	UBA	08.70	49.30	0160	1	PUR	DE
WEIL-AM-RHEIN	UBA	07.63	47.59	0250	1	PUR	DE
WELZHEIMER-WALD	UBA	09.57	48.89	0500	1	RUR	DE
SCHWARZWAL-SUED	UBA	07.77	47.81	0920	2	RUR	DE
FREUDENSTADT	UBA	08.41	48.47	0750	1	PUR	DE
KONSTANZ	UBA	09.17	47.67	0400	1	PUR	DE
BADEN-BADEN	UBA	08.22	48.77	0150	1	PUR	DE
SCHWAEBISCHE-ALB	UBA	09.21	48.35	0799	1	RUR	DE
BAD-WALDSEE	UBA	09.76	47.92	0610	1	PUR	DE
NEUSTADT-A. D. DONAU	UBA	11.78	48.85	0370	1	PUR	DE
BAD-REICHENHALL	UBA	12.86	47.72	0470	1	PUR	DE
GARM.-PART.-WANK	UBA	11.14	47.51	1776	3	RUR	DE
BREMERHAVEN	UBA	08.57	53.56	0003	1	PUR	DE
GREBENAU	UBA	09.47	50.76	0370	1	RUR	DE
WITZENHAUSEN	UBA	09.78	51.29	0605	1	RUR	DE
KOENIGSTEIN	UBA	08.44	50.20	0520	1	RUR	DE
SPESSART	UBA	09.40	50.16	0485	1	RUR	DE
FRANKENBERG	UBA	08.77	51.10	0395	1	RUR	DE
FUERTH-ODENWALD	UBA	08.82	49.65	0480	1	RUR	DE
BURG-HERZBERG	UBA	09.46	50.77	0509	1	RUR	DE
BAD-AROLSEN	UBA	08.93	51.43	0362	1	PUR	DE
LOECKNITZ	UBA	14.26	53.52	0021	1	RUR	DE
GOEHLEN	UBA	11.36	53.30	0025	1	RUR	DE
SOLLING-DASSEL	UBA	09.58	51.76	0500	1	RUR	DE
LINGEN	UBA	07.32	52.50	0030	1	PUR	DE
WURMBERG-BRAUNLAGE	UBA	10.62	51.76	0992	2	RUR	DE
NORDERNEY	UBA	07.16	53.70	0001	1	PUR	DE
CUXHAVEN	UBA	08.80	53.83	0003	1	RUR	DE
JORK	UBA	09.69	53.52	0003	1	RUR	DE
HERTEN	UBA	07.13	51.60	0102	1	PUR	DE
HUERTH	UBA	06.87	50.88	0090	1	PUR	DE
EGGEGBIRGE	UBA	08.95	51.83	0430	1	RUR	DE
EIFEL	UBA	06.28	50.65	0572	1	RUR	DE
ROTHAARGBIRGE	UBA	08.19	50.93	0635	1	RUR	DE
NETTETAL	UBA	06.20	51.33	0049	1	RUR	DE
NIEDERZIER	UBA	06.47	50.89	0105	1	RUR	DE
FINNENTROP	UBA	07.97	51.17	0310	1	RUR	DE
SOLINGEN	UBA	07.05	51.18	0207	1	PUR	DE
BORKEN-GEMEN	UBA	06.88	51.86	0045	1	PUR	DE
AACHEN-BURTSCHIED	UBA	06.09	50.76	0205	1	RUR	DE
WESTFALZ-DUNZWEILER	UBA	07.29	49.42	0455	1	RUR	DE
WESTEIFEL-WASCHEID	UBA	06.38	50.27	0680	1	RUR	DE
WESTERWALD-HERDORF	UBA	07.97	50.77	0480	1	RUR	DE
PFAELZERWALD-H	UBA	07.83	49.27	0606	1	RUR	DE
NEUSTADT-STROHMARKT	UBA	08.14	49.36	0138	1	PUR	DE
WESTERWALD-NEUHSL	UBA	07.73	50.43	0540	1	RUR	DE
ALTENDEICH	UBA	09.59	53.67	0001	1	RUR	DE
BRUNSBUETTTEL	UBA	09.23	53.91	0001	1	RUR	DE
FEHMARN	UBA	11.22	54.41	0002	1	RUR	DE
ST.-PETER-ORDING	UBA	08.62	54.33	0001	1	RUR	DE
GLUECKSBURG	UBA	09.57	54.84	0034	1	RUR	DE
SAARBRUECKEN	UBA	07.04	49.23	0300	1	PUR	DE
CARLSFELD	UBA	12.61	50.43	0896	2	RUR	DE
HOYERSWERDA	UBA	14.26	51.44	0117	1	PUR	DE
RADEBEUL-WAHNSDORF	UBA	13.68	51.12	0246	1	RUR	DE
ZINNWALD	UBA	13.75	50.73	0877	1	RUR	DE
FICHTELBERG	UBA	12.96	50.43	1214	2	RUR	DE
MITTELNDORF	UBA	14.21	50.94	0323	1	RUR	DE
COLLMBERG	UBA	13.01	51.30	0313	1	RUR	DE
WERNIGERODE	UBA	10.79	51.84	0230	1	PUR	DE
POUCH	UBA	12.39	51.63	0080	1	PUR	DE
HARZGERODE	UBA	11.15	51.66	0385	1	RUR	DE
NEUHAUS	UBA	11.14	50.50	0840	2	RUR	DE
POSSEN	UBA	10.87	51.35	0420	1	RUR	DE
WESTERLAND	UBA	08.31	54.93	0012	1	RUR	DE
DEUSELBACH	UBA	07.05	49.76	0480	1	RUR	DE
BROTHACKLRIEGEL	UBA	13.22	48.82	1016	2	RUR	DE
SCHAUINSLAND	UBA	07.91	47.91	1205	2	RUR	DE
WALDHOF	UBA	10.76	52.80	0074	1	RUR	DE
BASSUM	UBA	08.70	52.85	0052	1	RUR	DE
GITTRUP	UBA	07.67	52.04	0043	1	RUR	DE
REGNITZLOSAU	UBA	12.06	50.31	0595	1	RUR	DE
UECKERMUEENDE	UBA	14.07	53.75	0001	1	RUR	DE

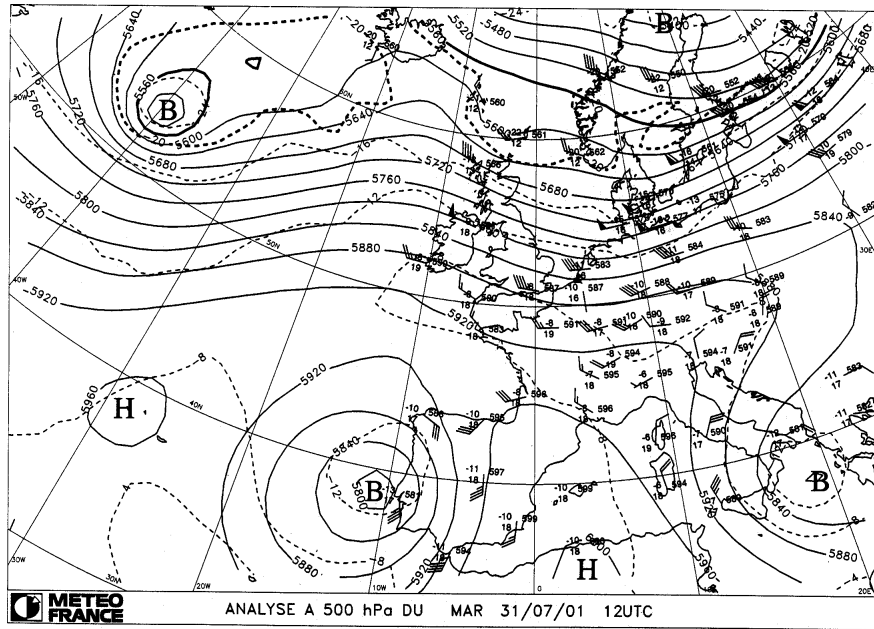
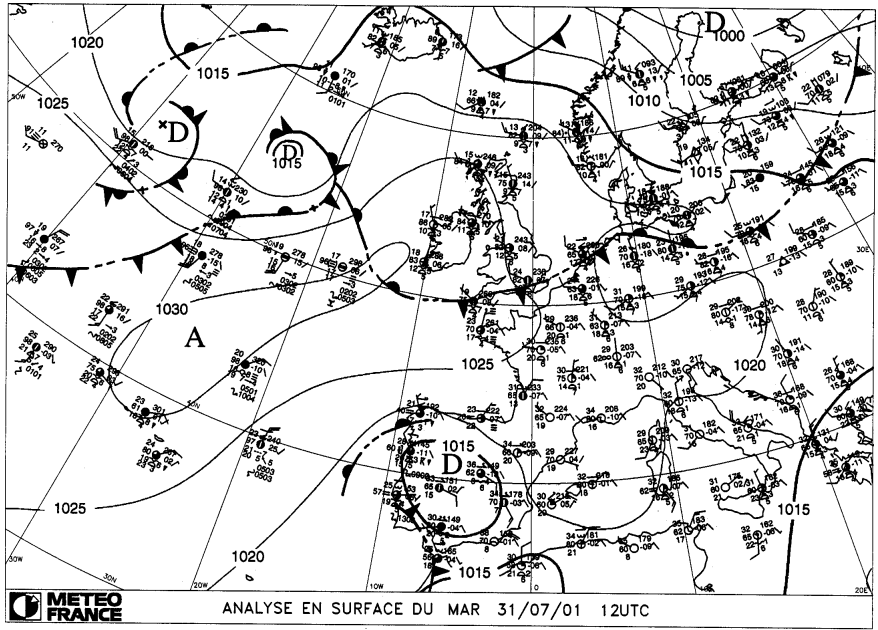
ZINGST	UBA	12.72	54.44	0001	1	RUR	DE
SCHMUECKE	UBA	10.77	50.66	0937	2	RUR	DE
NEUGLOBSOW	UBA	13.03	53.14	0065	1	RUR	DE
LUECKENDORF	UBA	14.79	50.83	0490	1	RUR	DE
LEINEFELDE	UBA	10.31	51.39	0356	1	RUR	DE
MELPITZ	UBA	12.93	51.53	0086	1	RUR	DE
HELGOLAND	UBA	07.88	54.19	0050	1	RUR	DE
LEHNMUEHLE	UBA	13.59	50.83	0527	1	RUR	DE
AUKRUG	UBA	09.79	54.07	0015	1	RUR	DE
FALKENBERG	UBA	14.12	52.17	0073	1	RUR	DE
SCHORFHEIDE	UBA	13.65	52.97	0070	1	RUR	DE
ASTON-HILL	DETR	-3.03	52.50	0370	1	RUR	UK
SOMERTON	DETR	-3.05	51.23	0055	1	RUR	UK
LADYBOWER	DETR	-1.75	53.40	0420	1	RUR	UK
LOUGH-NAVAR	DETR	-7.87	54.44	0130	1	RUR	UK
SIBTON	DETR	01.46	52.29	0046	1	RUR	UK
WICKEN-FEN	DETR	00.20	52.40	0005	1	RUR	UK
NARBERTH	DETR	-4.69	51.78	0160	1	RUR	UK
LULLINGTON-HEATH	DETR	00.18	50.79	0120	1	RUR	UK
HIGH-MUFFLES	DETR	-0.81	54.33	0267	1	RUR	UK
ESKDALEMUIR	DETR	-3.20	55.31	0269	1	RUR	UK
HARWELL	DETR	-1.32	51.57	0137	1	RUR	UK
YARNER-WOOD	DETR	-3.71	50.60	0119	1	RUR	UK
GLAZEBURY	DETR	-2.47	53.46	0021	1	PUR	UK
BUSH-ESTATE	DETR	-3.21	55.86	0180	1	RUR	UK
BOTTESFORD	DETR	-0.82	52.93	0032	1	RUR	UK
ROCHESTER	DETR	00.63	51.45	0001	1	RUR	UK
REDCAR	DETR	-1.07	54.61	0004	1	PUR	UK
NL-934	LML	06.28	53.33	0001	1	RUR	NL
NL-538	LML	05.05	52.81	0001	1	RUR	NL
NL-107	LML	06.04	51.12	0032	1	RUR	NL
NL-918	LML	05.57	52.92	0001	1	RUR	NL
NL-913	LML	06.80	53.14	0001	1	RUR	NL
NL-318	LML	03.75	51.30	0005	1	RUR	NL
NL-235	LML	04.36	51.44	0018	1	RUR	NL
NL-230	LML	05.15	51.52	0015	1	RUR	NL
NL-227	LML	05.56	51.27	0032	1	RUR	NL
NL-722	LML	06.65	52.10	0032	1	RUR	NL
NL-724	LML	05.67	51.97	0032	1	RUR	NL
NL-818	LML	05.90	52.65	0001	1	RUR	NL
NL-301	LML	03.93	51.66	0001	1	RUR	NL
NL-232	LML	05.67	51.65	0001	1	RUR	NL
NL-633	LML	04.85	52.23	0001	1	RUR	NL
PAYERNE	NABEL/SAEFL	06.94	46.80	0409	1	RUR	CH
CHAUMONT	NABEL/SAEFL	06.98	47.05	1140	3	RUR	CH
JUNGFRAUJOCH	NABEL/SAEFL	07.99	46.55	3580	5	RUR	CH
LAEGEREN	NABEL/SAEFL	08.35	47.48	0690	1	RUR	CH
RIGI	NABEL/SAEFL	08.47	47.07	1030	2	RUR	CH
TAENIKON	NABEL/SAEFL	08.90	47.47	0540	1	RUR	CH
MAGADINO	NABEL/SAEFL	08.93	46.16	0200	1	PUR	CH
DAVOS	NABEL/SAEFL	09.86	46.82	1640	3	RUR	CH
PARMA-CITTADELLA	ARPA/EMR	10.31	44.82	****	1	***	IT
MODENA-20-SETTEMBRE	ARPA/EMR	10.91	44.65	****	1	***	IT
BOLOGNA-CUCCOLINO	ARPA/EMR	11.34	44.45	****	1	***	IT
GHERARDI	ARPA/EMR	11.71	44.89	****	1	***	IT
COTIGNOLA	ARPA/EMR	12.25	44.50	****	1	***	IT
FORLI-GIARDINI	ARPA/EMR	12.02	44.22	****	1	***	IT
RIMINI-RICCIONE	ARPA/EMR	12.65	44.00	****	1	***	IT
CASTEL-SAN-GIOVANNI	ARPA/EMR	09.70	45.10	****	1	***	IT

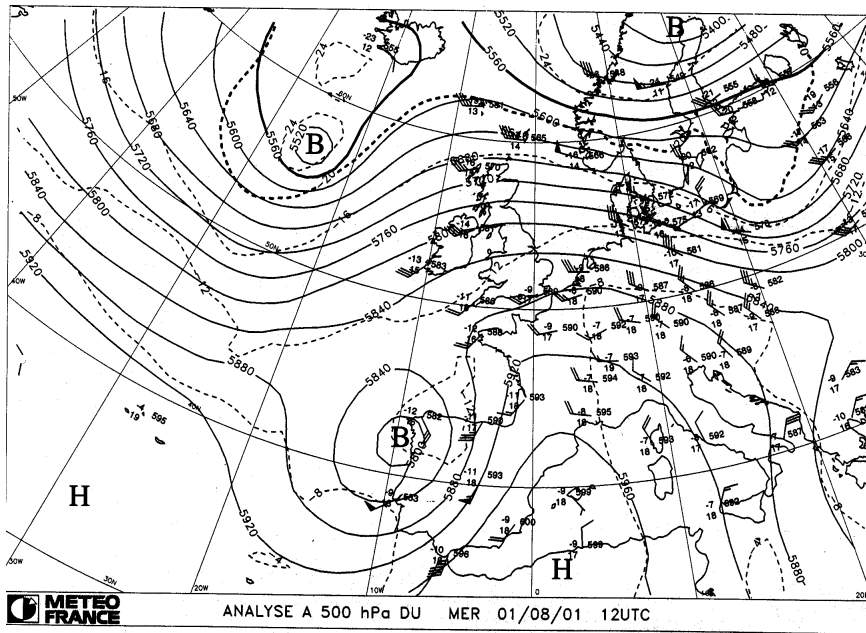
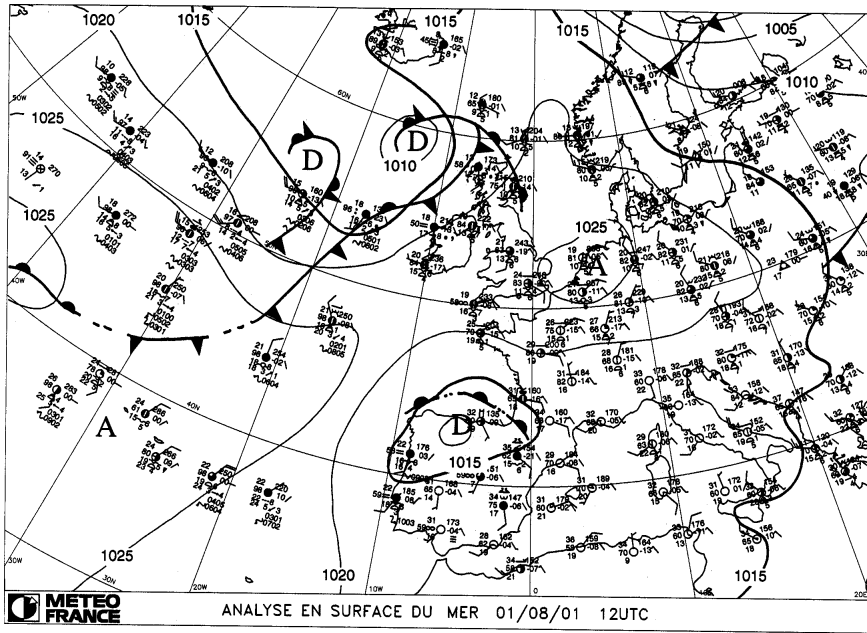
Annexe D

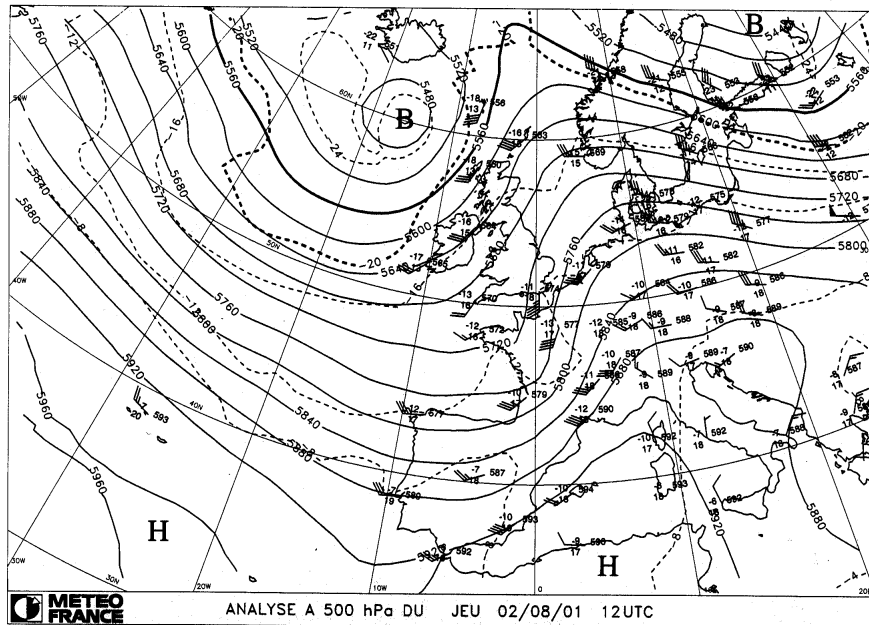
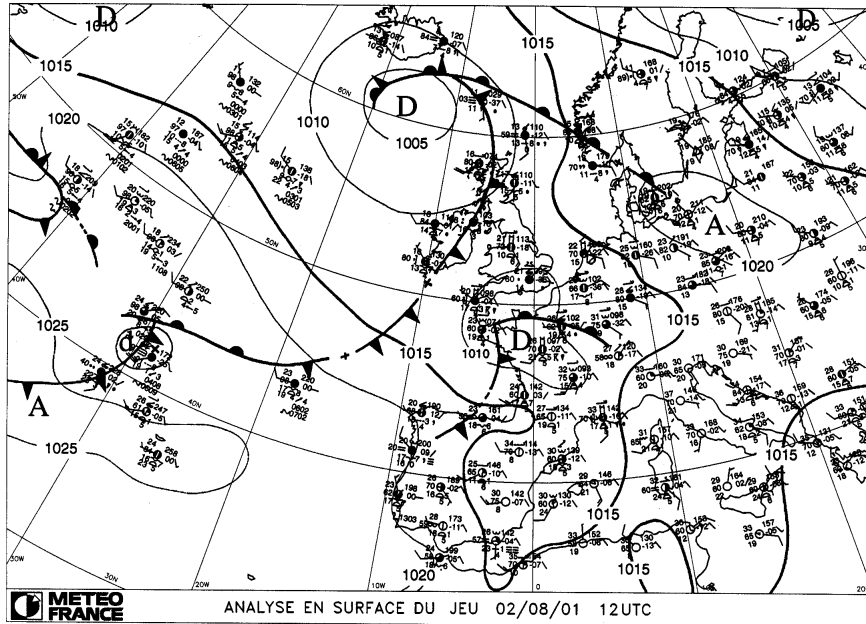
Analyses météorologiques

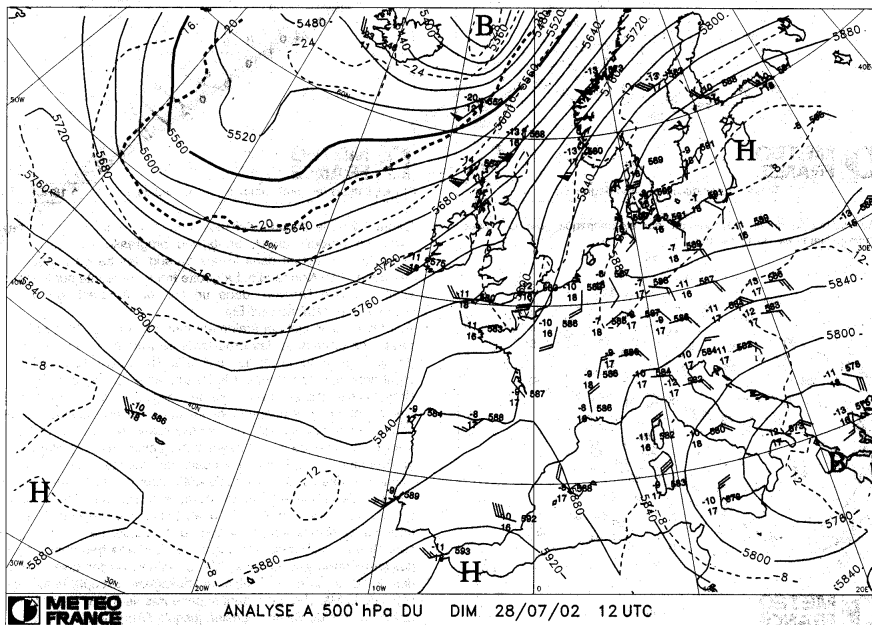
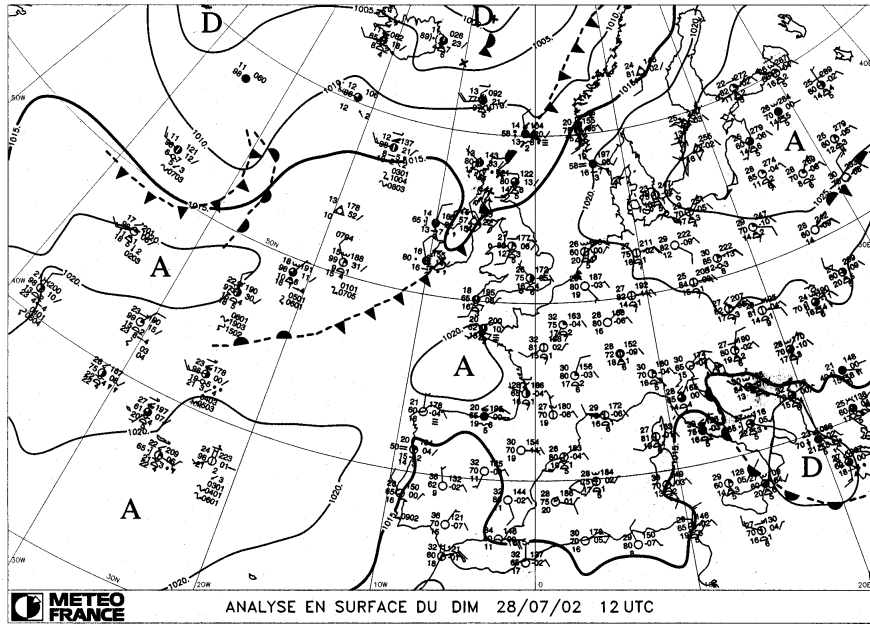
Les analyses météorologiques présentées dans cette annexe sont tirées du Bulletin hebdomadaire d'études et de renseignements *Météo.Hebdo*. Nous présentons les situations du 30 Juillet au 2 août 2001 [*Météo Hebdo*, 2001], et du 30 Juillet au 2 août 2002 [*Météo Hebdo*, 2002a, 2002b].

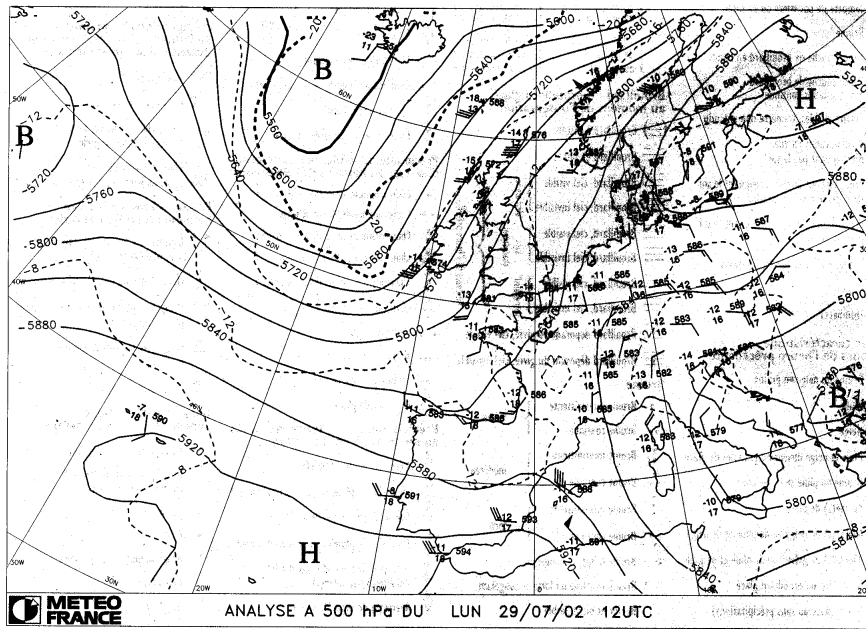
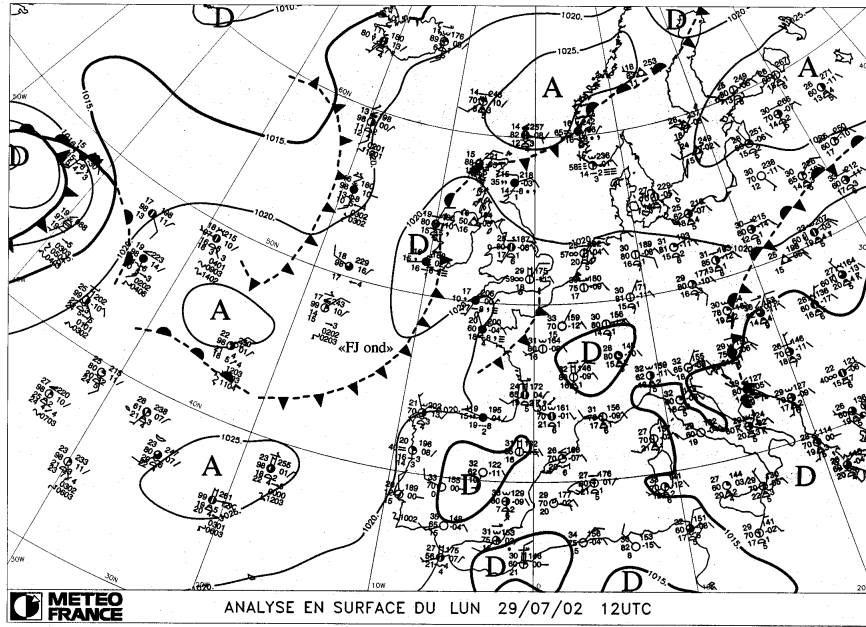


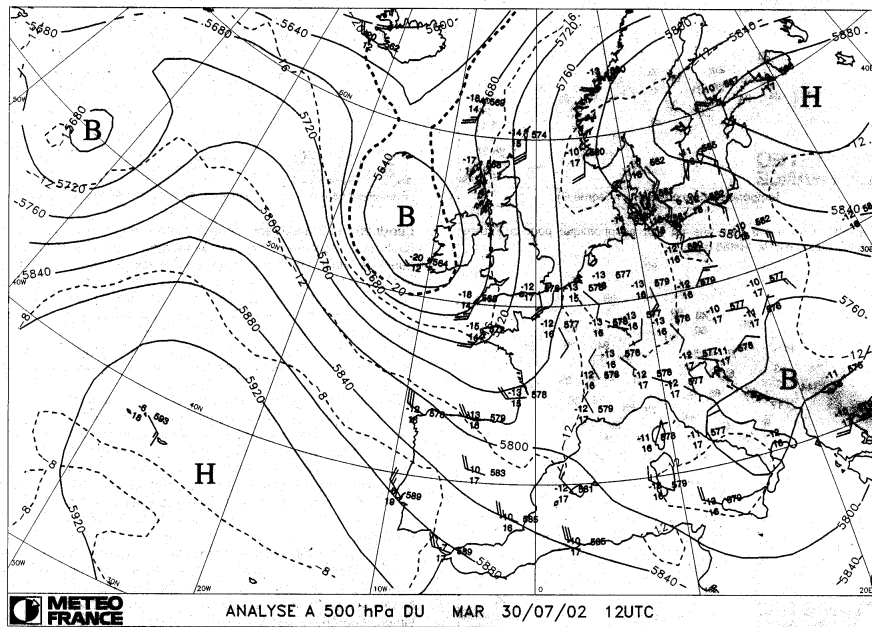
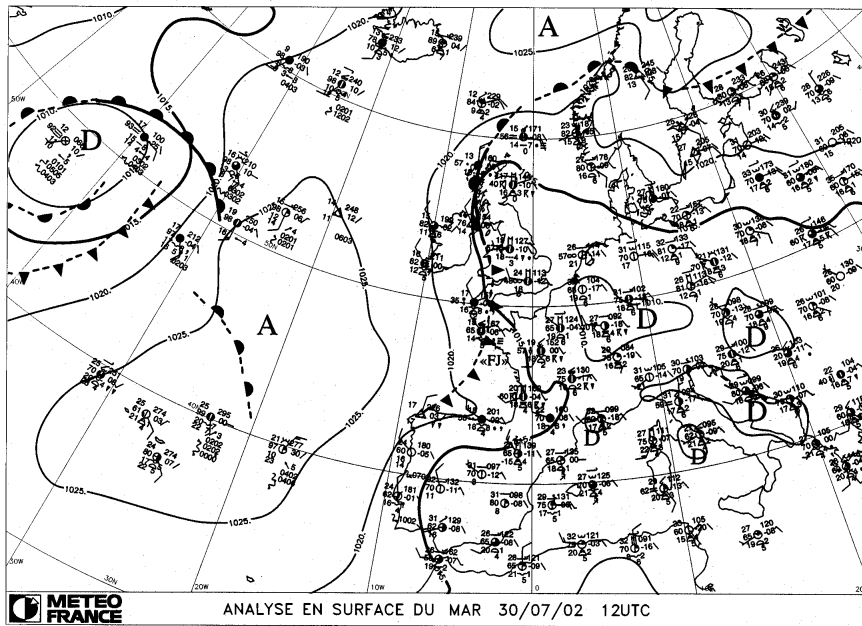


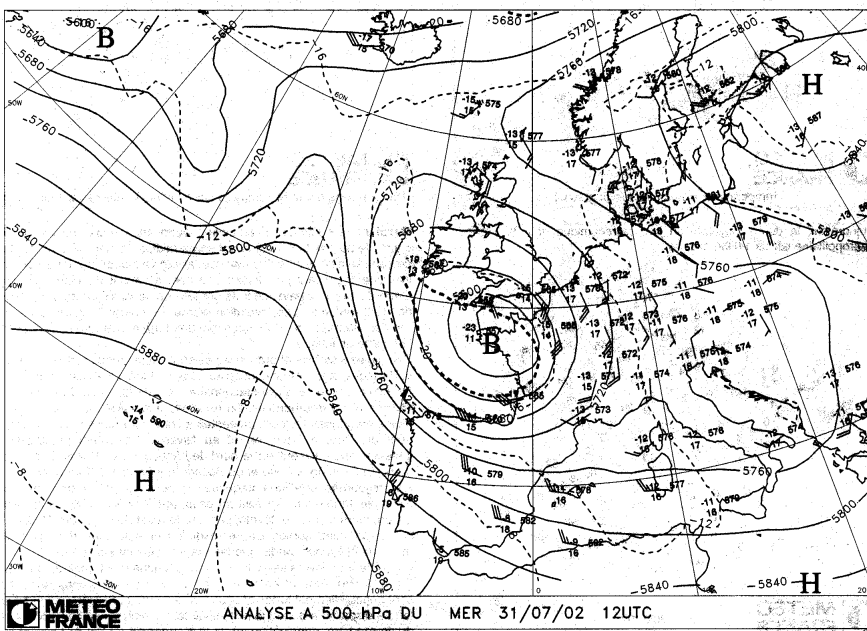
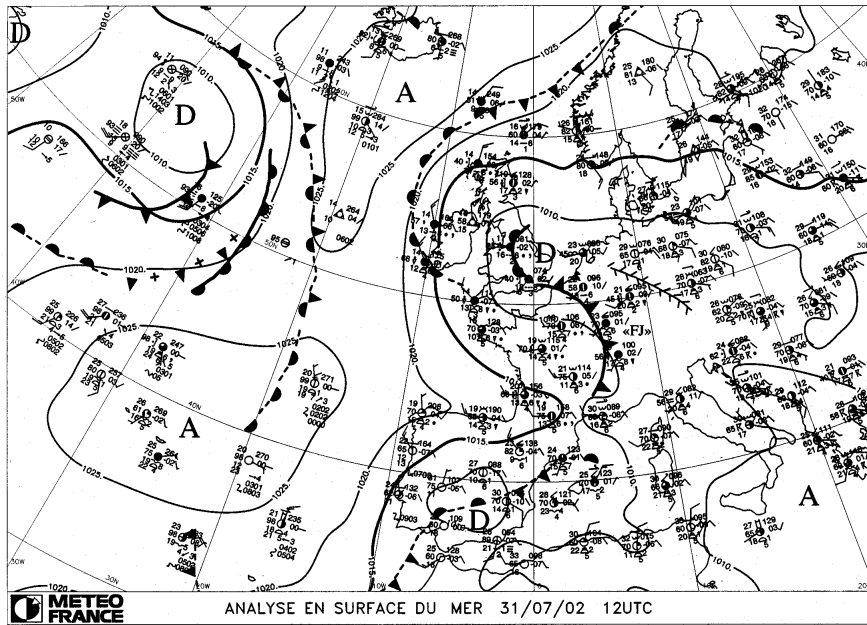


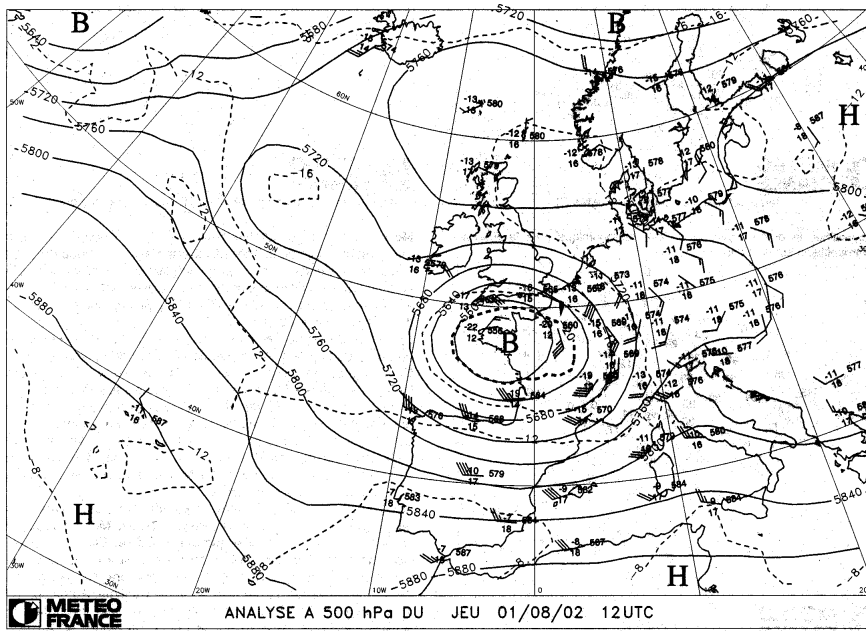
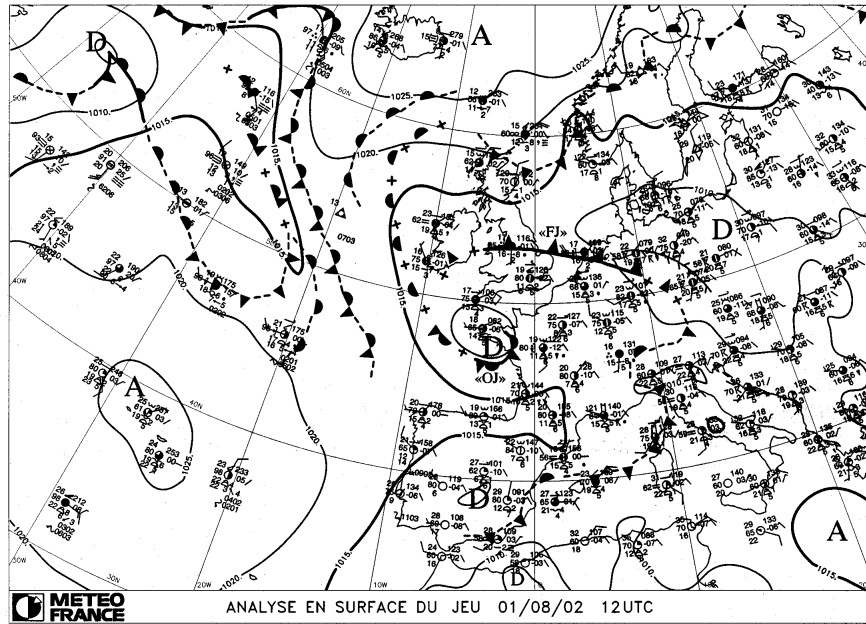


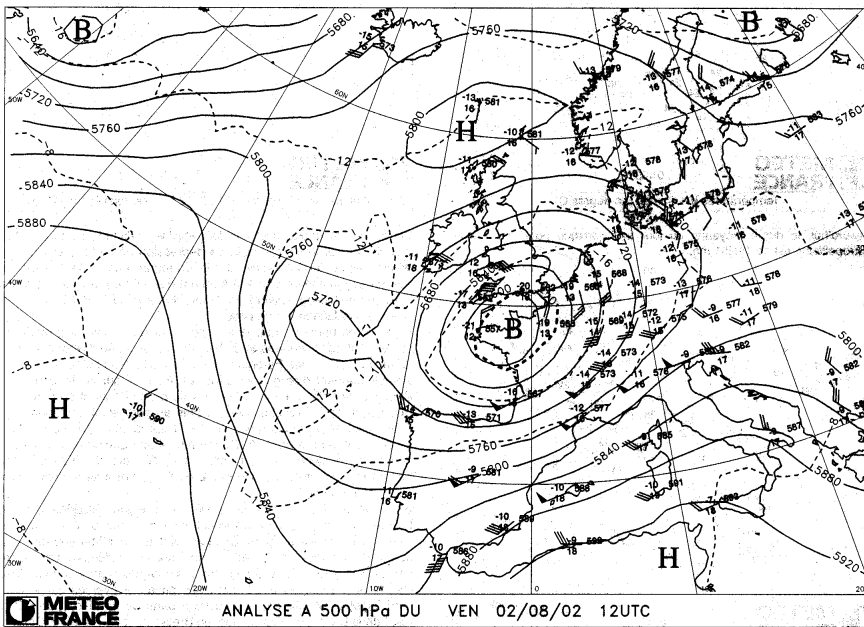
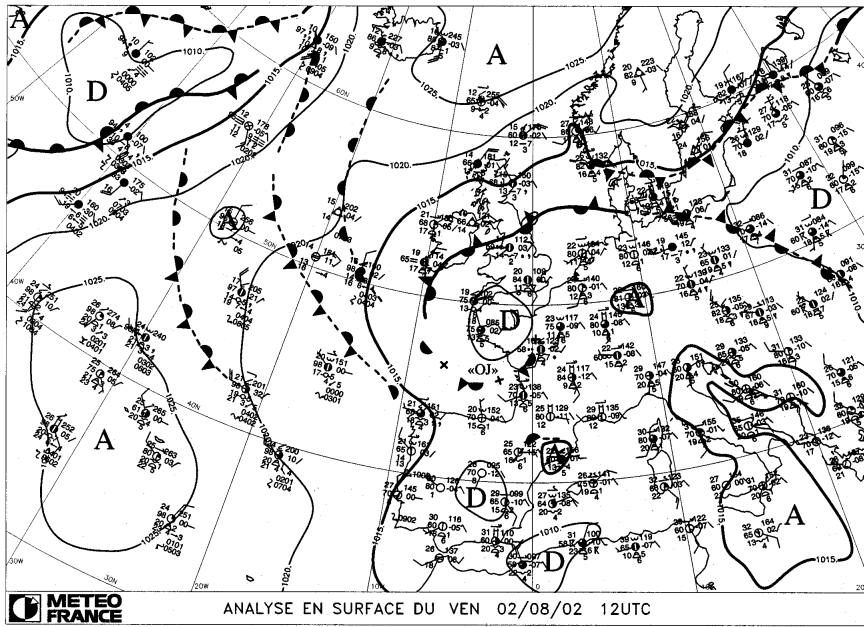




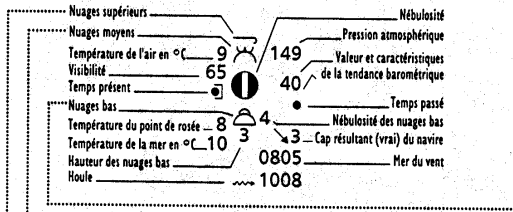








Notice explicative



Nuages de moyenne altitude

- ↳ **Altostratus** dont la majeure partie est semi-transparente et laisse apparaître faiblement le Soleil ou la Lune, comme au travers d'un verre dépoli.
- ↳ **Altostratus** dont la majeure partie est suffisamment dense pour masquer complètement le Soleil ou la Lune, ou nimbostratus.
- ~ **Alto cumulus** dont la majeure partie est semi-transparente : les divers éléments du nuage ne se modifient que lentement et sont tous situés à un seul niveau.
- ✓ **Alto cumulus en bancs** (souvent en forme de lentilles ou de poissons) dont la majeure partie est semi-transparente, ces bancs se présentent à un ou plusieurs niveaux et l'aspect de leurs éléments constitutifs se modifie constamment.
- ↳ **Alto cumulus semi-transparents en bandes**, ou alto cumulus en une ou plusieurs couches sensiblement continues (semi-transparentes ou opaques), envahissant progressivement le ciel : ces alto cumulus deviennent, en général, plus épais dans leur ensemble.
- ✕ **Alto cumulus** formés par l'étalement de cumulus (ou de cumulonimbus).
- ↳ **Alto cumulus** en deux ou plusieurs couches, généralement opaques par endroits et n'envahissant pas progressivement le ciel ; ou couche opaque d'alto cumulus, n'envahissant pas progressivement le ciel ; ou alto cumulus présents simultanément avec de l'altostratus ou du nimbostratus.
- ⌌ **Alto cumulus** présentant des bourgeonnements en forme de petites tours ou de créneaux, ou alto cumulus ayant l'aspect de flocons cumuliformes.
- ↳ **Alto cumulus** dans un ciel d'aspect chaotique ; ils sont généralement situés à plusieurs niveaux.

Nuages supérieurs

- ↳ **Cirrus** en forme de filaments, de brins ou de crochets, n'envahissant pas progressivement le ciel.
- ↳ **Cirrus denses**, en bancs ou en gerbes enchevêtrées qui, en général, n'augmentent pas et semblent être parfois les résidus de la partie supérieure d'un cumulonimbus ; ou cirrus présentant des bourgeonnements en forme de petites tours ou créneaux, ou cirrus ayant l'aspect de flocons cumuliformes.
- ↳ **Cirrus denses**, ayant souvent la forme d'enclume ; ces cirrus sont les résidus de parties supérieures de cumulonimbus.

- ↳ **Cirrus** en forme de crochets ou de filaments, ou les deux, envahissant progressivement le ciel ; ces nuages deviennent généralement plus denses dans leur ensemble.
- ↳ **Cirrus** (souvent en bandes convergeant vers un point ou vers deux points opposés de l'horizon) et cirrostratus, ou cirrostratus seul, dans les deux cas, ces nuages envahissent progressivement le ciel et deviennent généralement plus denses dans leur ensemble, mais le voile continu n'atteint pas 45 degrés au-dessus de l'horizon.
- ↳ **Cirrus** (souvent en bandes convergeant vers un point ou vers deux points opposés de l'horizon) et cirrostratus, ou cirrostratus seul, dans les deux cas, ces nuages envahissent progressivement le ciel et deviennent généralement plus denses dans leur ensemble ; le voile continu dépasse 45 degrés au-dessus de l'horizon, sans que le ciel soit totalement couvert.
- ↳ **Voile de cirrostratus** couvrant complètement la voûte céleste.
- ↳ **Cirrostratus** n'envahissant pas progressivement le ciel et ne couvrant pas complètement la voûte céleste.
- ↳ **Cirrocumulus seuls**, ou **cirrocumulus** coexistant avec des cirrus ou du cirrostratus, ou avec les deux, les cirrocumulus étant prédominants.

Nuages bas

- ↳ **Cumulus** à faible extension verticale et paraissant aplatis ou cumulus déchetés autres que de mauvais temps⁽¹⁾, ou les deux.
- ↳ **Cumulus** ayant une extension verticale modérée ou forte, avec généralement des protubérances en forme de dômes ou de tours, accompagnés ou non par d'autres cumulus ou par des stratocumulus, tous ayant leurs bases au même niveau.
- ↳ **Cumulonimbus** dont les sommets ont perdu, au moins partiellement, la netteté de leurs contours, mais qui ne sont ni nettement fibreux (cirriformes), ni en forme d'enclume : des cumulus, des stratocumulus ou des stratus peuvent également être présents.
- ↳ **Stratocumulus** formés par l'étalement de cumulus : des cumulus peuvent également être présents.
- ↳ **Stratocumulus** ne provenant pas de l'étalement de cumulus.
- ↳ **Stratus** en nappe ou en couche plus ou moins continue, ou en lambeaux déchetés, ou les deux, mais pas de stratus fractus de mauvais temps⁽¹⁾.
- ↳ **Stratus fractus** de mauvais temps ou cumulus fractus de mauvais temps⁽¹⁾, ou les deux (pannus), généralement au-dessous d'un altostratus ou d'un nimbostratus.
- ↳ **Cumulus et stratocumulus** autres que ceux formés par étalement de cumulus ; la base des cumulus est à un niveau différent de celle des stratocumulus.
- ↳ **Cumulonimbus**, dont la région supérieure est nettement fibreuse (cirriforme), souvent en forme d'enclume ; accompagné ou non de cumulonimbus sans enclume ou sans région supérieure fibreuse, de cumulus, de stratocumulus, de stratus ou de pannus.

(1) La locution « mauvais temps » se rapporte aux conditions qui existent généralement pendant les précipitations, et immédiatement avant ou après celles-ci.

- ∞ Visibilité réduite par fumée
- ∞ Brume sèche
- S Poussières en suspension dans l'air
- ⊗ Embruns
- ⊙ Poussières ou sable soulevés par le vent mais absence de tourbillon
- ⊗ Tourbillon(s) de poussière ou de sable
- (S) Tempête de poussière ou de sable
- ≡ Brume
- ≡ Mince couche de brouillard en bancs
- ≡ Mince couche de brouillard en bancs
- ≡ Mince couche de brouillard en bancs
- < Éclairs visibles, tonnerre non entendu
- Précipitations en vue, n'atteignant pas le sol
- () Précipitations en vue, atteignant le sol mais distantes de la station
- (•) Précipitations en vue, atteignant le sol près de la station
- ⊗ Orage, mais pas de précipitations
- ∇ Grains
- ⊗ Trombe(s)

Temps caractéristique au cours de l'heure précédente

- Bruine ou neige en grains
- Pluie
- Neige
- Pluie et neige mêlées ou granules de glace
- Bruine ou pluie se congelant
- Averse(s) de pluie
- Averse(s) de neige ou de pluie et de neige
- Averse(s) de grêle⁽¹⁾ ou de pluie et de grêle
- ≡ Brouillard ou brouillard glacé
- ⊗ Orage (avec ou sans précipitations)

⁽¹⁾ grêle, grésil ou neige roulée

Tempête de poussière, sable ou neige au moment de l'observation

- + Chasse-neige faible basse
- + Forte chasse-neige basse
- + Chasse-neige faible élevée
- + Forte chasse-neige élevée
- ⊗ Tempête de poussière ou de sable légère } - en diminution sans changement - en augmentation
- ⊗ Violente tempête de poussière ou de sable } - en diminution sans changement - en augmentation

Brouillard ou brouillard glacé au moment de l'observation

- ≡ Brouillard à distance
- ≡ Brouillard en bancs
- ≡ Brouillard, ciel visible } sans changement
- ≡ Brouillard, ciel invisible } sans changement
- ≡ Brouillard, ciel visible } sans changement
- ≡ Brouillard, ciel invisible } sans changement
- ≡ Brouillard déposé du givre, ciel visible
- ≡ Brouillard déposé du givre, ciel invisible

Bruine

- Bruine intermittente } faible
- Bruine continue } faible
- Bruine intermittente } modérée
- Bruine continue } modérée
- Bruine intermittente } forte
- Bruine continue } forte
- Bruine faible, se congelant
- Bruine modérée ou forte se congelant
- Bruine et pluie, faible

Bruine et pluie, modérée ou forte

Pluie non sous forme d'averses au moment de l'observation

- Pluie intermittente } faible
- Pluie continue } faible
- Pluie intermittente } modérée
- Pluie continue } modérée
- Pluie intermittente } forte
- Pluie continue } forte
- Pluie faible, se congelant
- Pluie modérée ou forte, se congelant
- Pluie et neige, faible
- Pluie et neige, modérée ou forte

Précipitations solides non sous forme d'averses au moment de l'observation

- * Chute intermittente de flocons de neige } faible
- ** Chute continue de flocons de neige } faible
- * Chute intermittente de flocons de neige } modérée
- ** Chute continue de flocons de neige } modérée
- * Chute intermittente de flocons de neige } forte
- ** Chute continue de flocons de neige } forte
- Poudrin de glace
- Neige en grains
- Étoiles de neige isolées
- △ Granules de glace

Précipitations sous forme d'averses au moment de l'observation

- ∇ Averse(s) de pluie faible(s)
- ∇ Averse(s) de pluie, modérée(s) ou forte(s)
- ∇ Averse(s) de pluie violente(s)
- ∇ Averse(s) de pluie et neige mêlées faible(s)
- ∇ Averse(s) de pluie et neige mêlées modérée(s) ou forte(s)
- ∇ Averse(s) de neige faible(s)

Averse(s) de neige modérée(s) ou forte(s)

- ∇ Averse(s) de grésil ou de neige roulée avec ou sans pluie ou pluie et neige mêlées } faible(s) modérée(s) ou forte(s)
- ∇ Averse(s) de grêle avec ou sans pluie ou pluie et neige mêlées sans tonnerre } faible(s) modérée(s) ou forte(s)

Précipitations au moment de l'observation après un orage durant l'heure précédente

- ⊗ Pluie faible
- ⊗ Pluie modérée ou forte
- ⊗ Faible chute de neige, ou pluie et neige mêlées, ou grêle⁽¹⁾
- ⊗ Chute modérée ou forte de neige, ou de pluie et neige mêlées, ou de grêle⁽¹⁾

Précipitations avec orage au moment de l'observation

- * Orage faible ou modéré, sans grêle⁽¹⁾ mais avec pluie et (ou) neige
- ⊗ Orage faible ou modéré, avec grêle⁽¹⁾
- * Orage fort sans grêle⁽¹⁾ mais avec pluie et (ou) neige
- ⊗ Orage avec tempête de poussière ou de sable
- ⊗ Orage fort avec grêle⁽¹⁾

Temps caractéristique pendant les 6 h précédentes

- + Tempête de : poussière, sable ou neige
- ≡ Brouillard ou brouillard glacé ou brume sèche épaisse
- Bruine
- Pluie
- * Neige ou pluie et neige mêlées
- ∇ Averse(s)
- ⊗ Orage(s)

Bibliographie

- Atkinson, R., Baulsch, D. L., Cox, R. A., Hampton, R. F., Kerr, J. A., Rossi, M. J., Troe J., Evaluated kinetics, photochemical and heterogeneous data, *J. Phys. Chem., Ref. Data*, 26, 3, 521-1012, 1997.
- Balgovind, R., A. Dalcher, M. Ghil, et E. Kalnay, A stochastic-dynamic model for the spatial structure of the forecast error statistics, *Mon. Wea. Rev.*, 111, 701-722, 1983.
- Bey, I., Contribution des processus nocturnes à la chimie troposphérique : modélisation des flux de radicaux et transformation des précurseurs d'ozone (COV, NOx), *Mémoire de thèse, Université Paris 12*, 1997.
- Bergman, K.H., Multivariate Analysis of Temperatures and Winds Using Optimum Interpolation, *Mon. Wea. Rev.*, 107, 1423-1444, 1979.
- Bergthorsson, P., et B.R., Döös, Numerical weather map analysis, *Tellus*, 7, 329-340, 1955.
- Bertino, L., Assimilation de données pour la prédiction de paramètres hydrodynamiques et écologiques : cas de la lagune de l'Oder, *Mémoire de thèse, Ecole des Mines de Paris, Fontainebleau*, 2001.
- Blond, N., L. Bel et R. Vautard, Three-dimensional ozone data analysis with an air quality model over Paris area, soumis à *J. Geophys. Res.*
- Bouttier, F., A dynamical Estimation of Forecast Error Covariances in an Assimilation System, *Mon. Wea. Rev.*, 122, 2376-2390, 1994.
- Buell, C., Correlation functions for wind and geopotential on isobaric surfaces, *J. Appl. Meteor.*, 11, 51-59, 1972.
- Builtjes, P. J. H., The LOTOS-long term ozone simulations project, *Summary Report, TNO-IM W Report R92/240, Delft, The Netherlands*, 1992.
- Burrows, W.R., M. Benjamin, S. Beauchamp, B. Thompson, CART decision - tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada, *J. of Applied Meteorology*, 34, 1848-1862, 1995.
- Brimblecombe, P., *The Big Smoke, Methuen*, 1987.
- Carpenter, jr. R. L., K. K. Droegemeier, P. R. Woodward, and C. E. Hane, Application of the piecewise parabolic method (PPM) to meteorological modeling, *Mon. Wea. Rev.*, 118, 586-612, 1990.
- Carter, W. P. L., A detail mechanism for the gas-phase atmospheric reactions of organic compounds, *Atmos. Environ.*, 24A, 481-518, 1990.
- Cathala, M.-L., J. Pailleux et V.-H. Peuch, Improving global simulations of UTLS ozone with assimilation of MOZAIC data, *Tellus*, sous presse 2002.

- Chang, J. S., Brost, R. A., Isaksen, I. S. A., Madronich, S., Middleton, P. Stockwell, W. R. et Walcek, C. J., A three-dimensional eulerian acid deposition model : physical concepts and formulation, *Journal of Geophysical Research*, 92, 14681-14700, 1987.
- Chauvet, P., Aide-mémoire de géostatistique linéaire, *Presses de l'Ecole des mines de Paris*, 2000.
- Chenevez J. et C.Ø. Jensen, Operational ozone forecasts for the region of Copenhagen by the Danish Meteorological Institute, *Atm. Environ.*, 35, 4567-4580, 2001.
- Cohn, S.E. et D.F. Parrish, The behaviour of Forecast Error Covariances for a Kalman Filter in Two Dimensions, *Mon. Wea. Rev.*, 119, 1757-1785, 1991.
- Colella, P., et Woodward, P. R., The piecewise parabolic method (PPM) for gas-dynamical simulations, *Journal of Computational Physics*, 11, 38-39, 1984.
- Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier et M. Fisher, The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part 1 : formulation. *Quart. J. Roy. Met. Soc.*, 124, 1783-1807, 1998.
- Cressman, G.P., An operational objective analysis system, *Mon. Wea. Rev.*, 87, 367-374, 1959.
- Cressie, N. A. C., Statistics for Spatial Data (revised edition), *Wiley, New York, NY*, 1993.
- Daley, R., Atmospheric data analysis, *Cambridge University Press, New York*, 1991.
- Daley, R., Atmospheric Data Assimilation, *Journal of the Meteorological Society of Japan*, 75(1B), 319-329, 1997.
- Dee, D.P., Simplification of the Kalman filter for meteorological data assimilation, *Q. J. R. Meteorol. Soc.*, 117, 365-384, 1991.
- De Moore, W. B., Sandetr, S. P., Golden, D. M., Hampton, R. F., Kurylo, M. J., Howard, C. J., Ravishankara, A. R., Kolb, C. E., et Molina, M. J., Chemical kinetics and photochemical data for use in stratospheric modelling evaluation, 11, JPL publication 94, 26, JPL, Pasadena, US, 1994.
- Desroziers, G., A coordinate change for data assimilation in spherical geometry of frontal structures, *Mon. Wea. Rev.*, 123, 1128-1145, 1997.
- Eskes, H.J., A.J.M. PETERS, P. F. Levelt, M. A. f. Alart, et H. M. Kelder, Variational assimilation of total-column ozone satellite data in a 2D lat-lon tracer-transport model, *J. Atmos. Sci.*, 56, 3560-3572, 1999.
- Eliassen, A., Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. Inst. Weather and Climate Res., *Acad. Sci. Oslo*, Rept. No. 5., 1954.
- Elbern, H., H. Schmidt, et A. Ebel, Adjoint of a gas phase mechanism and assimilation of air quality data for photo-oxidant modeling, in *Proceeding of the XVIII Quadrennial Ozone Symposium*, edited by R. D. Bojkov and G. Visconti, 703-706, L'Aquila, Italy, 1996.
- Elbern, H., H. Schmidt, et A. Ebel, Variational data assimilation for tropospheric chemistry modelling, *J. Geophys. Res.*, 102, 15,967-15,985, 1997.
- Elbern, H., et H. Schmidt, A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling, *J. Geophys. Res.*, 104, 18,583-18,598, 1999.
- Elbern, H., H. Schmidt, O. Talagrand, et A. Ebel, 4D-variational data assimilation with an adjoint air quality model for emission analysis, *Environ. Modell. Software*, 15,539-15,548, 2000.
- Elbern, H., et H. Schmidt, Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.*, 106, 3569-3590, 2001.
- EMEP, Transboundary photooxidant air pollution in Europe. calculations of tropospheric ozone and comparison with observations. *EMEP/MSC-W Report 2/98, Oslo, Norway*, 1998.

- Erismann, J. W., W. van Pul, et G. Wyers, Parametrization of surface resistance for the quantification of atmospheric deposition of acidifying pollutants and ozone, *Atm. Environ.* 28, 2595-2607, 1994.
- Erismann, J. W. et al., The Elspeetsche Veld experiment on surface exchanges of trace gases : Summary of results, *Atmospheric Environment*, 28, 487-496, 1994.
- Errico, R., T. Vukićević et K. Raeder, Examination of the accuracy of a tangent linear model, *Tellus*, 45A, 462-477, 1993.
- ESQUIF, Étude et Simulation de la QUALITÉ de l'air en Île-de-France, Institut Pierre Simon Laplace, Laboratoire Interuniversitaire des Systèmes Atmosphériques, Météo-France, Laboratoire d'Aérodynamique, Forschungszentrum Jülich, AIRPARIF, *Rapport final*, 2001.
- Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99(C5), 10,143-10,162, 1994.
- Fedorov, V., Kriging and other estimators of spatial field characteristics (with special reference to environmental studies), *Atmos. Environ.*, 23, 174-184, 1989.
- Finkelstein, P., The spatial analysis of acid precipitation data, *J. Clim. Appl. Meteorol.*, 23, 52-62, 1984.
- Fisher, M., et D.J. Lary, Lagrangian four-dimensional variational data assimilation of chemical species, *Q. J. R. Meteorol. Soc.*, 121, 1681-1704, 1995.
- Fuentes, M., A high frequency kriging approach for non-stationary environmental processes, *Environmetrics*, 12, 469-483, 2001.
- GENEMIS (Generation of European Emission Data for Episodes) project, *EUROTRAC annual report 1993, part 5, EUROTRAC international scientific secretariat, Garmisch-Partenkirchen, Germany*, 1994.
- Gandin, L.S., Objective analysis of meteorological fields, *Leningrad. Hydromet. Press.*, 1963.
- Gardner, M.W., et Dorling, S.R., Artificial neuronal networks (the multilayer perceptron) - a review of applications in the atmospheric sciences, *Atmospheric Environment*, 32(14-15), 2627-2636, 1998.
- Garrat, J. R., The Atmospheric Boundary Layer, Eds Houghton J. T., Rycroft M. J., Dessler A. J., *Cambridge atmospheric and space science series, New York*, 1992.
- Gaspari, G., et S. E. Cohn : Construction of correlation functions in two and three dimensions, *Quart. J. Roy. Meteor. Soc.*, 125, 723-757, 1999.
- Gauss, K.F., 1809, Theory of motion of the heavenly bodies, (Translation of Theorica Motus Corporum Coelestium) *Dover Press, New York*, 1963.
- Ghil, M. et P. Malanotte-Rizzoli, Data assimilation in meteorology and oceanography, *Adv. Geophys.*, 33, 141-266, 1991.
- Ghil, M., Meteorological Data Assimilation for Oceanographers. Part I : description and Theoretical Framework, *Dyn of Atm. and Oceans*, 13, 171-218, 1989.
- Gilchrist, B., et G.P. Cressman, An experiment in objective analysis, *Tellus*, 6, pp. 97-101, 1954.
- Günther A., Zimmermann P. et Wildermuth M., Natural volatile organic compound emission rate estimates for US woodland landscape, *Atmospheric Environment*, 28, 1197-1210, 1994.
- Günther A., Seasonal and spatial variations in the natural volatile organic compound emissions. *Ecol. Appl.*, 7(1), 34-45, 1997.

- Hauglustaine, D. A., G. P. Brasseur, S. Walters, P. J. Rasch, J.-F. Müller, L. K. Emmons, et M. A. Carroll, MOZART : A global chemical transport model for ozone and related chemical tracers, 2. Model results and evaluation, *J. Geophys. Res.* 103, 28291-28336, 1998.
- Hicks, B. B., Baldocchi, D. D., Myers, T. P., Hosker, R. P., et Matt, D. R., A preliminary multiple resistance routine for deriving dry deposition from measured quantities, *Water, air and soil pollution*, 36, 311-330, 1987.
- Hodzic, A., Pollution atmosphérique en Île-de-France, Mémoire de stage de recherche de l'Ecole Normale de la Météorologie (ENM), note de travail de l'ENM 817, 2002.
- Hoelzemann, J., H. Elbern, and A. Ebel, PSAS and 4D-var data assimilation for chemical state analysis by urban and rural observation sites, *Phys. and Chem. of the Earth*, 26, 807-812, 2001.
- Holtstag, A. A. M., Nieuwstadt, F. T. M., Scaling the atmospheric boundary layer, *Boundary-Layer Meteorology*, 36, 201-209, 1986.
- Hollingsworth, A., et Lönnberg, The statistical structure of short-range forecast errors as determined from radiosonde data. Part I : The wind field, *Tellus*, 38A, 111-136, 1986.
- Hollingsworth, A., Objective analysis for numerical weather prediction, *Short and medium range numerical weather prediction, collected papers presented at WMO/IUGG NWP Symposium, Tokyo, August 1986*, T. Matsuno, ed. *Special volume of the J. Meteor. Soc. Japan*, 11-59, 1987.
- Hov, O., Stordal, F., et Eliassen, A., Photochemical oxidant control strategies in Europe : a 19 day case study using a lagrangian model with chemistry, *NILU TR5/95*, 1985.
- Honoré, C., Vautard, R. et Beeckmann, M., Photochemical regimes in urban atmospheres : The influence of dispersion, *Geophysical Research letters*, 27, 13, 1895-1898, 2000.
- Honoré, C., La photochimie de l'ozone à l'échelle urbaine, un système dynamique non linéaire, *Mémoire de thèse, Université Paris 6*, Paris, 2000b.
- Ide, K., P. Courtier, M. Ghil, et A. Lorenc, Unified notation for data assimilation : Operational sequential and variational, *J. Meteorol. Soc. Jpn.*, 75, 181-189, 1997.
- Ionescu A., Y. Candau, E. Mayer, et I. Colda, Analytical determination and classification of pollutant concentration fields using air pollution monitoring network data, Methodology and application in the Paris area, during episodes with peak nitrogen dioxide levels, *Environmental Modelling and Software*, 15, 565-573, 2000.
- Jeuken, A. B. M., H. J. Eskes, P. F. J. van Velthoven, H. M. Kelder, et E. V. Hólm, Assimilation of total ozone satellite measurements in a three-dimensional tracer transport model, *J. Geophys. Res.*, 104, 5551-5563, 1999.
- Jakobs, H. J., H. Feldmann, H. Hass, et M. Memmesheimer, The use of nested models for air pollution studies : an application of the EURAD model to a SANA episode, *Journal of Applied Meteorology*, 34, 1301-1319, 1995.
- Julian, P., et H.J. Thiebaut, On some properties of correlation functions used in optimum interpolation schemes, *Mon. Wea. Rev.*, 103, 605-616, 1975.
- Kalman, A new approach to linear filtering and prediction problems, *Trans. ASME, Ser. D, J. Basic Eng.*, 83, 35-45, 1960.
- Kalman et Bucy, New results in linear filtering and prediction theory, *Trans. ASME, Ser. D, J. Basic Eng.*, 83, 95-108, 1961.
- Khattatov, B.V., J.-F. Lamarque, L.V. Lyjak, R. Ménard, P. Levelt, X. Tie, G.P. Brasseur, et J.C. Gille, Assimilation of satellite observations of long-lived chemical species in global chemistry transport models, *J. Geophys. Res.*, 105, 29, 135-29, 144, 2000.

- Kley, D., Geiss, Lehning, Neininger, Absolute and relative VOC and NO_x emissions of the Paris area, soumis à *J. Geophys. Res.*.
- Kleinman, L.I., Dependence of ozone production on NO and hydrocarbons in the troposphere, *Geophysical Research Letters*, 24, no. 18, 2299-2302, 1997.
- Lajaunie, C., Kriging with mass balance. E. Baafi et N. Schofield (Eds.), *Geostatistics Wollongong'96*, 1, Dordrecht, 80-91, Kluwer, 1997.
- Lattuati, M., Contribution à l'étude du bilan de l'ozone troposphérique à l'interface de l'Europe et de l'Atlantique Nord : Modélisation lagrangienne et mesures en altitude, *Mémoire de thèse, Université Paris 6, Paris*, 1997.
- Lacarra, J.-F., et O. Talagrand, Short-Range Evolution of Small Perturbations in Barotropic Model, *Tellus*, 40A, 81-95, 1988.
- Leighton, P.A., Photochemistry of Air Pollution, *Academic Press, New York*, 1961.
- Relievel, J., et F. Dentener, What controls Tropospheric Ozone?, *J. Geophys. Res.*, 3531-3551, 2000.
- Levelt, P.F., B.V. Khattatov, J.C. Gille, G.P. Brasseur, X.X. Tie, et J.W. Waters, Assimilation of MLS ozone measurements in the global three-dimensional chemistry transport model ROSE, *Geophys. Res. Lett.*, 25, 4493-4496, 1998.
- Lewis, J. M., et J. C. Derber, The use of adjoint equations to solve a variational adjustment problem with advective constraints, *Tellus*, 37A, 309-322, 1985.
- Lyster, P.M., S.E. Cohn, R. Ménard, L.-P. Chang, S.-J. Lin, et R. Olsen, An implementation of a two-dimensional filter for atmospheric chemical constituent assimilation on massively parallel computers, *Mon. Wea. Rev.*, 125, 1674-1686, 1997.
- Locher, F., L'Atmosphère, le Climat, et le Temps en France (1820-1880), Thèse en cours.
- Louis, J-F., A parametric model of vertical eddy fluxes in the atmosphere, *Boundary Layer Meteorology*, 48, 187-202, 1979.
- Lorenz, Analysis methods for numerical weather prediction. *Quart. J. Roy. Met. Soc.*, 112, 1177-1194, 1986.
- Lu, R., et R. P. Turco, An integrated air pollution modeling system : Application to the Los Angeles Basin, *Numerical Simulations in the Environmental and Earth Sciences, Cambridge University Press.*, 116-125, 1997.
- Matheron, G., Traité de géostatistique appliquée, *Editions Technip., Paris*, 1963.
- Matheron, G., La théorie des variables régionalisées et ses applications, *Fontainebleau, ENSMP, cahiers du centre de morphologie mathématique*, Fasc. 5. 210, 1970.
- Mc Rae, G. J., Goodin, W. R., et Seinfeld, J. H., Numerical solution of the atmospheric diffusion equation for chemically reacting flows, *Journal of computational Physics*, 45, 1-42, 1982.
- Memmesheimer M, H. Hass, J. Tippke, et A. Ebel, Modeling of episodic emission data for Europe with EURAD Emission Model EEM. *Proceedings of the International Speciality Conference 'Regional Photochemical Measurement and modeling studies*, Vol. 2, San Diego, CA, USA, AWMA, 1995.
- Ménard, R., S.E. Cohn, L.-P. Chang et P.M. Lyster, Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part I : Formulation, *Mon. Wea. Rev.*, 128, 2654-2686, 2000.

- Ménard, R., et L.-P. Chang, Stratospheric assimilation of chemical tracer observations using a Kalman filter, part II : Chi-square validated results and analysis of variance and correlation dynamics, *Mon. Wea. Rev.*, 128, 2672-2686, 2000.
- Menut L., Vautard R., Beeckmann M., et Honoré, C., Sensitivity of photochemical pollution using the adjoint of a simplified chemistry-transport model, *Journal of geophysical research*, 105, 15,379-15,402, 2000.
- Menut, L., R. Vautard, C. Flamant, C. Abonnel, M. Beekmann, P. Chazette, P. H. Flamant, D. Gombert, D. Guédalia, D. Kley, M. P. Lefebvre, B. Lossec, D. Martin, G. Mégie, P. Perros, M. Sicard, et G. Toupance, Measurements and modeling of atmospheric pollution over the Paris area : An overview of the ESQUIF project, *Annales Geophysicae*, 18, 1467-1481, 2000.
- Météo. Hebdo, Bulletin hebdomadaire d'études et de renseignements, Météo-France, 30, 2002a
- Météo. Hebdo, Bulletin hebdomadaire d'études et de renseignements, Météo-France, 31, 2002b
- Météo. Hebdo, Bulletin hebdomadaire d'études et de renseignements, Météo-France, 31, 2001
- Middleton, P., Stockwell, W. R., and Carter, W. P., Agregation and analysis of volatile organic compound emissions for regional modelling, *Atmos. Environ.*, 24, 1107-1133, 1990.
- Mylona, S., EMEP emission data, *Status report 1999, EMEP/MSC-W Note 1/99, Oslo, Norway, 1999.*
- O'Brien, J. J., A note on the vertical structure of the eddy exchange coefficient in the planetary boundary layer, *Journal of Atmospheric Sciences*, 27, 1213-1215, 1970.
- Olendrzynski, K., Operational EMEP Eulerian Acid Deposition Model, *EMEP/MSC-W Note 4/99, Oslo, Norway, 1999.*
- Panofsky, H.A., Objective weather map analysis, *Journal of Meteorology*, 6, 386-392, 1949.
- Parrish, D. et J. Derber, The national Meteorological Center's spectral statistical interpolation analysis system, *Mon. Wea. Rev.*, 120, 1747-1763, 1992.
- Perros, P., Costantini, et C. Jambert, Airborne study of the ozone urban plume of Paris during the summer 1999, soumis à *J. Geophys. Res.*.
- Peuch V.-H., M. Amodei, T. Barthet, M.-L. Cathala, B. Josse, M. Michou et P. Simon, MOCAGE : modèle de chimie atmosphérique à grande échelle, A.M.A 1999, Toulouse, France, 1999.
- Rabier, F., et P. Courtier, Four-dimensional assimilation in the presence of baroclinic instability, *Quart. J. Roy. Meteor. Soc.*, 118, 649-672, 1992.
- Peters, L.K., C.M. Berkowitz, G.R. Carmichael, R.C. Easter, G. Fairweather, S.J. Ghan, J.M. Hales, L.R. Leung, W.R. Pennell, F.A. Potra, R.D. Saylor, et T.T. Tsang, The current state and future directions of Eulerian models in simulating the tropospheric chemistry and transport of trace species : a review, *Atmos. Environ.*, 29, 189-222, 1995.
- Reynolds, S. D., Roth, P. M., et Seinfeld, J. H., Modeling of photochemical air pollution - I. Formulation of the model, *Atmos. Env.*, 1033-1061, 1973.
- Riishøjgaard, L.P., A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, *Tellus Ser. A*, 50, 42-57, 1998.
- Robeson, S.M., et D.G. Steyn, Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations, *Atmos. Environ.*, 24B, 303-312, 1990.
- Rutherford, I.D., Experiments on the updating of PE forecasts with real wind and geopotential data, *Preprints Third Conf, Probability and Statistics in Atmospheric Sciences, Boulder, Amer. Meteor. Soc.*, 198-201, 1973.

- Schlatter, T.W., Some experiments with a multivariate statistical objective analysis scheme, *Mon. Wea. Rev.*, 103, 246-257, 1975.
- Schmidt, H., Vierdimensionale Datenassimilation nach der Variationsmethode für ein mesoskaliges Chemietransportmodell, PhD-thesis, *Mitteilungen aus dem Institut für Geophysik und Meteorologie der Universität zu Köln* 132, Cologne, Germany, 1999.
- Schmidt, H., C. Derognat, R. Vautard, et M. Beekmann, A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in western Europe, *Atm. Environ.*, 35, 6277-6297, 2001.
- Seinfeld, J. H., et S. N. Pandis, *Atmospheric chemistry and physics*, Wiley-Interscience, New York, 1998.
- Silibello, C., G. Calori, G. Brusasca, G. Catenacci, et Finzi G., Application of a photochemical grid model to Milan metropolitan area, *Atmospheric Environment*, 32, 11, 2025-2038, 1998.
- Sillman, S., The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments, *Atmos. Envir.*, 33, 1821-1845, 1999.
- Simpson, D., Long period modelling of photochemical oxidants in Europe. Calculations for April-September 1985, April-October 1989, *EMEP/MSC-W Report 2/91*, Oslo, Norway, 1991.
- Simpson D., Long period modeling of photochemical oxidants in Europe, Calculations for July 1985, *Atmospheric Environment*, 26, 1609-1634, 1992.
- Simpson D., Inventorying emissions from nature in Europe, *Journal of Geophysical Research*, 104, 8113-8152, 1999
- Simpson, R. W. et A. P. Layton, Forecasting peak ozone levels, *Atmos. Environ.*, 17, 1649-1654, 1983.
- Sirois, A., J.A. Pudyklewicz, et A. Kallaur, A comparison between simulated and observed ozone mixing ratios in Eastern North America, *Journal of geophysical Research*, 104(D17), 21397-21423, 1999.
- Stohl, A., E. Williams, G. Wotawa et H. Kromp-Kolb, A european inventory of soil nitric oxide emissions and the effect of these emissions on the photochemical formation of ozone in Europe, *Atmos. Environ.*, 30, 3, 741-3,755, 1996.
- Stull R.B., An introduction to boundary layer meteorology, *Kluwer Academic Publishers, Reidel*, 666, 1988.
- Talagrand, O., et P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. I : Theory, *Q. J. R. Meteorol. Soc.*, 113, 1311-1328, 1987.
- Talagrand, O., Assimilation of Observations, an Introduction, *Journal of the Meteorology Society of Japan*, 75, 1B, 191-209, 1997.
- Talagrand, O., et Bouttier F., Internal Diagnostics of Data Assimilation Systems, *Proceedings of Seminar on Diagnosis of Models and Data Assimilation Systems, European Centre for Medium-range Weather Forecasts, Reading, England, September 1999*, 407-409, 2000.
- Toupance, G., Perros, P., Soedomo, M. et Tsalkani, N., Recherche sur la pollution oxydante à l'échelle régionale : le cas du bassin parisien, rapport du Centre d'Étude et de Recherche sur la Pollution Atmosphérique Régionale, Université Paris Val-de Marne, Avenue du Général de Gaulle, 94010 Créteil, 1988.
- Thépaut, J.-N. et P. Courtier, Four-dimensional data assimilation using the adjoint of a multilevel primitive-equation model, *Quart. J. Roy. Meteor. Soc.*, 117, 1225-1254, 1991.

- Thiebaut, H. J., Anisotropic correlation functions for objective analysis, *Mon. Wea. Rev.*, 104, 994-1002, 1976.
- Thiebaut, H. J., Extending analysis accuracy with anisotropic interpolation, *Mon. Wea. Rev.*, 105, 691-699, 1977.
- Thompson, P., A dynamical method of analyzing meteorological data, *Tellus*, 13, 334-349, 1961.
- Tilmes, S., et J. Zimmermann, investigation on the spatial scales of the variability in measured near-ground ozone mixing ratios, *Geophysical research Letters*, 25, 3827-3830, 1998.
- Tilmes, S., J. Brandt, F. Flato, J. Langner, R. Bergström, J.H. Christensen, A. Ebel, R. Friedrich, L.M. Frohn, A. Heidegger, ØHov, I. Jacobson, H. Jakobs, B. Wickert, et J. Zimmermann, Comparison of five Eulerian ozone prediction systems for summer 1999 using the German monitoring data, *Journal of Atmospheric Chemistry*, 2002.
- Toupance, G., L'ozone dans la basse troposphère. Théorie et pratique, *Pollution Atmosphérique*, 32-42, 1988.
- Tulet, P., Crassier V., et Rosset R., Air Pollution Modelling at a Regional Scale, *Environmental Modelling and Software*, 15, 693-696, 2000.
- van de Velde, R., Faber W., Katwijk V., Scholten H. J., Thewessen T. J. M., Verspuyl M., and Zevenbergen M., The preparation of a European land-use data base, *RIVM Report 712401001, Bilthoven, The Netherlands*, 1994.
- van Loon, M., et Heemink, A. W., Kalman Filtering for non-linear atmospheric chemistry models : first experiences, *Technical report, MAS-R9711, CWI, Amsterdam, The Netherlands*, March 1997.
- van Loon, M., Numerical smog prediction I : The physical and chemical model, *CWI report NM-R9411*, 1994.
- van Loon, P. J. H. Bultjes, et A. J. Segers, Data assimilation of ozone in the atmospheric transport chemistry model LOTOS, *Environmental Modelling Software*, 15, 603-609, 2000.
- Vautard, R., Beekmann M., honoré, C., et Deleuze, I., La pollution photochimique en région parisienne simulée par le modèle CHIMERE et l'influence du transport régional d'ozone. *Document non publié, LMD, Paris, France*, 1997.
- Vautard, R., M. Beekmann, J. Roux, et D. Gombert, Validation of a deterministic forecasting system for the ozone concentrations over the Paris area, *Atm. Environ.*, 35, 2449-2461, 2000.
- Vautard, R., D. Martin, M. Beekmann, R. Friedrich, A. Jaubertie, D. Kley, M. Iattuati, P. Moral, B. Neininger, et J. Theloke, Paris emission inventory diagnostic from the ESQUIF airborne measurements and a chemistry-transport model, soumis à *J. Geophys. Res.*.
- Vautard, R., D. Martin, M. Beekmann, R. Friedrich, A. Jaubertie, D. Kley, M. Iattuati, P. Moral, B. Neininger, et J. Theloke, Paris emission inventory diagnostic from the ESQUIF airborne measurements and a chemistry-transport model, soumis à *J. Geophys. Res.*.
- Verwer, J. G., Gauss-Seidel iteration for stiff ODEs from chemical kinetics, *SIAM Journal on Scientific Computing*, 15, 1243-1250, 1994.
- Wesely, M. L., et B. B. Hicks, Some factors that affect the deposition rates of sulfur dioxide and similar gases on vegetation. *Journal of Air Pollution Control Association*, 27, 1110-1116, 1977.
- Wesely, M. L., Parameterization of surface resistance to gaseous dry deposition, *Atmos. Environ.* 23, 1293-1304, 1989.

