Thèse de doctorat de l'Université Paris-Est

Présentée et soutenue publiquement le 15 décembre 2009 par

Irène Korsakissok

pour l'obtention du diplôme de docteur
de l'Université Paris-Est

Spécialité : Sciences et techniques de l'environnement

Changements d'échelle en modélisation de la qualité de l'air et estimation des incertitudes associées

Jury composé de

Pr Richard Perkins	Centrale Lyon	président et rapporteur
Pr Robert Rosset	Université P. Sabatier, CNRS	rapporteur
Pr Serge Guillas	University College London	examinateur
Dr Bruno Sportisse	INRIA	directeur de thèse
Dr Vivien Mallet	INRIA	co-directeur de thèse
Dr Laurence Rouïl	INERIS	invitée
Mr Philippe Lameloise	AIRPARIF	invité

Cette thèse est dédiée à mon père, Charles Korsakissok, qui m'a communiqué son enthousiasme pour les sciences et à mon frère Joël, qui depuis toujours me montre l'exemple.

Remerciements

Je me dois de remercier en tout premier lieu Vivien Mallet, qui a suivi mes travaux de thèse durant ces années passées au CEREA. Cela a été un plaisir de travailler avec lui, tant pour ses multiples qualités scientifiques et techniques que pour sa disponibilité, sa rigueur et sa patience. J'ai appris énormément de choses grâce à lui, et lui en suis très reconnaissante.

Je remercie également Bruno Sportisse qui m'a embauchée au CEREA, d'abord en qualité d'ingénieur de recherche, puis a su me convaincre de faire une thèse de doctorat. C'est de lui que vient l'impulsion de cette thèse et l'idée initiale du sujet. Je le remercie enfin d'avoir accepté de rester mon directeur de thèse malgré ses importantes responsabilités à l'INRIA et de m'avoir consacré un peu de son précieux temps.

Je remercie les rapporteurs, Richard Perkins et Robert Rosset, de même que les autres membres du jury, d'avoir accepté d'évaluer mon travail. Je remercie en particulier Serge Guillas qui m'a donné des pistes d'étude concernant la réduction d'échelle statistique.

Je remercie Christian Seigneur pour l'aide qu'il m'a apportée lors de la réalisation du modèle de panache sous-maille, ainsi que pour la relecture de certains chapitres de ma thèse.

Ces quelques années passées au CEREA ont été riches sur le plan scientifique, mais aussi sur le plan humain de par les multiples échanges et rencontres. Je me garderai de citer toutes les personnes dont j'ai pu apprécier la présence et l'amitié, de peur d'en oublier. Je remercie donc collectivement tous les membres du CEREA, passés et présents, pour leur humour, leurs discussions, leur aide occasionnelle et tous les agréables moments et gâteaux partagés avec eux. Je citerai en particulier mes différents collègues de bureau, Meryem Ahmed de Biasi avec qui j'ai pu faire un inoubliable voyage au Chili, Olivier Saunier et Megan Lebacque. Une mention spéciale à ceux avec qui les moments de détente ont pu se prolonger autour d'un verre ou dans un restaurant, Pierre Tran, Marilyne Tombette, Victor Winiarek, Elsa Real, Yelva Roustan. J'espère que ces moments se reproduiront à l'avenir ! Un petit encouragement aux nouveaux doctorants qui assurent la relève et en particulier à Régis Briant, dont j'ai eu le plaisir de co-encadrer le stage de fin d'étude. Une pensée particulière enfin à tous les bébés du CEREA, qui se sont multipliés ces dernières années, et surtout aux heureux parents.

Je ne peux enfin terminer sans adresser mes chaleureux remerciements aux membres de ma famille qui m'ont encouragée, et en particulier à mon mari Julien Bourdin qui a enduré stoïquement toutes les affres de la thèse et notamment les difficiles derniers mois de rédaction. Pour ce qu'il m'apporte, son soutien constant et inébranlable, et infiniment plus encore...

Table des matières

1 Changements d'échelle en qualité de l'air **11**
1.1 Différentes échelles . . . 13
1.1.1 Les émissions . . . 13
1.1.2 Le transport . . . 16
1.1.3 La turbulence . . . 18
1.1.4 Les processus de perte . . . 20
1.2 Une hiérarchie de données et de modèles . . . 25
1.2.1 Équation de dispersion réactive . . . 25
1.2.2 Différents types de modèles . . . 28
1.2.3 Données d'entrée . . . 30
1.2.4 Sensibilité des modèles aux données d'entrée . . . 32
1.3 Différentes méthodes de changements d'échelles . . . 32
1.3.1 Méthodes numériques et couplage de modèles . . . 33
1.3.2 Fermeture d'ordre supérieur . . . 36
1.3.3 Paramétrisations sous-mailles . . . 38
1.3.4 Méthodes stochastiques . . . 40
1.3.5 Prise en compte de la variabilité sous-maille a posteriori . . . 42
1.4 Conclusion . . . 43
1.4.1 Bilan . . . 43
1.4.2 Choix des méthodes . . . 44
1.4.3 Plan de la thèse . . . 45
1.4.4 Publications associées . . . 45

I Modèles pour l'échelle locale **47**

2 Description des modèles à l'échelle locale **49**
2.1 Les modèles gaussiens . . . 51
2.1.1 Lien avec la formulation eulérienne . . . 51
2.1.2 Modèle gaussien de panache . . . 53
2.1.3 Modèle gaussien à bouffées . . . 56
2.1.4 Modèle de source linéique . . . 58
2.2 Modélisation de la dispersion . . . 59
2.2.1 Variables de la dispersion turbulente . . . 59
2.2.2 Formules de Briggs . . . 63
2.2.3 Formule de Doury . . . 64
2.2.4 Théorie de la similitude . . . 65
2.3 Modélisation de la surhauteur . . . 68
2.3.1 Surhauteur de Briggs et HPDM . . . 69

2.3.2 Formule de Concawe . . . 70
2.3.3 Formule de Holland et Stümke . . . 70
2.3.4 Pénétration partielle dans la couche d'inversion . . . 71
2.4 Modélisation des processus de perte . . . 72
2.4.1 Décroissance radioactive . . . 72
2.4.2 Dépôt sec . . . 72
2.4.3 Lessivage . . . 74
2.5 Bilan : modèles de Polyphemus . . . 75

3 Evaluation des modèles à l'échelle locale 79
3.1 Critères d'évaluation . . . 81
3.1.1 Indicateurs statistiques . . . 81
3.1.2 Critères de performance . . . 81
3.2 Evaluation de la dispersion : Prairie Grass . . . 81
3.2.1 Présentation de l'expérience, et évaluation des écarts types . . . 82
3.2.2 Comparaison avec d'autres modèles gaussiens . . . 84
3.2.3 Statistiques par arc . . . 85
3.3 Evaluation avec surhauteur : Kincaid . . . 87
3.3.1 Présentation de l'expérience, et évaluation de la surhauteur . . . 87
3.3.2 Evaluation du modèle . . . 90
3.3.3 Comparaison avec d'autres modèles gaussiens . . . 95
3.4 Evaluation du dépôt : mesures de dépôt en bord de route . . . 96
3.4.1 Configuration . . . 98
3.4.2 Résultats moyens sur février 1997 . . . 98
3.4.3 Analyse des résultats sur certains jours . . . 100
3.5 Bilan : évaluation des modèles gaussiens et variabilité . . . 103
3.5.1 Performance des modèles . . . 104
3.5.2 Variabilité des paramétrisations et des résultats . . . 105

II Modélisation sous-maille des émissions ponctuelles : modèle de panache sous-maille 107

4 Modèle de panache sous-maille passif 109
4.1 Introduction . . . 111
4.1.1 Context . . . 111
4.1.2 Issues . . . 111
4.1.3 Methodology . . . 112
4.1.4 Outline . . . 112
4.2 Model description . . . 112
4.2.1 The Gaussian puff model . . . 113
4.2.2 Model coupling . . . 115
4.3 The ETEX case: impact of plume-in-grid at continental scale . . . 117
4.3.1 Modeling set-up . . . 117
4.3.2 Convergence with the time step between two puffs . . . 118
4.3.3 Comparison of results with/without the plume-in-grid approach . . . 118
4.4 Sensitivity to the local-scale dispersion . . . 122
4.4.1 Local-scale dispersion parameters . . . 123
4.4.2 Comparison between the Gaussian and Eulerian diffusion . . . 123
4.4.3 Vertical profiles . . . 125

4.5 Relevant timescales: a study on the injection time . . . 125
4.5.1 Impact on the concentrations . . . 125
4.5.2 Impact on the arrival times . . . 127
4.6 Influence of the grid resolution . . . 127
4.6.1 Impact of the initial volume of the emission . . . 128
4.6.2 Use of plume-in-grid with several grid resolutions . . . 130
4.7 Application to the Chernobyl disaster . . . 130
4.7.1 Modeling set-up . . . 130
4.7.2 Results . . . 133
4.8 Conclusions . . . 134
4.9 Appendix . . . 136
4.9.1 Coordinates . . . 136
4.9.2 Formulae to compute the puff integral over a volume . . . 136

5 Modèle de panache sous-maille réactif 139
5.1 Description du modèle . . . 141
5.1.1 Modèle de chimie pour les bouffées réactives . . . 141
5.1.2 Interaction entre deux bouffées superposées . . . 144
5.2 Application du panache sous-maille à la chimie de l'ozone . . . 149
5.2.1 Chimie de l'ozone . . . 149
5.2.2 Impact du panache sous-maille réactif . . . 152
5.2.3 Influence du nombre de sources . . . 157
5.3 Application à la qualité de l'air en Île-de-France . . . 157
5.3.1 Configuration . . . 157
5.3.2 Émissions . . . 158
5.3.3 Configuration de panache sous-maille . . . 160
5.3.4 Simulations . . . 160
5.4 Évaluation du modèle de panache sous-maille réactif . . . 160
5.4.1 Critères d'évaluation . . . 160
5.4.2 Impact du panache sous-maille sur les statistiques . . . 161
5.4.3 Répartition spatiale de l'impact . . . 163
5.5 Résultats pour certains jours . . . 165
5.5.1 Impact sur les polluants primaires . . . 166
5.5.2 Impact sur les polluants secondaires . . . 167
5.6 Analyse de sensibilité . . . 170
5.6.1 Influence de la diffusion verticale . . . 170
5.6.2 Influence de la modélisation locale . . . 173
5.7 Conclusions . . . 173
5.7.1 Chimie dans un panache . . . 173
5.7.2 Impact sur la photochimie en région parisienne . . . 175
5.7.3 Bilan et perspectives . . . 176

III Méthodes statistiques pour la réduction d'échelle et application à la qualité de l'air 177

6 Méthodes statistiques pour la réduction d'échelle 179
6.1 Introduction . . . 181
6.1.1 Historique . . . 181
6.1.2 Réduction d'échelle statistique et dynamique . . . 181

6.1.3 Une question d'échelles... . . . 181
6.2 Principes de la réduction d'échelle statistique . . . 182
6.2.1 Principe général . . . 182
6.2.2 Stationnarité . . . 183
6.2.3 Prédicteur et prédictand . . . 183
6.2.4 Représentativité des modèles . . . 184
6.2.5 Le problème de la variance . . . 184
6.3 Méthodes statistiques . . . 185
6.3.1 Pré-traitement des données . . . 185
6.3.2 Méthodes de régression . . . 187
6.3.3 Comparaison des méthodes . . . 190
6.4 Utilisation des méthodes statistiques pour la qualité de l'air . . . 190
6.4.1 Régression sur un ensemble de modèles . . . 191
6.4.2 Décomposition en composantes principales (PCA) . . . 192
6.4.3 Décomposition en composantes principales ajustées (PFC) . . . 193
6.5 Conclusion . . . 194
6.5.1 Méthodes de réduction d'échelle . . . 194
6.5.2 Questions . . . 194

7 Réduction d'échelle pour la qualité de l'air à l'échelle européenne 195
7.1 Application : qualité de l'air à l'échelle européenne . . . 197
7.1.1 Les réseaux de mesures . . . 197
7.1.2 Les simulations . . . 199
7.1.3 Méthode statistique utilisée : régression linéaire . . . 199
7.2 Résultats par type de station . . . 199
7.2.1 Régressions sur un seul modèle de référence (prévision à 1 jour) . . . 200
7.2.2 Résultats pour la prévision à N jours et stationnarité des coefficients . . . 202
7.3 Prise en compte de l'ensemble . . . 206
7.3.1 Performance en fonction du nombre de modèles . . . 206
7.3.2 Colinéarité . . . 210
7.4 Réduction de dimension . . . 213
7.4.1 Utilisation d'ensembles calibrés . . . 213
7.4.2 Réduction en composantes principales (PCA) . . . 214
7.4.3 Réduction en composantes principales ajustées (PFC) . . . 217
7.4.4 Comparaison des différentes méthodes . . . 223
7.5 Conclusion . . . 223
7.5.1 Résultats aux stations . . . 223
7.5.2 Représentativité des stations . . . 223

Conclusion 227

A Indicateurs statistiques 231
A.1 Indicateurs pour les modèles à l'échelle locale . . . 231
A.2 Indicateurs à l'échelle régionale et continentale . . . 231
A.3 Indicateurs utilisés pour l'ozone . . . 232

B Performance de réseaux de détection à différentes échelles 233
B.1 Introduction . . . 235
B.1.1 Context and objectives . . . 235
B.1.2 Methodology and outline . . . 235
B.2 Test scenarios . . . 236
B.2.1 Domains and computing strategy . . . 236
B.2.2 Sources . . . 238
B.2.3 Meteorological scenarios . . . 239
B.2.4 Simulations . . . 239
B.3 Sensor and network statistics . . . 241
B.3.1 Statistical indicators for sensors . . . 241
B.3.2 Constraints on the sensors . . . 242
B.3.3 Statistical indicators for networks . . . 243
B.4 Sensors performance . . . 243
B.4.1 Influence of scenarios and meteorology . . . 243
B.4.2 Emitted quantities and loss processes . . . 246
B.5 Networks performance . . . 246
B.5.1 Networks defined with respect to detection probability . . . 246
B.5.2 Networks defined as concentric circles . . . 249
B.5.3 Networks sensitivity to wind variations . . . 253
B.6 Conclusions . . . 253

Chapitre 1

Changements d'échelle en qualité de l'air

L'évolution des polluants dans l'atmosphère dépend de phénomènes variés, tels que les émissions, la météorologie, la turbulence ou les transformations physico-chimiques, qui ont des échelles caractéristiques spatiales et temporelles très diverses. Il est très difficile, par conséquent, de représenter l'ensemble de ces échelles dans un modèle de qualité de l'air. Les modèles de chimie-transport, couramment utilisés, ont une résolution bien supérieure à la taille des plus petites échelles (de quelques kilomètres à quelques dizaines de kilomètres). Tout processus de taille caractéristique inférieure à celle de la maille est donc mal représenté par ces modèles. On parle alors de « processus sous-maille », que les méthodes de changement d'échelle, abordées dans ce chapitre, visent à mieux représenter. La variabilité sous-maille liée aux processus non résolus est en effet une source d'erreur, en particulier lorsqu'il s'agit de comparer les résultats d'un modèle à des observations effectuées en des points de mesure. L'échelle représentative des stations d'observation est en effet typiquement inférieure à la taille d'une maille.

Ce chapitre présente les différentes échelles en jeu en modélisation de la qualité de l'air. Il s'agit de définir quelles échelles sont mal représentées par les modèles, et d'effectuer une revue des différentes méthodes possibles pour quantifier ces limitations et mieux prendre en compte les petites échelles :
- *les méthodes de raffinement de modèles, qu'il s'agisse de descendre en résolution ou de coupler des modèles à deux échelles différentes,*
- *les paramétrisations sous-mailles qui visent à mieux représenter un phénomène sous-maille par des paramétrisations physiques,*
- *les méthodes stochastiques, qui décrivent la variabilité sous-maille comme un processus stochastique représenté par une densité de probabilité.*

Les choix effectués lors de la thèse sont enfin explicités en étant remis dans le contexte très large des changements d'échelle en qualité de l'air.

Sommaire

1.1 Différentes échelles **13**
1.1.1 Les émissions 13
1.1.2 Le transport 16
1.1.3 La turbulence 18

1.1.4 Les processus de perte 20
1.2 Une hiérarchie de données et de modèles 25
1.2.1 Équation de dispersion réactive 25
1.2.2 Différents types de modèles 28
1.2.3 Données d'entrée 30
1.2.4 Sensibilité des modèles aux données d'entrée 32
1.3 Différentes méthodes de changements d'échelles 32
1.3.1 Méthodes numériques et couplage de modèles 33
1.3.2 Fermeture d'ordre supérieur 36
1.3.3 Paramétrisations sous-mailles 38
1.3.4 Méthodes stochastiques 40
1.3.5 Prise en compte de la variabilité sous-maille a posteriori 42
1.4 Conclusion 43
1.4.1 Bilan 43
1.4.2 Choix des méthodes 44
1.4.3 Plan de la thèse 45
1.4.4 Publications associées 45

La description des processus physiques et chimiques et des équations associées s'inspire largement de
– Sportisse [2007a],
– Seinfeld et Pandis [1998],
– Jacob [1999]

La modélisation de la qualité de l'air a pour enjeu de prévoir la concentration de polluants dans l'atmosphère. Il peut s'agir par exemple de prévoir le dépassement de seuils de concentration au-delà desquels les polluants sont potentiellement dangereux, ou de modéliser l'impact d'installations industrielles ou encore du trafic routier. Les modèles peuvent servir d'aide à la décision, en permettant de simuler l'effet des mesures de réduction d'émission. Enfin, ils peuvent être utilisés en situation de crise, en cas de rejet accidentel de composés dangereux dans l'atmosphère, pour prendre des mesures appropriées. La diversité de ces applications montre déjà qu'il existe différentes échelles spatiales et temporelles en jeu. En effet, l'étude de l'impact d'une installation industrielle par exemple, pourra se focaliser sur un champ relativement proche (quelques kilomètres). Pour modéliser un polluant tel que l'ozone, ou un rejet accidentel de radionucléides, on se place en revanche à l'échelle continentale. Il est donc nécessaire de disposer d'une variété de modèles afin de répondre aux différents besoins, et de connaître leurs domaines d'application et leurs limitations. En d'autres termes, pour chaque type de modèle, il y aura une « échelle bien représentée », que l'on peut essayer de définir. De même, pour un point de mesure, on pourra définir une « échelle de représentativité », qui correspond au domaine où l'évolution des concentrations est bien représentée par les observations à la station.

Il n'est pas nécessaire, ni forcément souhaitable, de n'avoir qu'un seul modèle performant à toutes les échelles. Toutefois, on peut améliorer les modèles les plus couramment utilisés, et leur représentation des petites échelles. Mieux représenter certaines échelles n'aura pas toujours un impact sur les résultats du modèle. Par exemple, mieux représenter les petites échelles de la turbulence dans un panache en sortie de cheminée n'aura pas d'effet sur les concentrations observées à une station de mesure, si celui-ci est placé à 10 km de la source. Il s'agit donc d'estimer à la fois l'échelle d'un processus et celle de son impact sur les résultats du modèle.

1.1 Différentes échelles

La concentration des polluants dans l'atmosphère dépend de nombreux processus physiques, qui ont des échelles caractéristiques variées. Il s'agit de modéliser l'émission des polluants, leur transport dans l'atmosphère (nécessitant la connaissance des champs météorologiques), leurs transformations chimiques ainsi que l'interaction avec le sol, notamment au travers de l'absorption par les végétaux et du dépôt sur le sol et les bâtiments. Les nuages et la pluie peuvent également interagir avec le transport des polluants, notamment au travers du lessivage (entraînement des polluants par la pluie, qui se déposent alors au sol). Dans tous ces processus, les échelles de temps et/ou d'espace sont très diverses et certaines sont en général mal résolues par les modèles.

1.1.1 Les émissions

Parmi les espèces participant à la qualité de l'air, on distingue les *espèces primaires*, directement émises, et les *espèces secondaires* qui sont formées dans l'atmosphère par le biais de transformations physico-chimiques (par exemple l'ozone). Il existe principalement deux types d'émissions :

- Les *émissions biogéniques*, provenant de sources naturelles telles que les volcans, les feux de forêts, l'émission de sels de mer ou l'érosion du sable, ou encore les composés organiques volatiles (COV) émis en particulier par la photosynthèse,
- Les *émissions anthropiques*, dues à l'activité humaine, c'est-à-dire au transport (routier et aérien), aux émissions industrielles, au chauffage et à l'agriculture.

La part des différents types d'émissions dans les concentrations d'espèces primaires dépend du type de polluant considéré [Sportisse, 2007a]. On se focalise dans cette partie sur les émissions anthropiques, dont la variabilité spatiale est importante en comparaison des émissions biogéniques, plus diffuses (à l'exception des feux de forêt).

Variabilité horizontale

La répartition spatiale des émissions anthropiques est très hétérogène et dépend de la disposition du réseau routier et des installations industrielles, en particulier. Les concentrations de polluants en champ proche de ces sources (typiquement à une centaine de mètres d'une route ou d'une cheminée) sont beaucoup plus élevées que les concentrations moyennes. Par exemple, les concentrations de dioxyde d'azote NO_2 sont environ deux fois plus importantes dans les 50 m autour des routes que les concentrations mesurées par les stations de fond [Hewitt, 1991]. De même, les concentrations de soufre sont divisées par deux entre une distance de 50 m et 150 m d'une route [Reponen *et al.*, 2003]. Ce rapport est encore plus important si l'on considère les particules ultra-fines, dont les concentrations diminuent très rapidement lorsque la distance à la source augmente, notamment à cause de la coagulation [Zhu *et al.*, 2002]. La figure 1.1 illustre cette variabilité spatiale des concentrations de NO_2 due au réseau routier. Il s'agit de l'Île-de-France, pour l'année 2007. Les concentrations en bord de route sont effectivement deux à trois fois plus importantes que les concentrations de fond, et dépassent le seuil réglementaire pour les concentrations de fond, qui est de 40 $\mu g\,m^{-3}$. Ces dépassements de seuil sont très localisés (environ 2% de la surface de la région) mais peuvent poser des problèmes sanitaires importants pour les personnes exposées.

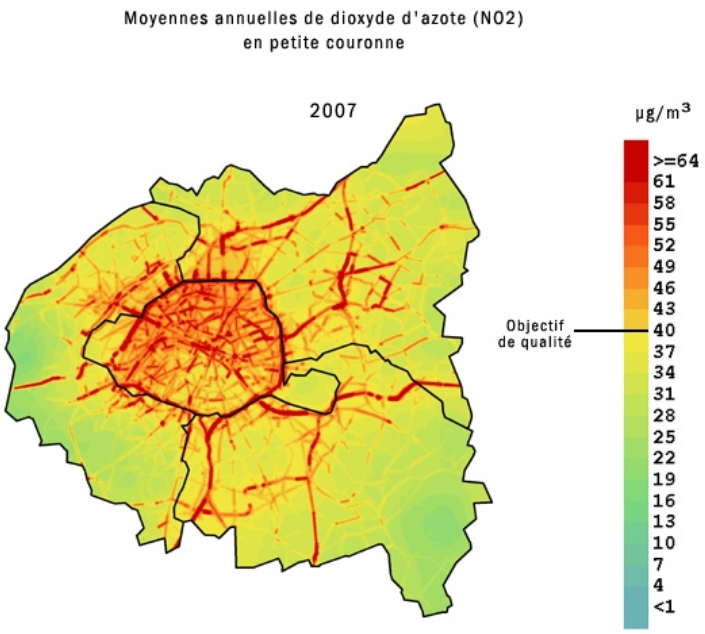


FIG. 1.1 – Concentrations moyennes de NO_2 sur l'Île-de-France (petite couronne) pour l'année 2007. Source : Airparif.

Une autre illustration du type de problèmes concrets que peut poser cette variabilité spatiale est une discussion récente ayant eu lieu à l'EPA (Environment Protection Agency) aux États-Unis [EPA, 2009]. Il s'agissait de modifier les normes utilisées pour réglementer le dioxyde

d'azote (NO_2), en passant d'un objectif annuel à une réglementation basée sur les concentrations horaires. Le comité scientifique chargé de faire des propositions concernant le réseau de mesures utilisé (jusqu'alors, uniquement des stations de fond) s'est divisé autour de deux propositions :

1. L'installation d'un réseau complémentaire de l'existant, comprenant des stations de mesure en bord de route utilisées pour mesurer les dépassements de seuil,
2. L'utilisation du seul réseau de stations de fond pour les dépassements de seuils, les stations de trafic étant uniquement pour la recherche.

Les arguments en faveur de la seconde proposition étaient basés d'une part sur l'absence d'études épidémiologiques basées sur les concentrations en bord de route, et d'autre part sur la difficulté de choisir des emplacements de capteurs représentatifs de ces concentrations. Cette difficulté est due à la variabilité spatiale induite par les bâtiments, mais aussi par le type de trafic, par exemple à cause des filtres à particules pour véhicules diesel, qui peuvent augmenter la fraction de NO_2 dans les NOx[1] émis. Il peut donc y avoir localement des pics de NO_2 dans les endroits où des véhicules plus « propres » sont utilisés. La question se posait donc de définir l'échelle de représentativité des capteurs utilisés pour mesurer les pics de concentration et les dépassements de seuil. Il était notamment proposé d'utiliser la modélisation afin de déterminer les emplacements de mesure adéquats, ainsi que les liens entre les concentrations de fond et les concentrations en bord de route.

Cet exemple permet de montrer que la variabilité spatiale des émissions pose des problèmes non seulement de modélisation, mais aussi de mesure, autour de la question d'« échelle de représentativité ». Cela permet également d'illustrer l'importance d'utiliser des modèles permettant de résoudre les échelles caractéristiques des émissions, notamment pour l'impact sanitaire des routes.

Variabilité verticale

La répartition spatiale horizontale des émissions est donc très hétérogène et son échelle caractéristique est de quelques centaines de mètres, voire moins. De plus, il existe également une hétéorogénéité sur la verticale. Ainsi, pour les sources au sol telles que les routes, les fortes concentrations sont principalement localisées dans les quelques premiers mètres au-dessus du sol, alors que de nombreux modèles supposent un mélange homogène sur une hauteur de quelques dizaines de mètres. De même, le panache émis par une cheminée n'est pas mélangé sur la verticale avant une certaine distance en aval de la source. Pour les sources en hauteur, le panache ne touche donc pas le sol immédiatement, contrairement à ce qui se passerait si le panache était mélangé immédiatement sur la verticale. Cette variabilité verticale est importante à proximité des sources, car le panache n'est mélangé sur la verticale qu'au bout de quelques kilomètres (la distance dépendant de l'efficacité du mélange, donc des conditions météorologiques).

Autres processus

Si l'échelle caractéristique des émissions est mal représentée, une partie de l'information sur leur répartition spatiale est perdue. Cela a des conséquences directes sur la modélisation des concentrations de polluants, mais aussi indirectes. Ainsi, les transformations chimiques des polluants dépendent des concentrations des différentes espèces en jeu. Si les espèces émises sont plus concentrées et moins bien mélangées aux espèces de fond, cela peut modifier les réactions et par exemple retarder la formation de certains composés secondaires (voir partie 1.1.4). De plus, le vent varie sur la verticale, et sa direction n'est pas forcément la même à la hauteur d'une cheminée de 100 m, ou proche du sol. Modéliser la hauteur exacte du panache permet d'y associer

[1] Dans la suite, on appellera NOx les oxydes d'azote, qui regroupent NO et NO_2.

les conditions météorologiques prises à la bonne hauteur, correspondant à son transport et à sa diffusion. La figure 1.2 donne un exemple (un peu extrême) de l'importance de la prise en compte de la météorologie à la hauteur exacte d'émission d'un panache. Cependant, il serait très difficile en pratique de bien modéliser une telle situation : cela ne nécessite pas seulement de connaître les hauteurs exactes des cheminées, mais également de disposer de données météorologiques précises à l'emplacement de la source.

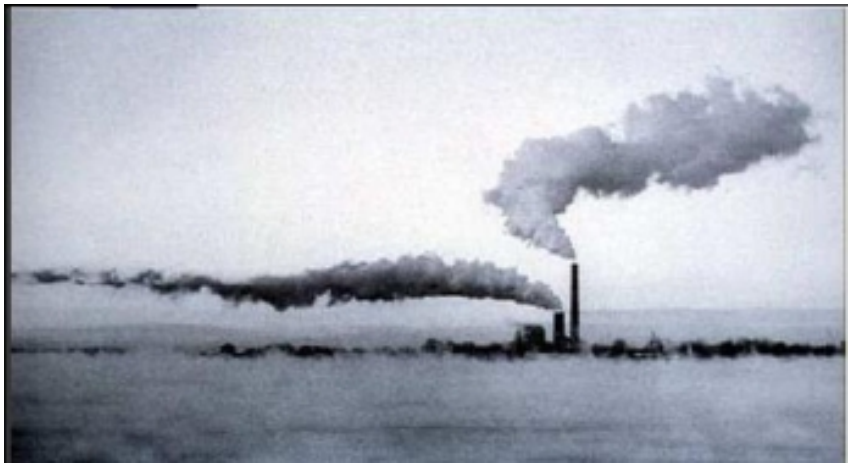


Fig. 1.2 – Illustration de l'importance de prendre en compte la météorologie à la hauteur exacte d'émission, et non à une hauteur moyenne : deux panaches émis à des hauteurs différentes partent dans des directions opposées.

1.1.2 Le transport

Dans cette partie, on décrit le transport (modélisé par l'*advection*), c'est-à-dire le déplacement des polluants dans l'atmosphère, sans modifier les concentrations.

Échelles des phénomènes atmosphériques

Structure verticale de l'atmosphère L'atmosphère est divisée en plusieurs couches, définies à partir du profil vertical de température. La première partie de l'atmosphère (la plus proche du sol) est la *troposphère*, d'une hauteur de 8 à 18 km. Cette partie est caractérisée par une décroissance de la température avec l'altitude. Au-dessus, se situe la *stratosphère*, jusqu'à environ 50 km, où le gradient de température s'inverse et devient positif, notamment grâce à l'absorption du rayonnement solaire par l'ozone. Ces deux couches et les suivantes sont représentées sur la figure 1.3.

En qualité de l'air, les polluants émis dans l'atmosphère sont transportés principalement à l'intérieur de la *couche limite atmosphérique* (appelée ensuite CLA ou juste « couche limite »). Il s'agit de la couche la plus basse de la troposphère, qui est directement influencée par le sol, et dont la hauteur varie entre quelques centaines de mètres et 2 km environ (cf partie 2.2.1). La partie de l'atmosphère située au-dessus de la couche limite est appelée « troposphère libre » (libre de l'influence du sol).

Mouvements horizontaux dans l'atmosphère Le transport horizontal des polluants dépend avant tout de la météorologie, et en particulier du vent (pour le transport horizontal) et de la

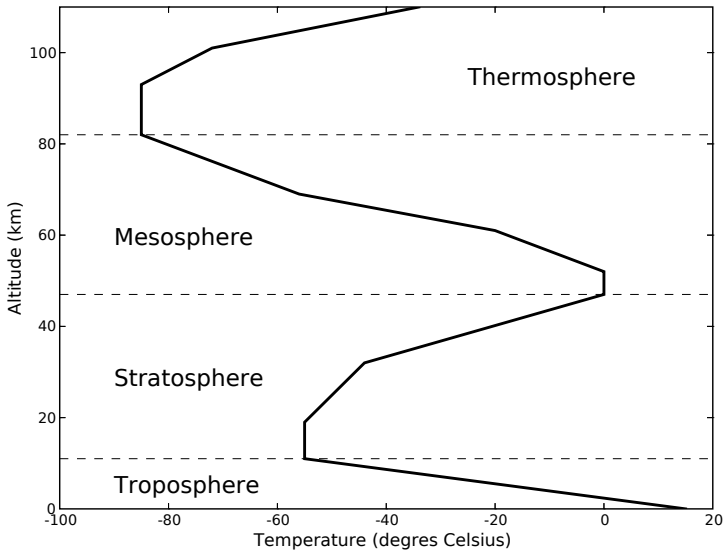


FIG. 1.3 – Profil vertical de température typique, et principales couches de l'atmosphère. Source : Mallet [2005].

convection (pour le transport vertical). En météorologie, on peut distinguer des échelles spatiales et temporelles qui sont caractéristiques de différents phénomènes atmosphériques (tableau 1.1). La micro-échelle et la méso-échelle correspondent aux phénomènes atmosphériques prépondérants dans le transport des polluants aux échelles considérées ici. L'échelle synoptique est définie comme l'échelle à laquelle les phénomènes atmosphériques, et en particulier le vent, sont indépendants de la surface terrestre, et ne dépendent plus que de la force de Coriolis et du gradient de pression (*approximation géostrophique*). Les mouvements synoptiques se situent au-dessus de la couche limite atmosphérique, et correspondent typiquement au transport de polluants à l'échelle intercontinentale.

Échelle	Taille caractéristique	Phénomène atmosphérique
Micro-échelle	< 1 km	Météorologie locale, turbulence
Méso-échelle	1 km–1000 km	Mouvements thermiques ascendants, orage
Échelle synoptique	10^4 km	Mouvements synoptiques

TAB. 1.1 – Échelles spatiales caractéristiques des phénomènes atmosphériques.

Transport horizontal

Le transport horizontal est principalement dû au vent. L'ordre de grandeur de la vitesse du vent zonal (est/ouest) est $U = 10\,\mathrm{m\,s^{-1}}$, tandis que le vent méridional (sud/nord) a une vitesse typique plus faible, de l'ordre de 1 ou $2\,\mathrm{m\,s^{-1}}$. Ces ordres de grandeurs permettent de caractériser les échelles temporelles typiques associées à l'échelle spatiale ciblée pour une application. Ainsi, le temps de transport continental (est/ouest) est de quelques jours, et d'environ 2 semaines pour un mélange intercontinental et 2 mois pour le mélange hémisphérique dans la

direction sud/nord [Jacob, 1999].

La description du transport des polluants nécessite une bonne connaissance des champs météorologiques, avec une résolution suffisante pour décrire la variabilité horizontale et verticale. En effet, la vitesse du vent dépend fortement de l'altitude, étant donné la présence de cisaillement dû au frottement au sol. De plus, il existe également une variabilité spatiale importante à micro-échelle, due notamment à la présence d'obstacles et à l'influence du relief. En effet, la proximité du relief ou encore de côtes modifie l'écoulement de différentes façons [Perkins *et al.*, 2002]. :

- La déviation de l'écoulement, qui passe au-dessus ou éventuellement contourne un relief,
- Les écoulements thermiques forcés par le relief (brise de vallée ou brise de mer et de terre),
- Le sillage aérodynamique en aval du relief.

De même, le climat urbain est spécifique, d'une part à cause de la présence des bâtiments qui constituent des obstacles à l'écoulement, mais également du fait qu'une ville a une température plus élevée de quelques degrés que le milieu rural qui l'entoure, notamment de nuit (*îlot de chaleur urbain*). Cela provient des sources de chaleurs anthropiques (chauffage par exemple), ainsi que du comportement radiatif spécifique du milieu bâti.

Transport vertical et convection

Le vent vertical étant négligeable, ce sont d'autres mécanismes qui se chargent du mélange des polluants sur la verticale. Il s'agit principalement de la turbulence, c'est-à-dire des mouvements tourbillonnaires balayant un large spectre d'échelles. Ce point est détaillé dans la partie 1.1.3. Le transport vertical turbulent a une origine thermique liée à l'effet de *flottabilité* (mouvements turbulents ascendants dus à un gradient de température, voir partie 2.2.1). Dans la couche limite convective, on peut alors distinguer des zones de mouvements d'air ascendants, entourées de zones plus grandes de subsidence, où l'air redescend. A l'intérieur d'une « colonne » ascendante (« updraft ») ou descendante (« downdraft »), les polluants sont transportés très rapidement sur toute la hauteur de la couche limite. Cela contribue à la ségrégation des polluants, car l'échange entre les mouvements ascendant et descendant se fait avec un temps caractéristique beaucoup plus long que le mélange vertical à l'intérieur d'une colonne.

Lorsque les mouvements convectifs ont lieu à l'échelle synoptique (passage d'une masse d'air froid au-dessus de l'air plus chaud), on parle de « convection profonde ». Ces mouvements peuvent être à l'origine d'orages si le contenu en eau de l'air ascendant est important, voire de tornades. Ce phénomène s'apparente à la turbulence d'origine thermique, mais il s'agit ici des plus grandes échelles de la turbulence, à savoir les tourbillons qui, en raison de leur taille, sont responsables de l'advection des polluants et non de leur diffusion [Corrsin, 1974].

1.1.3 La turbulence

La couche limite atmosphérique est un milieu extrêmement turbulent. Cela signifie que des mouvements à l'intérieur de l'air, qui peuvent être visualisés comme des tourbillons (figure 1.4), dissipent l'énergie cinétique de l'écoulement. Ces tourbillons participent à la dispersion des polluants dans l'atmosphère, notamment sur la verticale. L'une des caractéristiques d'un écoulement turbulent est la grande variété d'échelles spatiales et temporelles mises en jeu. L'influence de la turbulence sur la dispersion des polluants est représentée en général dans les modèles par le mécanisme de *diffusion*. C'est ce mécanisme qui est responsable de la dilution des polluants et donc de la baisse des concentrations en aval des sources.

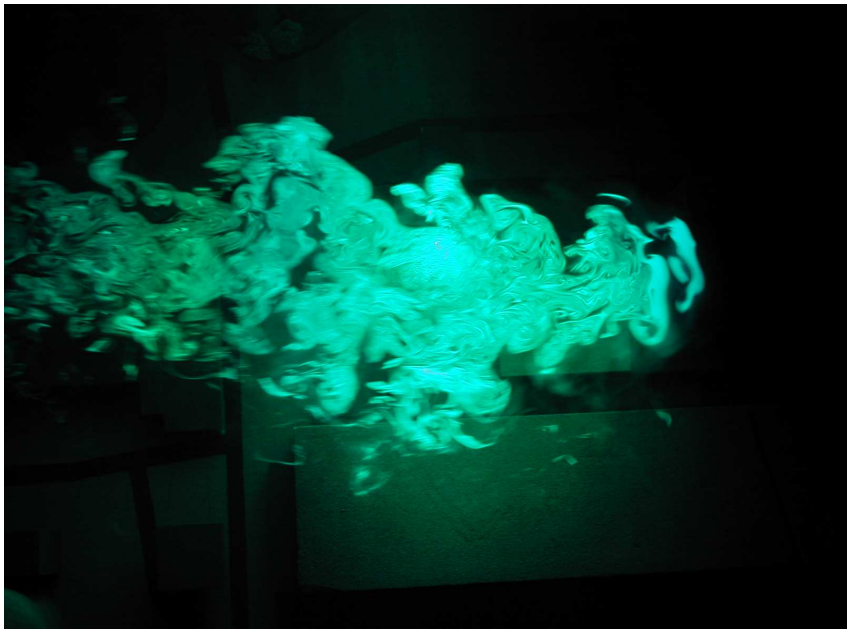


FIG. 1.4 – Visualisation de la turbulence dans un écoulement en soufflerie. Soufflerie de Centrale Lyon, écoulement autour de la maquette de la centrale nucléaire du Bugey. Crédit : Centrale Lyon et Institut de radioprotection et de sûreté nucléaire (O. Isnard).

Échelles de la turbulence

Un processus classiquement mis en avant dans les écoulements turbulents est la *cascade d'énergie* : l'énergie cinétique est transférée des grandes échelles aux les petites échelles lorsque les grands tourbillons se divisent en tourbillons plus petits. A la plus petite échelle, appelée *échelle de Kolmogorov*, l'énergie est dissipée par la diffusion moléculaire [Kolmogorov, 1991]. Ce phénomène de cascade contribue au brassage des polluants dans l'atmosphère. Il a été décrit de façon théorique par Richardson, puis Kolmogorov, et vérifié expérimentalement pour les écoulements atmosphériques [Lumley et Panofsky, 1964]. Les petites échelles de la turbulence sont caractérisées uniquement par le taux de dissipation de l'énergie turbulente ϵ_k et par la viscosité dynamique de l'air ν. On définit alors les échelles de Kolmogorov, de longueur η et de vitesse v_k

$$\eta \simeq \nu^{3/4}\epsilon_k^{-1/4} \qquad v_k \simeq \nu^{1/4}\epsilon_k^{1/4}. \tag{1.1}$$

L'analyse de Kolmogorov donne une estimation du ratio entre l'échelle spatiale représentative des grandes échelles de l'écoulement L, et l'échelle de Kolmogorov η

$$\frac{L}{\eta} \simeq Re^{3/4}, \tag{1.2}$$

avec Re le nombre de Reynolds caractéristique de l'écoulement

$$Re = \frac{UL}{\nu}, \tag{1.3}$$

U étant la vitesse du vent. Pour des valeurs typiques de la couche limite atmosphérique ($U = 10\,\mathrm{m\,s^{-1}}$ et $\nu \simeq 10^{-5}\,\mathrm{m^2\,s^{-1}}$), on aura $Re \simeq 10^9$, ce qui est représentatif d'un écoulement très turbulent. Dans un tel milieu, le rapport entre les plus grandes et les plus petites échelles de la turbulence donné par l'équation 1.2 est donc très grand, et il est donc très difficile pour un

modèle de représenter l'ensemble des échelles. L'échelle « bien représentée » par les modèles est donc liée notamment à l'échelle de la turbulence modélisée.

Impact de la turbulence sur la dispersion des polluants

Toutes les échelles de la turbulence n'auront pas le même impact sur la dispersion des polluants. Une façon de visualiser l'impact de la turbulence sur la dispersion est de prendre l'exemple d'un nuage de polluants (ou « bouffée »), émis instantanément par une source ponctuelle (figure 1.5). Les tourbillons dont la taille caractéristique est beaucoup plus grande que la taille de la bouffée ne participent pas à sa diffusion, ils « voient » la bouffée comme une particule de fluide qui est transportée comme le reste de l'écoulement. Les tourbillons de taille beaucoup plus petite que la bouffée participent également très peu au mélange. Les petits tourbillons situés à la frontière de la bouffée entraînent de l'air frais petit à petit, induisant une dispersion uniforme du panache. Les tourbillons dont la taille caractéristique est du même ordre de grandeur que celle du panache ont la plus grande contribution à la dispersion du panache, qui est déformé sous l'effet de la turbulence non homogène. En pratique, un nuage de polluants est toujours dispersé dans l'atmosphère, car il existe toujours des tourbillons de taille inférieure à celle du nuage. L'efficacité du mélange turbulent dépend donc à la fois de la turbulence et de l'échelle caractéristique du nuage de polluant, qui varie en fonction de la distance à la source.

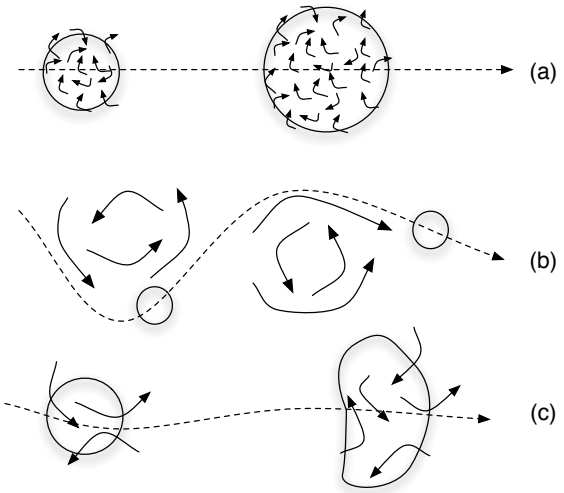


Fig. 1.5 – Dispersion d'une bouffée de polluants, dans trois situations : (a) la taille caractéristique des tourbillons est beaucoup plus petite que celle de la bouffée, (b) la taille caractéristique des tourbillons est beaucoup plus grande que celle de la bouffée, et (c) les tailles caractéristiques des tourbillons et de la bouffée sont comparables. D'après Seinfeld et Pandis [1998].

1.1.4 Les processus de perte

En plus des émissions et du transport dans l'atmosphère, de nombreuses espèces subissent des transformations physico-chimiques ainsi que d'autres processus de pertes :

- la filiation radioactive (pour les radioéléments),
- la destruction (ou production) chimique ou photochimique (pour les espèces réactives),
- la condensation/évaporation, et la coagulation (pour les particules),

- le dépôt sec à la surface du sol (dû notamment à l'absorption par les végétaux),
- la sédimentation gravitationnelle (pour les plus grosses particules),
- le lessivage humide par la pluie ou à l'intérieur des nuages.

Dépôt et lessivage

La plupart des processus de perte ont des échelles caractéristiques plus petites que celles représentées par les modèles. Il convient de citer notamment le dépôt sec et le lessivage, qui dépendent respectivement du type de sol (bâtiments, végétaux) et de la présence de nuages. Dans les deux cas, l'échelle caractéristique de la variabilité spatiale peut être de quelques mètres seulement. L'information sur l'occupation des sols est relativement facile à obtenir, mais contient des incertitudes, notamment car elle varie au cours des saisons. La présence de nuages et de pluie en un endroit précis est plus difficile à diagnostiquer. Par ailleurs, la présence des nuages a également un impact indirect sur la photochimie, ainsi que sur la turbulence d'origine thermique, puisque le rayonnement solaire est diminué. L'impact des nuages sur les concentrations de polluants est donc typiquement un phénomène mal représenté dans les modèles.

Chimie de l'atmosphère

Parmi les réactions chimiques ayant lieu entre les espèces présentes dans l'atmosphère, on distingue principalement deux types de réactions :

1. Les réactions *photochimiques*, qui représentent l'interaction entre une molécule et un photon. La plus notable est la dissociation photolytique

$$\mathrm{X} + \mathrm{h}\nu \rightarrow \mathrm{X}^{*}. \tag{R 1.1}$$

   Dans cette réaction, hν représente le rayonnement absorbé. On peut alors calculer le taux de disparition de l'espèce en jeu, de concentration c_X :

$$\frac{\mathrm{d}c_X}{\mathrm{dt}} = -J_r c_X, \tag{1.4}$$

   où J_r est la *constante photolytique* de la réaction, qui dépend de l'espèce et du rayonnement solaire. Ce type de réaction est dit du *premier ordre*.

2. Les réactions chimiques du *second ordre*, qui font intervenir deux espèces A et B dans une réaction du type

$$\mathrm{A} + \mathrm{B} \rightarrow P. \tag{R 1.2}$$

   Dans ce cas, la vitesse de disparition d'une espèce (ou taux de réaction) dépendra des concentrations des espèces A et B, ainsi que de la *constante de réaction* k :

$$\frac{\mathrm{d}c_A}{\mathrm{dt}} = -k c_A c_B. \tag{1.5}$$

Notons également que la filiation radioactive peut, en première approximation, s'apparenter à une chimie du premier ordre, où le taux de disparition de l'espèce radioactive X s'écrit

$$\frac{\mathrm{d}c_X}{\mathrm{dt}} = -\lambda c_X, \tag{1.6}$$

avec λ le coefficient de décroissance radioactive, qui dépend de l'espèce.

Le temps de vie chimique τ_{chem} d'une espèce dépend donc de l'ensemble des réactions de production et de consommation de cette espèce. Cela fait intervenir la réactivité propre de l'espèce, mais également la température (dont dépendent les constantes de réaction), ainsi que les concentrations des autres espèces, donc le milieu environnant. Les réactions chimiques faisant intervenir les différentes espèces de l'atmosphère, et notamment l'ozone et les NOx, sont développées dans le chapitre 5.

La cinétique chimique dépendant des concentrations de polluants dans l'atmosphère, les hétérogénéités de concentrations influent donc sur le taux de production et de consommation des espèces. Ces hétérogénéités proviennent principalement

1. de la variabilité spatiale des émissions (partie 1.1.1),
2. de la turbulence (partie 1.1.3).

Influence des émissions sur la chimie La variabilité spatiale des émissions a une influence directe sur les concentrations, et la chimie est donc différente à proximité des sources. Ainsi, la chimie dans un panache émis par une cheminée est très différente de la chimie « de fond » : proche de la source, les espèces émises sont peu mélangées avec les espèces de fond, et la composition du panache est différente de celle de l'air ambient. Les réactions chimiques prépondérantes ne sont donc pas les mêmes que lorsque le panache est davantage dispersé, en aval de la source [Karamchandani *et al.*, 1998].

Influence de la turbulence sur la chimie L'influence de la turbulence sur les concentrations provient du mélange effectué par la turbulence, crucial dans le cas où les espèces ne sont pas mélangées au départ. Typiquement, il s'agit de la réaction entre une espèce émise au sol, à faible ou moyenne durée de vie, et une espèce au temps de vie plus long, présente dans l'atmosphère. Si le mélange dû à la turbulence n'est pas suffisamment efficace, les réactions chimiques peuvent être retardées par un effet dit de *ségrégation* [Vilà-Guerau de Arellano *et al.*, 2004]. On peut définir un nombre de Damköhler, noté Da, qui donne une indication de l'effet de ségrégation :

$$Da = \frac{\tau_{\text{turb}}}{\tau_{\text{chem}}}, \tag{1.7}$$

avec τ_{turb} le temps caractéristique de la turbulence, et τ_{chem} celui de la chimie. Si $Da \ll 1$, le mélange turbulent est beaucoup plus rapide que la cinétique chimique, et le mélange peut être considéré comme homogène au regard de la chimie. Inversement, si $Da \gg 1$, le mélange turbulent est très lent, et la cinétique chimique aura atteint un équilibre avant que la turbulence ne redistribue les polluants. Finalement, lorsque $Da \simeq 1$, la chimie et la turbulence ont le même ordre de grandeur et leurs interactions sont importantes.

En réalité, comme le montre la partie 1.1.3, il n'existe pas une seule échelle caractéristique de la turbulence τ_{turb} , mais une cascade d'échelles. Ainsi, on peut définir deux nombres de Damköhler (Molemaker et Vilà-Guerau de Arellano [1998], Vilà-Guerau de Arellano *et al.* [2004]) :

– le nombre de Damköhler caractéristique des grandes échelles de la turbulence

$$Da_t = \frac{\tau_t}{\tau_{\text{chem}}}, \tag{1.8}$$

avec τ_t représentant le temps caractéristique des grands tourbillons générés par la convection. Ce temps peut être défini, dans une couche limite convective, comme le rapport entre la hauteur de couche limite h et l'échelle de vitesse convective w_* [Deardorff, 1970],

– le nombre de Damköhler caractéristique des petites échelles de la turbulence, donc des échelles de Kolmogorov

$$Da_k = \frac{\tau_k}{\tau_{\text{chem}}}, \tag{1.9}$$

avec, d'après l'équation 1.1,

$$\tau_k = \frac{\eta}{v_k} = \left(\frac{\nu}{\epsilon_k}\right)^{1/2}. \tag{1.10}$$

Suivant l'ordre de grandeur de ces deux nombres de Damköhler, les réactions chimiques peuvent être limitées soit par toutes les échelles de la turbulence (chimie très rapide, $Da_k \gg 1$), soit uniquement par les petites échelles ($Da_k < 1 < Da_t$). Dans le cas d'une chimie très lente, le mélange a le temps de se faire par toutes les échelles de la turbulence ($Da_t \ll 1$).

Temps caractéristique (s)	Espèce chimique	Phénomène atmosphérique
10^{-1}–1	OH, HO_2	Turbulence
10^2–10^3	NO, NO_2	Mouvements thermiques ascendants
10^3–10^4	CH_2O, Isoprène	Convection, orage
10^5–10^6	H_2O, SO_2	Mouvements synoptiques
10^7	O_3 troposphérique, CO	Circulation générale
10^8–10^9	CH_4	Changement climatique

Tab. 1.2 – Échelles temporelles caractéristiques des transformations chimiques de certaines espèce chimiques, et de certains phénomènes atmosphériques susceptibles d'affecter le mélange de ces espèces. Source : Vilà-Guerau de Arellano *et al.* [2004].

Il convient enfin de noter que l'efficacité des réactions chimiques dépend non seulement du mélange turbulent mais également des proportions respectives des réactifs (notamment si l'une des deux espèces est en excès), ainsi que de l'équilibre chimique. Si l'équilibre est atteint, on peut considérer que les espèces apparaissent et disparaissent à la même vitesse. En pratique, l'équilibre chimique est perturbé notamment par les émissions [Krol *et al.*, 2000], mais aussi par exemple par la présence de nuages [Vilà-Guerau de Arellano et Cuijpers, 2000]. Enfin, différentes études, notamment Auger et Legras [2007], tendent à montrer que l'hétérogénéité spatiale des émissions a un impact prépondérant sur la chimie, en regard duquel la ségrégation due à la turbulence est beaucoup plus faible.

Temps de résidence des espèces

Pour déterminer le temps de résidence dans l'atmosphère d'une espèce, il faut prendre en compte non seulement la chimie, mais également l'ensemble de ses processus de perte. Le temps caractéristique de résidence d'une espèce peut se calculer assez facilement, connaissant les vitesses de disparition caractéristiques de chacun des processus de perte. L'ordre de grandeur des temps de résidence dans l'atmosphère pour les principales espèces est donné par la figure 1.6. Ces temps caractéristiques sont importants car ils déterminent l'échelle que l'on doit considérer pour étudier un polluant donné. Ainsi, l'ozone troposphérique a un temps de vie de plusieurs jours, ce qui fait de lui un polluant continental. Ce temps est plus faible pour les NOx et le SO_2, qui ont donc un impact essentiellement local autour des sources.

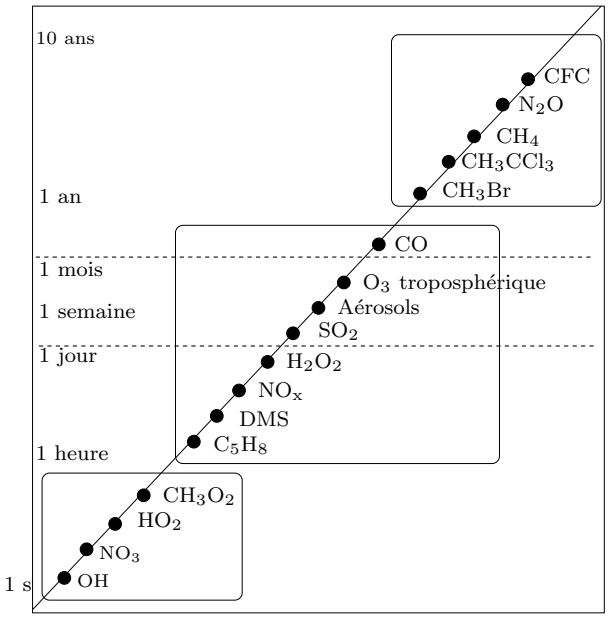


FIG. 1.6 – Ordre de grandeur du temps de résidence des principales espèces atmosphériques. Source : Sportisse [2007a].

1.2 Une hiérarchie de données et de modèles

1.2.1 Équation de dispersion réactive

Dans cette partie, on considère l'évolution d'une espèce X_i, de concentration c_i dans l'atmosphère. Cette espèces est soumise à tous les processus détaillés dans la partie précédente, et résumés par la figure 1.7 : émissions, transport, diffusion, chimie, dépôt et lessivage.

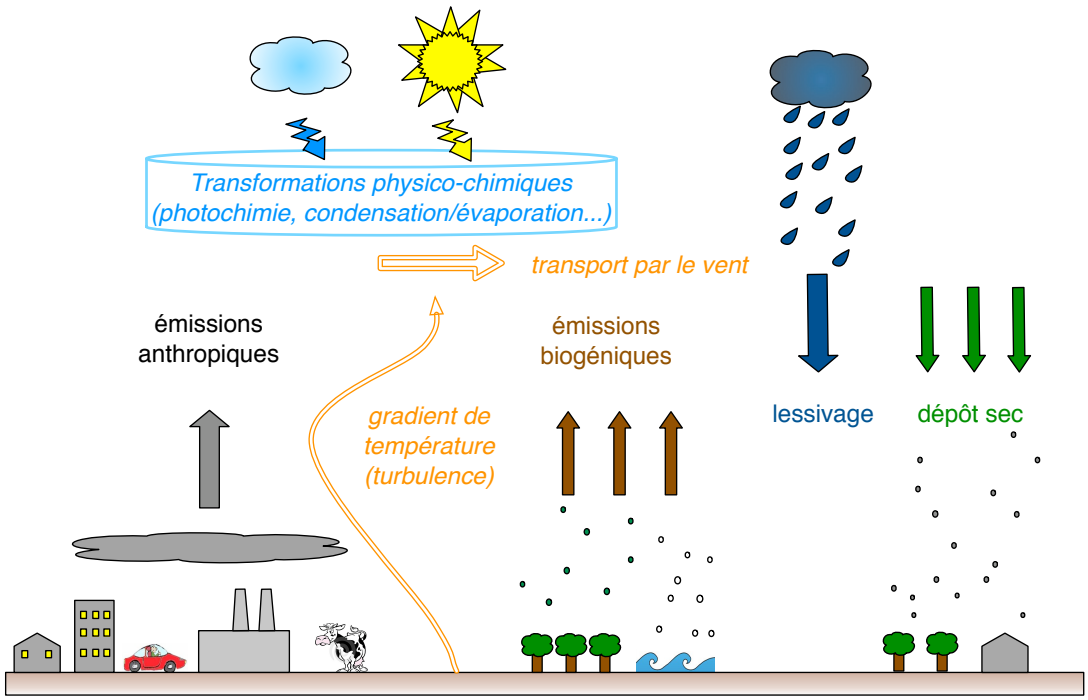


FIG. 1.7 – Résumé de l'ensemble des processus intervenant dans l'étude des polluants de l'atmosphère, qu'un modèle de qualité de l'air doit modéliser.

Il s'agit à présent de traduire cette évolution en équation. On fait pour cela l'hypothèse que les champs météorologiques (vent, température) sont connus, et que la dynamique de l'écoulement n'est pas influencée par les polluants.

Équation d'advection-diffusion

L'évolution de la concentration c_i est donnée par l'équation d'advection-diffusion, qui découle du principe de conservation de la masse :

$$\underbrace{\frac{\partial c_i}{\partial t}}_{(1)} + \underbrace{\mathrm{div}(U c_i)}_{(2)} = \underbrace{\mathrm{div}\left(D_{m_i} \nabla c_i\right)}_{(3)} + \chi_i\left(\mathbf{c}\right) + S_i \tag{1.11}$$

avec :

- (1) le terme instationnaire,
- (2) le terme d'advection, c'est-à-dire le transport par l'écoulement, dépendant du champ de vent $U = (u, v, w)$,
- (3) le terme de diffusion moléculaire, régie par une loi de Fick, avec une matrice diagonale D_{m_i} contenant les coefficients de diffusion moléculaire, et dépendant de l'espèce X_i,

- χ_i le terme de production ou de destruction de X_i par réactions chimiques, dépendant de l'ensemble des concentrations $\boldsymbol{c} = (c_1, ..., c_n)$ pour n espèces réactives, ainsi que du champ de température,
- S_i représentant le terme source (positif ou négatif), c'est-à-dire les émissions ainsi que les pertes par dépôt et lessivage.

Une description alternative est donnée par l'évolution du *rapport de mélange* C_i, c'est-à-dire du rapport de la concentration c_i à la densité du fluide porteur (ici, l'air) ρ. La densité du fluide porteur vérifie l'équation de continuité

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho U) = 0, \tag{1.12}$$

ce qui conduit alors à une équation d'advection-diffusion faisant intervenir C_i

$$\frac{\partial C_i}{\partial t} + U \cdot \nabla C_i = \frac{1}{\rho}\left(\operatorname{div}\left(D_{m_i} \nabla (\rho C_i)\right) + \chi_i(\rho \boldsymbol{C}) + S_i\right). \tag{1.13}$$

Représentation de la turbulence

L'équation d'advection-diffusion donne une description exacte de l'évolution des concentrations au niveau « microscopique ». Cependant, il est difficile de prendre en compte l'ensemble des échelles de la turbulence décrites en partie 1.1.3 dans les modèles. On utilise donc dans les équations une approche appelée la *décomposition de Reynolds*. Si Ψ représente une variable liée à l'écoulement turbulent (température, vitesse...), on décompose ce champ entre une valeur moyenne, notée $\langle \Psi \rangle$, et une fluctuation turbulente notée Ψ', selon

$$\Psi = \langle \Psi \rangle + \Psi'. \tag{1.14}$$

Naturellement, la moyenne des fluctuations est nulle : $\langle \Psi' \rangle = 0$. Les équations décrivant l'évolution de Ψ sont alors écrites avec cette décomposition, puis moyennées, sachant que l'opérateur $\langle \cdot \rangle$ est linéaire. Si l'on veut obtenir l'évolution des grandeurs moyennes $\langle \Psi \rangle$, il faut ajouter à cela un modèle de « fermeture » pour les termes non linéaires de fluctuations. En effet, les termes non linéaires de l'équation initiale font apparaître, avec la décomposition de Reynolds, des produits de fluctuations

$$\Psi_1 \Psi_2 = \left(\langle \Psi_1 \rangle + \Psi_1'\right) \times \left(\langle \Psi_2 \rangle + \Psi_2'\right), \tag{1.15}$$

ce qui, une fois moyenné, étant donné que les moyennes des fluctuations sont nulles, donne

$$\langle \Psi_1 \Psi_2 \rangle = \underbrace{\langle \Psi_1 \rangle \langle \Psi_2 \rangle}_{(1)} + \underbrace{\langle \Psi_1' \Psi_2' \rangle}_{(2)}. \tag{1.16}$$

Dans l'équation 1.16, le terme (2) représente un terme de corrélation que l'on ne connaît pas. Il est donc nécessaire, pour fermer les équations, d'exprimer ce terme en fonction des autres variables du problème.

En pratique, la grandeur Ψ peut être une composante du champ de vitesse de vent $U = (u, v, w)$

$$u = \langle u \rangle + u', \quad v = \langle v \rangle + v', \quad w = \langle w \rangle + w', \tag{1.17}$$

ou une concentration (ou un rapport de mélange)

$$c_i = \langle c_i \rangle + c_i'. \tag{1.18}$$

On peut alors appliquer cette décomposition à l'équation 1.11, à laquelle on applique ensuite l'opérateur $\langle\cdot\rangle$ afin de trouver l'évolution des quantités moyennes. On trouve alors, d'après l'équation 1.16 pour le calcul de la moyenne du produit de deux variables,

$$\frac{\partial\langle c_i\rangle}{\partial t} + \mathrm{div}\left(\langle U\rangle\langle c_i\rangle\right) = -\mathrm{div}\left(\langle U'c_i'\rangle\right) + \mathrm{div}\left(D_{m_i}\nabla\langle c_i\rangle\right) + \langle\chi_i\rangle + \langle S_i\rangle, \tag{1.19}$$

L'équation 1.19 donne alors l'évolution des quantités moyennes (moment d'ordre 1) pour les concentrations. Cette approche, largement utilisée en modélisation, s'appelle RANS (Reynolds Averaged Navier Stokes). Pour résoudre l'équation, on rajoute des conditions aux limites, données par les flux entrants aux frontières du domaine simulé, par le dépôt sec et les émissions au sol, et par un flux nul à la frontière entre la couche limite et la troposphère libre (pas d'échanges entre ces deux milieux, aux échelles considérées).

Fermeture des flux turbulents

Dans l'équation 1.19, il reste à exprimer les flux turbulents $\langle U'c_i'\rangle$ (ou plus précisément $\langle u'c_i'\rangle$, $\langle v'c_i'\rangle$ et $\langle w'c_i'\rangle$). Une approche classique est la fermeture d'ordre 1, ou théorie-K, [Panofsky et Dutton, 1984]) qui exprime les flux turbulents d'une variable en fonction du gradient de la variable moyennée, à l'instar de la loi de Fick pour la diffusion moléculaire. On obtient alors

$$\langle\Psi'V'\rangle = -K_\Psi\nabla\langle\Psi\rangle, \tag{1.20}$$

ce qui donne pour la fraction molaire C_i

$$\langle C_i'V'\rangle = -K\nabla\langle C_i\rangle, \tag{1.21}$$

et donc pour la concentration, en négligeant les fluctuations de densité ($\rho \simeq \langle\rho\rangle$),

$$\langle c_i'V'\rangle = -\rho K\nabla\frac{\langle c_i\rangle}{\rho}. \tag{1.22}$$

La diffusion turbulente K est supposée être la même pour toutes les espèces. En pratique, on a de plus $K \gg D_m$: la diffusion turbulente est très grande devant la diffusion moléculaire, et celle-ci est généralement négligée.

Effet de ségrégation

Dans l'équation 1.19 , le terme $\langle\chi_i\rangle$ représente la production-consommation de l'espèce $\mathrm{X_i}$ par les réactions chimiques. Or, comme décrit dans la partie 1.1.4, il existe des termes de chimie non linéaires, où les produits de concentrations de deux espèces X_i et X_j apparaissent. Comme pour les flux turbulents, on a alors un problème de fermeture

$$\langle c_ic_j\rangle = \langle c_i\rangle\langle c_j\rangle + \langle c_i'c_j'\rangle, \tag{1.23}$$

de façon similaire à l'équation 1.16. Le terme du second ordre est en général négligé, mais l'approximation n'est plus valide lorsque les temps caractéristiques de la chimie sont plus petits que ceux de l'homogénéisation turbulente ($Da \gg 1$). Un terme supplémentaire appelé *intensité de ségrégation* apparaît alors

$$\langle c_ic_j\rangle = \langle c_i\rangle\langle c_j\rangle\left(1 + I_s\right) \quad \text{avec} \quad I_s = \frac{\langle c_i'c_j'\rangle}{\langle c_i\rangle\langle c_j\rangle}. \tag{1.24}$$

Une façon de prendre en compte les effets de la turbulence sur la chimie est donc de paramétrer l'intensité de ségrégation I_s qui est habituellement négligée [Vinuesa et Vilà-Guerau de Arellano, 2005]. Cette paramétrisation est décrite en partie 1.3.3.

1.2.2 Différents types de modèles

On peut principalement distinguer deux catégories de modèles :

- les modèles *on-line* qui résolvent les équations décrivant l'écoulement (équations de Navier-Stokes) en même temps que l'évolution des concentrations de polluants,
- les modèles *off-line* qui considèrent que l'influence des polluants sur les champs météorologiques est négligeable. Les champs météorologiques (vent, température, pression) sont alors pré-calculés, par exemple par un code météorologique, et utilisés comme données d'entrée du modèle.

Les équations décrites dans la partie précédente se placent dans le cadre du couplage off-line, et c'est également dans ce cadre que se situe la suite de ce travail. Il est toutefois utile de dire quelques mots sur les modèles on-line.

Modèles de mécanique des fluides (on-line)

On ne décrira pas ici les équations de la mécanique des fluides, ni les détails des modèles qui les résolvent. Des informations détaillées sur le sujet se trouvent dans nombre d'ouvrages [Sagaut, 2002]. On s'intéresse ici à l'utilisation possible de ces modèles pour décrire la dispersion de polluants, éventuellement réactifs, par un écoulement. La problématique essentielle de ce type de modèle est de savoir quelles échelles de la turbulence sont résolues. Les modèles de mécanique des fluides (CFD, Computational Fluid Dynamics) se divisent à ce titre en trois catégories :

1. *Les modèles de simulation numérique directe* (DNS, Direct Navier-Stokes) qui résolvent directement les équations de la mécanique des fluides sans approximation. Cela revient à dire que toutes les échelles de la turbulence sont résolues, et ces modèles sont donc extrêmement coûteux en temps de calcul : cela nécessite d'avoir un rapport de taille entre les plus petites et les plus grandes mailles de l'ordre du rapport entre les plus grandes échelles de turbulence et l'échelle de Kolmogorov (équation 1.2).
2. *Les modèles de simulation des grandes échelles* (LES, Large-Eddy Simulation) qui résolvent les grandes échelles de la turbulence. Cela nécessite de filtrer le spectre d'énergie turbulente, pour séparer les grandes échelles (résolues) et les petites échelles (paramétrées par un modèle statistique appelé « modèle sous-maille »). Le filtre est choisi de façon à ce que les échelles qui influent le plus sur le transport des polluants dans l'atmosphère soient simulées.
3. *Les modèles RANS*, pour Reynolds Averaged Navier-Stokes , qui sont les plus couramment utilisés, et donnent l'évolution des grandeurs moyennes de l'écoulement. Ils nécessitent une fermeture des flux turbulents. La méthode la plus courante est la fermeture d'ordre 1, mais une méthode également très utilisée est la fermeture k–ϵ.

L'avantage de ces modèles, par rapport aux modèles off-line, est de résoudre explicitement l'écoulement, et donc en particulier de pouvoir prendre en compte l'influence du relief et des obstacles. De plus, l'utilisation de simulations LES, voire DNS, permet de quantifier l'influence de la turbulence sur les polluants et notamment sur la chimie, comme dans Molemaker et Vilà-Guerau de Arellano [1998] (DNS) ou Vilà-Guerau de Arellano *et al.* [2004] et Auger et Legras [2007] (LES) par exemple. Ce type de modèle est en général utilisé pour déterminer des paramétrisations sous-mailles utilisables ensuite dans les modèles de chimie-transport.

Modèles de dispersion (off-line)

Différentes approches peuvent être utilisées pour résoudre l'équation 1.19. Une synthèse plus complète des différents types de modèles pourra être trouvée dans Perkins *et al.* [2002] et Sportisse [2007a] par exemple. On distingue principalement deux approches :

1. *Les modèles lagrangiens particulaires*, qui consistent à suivre des particules dans l'écoulement. Pour un champ de vitesse U et un champ de diffusion K connus, on peut écrire les équations du mouvement pour chaque particule, en rajoutant une composante stochastique pour décrire la variabilité due à la turbulence. Ces modèles sont toutefois difficilement généralisables au cas réactif, car il faut repasser de la représentation particulaire (chaque particule est « suivie » indépendamment) aux concentrations dans un volume donné.
2. *Les modèles eulériens de chimie-transport* (CTM, Chemistry-Transport Model) qui résolvent directement l'équation 1.19 sur un maillage donné, par des méthodes numériques appropriées. Ce type de modèle est très utilisé en prévision ainsi qu'en étude d'impact. La principale limitation provient de l'utilisation d'un maillage : tout processus d'échelle caractéristique inférieure à celle de la maille est, par définition, mal représenté. L'ensemble des techniques permettant de mieux représenter ces processus fait l'objet de la partie 1.3.

Par ailleurs, il existe également des *modèles gaussiens*, qui peuvent se déduire des équations d'advection-diffusion sous certaines hypothèses simplificatrices. Ces modèles sont utilisés pour l'échelle locale, c'est-à-dire l'échelle à laquelle on peut considérer que les champs météorologiques sont homogènes, autour d'une source donnée (de l'ordre de quelques kilomètres, voire quelques dizaines, autour de la source). Ces modèles, qui représentent les sources ponctuelles sous forme de *panache* ou de *bouffées*, sont décrits dans le chapitre 2. Enfin, on peut signaler les *modèles de boîte*, qui sont définis en considérant des domaines de grande dimension (par rapport à une maille classique) où les concentrations sont homogènes. Ce type de modèle, très utilisés dans les années 90 en raison du faible coût calcul, n'est pratiquement plus utilisé aujourd'hui.

Des modèles couplés météorologie-dispersion existent également à de plus grandes échelles. Il s'agit principalement de coupler un code de météorologie à méso-échelle avec un CTM. Ce type de couplage existe également à encore plus grande échelle pour les modèles de climat (GCM, General Circulation Model), où la modélisation de la photochimie est beaucoup moins détaillée que dans un CTM classique.

Une hiérarchie de modèles et d'échelles

On définit à présent différentes échelles, qui représentent la taille du domaine sur lequel on veut effectuer les calculs de dispersion. On distingue principalement l'**échelle locale** (quelques dizaines de kilomètres), l'**échelle régionale** (quelques centaines de kilomètres) et l'**échelle continentale** (quelques milliers de kilomètres). Enfin, on peut également définir l'**échelle globale**, caractéristique du transport intercontinental, voire hémisphérique. L'ensemble de ces échelles est représenté figure 1.8. La définition de ces échelles est à rapprocher des temps de vie caractéristiques des différents polluants dans l'atmosphère (figure 1.6). Ainsi, l'étude de l'ozone nécessitera une modélisation à l'échelle continentale, celle des NOx une modélisation à l'échelle régionale ou locale.

Le tableau 1.3 récapitule les différentes échelles définies ci-dessus, ainsi que les modèles pouvant être utilisés pour chacune des échelles et les applications associées. Dans la suite, on s'intéresse aux modèles off-line, en particulier gaussiens et eulériens, pour les échelles allant de l'échelle locale à l'échelle continentale.

Enfin, il convient de distinguer l'échelle modélisée (représentée en pratique par le domaine de calcul sur lequel la simulation est effectuée), et l'échelle « bien représentée » par le modèle. Celle-ci n'est pas clairement définie, en particulier car elle dépend du processus considéré. Cependant, pour les modèles eulériens, on peut considérer que cette échelle correspond à la taille d'une maille, car tout processus dont l'échelle caractéristique est inférieure à la taille d'une maille est mal représenté. La taille d'une maille varie de ~ 1 km (échelle locale) à ~ 100 km (échelle

continentale) environ. De même, pour les modèles gaussiens à bouffées, on peut considérer que la plus petite échelle bien représentée correspond à la taille caractéristique d'une bouffée, car les hétérogénéités au sein d'une bouffée sont peu ou mal modélisées.

Type de modèle	Domaine de calcul	Application
CFD réactive (on-line)		
DNS	échelle très locale $\sim$ 1 km	Recherche (turbulence)
LES	échelle très locale $\sim$ 1 km	Recherche (turbulence)
RANS	échelle locale $\sim$ 10 km	Environnement complexe (bâti...)
Modèles off-line		
Modèles gaussiens	échelle locale $\sim$ 10 km	Risque industriel (impact local)
Modèles lagrangiens	échelle locale $\sim$ 10 km échelle régionale $\sim$ 100 km et échelle continentale $\sim$ 1000 km	Risque industriel (radionucléides), biologique
Modèles de chimie-transport	échelle régionale $\sim$ 100 km et échelle continentale $\sim$ 1000 km	Prévision et impact (photochimie)
Modèles globaux	échelle globale $\sim$ 10 000 km	Transport intercontinental
Modèles on-line		
CTM couplé	échelle continentale	Recherche (photochimie)
GCM	échelle globale	Impact climatique

Tab. 1.3 – Échelles spatiales dans la modélisation, et modèles associés. Pour chaque modèle, l'échelle à laquelle il est utilisé (domaine de calcul) est indiquée, ainsi que l'application la plus courante associée.

1.2.3 Données d'entrée

On se place à présent dans le cadre d'un modèle de chimie-transport, ou d'un modèle plus simple de type gaussien. Pour résoudre l'équation d'advection-diffusion, le modèle doit disposer d'un certain nombre de données d'entrées. Ces données sont fournies la plupart du temps sous la forme d'un inventaire (pour les émissions et les données de sol), ou de sorties de modèles (pour les champs météorologiques et les conditions aux limites). Pour l'échelle locale, il s'agit en général de données d'observation.

Les émissions

Inventaires d'émissions spatialisés Les données d'émissions anthropiques pour les modèles de chimie-transport sont en général fournies sous la forme d'inventaires, par classe d'activité économique. La distribution temporelle des émissions est donnée par des profils types, afin d'atteindre le niveau de discrétisation temporelle voulu dans les modèles (de l'ordre de l'heure). Enfin, les émissions sont données pour des groupements d'espèces (par exemple, pour l'ensemble des COV) qu'il convient de répartir ensuite en fonction des espèces utilisées dans le modèle. La résolution spatiale des inventaires d'émissions utilisés varie, dans les cas présentés dans cette thèse (pour l'Europe et/ou la région parisienne) entre 50 km pour l'inventaire EMEP[2], et une

[2] `http://webdab.emep.int/`

résolution de l'ordre du kilomètre (inventaire Airparif). Les émissions biogéniques, elles, sont calculées en fonction de l'occupation des sols (végétation), du rayonnement solaire et de la température.

Données de sources ponctuelles Dans le cas de sources ponctuelles, et notamment si l'on utilise des modèles gaussiens pour les représenter, il n'est pas nécessaire de disposer d'un inventaire spatialisé. La source est représentée de façon plus précise, à partir de quelques données telles que sa hauteur, son débit, sa température et sa vitesse d'éjection.

Les données météorologiques

Champs météorologiques 3D Les champs météorologiques (vent, température, hauteur de couche limite, intensité du rayonnement solaire) proviennent de modèles météorologiques à méso-échelle, ou d'analyses basées sur des observations. Les champs, donnés en sortie de ces modèles, sont ensuite projetés sur les grilles de calcul des CTM, par des procédures de projection et d'interpolation. Dans les modèles utilisés dans cette thèse, ils proviennent d'ECMWF (European Centre for Medium-Range Weather Forecasts) ou du modèle MM5 (Fifth-Generation NCAR/Penn State Mesoscale Model). Étant donné que les champs météorologiques sont donnés en général avec une résolution assez grande (souvent entre 50 et 150 km pour les applications présentées ici), leur interpolation ne permet pas de reconstituer la variabilité locale des champs météorologiques.

Données météorologiques locales Pour les modèles à l'échelle locale, on utilise en général des mesures de météorologie locales (par exemple provenant d'un mât à proximité de l'installation industrielle). Il est possible également d'utiliser des modèles « diagnostiques » qui permettent de reconstruire des champs météorologiques à partir de quelques mesures locales, en prenant mieux en compte le relief et le bâti environnants.

L'occupation des sols

Champs 2D d'occupation des sols L'occupation des sols (LUC, Land Use Coverage) est donnée par des *classes* décrivant le type de terrain : étendue d'eau, forêt, champs, milieu urbain... Dans cette thèse, deux descriptifs d'occupation des sols sont utilisés :

- les données USGS (U.S. Geological Survey), d'une résolution de 1 km^2, avec 24 classes,
- les données GLCF (Global Land Cover Facility), d'une résolution de 1 km^2, avec 13 classes.

Ces données proviennent d'observations satellitaires. La résolution signifie qu'un seul type de terrain est associé à chaque kilomètre carré. Pour utiliser ces données dans un modèle de dispersion dont les mailles sont plus grandes, on peut décrire l'occupation des sols dans chaque maille par la proportion de chacune des classes présentes dans la maille.

Échelle locale Pour les simulations à l'échelle locale, les modèles gaussiens ne permettent pas une modélisation détaillée du relief. Cependant, il peut être possible de tenir compte de certaines caractéristiques (bâtiments, surélévation...). Les vitesses de dépôt sont calculées en entrée, en prenant des valeurs typiques des espèces considérées pour le type de terrain dominant autour de la source.

Les conditions initiales et aux limites

Échelles continentale et régionale Il s'agit des espèces transportées depuis l'extérieur du domaine de la simulation. Dans le cas des simulations à l'échelle européenne présentées dans

cette thèse, ces données proviennent de simulations par le modèle global Mozart 2 [Horowitz *et al.*, 2003]. Dans le cas d'une simulation à l'échelle régionale, ces données peuvent également provenir d'une première simulation à l'échelle continentale, à plus grande résolution (one-way nesting).

Échelle locale Dans le cas d'une simulation à l'échelle locale, il n'y a pas à proprement parler de conditions aux limites. Les concentrations de fond sont supposées constantes et sont en général fournies par la station de mesure la plus proche.

1.2.4 Sensibilité des modèles aux données d'entrée

Variabilité sous-maille et incertitude

La partie précédente donne une idée de la résolution typique des données d'entrée utilisées par les modèles eulériens. Il est important de remarquer que *l'échelle bien représentée d'un modèle ne pourra pas être inférieure à celle des données d'entrée qui lui sont fournies*, sans ajout d'informations supplémentaires. Ainsi, la description de l'occupation des sols est en général une donnée sous-maille, pour laquelle on perd en résolution lorsque l'on passe de la résolution des données à celle du modèle eulérien. Inversement, les données météorologiques sont presque toujours fournies à plus grande résolution. Une interpolation bilinéaire permet de reconstituer des champs météorologiques sur la grille du modèle eulérien, mais la variabilité des champs météorologiques n'est pas bien représentée. Descendre en résolution avec le modèle eulérien seul ne permettra pas de reconstituer cette variabilité.

Il est important de noter également qu'il s'agit ici de quantifier les erreurs des modèles dues à la *variabilité* des données, c'est-à-dire aux hétérogénéités spatiales et temporelles, et non aux incertitudes sur ces données. Il est toutefois difficile de séparer totalement les deux aspects du problème. En particulier, plus la description des données est fournie à de petites échelles, plus les incertitudes (relatives) sur ces données pourront être élevées [Stohl *et al.*, 1995].

Impact des données d'entrée

De nombreuses études de sensibilité montrent l'impact des données d'entrée dans les performances des modèles de chimie-transport. L'influence des données d'émissions, et des incertitudes associées, sur les concentrations d'ozone, a été largement étudiée (Schneider *et al.* [1997], Pryor [1998], Hanna *et al.* [2001], Mallet et Sportisse [2005]).

Concernant en particulier l'influence de la *résolution* des données, de nombreuses études, menées notamment sur les expériences ETEX (European Tracer EXperiment), ont montré l'importance d'utiliser des données météorologiques méso-échelle à résolution suffisante. Cela permet de modéliser certains phénomènes tels que des tourbillons anti-cycloniques horizontaux, non résolus par les champs ECMWF à faible résolution (sans ré-analyse). On peut citer notamment Sorensen [1998] et Nasstrom et Pace [1998] pour les modèles lagrangiens (à particules ou à bouffées), et Brandt *et al.* [1998] pour un modèle hybride bouffées-eulérien.

Enfin, l'influence respective de la résolution des émissions et des champs météorologiques est étudiée dans Valari et Menut [2008], pour montrer que l'impact des émissions est prépondérant. De plus, l'existence d'une résolution au-dessous de laquelle le raffinement des émissions peut dégrader les résultats est mise en évidence, ce qui souligne le problème des incertitudes sur les données d'entrée à haute résolution.

1.3 Différentes méthodes de changements d'échelles

On s'intéresse à présent aux méthodes permettant de mieux représenter les changements d'échelles dans les modèles couramment utilisés en qualité de l'air, à savoir les modèles de chimie-transport. L'utilisation de données d'entrée fournissant des informations à haute résolution peut être une première approche, cependant elle n'est pas toujours possible en pratique. De plus, cela ne permet pas de s'abstraire de toutes les limitations inhérentes aux modèles. On se focalise, par conséquent, sur les moyens d'améliorer les modèles eux-mêmes, et/ou de représenter leur variabilité sous-maille. On distingue pour cela plusieurs approches :

1. La première méthode consiste à utiliser localement (dans des zones à forte variabilité) un modèle plus fin, soit en raffinant le maillage du modèle eulérien, soit en le couplant avec un modèle pour l'échelle locale.
2. La seconde possibilité, moins coûteuse en temps de calcul, consiste à utiliser des *paramétrisations sous-maille*. Ces paramétrisations sont en général issues de simulations faites avec des modèles à petite échelle (LES par exemple), très coûteux en temps de calcul, ou encore déterminées de façon empirique. Il s'agit de modifier le modèle de façon à mieux représenter les effets sous-maille d'un processus donné, tel que les nuages, la turbulence ou les émissions.
3. Il existe également des *méthodes stochastiques*, qui consistent à représenter la variabilité sous-maille des concentrations et/ou d'un processus physique par une densité de probabilité.
4. Une variante est de déterminer des relations statistiques entre les concentrations à l'échelle de la maille et les concentrations aux stations, pour tenir compte de l'erreur de représentativité commise par les modèles.

1.3.1 Méthodes numériques et couplage de modèles

Méthode numériques

Raffinement de maillage et maillages imbriqués (« nesting ») Une première approche, pour réduire l'erreur sous-maille commise dans les modèles eulériens, peut être basée simplement sur la diminution de la taille des mailles utilisées. Cela doit permettre de réduire le nombre de phénomènes mal résolus, ainsi que la diffusion numérique induite par l'hypothèse de concentrations homogènes dans une maille. Cette approche est cependant rapidement coûteuse en temps de calcul. Une technique courante est l'utilisation de différents maillages imbriqués, de résolution de plus en plus fine (« nesting »). L'interaction entre les différents domaines peut être à sens unique (le grand domaine fournit les conditions aux limites du petit), ou prendre en compte une rétroaction : on parle de « one-way nesting » et « two-way nesting » respectivement. Les données d'entrée sont soit interpolées, soit recalculées sur le maillage plus fin (avec le modèle météorologique MM5 par exemple). La figure 1.8 montre un exemple de « one-way nesting » (simulations présentées au chapitre 5) : un modèle global est utilisé pour générer les conditions aux limites de la simulation européenne, à une résolution 50 km. Celle-ci est utilisée pour générer les conditions aux limites pour la simulation à l'échelle de la région parisienne (résolution de 5 km). Pour mieux représenter les sources ponctuelles et descendre encore en échelle, il est possible d'utiliser un modèle gaussien pour l'échelle locale.

L'impact de la résolution du maillage a été largement étudié en photochimie, notamment concernant l'ozone (Gillani [1986], Liang et Jacobson [2000], Chock *et al.* [2002]). De plus, cet impact a été évalué récemment non seulement du point de vue des performances du modèle, mais aussi pour la sensibilité de l'ozone aux mesures réductions d'émissions (Cohan *et al.* [2006],

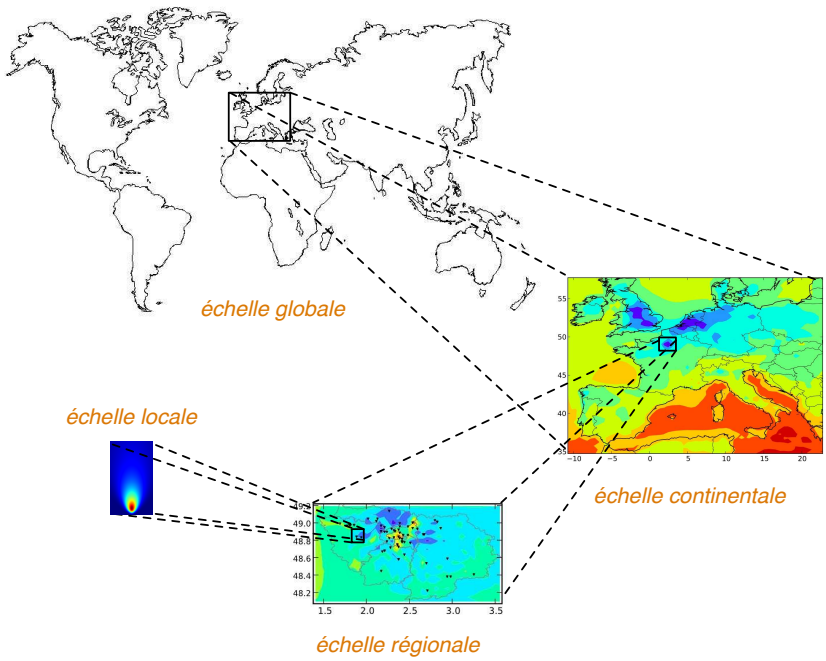


FIG. 1.8 – Illustration de l'utilisation de modèles à différentes échelles, par une méthode de « one-way nesting ». La figure à l'échelle européenne représente les résultats d'une simulation effectuée avec Polair3D, pour l'ozone. La figure à l'échelle régionale montre un résultat de simulation à l'échelle de la région parisienne, et enfin l'échelle locale illustre l'utilisation d'un modèle gaussien de panache pour représenter le panache issu d'une source ponctuelle.

Arunachalam *et al.* [2006]). On constate en général une légère amélioration des performances des modèles avec une résolution plus fine. L'utilisation de maillages trop grossiers implique notamment une sur-estimation des concentrations d'ozone, due à une dilution des émissions de NOx. Raffiner le maillage au-dessous de 12 km environ a cependant peu d'impact sur les résultats des modèles (Cohan *et al.* [2006] par exemple), et permet principalement de mieux représenter les structures spatiales et la variabilité sous-maille. L'utilisation de maillages imbriqués a également été évaluée pour d'autres polluants que l'ozone (Jakobs *et al.* [1995] pour le SO_2, Wang *et al.* [2004] pour le CO).

Un raffinement supplémentaire de maillage, fait localement autour des sources, a également des limitations : (1) il est difficile de connaître a priori les bons emplacements pour raffiner le maillage, et de le faire évoluer en temps (avec le déplacement d'un panache par exemple), et (2) il reste une perte d'information à l'interface avec le maillage grossier [Odman et Russel, 1991]. De plus, réduire la taille des mailles au-delà de 1 km est possible [Jacobson et Seinfeld, 2004], mais reste techniquement difficile et très coûteux.

Maillage adaptatif (« gridding ») Une variante des méthodes numériques précédentes est le maillage adaptatif, qui est moins coûteux en temps de calcul, car le maillage est raffiné uniquement autour des zones à fortes variabilités (les sources, ou les endroits à orographie complexe). Ce maillage peut être dynamique, c'est-à-dire adapté en temps pour correspondre aux variations dues par exemple à la météorologie. Cette technique a été utilisée aussi bien pour l'ozone (Ghorai *et al.* [2000]; Srivastava *et al.* [2001]; Constantinescu *et al.* [2008]; Lagzi *et al.* [2009]) que pour la dispersion de radionucléides [Lagzi *et al.*, 2004]. Dans les applications, les résultats du maillage adaptatif sont comparés à ceux donnés par un maillage fin (uniforme), et le résultat est en général comparable avec un temps de calcul plus faible. L'un des paramètres importants est le critère permettant de déterminer les zones à raffiner ; pour l'ozone, il est important de choisir un critère de raffinement de maillage sur les précurseurs tels que les NOx (Constantinescu *et al.* [2008]). Les techniques de raffinement de maillage peuvent permettre de descendre localement jusqu'à des mailles d'une centaine de mètres (autour de sources ponctuelles), ce qui conduit toutefois à une augmentation du temps de calcul car le pas de temps utilisé est alors plus faible [Odman et Khan, 2004].

Couplage de modèles

Les méthodes numériques basées sur le raffinement de maillage restent subordonnées aux limites du modèle utilisé, et de nombreux processus sous-maille restent mal représentés. De plus, les hypothèses utilisées dans un modèle ne sont valides que pour une certaine échelle. Il est donc préférable d'utiliser des modèles spécifiques à l'échelle locale (voir tableau 1.3), ce qui suppose toutefois de disposer de données d'entrée adéquates, et en particulier d'inventaires d'émissions suffisamment précis.

Panache sous-maille Les modèles de panache sous-maille consistent à coupler un modèle gaussien (de panache ou à bouffées), utilisé autour des sources ponctuelles, avec un modèle eulérien (figure 1.8). Lorsque le panache n'est plus modélisé correctement par le modèle gaussien et/ou que l'erreur commise par le modèle eulérien est suffisamment faible, le panache est transféré dans le modèle eulérien. Ce type de méthode est utilisé en photochimie (Seigneur *et al.* [1983]; Sillman *et al.* [1990]; Morris *et al.* [1991]; Kumar et Russell [1996]; Byun et Schere [2006]; Karamchandani *et al.* [2002]), mais aussi pour la dispersion de radionucléides [Brandt, 1998]. De tels couplages sont également décrits dans Lipphardt [1997] et Gauthier [1999]. Bien que

le champ d'application de ces modèles soit principalement centré sur les sources ponctuelles, l'application à la modélisation des routes est également possible [Karamchandani *et al.*, 2009]. De plus, les modèles de panache sous-maille en photochimie ont été récemment étendus à la modélisation des particules [Karamchandani *et al.*, 2006]. L'utilisation d'un modèle de panache sous-maille donne de bons résultats et permet une meilleure représentation des concentrations autour des sources et en aval, dans le panache. Les performances globales des modèles, basées sur des stations « de fond » souvent éloignées des sources, ne sont cependant pas toujours impactées. Le temps de calcul peut être important si de nombreuses sources réactives sont traitées en sous-maille (notamment pour la chimie). Par ailleurs, il est également possible d'utiliser un modèle lagrangien particulaire à la place du modèle gaussien. Chock *et al.* [2002] et Weinroth *et al.* [2008] utilisent un modèle lagrangien pour simuler la dispersion des polluants, tout en simulant les réactions chimiques sur les concentrations moyennes dans les mailles eulériennes.

Une alternative (moins coûteuse) est le couplage « off-line » entre un modèle eulérien et un modèle gaussien. Le premier est utilisé pour donner une estimation des espèces de fond et notamment de la chimie, et le second (sans chimie) est utilisé pour interpoler a posteriori les résultats en donnant une meilleure répartition des concentrations près des sources. Cette méthode est utilisée principalement pour la modélisation des routes (Isakov et Venkatram [2006]; Isakov *et al.* [2009]; Stein *et al.* [2007]; Venkatram *et al.* [2009]). Cependant, elle ne tient pas compte des non-linéarités dues à la chimie.

Autres couplages D'autres couplages sont possibles, en utilisant différents types de modèles. Ainsi, Soulhac *et al.* [2003] propose une imbrication (« one-way nesting ») de plusieurs modèles : (1) un modèle eulérien de chimie-transport pour l'échelle régionale (208 km × 208 km), (2) un modèle eulérien de CFD à l'échelle urbaine (32 km × 32 km), (3) un modèle de réseau routier à l'échelle locale (500 m × 500 m), et (4) un modèle gaussien pour les sources ponctuelles. Une telle approche permet une amélioration significative des performances aux capteurs proches des sources. Cependant, elle nécessite de disposer de données d'entrée détaillées pour les émissions et la topographie à l'échelle locale.

Il est également possible d'utiliser un modèle météorologique local diagnostique ou pronostique, et de le coupler avec un modèle eulérien raffiné [Brucher *et al.*, 2000], ou avec un modèle à l'échelle locale de type gaussien.

1.3.2 Fermeture d'ordre supérieur

L'équation 1.19 donne l'évolution des concentrations moyennes, et pose un problème de fermeture des termes du second ordre, notamment pour les flux turbulents $\langle c_i' V' \rangle$. D'autres termes peuvent également poser un problème de fermeture. On écrit tout d'abord de manière plus explicite le terme source S_i :

$$S_i = E_i - \Lambda_i c_i, \tag{1.25}$$

avec E_i le terme d'émission, et Λ_i le coefficient de lessivage de l'espèce. Ces deux termes ont également une variabilité sous-maille modélisable par des fluctuations. Dans cette équation, les émissions E_i et le coefficient de lessivage Λ_i sont deux termes pour lesquels on peut utiliser la décomposition de Reynolds, afin de modéliser les fluctuations sous-maille. De même, la vitesse de dépôt v_{di} de l'espèce i intervient dans les conditions aux limites au sol, et peut être décomposée pour représenter la variabilité sous-maille des types de sol. Enfin, dans l'équation 1.19, la matrice des coefficients de diffusion K et le taux de réaction k peuvent également être décomposés de cette façon pour tenir compte des fluctuations des vitesses de vent.

En général, ces fluctuations sont négligées, ou paramétrées en fonction des grandeurs moyennes de l'écoulement (théorie-K pour la diffusion). Il existe différentes façon de prendre en compte les fluctuations sous-maille dans les modèles :

1. les modéliser explicitement, en utilisant une fermeture d'ordre supérieur,
2. les exprimer analytiquement en fonction d'autres variables connues, telles que les grandeurs caractéristiques de la couche limite et les variables moyennes : il s'agit de paramétrisations sous-maille,
3. les représenter de façon stochastique par une fonction de densité de probabilité (PDF pour « probability density function »).

Dans cette partie, on décrit la première solution. Les paramétrisations sous-maille sont décrites en partie 1.3.3, et la modélisation stochastique est détaillée dans la partie 1.3.4.

Fermeture d'ordre supérieur et évolution de la variance

Une façon de tenir compte de ces fluctuations est de les modéliser explicitement en utilisant une fermeture d'ordre supérieur. Cela signifie en pratique qu'il faut connaître les fluctuations de vent, et les intégrer à l'équation d'advection-diffusion. Cela permet alors de connaître non seulement l'évolution de la moyenne $\langle c_i \rangle$ mais aussi l'évolution de la variance, ou moment d'ordre 2, $\langle c_i'^2 \rangle$. Cette évolution est donnée par une équation supplémentaire [Garratt, 1992]

$$\frac{\partial \langle c_i'^2 \rangle}{\partial t} + \mathrm{div}\left(\langle U \rangle \langle c_i'^2 \rangle\right) = \mathrm{div}\left(D_{m_i} \nabla \langle c_i'^2 \rangle - \langle U' c_i'^2 \rangle\right) + 2\,\langle c_i' \chi\left(\boldsymbol{c}\right) \rangle + P_c - \epsilon_c, \tag{1.26}$$

avec P_c le terme de production turbulente

$$P_c = -2 \langle U' c' \rangle \nabla \langle c_i \rangle, \tag{1.27}$$

et ϵ_c le terme de dissipation turbulente

$$\epsilon_c = 2 D_{m_i} \langle |\nabla c_i'|^2 \rangle. \tag{1.28}$$

Représentation de la turbulence

Si l'on ne dispose que des fluctuations du vent, sans les fluctuations d'émissions (ou de pertes), l'équation 1.26 permet de modéliser la turbulence et son impact sur la ségrégation chimique. Ce type d'approche est utilisé par exemple pour certains modèles à bouffées [Sykes, 2000]. Cela permet alors de tenir compte explicitement de l'intensité de ségrégation due à la turbulence dans les réactions chimiques (équation 1.24). Karamchandani *et al.* [2000] introduit ce type de modélisation dans un modèle de panache sous-maille (pour la chimie dans le panache). L'effet de la ségrégation est importante près du centre du panache, pour O_3 et NOx, mais diminue rapidement avec la distance à la source.

Représentation des émissions

On peut étendre cette fermeture d'ordre supérieur à une modélisation des fluctuations d'émission. Les émissions pour l'espèce modélisée s'écrivent alors

$$E_i = \langle E_i \rangle + E_i', \tag{1.29}$$

et il s'agit de modéliser l'apport des fluctuations d'émissions E_i' sur les fluctuations de concentration. Galmarini *et al.* [2008] fait l'hypothèse que les émissions moyennes $\langle E_i \rangle$ contribuent à

la concentration moyenne $\langle c_i \rangle$, en tant que terme source de l'équation d'advection-diffusion, et que les fluctuations d'émissions sont un terme source des fluctuations de concentration. L'équation 1.26 est alors pourvue d'un terme supplémentaire correspondant aux fluctuations du terme source, de la forme $2\langle c_i' E_i' \rangle$, qu'il convient de paramétrer pour fermer l'équation. On remarque, au passage, que l'hypothèse faite ici ne tient pas compte des non-linéarités, et néglige l'apport des fluctuations d'émissions à des espèces secondaires via la chimie (et notamment la ségrégation turbulente). La fermeture du terme de covariance concentration-émission peut s'écrire

$$\langle c_i' E_i' \rangle = r \sigma_E \sigma_c = r \left(\langle E_i'^2 \rangle \, \langle c_i'^2 \rangle \right)^{1/2}, \tag{1.30}$$

avec r un coefficient de corrélation qu'il convient de déterminer. Enfin, dans Galmarini *et al.* [2008], les fluctuations d'émission (et donc la moyenne et l'écart type) sont définis en découpant une maille de surface A en N petites surfaces élémentaires a_k ayant chacune un taux d'émission E_k. On a alors une émission moyenne donnée par la moyenne des émissions surfaciques « élémentaires »

$$\langle E \rangle = \frac{1}{A} \sum_{k=1}^{N} a_k E_k, \tag{1.31}$$

et un écart type donné par

$$\sigma_E^2 = \frac{1}{A} \sum_{i=1}^{N} a_k \left(\langle E \rangle - E_k \right)^2, \tag{1.32}$$

Cette modélisation revient à supposer que les concentrations dans une maille ont une distribution gaussienne autour de la moyenne. De plus, la détermination des paramètres de cette gaussienne suppose de disposer d'un inventaire suffisamment fin, par rapport à la taille d'une maille, pour que le nombre de surfaces élémentaires soit grand et que la modélisation empirique soit pertinente.

1.3.3 Paramétrisations sous-mailles

Les paramétrisations sous-mailles sont des modifications des équations et/ou des paramètres utilisés dans les équations d'avection-diffusion, afin de mieux représenter l'effet de certains processus sous-maille. Ces paramétrisations se déduisent en général de calculs effectuées à l'aide de modèles à petite échelle (LES, DNS), ou encore de relations empiriques basées sur des observations. On donne ici quelques exemples de paramétrisations sous-maille, la liste n'étant pas exhaustive.

Paramétrisation des émissions

Paramétrisation des routes De façon théorique, Arya [1999] donne une expression de la concentration en fonction de la distance à la source linéique, pour un traceur passif (sans chimie), de la forme $c(x) \propto x^{-\beta}$, où β dépend des conditions météorologiques et du type de sol. Dans Rodes et Holland [1981], l'estimation $\beta = 0.63$ est fournie pour les NOx. Des études relativement récentes ont porté sur le développement d'une relation avec la concentration de fond (dans la maille) c_b, la concentration au point d'émission (par une route) c_0, et la concentration à une distance x de la route. Par exemple, Gilbert *et al.* [2003] et Pleijel *et al.* [2004] donnent une relation logarithmique pour le NO_2, de type

$$c(x) = c_b + c_0 - k \ \log(x). \tag{1.33}$$

Barzyk *et al.* [2009] développe une relation pour les COV, en fonction d'une « distance effective » qui dépend des conditions météorologiques. Ce type de modèle peut remplacer un couplage de

modèles plus coûteux en temps de calcul. Il est employé soit en paramétrisation sous-maille (dépendant de la météorologie) soit en post-traitement.

Paramétrisation des émissions ponctuelles L'enjeu est d'éviter de représenter explicitement les sources ponctuelles avec un modèle de panache sous-maille, ce qui peut être coûteux pour de nombreuses sources réactives. Esler [2003] propose d'utiliser une émission « équivalente », corrigée pour représenter la perturbation globale due à la dilution plus lente du panache. Cette émission serait immédiatement diluée dans la maille eulérienne, mais donnerait des résultats proches de ceux obtenus par le modèle de panache sous-maille.

Paramétrisation des mouvements convectifs

Il s'agit de représenter les mouvements turbulents sous-maille, qui se traduisent par des colonnes d'air ascendantes (« updraft ») et descendantes (« updraft »), entraînant un mélange rapide des polluants sur la verticale (décrit en partie 1.1.2). Différentes modélisations de l'impact de ce phénomène sur le transport des polluants ont été proposées. En particulier, Han et Byun [2005] proposent, à partir de Chatfield et Brost [1987], un schéma appelé « two-stream model » (TSM) où les mouvements ascendants et descendants sont traités séparément. Ce schéma, calibré et testé sur des expériences de LES, donne des résultats nettement supérieurs aux autres types de modélisation, ainsi qu'à la modélisation classique (sans prise en compte de la convection). Le principe est de décrire les équations de chimie-transport pour chacun des deux flux verticaux. On suppose que, dans chaque maille, une fraction α_u est occupée par les flux ascendants, et une fraction $\alpha_d = 1 - \alpha_u$ par les flux descendants. La concentration moyenne totale dans la maille à une hauteur z est alors donnée par

$$\langle c \rangle(z) = \alpha_u(z)\langle c_u \rangle(z) + \alpha_d(z)\langle c_d \rangle(z), \tag{1.34}$$

avec $\langle c_u \rangle(z)$ et $\langle c_d \rangle(z)$ les concentrations moyennes sur les flux ascendants (resp. descendants) de la maille. L'évolution de ces concentrations est alors donnée par deux équations d'advection-diffusion

$$\alpha_u \frac{\partial \langle c_u \rangle}{\partial t} = -\frac{\partial}{\partial z}\left(\alpha_u \langle w_u \rangle \langle c_u \rangle\right) + \frac{\partial}{\partial z}\left(\alpha_u K_u \frac{\partial \langle c_u \rangle}{\partial z}\right) - r_u \langle c_u \rangle + r_d \langle c_d \rangle, \tag{1.35}$$

pour les flux ascendants, et

$$\alpha_d \frac{\partial \langle c_d \rangle}{\partial t} = -\frac{\partial}{\partial z}\left(\alpha_d \langle w_d \rangle \langle c_d \rangle\right) + \frac{\partial}{\partial z}\left(\alpha_d K_d \frac{\partial \langle c_d \rangle}{\partial z}\right) - r_d \langle c_d \rangle + r_u \langle c_u \rangle \tag{1.36}$$

pour les flux descendants. Dans les équations 1.35 et 1.36, les vitesses verticales de vent sont notées w_u (resp. w_d), et les fermetures turbulentes associées sont $K_u = \langle w'_u c'_u \rangle$ (resp. K_d). Enfin, les paramètres r_u et r_d sont les termes horizontaux de divergence, qui garantissent la conservation de la masse dans la maille par des échanges entre les colonnes, en supposant que la moyenne horizontale des vents verticaux est nulle ($\alpha_u \langle w_u \rangle + \alpha_d \langle w_d \rangle = 0$). Il reste donc à paramétrer les coefficients K, α et w.

Utiliser ce type de schéma permet une meilleure représentation de la dilution des polluants sur la hauteur de la couche limite, qui peut être sous-estimée par le schéma d'advection-diffusion classique. L'impact de ce schéma sur la photochimie est estimé dans Vautard *et al.* [2007].

Paramétrisation de la ségrégation chimique

Dans la partie 1.1.4, il est question de l'influence de la turbulence sur les réactions chimiques. Cette influence n'est pas modélisée dans les modèles de chimie-transport (fermeture du premier

ordre) et représente typiquement un phénomène sous-maille. La paramétrisation des flux verticaux ascendants et descendants décrite précédemment est une solution possible pour modéliser ce phénomène. Dans Vinuesa et Vilà-Guerau de Arellano [2003], une paramétrisation sous-maille est proposée afin de modéliser l'influence de cette ségrégation chimique sur les réactions. A partir d'expériences de LES, l'effet de cette paramétrisation sur la photochimie est étudié dans Vinuesa et Vilà-Guerau de Arellano [2005]. D'après les équations 1.5 et 1.24, le taux de réaction pour une réaction du second ordre de type

$$\mathrm{A} + \mathrm{B} \rightarrow P \tag{R 1.3}$$

est donné par

$$\begin{aligned} \langle R \rangle &= -k\langle c_A c_B \rangle \\ &= -k\left(\langle c_A \rangle \langle c_B \rangle + \langle c'_A c'_B \rangle\right) \\ &= -k\langle c_A \rangle \langle c_B \rangle \left(1 + I_s\right), \end{aligned} \tag{1.37}$$

avec I_s l'intensité de ségrégation donnée par

$$I_s = \frac{\langle c'_A c'_B \rangle}{\langle c_A \rangle \langle c_B \rangle}. \tag{1.38}$$

On peut alors définir un taux de réaction effectif k_{eff}, que l'on utiliserait à la place du taux de réaction k, pour tenir compte de la ségrégation

$$k_{\mathrm{eff}} = k\left(1 + I_s\right). \tag{1.39}$$

Il reste à présent à paramétrer I_s. On peut la réécrire en fonction de la corrélation moyenne entre les espèces A et B (de façon similaire à la fermeture utilisée pour les émissions dans l'équation 1.30)

$$r_{AB} = \frac{\langle c'_A c'_B \rangle}{\sigma_A \sigma_B}, \quad \text{donc} \quad I_s = r_{AB} \frac{\sigma_A \sigma_B}{\langle c_A \rangle \langle c_B \rangle}, \tag{1.40}$$

avec σ_A et σ_B les écarts types de c_A et c_B respectivement. Ces écarts types peuvent être paramétrés, pour des scalaires inertes (sans chimie), en fonction des flux de surface et d'échange avec la troposphère libre, ainsi que des paramètres h (hauteur de couche limite) et w_* (échelle de vitesse convective) [Moeng et Wyngaard, 1984]. Pour des scalaires réactifs, Vinuesa et Vilà-Guerau de Arellano [2003] ajoutent un terme dépendant du nombre de Damköhler ainsi que d'échelles de concentrations c_A* et c_B*, définies d'après Cuijpers et Holtslag [1998]. Le nombre de Damköhler est défini de façon similaire à l'équation 1.8, pour l'espèce A,

$$Da_A = \frac{\tau_t}{\tau_c} = \frac{h}{w*} k\langle c_B \rangle. \tag{1.41}$$

Les valeurs obtenues en pratique pour les coefficients de corrélations sont $r_{AB} = -0.75$ pour deux espèces entraînées dans des directions opposées (typiquement, l'une émise au sol, l'autre entraînée par la couche limite), et $r_{AB} = 0.9$ pour deux espèces mélangées dans la même direction. Dans ce dernier cas, I_s peut être positive car la turbulence a pour effet de rendre le mélange entre les espèces plus efficace. Dans Vinuesa et Vilà-Guerau de Arellano [2005], les valeurs obtenues pour le rapport k_{eff}/k sont données pour des réactions impliquées dans la formation de l'ozone, pour une couche limite convective typique en été, sur une moyenne de 6 heures. Les valeurs sont comprises entre 0.595 et 1.005, et la plupart sont au-dessus de 0.98, ce qui implique un impact relativement faible de la ségrégation. L'impact de la turbulence sur les taux de réaction peut être notable de façon très locale, mais est probablement négligeable en général à l'échelle d'un modèle de chimie-transport, en particulier en comparaison d'autres facteurs tels que les hétérogénéités des émissions.

1.3.4 Méthodes stochastiques

Dans les sections précédentes, il s'agissait de raffiner les modèles, pour diminuer le nombre de processus sous-maille mal représentés. Il s'agit ici de représenter les concentrations et/ou certains processus de façon stochastique. Ces méthodes sont en général coûteuses en temps de calcul. De plus, l'un des points cruciaux, commun à tous les modèles présentés ici, est de bien choisir la fonction de densité de probabilité (PDF) pour représenter les processus stochastiques.

Modèles micro-macro

Il s'agit de représenter de façon stochastique la variabilité sous-maille de certains paramètres (modèle « micro »), et d'utiliser une méthode de type Monte Carlo afin générer la paramétrisation sous-maille à introduire dans le modèle « macro » déterministe (ici, l'équation d'advection-diffusion).

Par exemple, on suppose que $\Lambda(p)$ est le coefficient de lessivage donné par une paramétrisation physique, dépendant d'un paramètre p (éventuellement un vecteur) tel que la taille des nuages, ou le contenu en eau liquide. Si l'on connaît parfaitement la forme analytique de $\Lambda(p)$, l'approche classique est de calculer, pour une maille donnée, la valeur $\Lambda(\langle p\rangle)$ en fonction des quantités moyennes sur la maille. Comme le modèle est en général non linéaire, cette approximation ne tient pas compte des fluctuations p'. Le modèle micro/macro consiste à donner une fonction de densité de probabilité (PDF) à la variable p, $P(p)$, pour représenter sa variabilité sous-maille. Le problème revient alors à approcher l'intégrale

$$\langle\Lambda(p)\rangle = \int P(p)\,\Lambda(p)\,\mathrm{d}p, \tag{1.42}$$

par une méthode de Monte Carlo par exemple. L'un des principaux problèmes est de fixer la forme de la PDF à utiliser pour le paramètre physique p. Dans le cas de la forme des gouttes de pluie, il existe des estimations de cette PDF décrites dans Sportisse [2007b] par exemple.

Le modèle micro peut également être la paramétrisation de l'intensité de ségrégation turbulente, donnée par l'équation 1.40, où l'on fait varier r_{AB} et les paramètres intervenant dans la formulation de σ_A et σ_B (paramètres de couche limite : hauteur, flux de surface, vitesse convective). De même, le modèle micro peut être le modèle de fluctuation des émissions donné par l'équation 1.30, où l'on peut faire varier notamment le coefficient de corrélation, et d'autres paramètres intervenant dans la définition des écarts types. Enfin, le modèle micro pourrait être le coefficient de diffusion turbulente K qui dépend lui aussi de paramètres de la couche limite. L'avantage d'une telle approche est de définir la variabilité sous-maille par une méthode stochastique. L'inconvénient principal est le temps de calcul associé aux simulations Monte Carlo. Cependant, de telles simulations pourraient éventuellement servir à former des paramétrisations sous-mailles représentatives de certaines situations typiques.

Micro-mélange

Il s'agit ici de représenter le champ de concentration comme un processus stochastique, avec une PDF associée. Ce type d'approche est utilisé en combustion (voir Pope [2000] par exemple). Le champ de concentrations $c(x,t)$ a une PDF de la forme $P(\Psi,x,t)$ avec Ψ une valeur possible de $c(x,t)$. On écrit alors l'équation d'évolution de P

$$\frac{\partial\langle P\rangle}{\partial t} + \mathrm{div}\left(P\left[\langle U\rangle + \langle U'|c(\cdot)=\Psi\rangle\right]\right) = -\frac{\partial}{\partial\Psi}\left(P\left[\langle D\Delta c|c(\cdot)=\Psi\rangle + \langle\chi(\Psi)\rangle\right]\right). \tag{1.43}$$

Dans l'équation 1.43, à gauche, le terme de diffusion peut être paramétré de façon classique avec une fermeture de type théorie-K. A droite, le terme lié aux réactions chimiques $\langle\chi(\Psi)\rangle$

ne nécessite pas de fermeture, contrairement à $\langle D\Delta c|c(\cdot)=\Psi\rangle$ qui correspond à la dissipation visqueuse. Les paramétrisations liées à la fermeture de ce terme sont appelées modèles de micro-mélange. Un modèle simple, appelé IEM (Interaction by Exchange with the Mean), a déjà été appliqué à la qualité de l'air [Fraigneau *et al.*, 1996; Gonzalez, 1997; Luhar et Sawford, 2005]. L'idée est de supposer que le modèle « micro » de dissipation visqueuse échange avec le modèle « macro » (donnant l'évolution des concentrations moyennes $\langle c\rangle$) avec un temps caractéristique noté τ_m :

$$\langle D\Delta c|c(\cdot)=\Psi\rangle = -\frac{1}{\tau_m}\left(\Psi - \langle c\rangle\right). \tag{1.44}$$

Dans ce cas, le temps caractéristique τ_m doit être paramétré, par exemple à l'aide de modèles de LES. Une approche de type Monte Carlo est également possible.

Scalaire conservatif

Une autre méthode héritée du domaine de la combustion est appelée CMC (Conditional Moment Closure), et consiste à relier les concentrations des espèces réactives avec celle d'un scalaire non réactif (conservatif). Cette méthode appliquée aux problèmes de qualité de l'air est présentée dans Mastorakos [2003]. Si ξ est le scalaire conservatif, et c_i la concentration de l'espèce i réactive que l'on cherche à relier à ξ, on s'intéresse à l'évolution de

$$q_i(\eta) = \langle c_i|\xi=\eta\rangle, \tag{1.45}$$

que l'on appelle parfois la fraction de mélange. De plus, on donne au scalaire ξ une PDF $P(\eta)$, qui dépend de la moyenne $\mu=\langle\xi\rangle$ et de la variance $\sigma^2=\langle(\xi-\mu)^2\rangle$ du scalaire. Encore une fois, l'élément clé est donc de spécifier la forme de la PDF. La concentration moyenne de c_i est alors donnée par

$$\langle c_i\rangle = \int_0^1 q_i(\eta)\,P(\eta)\,\mathrm{d}\eta, \tag{1.46}$$

en supposant que ξ a été normalisé (d'où les bornes de l'intégrale). On utilise alors l'équation d'évolution de q_i qu'il est possible de déterminer en fonction des autres valeurs (μ, σ et $P(\eta)$), afin d'en déduire la concentration moyenne. Ce type de méthode est utilisé à l'échelle très locale, notamment dans un panache à proximité d'une cheminée, pour les espèces à chimie rapide (NOx, OH, O_3).

1.3.5 Prise en compte de la variabilité sous-maille a posteriori

PDF des concentrations dans une maille

Il s'agit, là encore, de représenter la variabilité sous-maille des concentrations par une densité de probabilité (PDF). Il s'agit de la variabilité totale, due à l'ensemble des sources de variabilité sous-maille décrites en partie 1.1. La distribution des concentrations à l'intérieur d'une maille est alors représentée par une variable aléatoire, dont la concentration à un capteur n'est qu'une réalisation. Cette approche est faite « off-line », et les équations d'advection-diffusion ne sont pas modifiées. Il s'agit donc plutôt de quantifier a posteriori la variabilité sous-maille, pour en déduire par exemple un intervalle de confiance.

Soit une maille i contenant n_i stations (n_i étant grand). Le modèle prévoit une concentration $\langle x\rangle_i$ dans cette maille ($\langle\cdot\rangle_i$ désigne la moyenne spatiale sur la maille i). La concentration moyenne observée dans la maille (supposée parfaite) à un instant t est

$$\langle y(t)\rangle_i \simeq \frac{1}{n_i}\sum_{k=1}^{n_i} y_k(t), \tag{1.47}$$

avec $y_k(t)$ la concentration observée en une station k de la maille i. On peut alors définir l'erreur modèle, comme l'erreur faite par le modèle à l'échelle où les concentrations sont modélisées (la maille)

$$\epsilon_m(t) = \langle x(t) \rangle_i - \langle y(t) \rangle_i, \tag{1.48}$$

et l'erreur de représentativité à une station comme l'écart entre la moyenne dans une maille, et cette station à l'intérieur de la maille :

$$\epsilon_r(t) = \langle y(t) \rangle_i - y_k(t). \tag{1.49}$$

L'erreur totale du modèle à la station k est alors

$$\epsilon(t) = \epsilon_r(t) + \epsilon_m(t). \tag{1.50}$$

La variabilité sous-maille serait alors donnée par l'ensemble des stations, chacune donnant une réalisation de la variable aléatoire qui représente concentrations. Le problème de cette méthode empirique est qu'en pratique, le nombre de stations par maille est en général trop faible, et la répartition des stations n'est pas uniforme. Une alternative, utilisée dans Ching *et al.* [2006], est d'utiliser un modèle eulérien à une résolution de plus en plus fine, à la façon du « nesting », et de représenter la variabilité sous-maille des grandes mailles grâce au modèle le plus fin. Cette méthode présente également des inconvénients, le principal étant que, comme pour le nesting, il reste des phénomènes sous-maille non résolus. Ching *et al.* [2006] propose d'utiliser des modèles de CFD ou de LES pour descendre à plus fine résolution. L'idée sous-jacente étant de produire des PDF « typiques » de différentes configurations, représentatives de la variabilité sous-maille des concentrations. Cela peut permettre d'évaluer des modèles en comparant les observations avec un intervalle de confiance donné par la PDF, et non avec une seule valeur prise dans la maille. Cependant, une paramétrisation de ces PDF est difficile à obtenir, car les variations entre différentes mailles et en temps sont très importantes.

Réduction d'échelle statistique

Là encore, on considère que l'observation dans une maille est une réalisation particulière d'une variable aléatoire, et diffère de la concentration donnée par le modèle à cause de l'erreur de représentativité. L'idée de variabilité sous-maille n'est plus utilisée explicitement, car on considère simplement les relations statistiques entre une variable à grande échelle (donnée par le modèle), et une variable à petite échelle (donnée par les observations au point de mesure). La variable à grande échelle peut être la concentration dans une maille, voire dans plusieurs mailles adjacentes, par exemple. La relation est calculée via des méthodes statistiques (de régression notamment), sur un temps suffisamment long pour représenter une erreur systématique attribuée à l'erreur de représentativité. Cette méthode, essentiellement utilisée pour la modélisation du climat, a été adaptée à la qualité de l'air dans Guillas *et al.* [2006] et Guillas *et al.* [2008]. L'avantage de cette technique est, d'une part, son faible coût calcul (il s'agit d'un post-traitement), et d'autre part la prise en compte de l'ensemble de l'erreur de représentativité. Cependant, elle ne permet pas de distinguer l'apport de la variabilité des différents processus dans l'erreur de représentativité, ni de distinguer celle-ci par rapport à l'erreur modèle, ou l'erreur dans les observations.

1.4 Conclusion

1.4.1 Bilan

Il existe donc de nombreuses échelles spatiales et temporelles intervenant dans les modèles de qualité de l'air. Elles proviennent en particulier de la variabilité spatiale des émissions, mais

aussi de la turbulence, de l'hétérogénéité des champs météorologiques et de l'occupation des sol, qui ne sont pas prises en compte par les modèles eulériens à une échelle inférieure à celle de la maille. Afin de mieux représenter ces phénomènes sous-maille, la première possibilité est de raffiner les modèles, soit en raffinant le maillage, soit en utilisant des modèles plus précis à échelle locale. Cette solution est très utilisée, mais souvent coûteuse en temps de calcul. De plus, il reste en général toujours une partie des phénomènes sous-maille non représentés, et la variabilité sous-maille n'est pas quantifiée. La seconde possibilité consiste à paramétrer les processus sous-maille afin d'améliorer leur modélisation. Cette solution est en général bonne du point de vue du temps de calcul, et permet de quantifier l'impact du traitement sous-maille pour chacun des processus physiques en jeu. Cependant, ces paramétrisations sont elles-mêmes des approximations, qui doivent être déterminées de façon empirique par une modélisation (LES par exemple) et/ou des observations. Une alternative est d'utiliser des méthodes stochastiques, qui représentent une variable (concentration, coefficient de lessivage, autre paramètre...) par une variable aléatoire. L'un des points clés de ces méthodes est alors de déterminer les formes des PDF utilisées. Ces méthodes sont généralement coûteuses en temps de calcul car elles font souvent intervenir des méthodes de type Monte Carlo. Représenter les concentrations par une densité de probabilité dans une maille revient à dire que l'observation n'est qu'une réalisation particulière d'une variable aléatoire, et que l'on peut déterminer une relation systématique entre les concentrations dans une maille et au point de mesure, par des méthodes statistiques. Cela peut être fait a posteriori, sans modifier le modèle lui-même, notamment par des méthodes statistiques de réduction d'échelle. Ces méthodes présentent l'avantage d'un faible coût calcul et d'une prise en compte de l'ensemble des processus sous-maille, mais ne permettent pas une description détaillée de ces processus.

1.4.2 Choix des méthodes

L'ensemble des méthodes liées aux changements d'échelle est trop large pour être traité de façon exhaustive dans cette thèse. Les choix effectués pour la suite de ce travail sont guidés par un certain nombre de considérations :

1. La plus forte variabilité se concentre autour des zones d'émission, y compris quand il s'agit de quantifier l'impact des fluctuations turbulentes ou de la ségrégation chimique. Se focaliser sur une zone d'émission permet également de définir la notion d'échelle locale (par rapport au point d'émission),
2. Dans la littérature existante, l'impact des hétérogénéités d'émissions est en général beaucoup plus important que celui de la turbulence et de la ségrégation chimique, à l'échelle d'un modèle de chimie-transport.
3. Descendre en résolution ne permet pas nécessairement une amélioration de la performance des modèles car l'amélioration est locale, et n'affecte pas les statistiques globale des modèles,
4. La notion d'échelle de représentativité est également importante ; elle inclut la somme de tous les effets sous-maille mal représentés, et leur effet sur la représentation des concentrations par des mesures ponctuelles.

Les deux premiers points indiquent qu'il est important de se focaliser sur une meilleure représentation sous-maille des émissions. De plus, au travers de cette étude, il est possible d'étudier la représentation de la diffusion et de la chimie à l'échelle locale, et d'en déduire l'impact des hétérogénéités spatiales sur ces deux processus. Le second point indique qu'il n'est pas forcément nécessaire de représenter très finement les processus, si l'on considère seules les performances des modèles. Enfin, le dernier point indique qu'une méthode alternative, focalisée sur les points

d'observation et non sur les sources, peut être un bon complément afin de prendre en compte l'ensemble des erreurs dues à la variabilité sous-maille. Par conséquent, deux méthodes seront étudiées dans cette thèse :

- La représentation sous-maille des émissions par un couplage de type panache sous-maille,
- Les méthodes statistiques de changement d'échelle, basées sur une représentation statistique des liens entre les concentrations à l'échelle d'une maille et les concentrations aux stations de mesure.

Ces deux méthodes ont un coût calcul bien maîtrisé, qui permet de les mettre en place et de les utiliser de façon opérationnelle en prévision de la qualité de l'air. De plus, elles fonctionnent avec relativement peu d'informations supplémentaires, et ne nécessitent pas de description détaillée des émissions ou de l'orographie, qui ne sont pas toujours disponibles. La première méthode permet d'étudier en détail l'impact de l'hétérogénéité des émissions sur les performances des modèles, mais aussi sur la diffusion et la chimie, tandis que la seconde méthode permet de traiter l'erreur de représentativité, notamment en fonction du type de station (et donc de la distance aux principales sources d'émission). A travers l'étude du modèle de panache sous-maille, d'autres questions seront également abordées :

- Quelle est l'échelle spatiale et temporelle de l'impact de la représentation sous-maille des émissions ?
- Quel est l'impact des paramétrisations locales sur cette modélisation ?
- Quelle est l'influence de la résolution du maillage, avec et sans la représentation sous-maille ?

1.4.3 Plan de la thèse

Cette thèse se divise en trois partie. Dans la première, les modèles gaussiens qui ont été implémentés dans la plate-forme de modélisation Polyphemus sont présentés, et évalués par comparaison à des observations ainsi qu'à d'autres modèles gaussiens largement utilisés. Ces modèles sont utilisés pour les sources ponctuelles mais aussi pour les routes. Dans la deuxième partie, le modèle gaussien à bouffées est couplé au modèle eulérien Polair3D, pour le traitement sous-maille des sources ponctuelles. Ce couplage est évalué à l'échelle continentale pour un traceur passif, ainsi qu'à l'échelle régionale pour la photochimie. Enfin, en troisième partie, les méthodes statistiques pour la réduction d'échelle sont présentées. Différentes méthodes statistiques sont évaluées, pour un seul modèle ou pour un ensemble de modèles, pour la photochimie à l'échelle européenne.

1.4.4 Publications associées

Les résultats présentés dans cette thèse ont fait, ou feront, l'objet des publications suivantes :

1. Chapitres 2 et 3 :
   Korsakissok et Mallet, 2009. Comparative study of Gaussian dispersion formulas within the Polyphemus platform : evaluation with Prairie Grass and Kincaid experiments. *J. Applied Meteor.*, 48(12) :2459–2473.
2. Chapitre 4 :
   Korsakissok et Mallet, 2010b. Subgrid-scale treatment for major point sources in an Eulerian model : A sensitivity study on the European Tracer Experiment (ETEX) and Chernobyl cases. *J. Geophys. Res.*, 115(D03303).
3. Chapitre 5 :
   Korsakissok et Mallet, 2010a. Development and application of a reactive plume-in-grid model : evaluation over Greater Paris. *Atmos. Chem. Phys. Discuss.*, 10(2) :5091–5134.

4. Chapitre 7 :
   Korsakissok et Mallet, 2010c. Statistical downscaling of ensemble simulations for air quality modeling. En préparation.
5. Annexe B :
   Korsakissok *et al.*, 2010 Assessment of the detection abilities of monitoring networks for passive tracers at local and regional scales. *Atmos. Env.*, 44(3) :408–420.

Dans le cadre de cette thèse, de nombreux développements ont été intégrés dans la plateforme de modélisation Polyphemus. Ces développements sont librement distribués, documentés et déjà diffusés au-delà du laboratoire d'accueil de la thèse. Ces développements ont été l'occasion de contribuer aux articles suivants :

– Mallet, V., Quélo, D., Sportisse B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M. et Foudhil, H. (2007). Technical Note : The air quality modeling system Polyphemus. *Atmos. Chem. Phys.*. 7(20) :5479–5487

– Mallet, V., Quélo, D., Sportisse B., et Korsakissok, I. (2007). Polyphemus : une plateforme multimodèles pour la pollution atmosphérique et l'évaluation des risques. *Pollution Atmosphérique n°*196.

Première partie

Modèles pour l'échelle locale

Chapitre 2

Description des modèles à l'échelle locale

Les modèles gaussiens permettent de représenter la dispersion à l'échelle locale, dans les premières dizaines de kilomètres autour d'une source. Ces modèles sont très utilisés car ils donnent de bons résultats à cette échelle, tout en fournissant une formulation analytique de la dispersion qui est conceptuellement simple et peu coûteuse à mettre en place et à utiliser. C'est pourquoi la littérature sur le sujet abonde depuis au moins trente ans, et l'utilisation des modèles gaussiens est toujours très répandue malgré les progrès dans des domaines de modélisation plus complexes ainsi que dans les performances de calcul. Dans ce chapitre, il s'agit de faire un bilan de l'état de l'art sur les modèles gaussiens, et de présenter les modèles implémentés dans la plate-forme de modélisation Polyphemus en décrivant l'ensemble des paramétrisations empiriques disponibles dans ces modèles.

Sommaire

2.1 Les modèles gaussiens . . . 51
- 2.1.1 Lien avec la formulation eulérienne . . . 51
- 2.1.2 Modèle gaussien de panache . . . 53
- 2.1.3 Modèle gaussien à bouffées . . . 56
- 2.1.4 Modèle de source linéique . . . 58

2.2 Modélisation de la dispersion . . . 59
- 2.2.1 Variables de la dispersion turbulente . . . 59
- 2.2.2 Formules de Briggs . . . 63
- 2.2.3 Formule de Doury . . . 64
- 2.2.4 Théorie de la similitude . . . 65

2.3 Modélisation de la surhauteur . . . 68
- 2.3.1 Surhauteur de Briggs et HPDM . . . 69
- 2.3.2 Formule de Concawe . . . 70
- 2.3.3 Formule de Holland et Stümke . . . 70
- 2.3.4 Pénétration partielle dans la couche d'inversion . . . 71

2.4 Modélisation des processus de perte . . . 72
- 2.4.1 Décroissance radioactive . . . 72
- 2.4.2 Dépôt sec . . . 72
- 2.4.3 Lessivage . . . 74

2.5 Bilan : modèles de Polyphemus . . . 75

- Les modèles présentés ici sont décrits et évalués dans Korsakissok et Mallet [2009]
- La modélisation des sources linéiques est décrite et analysée dans Briant [2009], rapport de stage co-encadré par Irène Korsakissok et Christian Seigneur,
- Les paramétrisations pour le dépôt et le lessivage dans les modèles gaussiens sont présentées dans Korsakissok *et al.* [2006].

Pour approfondir la modélisation gaussienne de la dispersion, on pourra consulter

- Arya [1999],
- Hanna *et al.* [1982].

2.1 Les modèles gaussiens

Il existe deux types de modèles gaussiens. D'une part, le modèle de panache, ou gaussien stationnaire, modélise le panache émis par une source ponctuelle par une distribution gaussienne dans deux directions (horizontale perpendiculaire au vent, et verticale), et suppose une météorologie stationnaire. D'autre part, le modèle à bouffées modélise une émission instantanée par une bouffée gaussienne dans les trois directions. Les différences entre ces deux types de modèles, les hypothèses faites dans les deux cas, ainsi que les modifications des formules pour prendre en compte les réflexions et le champ lointain, sont détaillées dans cette partie. Dans la plate-forme Polyphemus [Mallet *et al.*, 2007] , un modèle gaussien de panache, « plume », ainsi qu'un modèle à bouffées, « puff », sont disponibles.

2.1.1 Lien avec la formulation eulérienne

Il existe deux approches pour modéliser le comportement d'une espèce trace dans l'atmosphère, c'est-à-dire sa distribution spatiale et temporelle. L'approche *eulérienne* consiste à décrire cette distribution dans un référentiel fixe, en fonction des caractéristiques du fluide porteur en un point donné. L'approche *lagrangienne* se place dans le référentiel du fluide en mouvement, et décrit le comportement statistique d'un groupe de particules en déplacement. Ces deux approches donnent des formulations différentes, qui peuvent être reliées entre elles. Le modèle gaussien à bouffées peut être considéré comme un modèle lagrangien simplifié, dans la mesure où l'on « suit » le polluant émis sur sa trajectoire : la distribution statistique d'un grand nombre de particules est simplifiée et modélisée par une distribution gaussienne appelée bouffée. On présente ici la description eulérienne de la dispersion atmosphérique d'un polluant, et la façon dont la description gaussienne des sources ponctuelles s'en déduit. Pour plus de détails sur les calculs décrits ci-dessous, on se référera à Seinfeld et Pandis [1998] dont cette partie s'inspire largement.

Approche eulérienne

Dans le cas général d'une approche eulérienne, la concentration c d'une espèce non réactive dans l'atmosphère est représentée par l'équation d'advection-diffusion déjà présentée au chapitre 1 (équation 1.19)

$$\frac{\partial c}{\partial t} = -\mathrm{div}(Vc) + \mathrm{div}\left(\rho K \nabla \frac{c}{\rho}\right) + S, \tag{2.1}$$

avec ρ la densité du fluide porteur, V le vent moyen et K la matrice de diffusion turbulente. Enfin, S est le terme représentant les sources et les puits de l'espèce. Ici, la matrice de diffusion turbulente K a été utilisée pour fermer l'équation et représenter les termes du second ordre (fermeture du premier ordre appelée « théorie-K »). On a supposé de plus que cette diffusion turbulente est très grande devant la diffusion moléculaire, qui est donc négligée. La matrice K n'est pas connue, et doit être estimée par des paramétrisations empiriques, où l'on suppose que les termes extra-diagonaux sont négligeables. Elle s'écrit alors

$$K = \begin{pmatrix} K_x & 0 & 0 \\ 0 & K_y & 0 \\ 0 & 0 & K_z \end{pmatrix}.$$

Représentation eulérienne d'une source instantanée

On se place à présent dans le cas d'une seule source ponctuelle, de coordonnées $(0, y_s, z_s)$, émettant une masse totale Q de façon instantanée à $t = 0$, et sans processus de pertes. On

considère une situation météorologique constante et homogène, avec un vent moyen $V = (\bar{u}, 0, 0)$. On considère enfin que les coefficients de diffusion K_x, K_y et K_z sont constants. L'équation 2.1 s'écrit alors

$$\frac{\partial c}{\partial t} + \bar{u}\frac{\partial c}{\partial x} = K_x \frac{\partial^2 c}{\partial x^2} + K_y \frac{\partial^2 c}{\partial y^2} + K_z \frac{\partial^2 c}{\partial z^2}, \tag{2.2}$$

avec, pour condition initiale

$$c(x, y, z, 0) = Q\delta(x)\delta(y - y_s)\delta(z - z_s), \tag{2.3}$$

et pour conditions aux limites

$$c(x, y, z, t) = 0 \qquad x, y, z \to \pm\infty. \tag{2.4}$$

Pour résoudre cette équation, on écrit la concentration sous la forme

$$c(x, y, z, t) = Q\ G_x(x, t) \times G_y(y, t) \times G_z(z, t) \quad \text{avec} \tag{2.5}$$

$$G_\alpha(\alpha, 0) = \delta(\alpha - \alpha_s) \qquad \alpha \in \{x, y, z\}. \tag{2.6}$$

Cette équation a pour solution analytique

$$G_\alpha(\alpha, t) = \frac{1}{2(\pi t K_\alpha)^{1/2}} \times \exp\left(-\frac{(\alpha - \alpha_s - V_\alpha t)^2}{4K_\alpha t}\right), \qquad \alpha \in \{x, y, z\}. \tag{2.7}$$

avec $x_s = 0$, $V_x = \bar{u}$ et $V_y = V_z = 0$. On a donc finalement une expression de la concentration

$$c(x, y, z, t) = \frac{Q}{8(\pi t)^{3/2}\,(K_x K_y K_z)^{1/2}} \times \exp\left(-\frac{(x - \bar{u}t)^2}{4K_x t}\right) \exp\left(-\frac{(y - y_s)^2}{4K_y t}\right) \exp\left(-\frac{(z - z_s)^2}{4K_z t}\right) \tag{2.8}$$

Représentation eulérienne d'une source continue

On se place à présent dans le cas d'une seule source ponctuelle, de coordonnées $(0, y_s, z_s)$, ayant un débit massique constant Q_s, et sans pertes processus de pertes. On considère une situation météorologique constante et homogène, avec un vent moyen $V = (\bar{u}, 0, 0)$. Pour simplifier, on considère pour le moment que la turbulence K est constante et homogène. L'équation 2.1 devient donc stationnaire et s'écrit alors

$$\bar{u}\frac{\partial c}{\partial x} = K\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2}\right) + Q_s\delta(x)\delta(y - y_s)\delta(z - z_s) \tag{2.9}$$

avec, comme condition aux limites

$$c(x, y, z, t) = 0 \qquad x, y, z \to \pm\infty. \tag{2.10}$$

Cette équation peut se résoudre en utilisant la variable $r^2 = x^2 + (y - y_s)^2 + (z - z_s)^2$. On obtient alors

$$c(x, y, z) = \frac{Q_s}{4\pi|K|r} \times \exp\left(-\frac{\bar{u}(r - x)}{2K}\right). \tag{2.11}$$

Il est nécessaire de rajouter ici une hypothèse supplémentaire, qui suppose que le panache est étroit dans la direction perpendiculaire au vent y (« slender plume approximation »). On a alors $x^2 \gg (y - y_s)^2 + (z - z_s)^2$, et la distance r peut s'écrire

$$r \simeq x\left(1 + \frac{(y - y_s)^2 + (z - z_s)^2}{2x^2}\right). \tag{2.12}$$

L'approximation de panache étroit revient à supposer que le vent (dans la direction x) est suffisamment important par rapport à la turbulence. De ce fait, on peut négliger la diffusion turbulente dans cette direction. Dans l'équation 2.11, on remplace alors r par x et $r - x$ par $\frac{(y-y_s)^2+(z-z_s)^2}{2x}$. Enfin, la procédure précédemment décrite se généralise aisément au cas de turbulence non homogène $K_x \neq K_y \neq K_z$. L'équation devient alors

$$c(x,y,z,t) = \frac{Q_s}{4\pi(K_yK_z)^{1/2}x} \times \exp\left(-\frac{\bar{u}(y-y_s)^2}{4K_yx}\right)\exp\left(-\frac{\bar{u}(z-z_s)^2}{4K_zx}\right). \tag{2.13}$$

2.1.2 Modèle gaussien de panache

En sortant du cadre eulérien, on cherche à représenter simplement une source ponctuelle continue, en supposant que la distribution des concentrations est gaussienne. On fait les hypothèses suivantes :

1. Une émission ponctuelle continue (donc active pendant un temps assez long pour avoir un panache stabilisé entre la source et le point observé le plus lointain), de débit constant Q_s,
2. Des vents suffisamment importants pour que la diffusion turbulente dans la direction du vent soit négligeable en comparaison de l'advection (approximation de panache étroit),
3. Conditions météorologiques uniformes et constantes (obtention d'un panache stable avant que la situation météorologique n'évolue).

Formule de panache gaussien

L'hypothèse (1) suggère que la concentration en un point ne varie pas au cours du temps ; en pratique, il s'agit donc de considérer des concentrations moyennées sur un temps suffisamment long (correspondant souvent au temps d'intégration d'un instrument de mesure). Du fait de la météorologie constante (hypothèse (3)), il n'y a donc plus de dépendance explicite au temps (on parle parfois de « gaussien stationnaire »). De plus, l'hypothèse (2) permet de négliger la turbulence dans la direction x (direction du vent). Si l'on fait l'hypothèse que le panache est représenté par une distribution gaussienne dans les deux directions y et z, centrée sur les coordonnées de la source, et d'écarts types (σ_y, σ_z), on peut alors écrire

$$\begin{aligned} c(y,z) &= \frac{Q_s}{\bar{u}}\, G_y(y-y_s)\, G_z(z-z_s) \\ &= \frac{Q_s}{\bar{u}} \left[\frac{1}{\sqrt{2\pi}\sigma_y}\exp\left(-\frac{(y-y_s)^2}{2\sigma_y^2}\right)\right]\left[\frac{1}{\sqrt{2\pi}\sigma_z}\exp\left(-\frac{(z-z_s)^2}{2\sigma_z^2}\right)\right] \end{aligned} \tag{2.14}$$

avec Q_s le débit de la source (en unité de masse par seconde), et $\bar{u}$ la vitesse moyenne du vent. Les fonctions G_α sont celles définies par l'équation 2.7, où la dépendance à t a été éliminée en supposant que $t = x/\bar{u}$ (hypothèse de météorologie stationnaire). L'équation 2.14 donne une expression analytique de la concentration émise par une source ponctuelle continue en tout point de l'espace (figure 2.1).

L'équation 2.14 peut donc être reliée à la solution de l'équation eulérienne obtenue sous les mêmes hypothèses (équation 2.13), en supposant $t = x/\bar{u}$ et en reliant les écarts types gaussiens au coefficient de diffusion turbulente de l'équation 2.1 par la relation

$$\sigma_\alpha = \sqrt{2K_\alpha t}, \qquad \alpha \in \{x,y,z\}. \tag{2.15}$$

Il convient de noter cependant que cette forme en $\sigma_\alpha \propto t^{1/2}$ ne s'applique que très loin de la source, une fois que la taille du panache couvre l'ensemble du spectre de taille des tourbillons

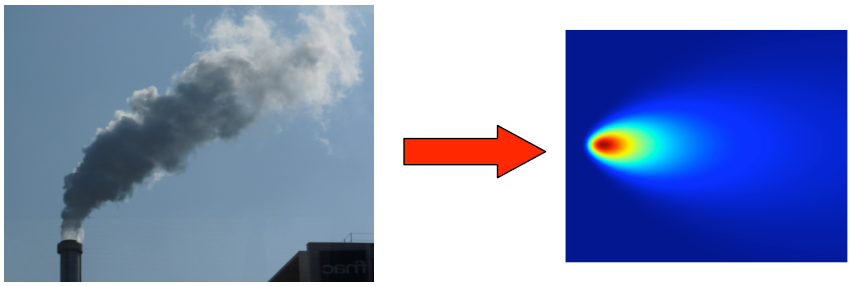


FIG. 2.1 – Modèle gaussien de panache stationnaire : le panache émis par une cheminée est représenté par une distribution gaussienne dans deux directions. A gauche : photo de panache issu d'une cheminée (crédit : Yelva Roustan). A droite : exemple de sortie du modèle de panache gaussien de Polyphemus.

de la turbulence. Proche de la source, σ_α croît plus rapidement. Les formules donnant les écarts types d'un panache atmosphérique en fonction de la distance à la source (ou du temps de trajet de la bouffée) sont, en pratique, déterminées de façon empirique sur des expériences de dispersion. Dans la plupart de ces formules, on a en général $\sigma_\alpha \propto t^{1/2}$ lorsque t est suffisamment grand. Les formulations utilisées dans les modèles développés dans cette thèse sont détaillées en partie 2.2.

Notons par ailleurs que l'on a

$$\int_{-\infty}^{\infty} G_\alpha(\alpha)\mathrm{d}\alpha = 1, \tag{2.16}$$

ce qui permet donc de s'assurer de la conservation de la masse

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \bar{u}c(y,z)\mathrm{d}y\ \mathrm{d}z = Q. \tag{2.17}$$

Conditions aux limites

Il n'est pas exact de supposer un panache s'étend dans tout l'espace. Il existe en effet des conditions aux limites au sol ($z = 0$), et éventuellement à la hauteur d'inversion ($z = z_i$). La hauteur d'inversion est la hauteur à laquelle la température de l'atmosphère augmente avec l'altitude au lieu de diminuer, empêchant ainsi l'air chargé de polluants de s'élever davantage. Il s'agit donc de la hauteur maximale de mélange : elle est modélisée comme un « plafond », de façon similaire au sol. Comme cette hauteur varie (de quelques dizaines de mètres la nuit, jusqu'à 1 ou 2 km), des phénomènes d'entraînement se produisent en pratique, mais ils sont négligés dans le cadre de l'échelle locale et de la météorologie stationnaire. Dans les modèles de Polyphemus, cette variation est décrite de façon binaire. Pendant la nuit, l'inversion est supposée être proche du sol, et la plupart des polluants sont au-dessus de cette inversion. Pendant le jour, la hauteur d'inversion est supposée élevée, c'est-à-dire de l'ordre de plusieurs centaines de mètres (et non quelques dizaines, comme de nuit), et donc au-dessus des sources : les réflexions sur la couche d'inversion sont donc modélisées. La définition de la couche d'inversion et ses implications pratiques sont détaillées en partie 2.2.1, notamment avec la figure 2.4.

Le panache n'est pas absorbé par le sol et la couche d'inversion, mais s'y réfléchit : si l'on n'a aucun processus de perte de type dépôt, et aucun échange avec la troposphère libre, les réflexions sont parfaites. La prise en compte du sol s'effectue en modélisant une source virtuelle

« sous » le sol, à la distance $-z_s$. L'équation 2.14 est alors modifiée et devient :

$$C(x,y,z) = \frac{Q_s}{2\pi\sigma_y\sigma_z\bar{u}} \exp(-\frac{(y-y_s)^2}{2\sigma_y^2}) \left[\exp\left(-\frac{(z-z_s)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+z_s)^2}{2\sigma_z^2}\right)\right]. \tag{2.18}$$

Dans le cas où une couche d'inversion est présente à la hauteur z_i, les réflexions sur celle-ci sont prises en compte de façon similaire. La concentration est alors donnée par l'équation 2.19

$$\begin{aligned} C(x,y,z) &= \frac{Q_s}{2\pi\sigma_y\sigma_z\bar{u}} \exp\left(-\frac{(y-y_s)^2}{2\sigma_y^2}\right) \times \\ & \sum_{N=-N_r}^{+N_r} \left[\exp\left(-\frac{(z-z_s+2Nz_i)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+z_s+2Nz_i)^2}{2\sigma_z^2}\right)\right], \end{aligned} \tag{2.19}$$

où N_r réflexions sur la couche d'inversion sont considérées. Dans Arya [1999], la valeur $N_r = 5$ est présentée comme le maximum au-delà duquel les termes de la somme sont négligeables. En pratique, N_r est souvent pris égal à 1, comme pour le sol. Les formulations précédentes présentent l'avantage de la simplicité, mais ne sont valides que pour $0 \leq z \leq z_i$. En effet, si l'on supposait que ces formules donnent la concentration dans tout l'espace, l'intégrale sur tout l'espace dans la direction z s'écrirait

$$\int_{-\infty}^{\infty} \left[\exp\left(-\frac{(z-z_s)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+z_s)^2}{2\sigma_z^2}\right)\right] \mathrm{d}z = 2 \tag{2.20}$$

avec

$$\begin{aligned} & \int_{-\infty}^{\infty} \left[\exp\left(-\frac{(z-z_s)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+z_s)^2}{2\sigma_z^2}\right)\right] \mathrm{d}z \\ = & \int_{-z_i}^{z_i} \left[\exp\left(-\frac{(z-z_s)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+z_s)^2}{2\sigma_z^2}\right)\right] \mathrm{d}z \\ = & \; 2\int_{0}^{z_i} \left[\exp\left(-\frac{(z-z_s)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+z_s)^2}{2\sigma_z^2}\right)\right] \mathrm{d}z \end{aligned} \tag{2.21}$$

Pour assurer la conservation de la masse, il faut donc mettre les concentrations à zéro en dehors de la couche limite. L'équation 2.17 de conservation de la masse devient donc

$$\int_0^{z_i} \int_{-\infty}^{\infty} c\bar{u} \; \mathrm{d}y \; \mathrm{d}z = Q_s. \tag{2.22}$$

Notons que les réflexions ne sont pas prises en compte systématiquement dans les modèles développés pour cette thèse, mais seulement lorsque le panache est suffisamment proche du sol ou de la couche d'inversion

- Il y a réflexion au sol si $\sigma_z > z_s$
- Il y a réflexion sur la couche d'inversion si $z_s + \sigma_z > z_i$.

Enfin, lorsque le panache est suffisamment étendu sur la verticale ($\sigma_z > 1.5\ z_i$), on considère que les différentes réflexions et le mélange turbulent l'ont rendu homogène sur la verticale. La concentration est alors donnée par la formule de champ lointain 2.23

$$C(x,y,z) = \frac{Q_s}{2\pi\sigma_y z_i\bar{u}} \exp\left(-\frac{(y-y_s)^2}{2\sigma_y^2}\right). \tag{2.23}$$

2.1.3 Modèle gaussien à bouffées

Formule de bouffée gaussienne

Dans le cas d'un modèle instationnaire, le panache est représenté sous la forme d'une série de bouffées gaussiennes dans les trois directions. Les hypothèses présentées en partie 2.1.2 sont alors moins contraignantes : les hypothèses 1 et 2 n'ont plus lieu d'être. De plus, on suppose simplement que la météorologie est uniforme à l'intérieur d'une même bouffée (hypothèse 3), mais elle peut varier d'une bouffée à l'autre, et dans le temps. Dans le cas d'une seule bouffée émise, de quantité totale Q, La concentration de la bouffée supposée gaussienne dans les trois directions peut s'écrire

$$C(x,y,z) = \frac{Q}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z} \exp\left(-\frac{(x-x_c)^2}{2\sigma_x^2}\right) \exp\left(-\frac{(y-y_c)^2}{2\sigma_y^2}\right) \exp\left(-\frac{(z-z_c)^2}{2\sigma_z^2}\right). \quad (2.24)$$

Les coordonnées du centre de la bouffée sont x_c dans la direction du vent, y_c et z_c dans les directions horizontale (perpendiculaire au vent) et verticale. Dans le cas d'un vent constant u, on a donc $x_c = u\ t_c$ avec t_c le temps écoulé depuis que la bouffée a été émise (l'« âge » de la bouffée). Les trois écarts types de la gaussienne sont σ_x et σ_y dans la direction horizontale (direction du vent et perpendiculaire au vent respectivement), et σ_z dans la direction verticale. De même que dans le cas du panache stationnaire, l'équation 2.24 est identique à la solution eulérienne 2.8 en supposant que les écarts types gaussiens et la diffusion eulérienne sont reliés par l'équation 2.15. En général, les formules de dispersion pour σ_y et σ_z sont les mêmes que dans le cas du modèle de panache. Pour les paramétrisations qui ne donnent pas de formulation spécifique de σ_x, celle de σ_y est utilisée.

Les réflexions sur le sol et la couche d'inversion ainsi que le champ lointain sont pris en compte de manière similaire au modèle de panache (équations 2.19 et 2.23 respectivement).

Liens entre les formulations de panache et de bouffée

Dans un cas de source continue et de météorologie uniforme, la moyenne sur un temps suffisamment long du modèle à bouffées permet de retrouver la solution stationnaire du modèle de panache (figure 2.1.3).

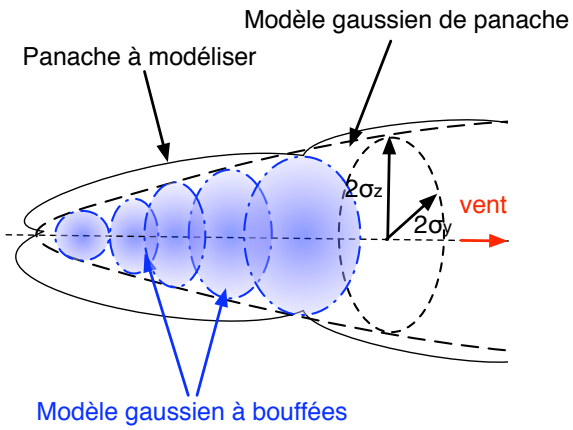


FIG. 2.2 – Modèle gaussien instationnaire à bouffées : le panache est discrétisé en une série de bouffées, gaussiennes dans les trois directions.

Si la source continue modélisée en partie 2.1.2, de débit Q_s, est discrétisée en un certain nombre de bouffées, à un pas de temps Δt_{puff}, alors chaque bouffée contiendra la quantité

$Q = \Delta t_{\text{puff}} \times Q_s$. La concentration en un point est donnée par la somme de la contribution de toute les bouffées

$$C(x,y,z) = \sum_{i=1}^{N_{\text{puff}}} \frac{Q_s \times \Delta t_{\text{puff}}}{(2\pi)^{3/2}\sigma_x^i\sigma_y^i\sigma_z^i} \exp\left(-\frac{(x-x_c^i)^2}{2{\sigma_x^i}^2}\right) \exp\left(-\frac{(y-y_c^i)^2}{2{\sigma_y^i}^2}\right) \exp\left(-\frac{(z-z_c^i)^2}{2{\sigma_z^i}^2}\right) \tag{2.25}$$

Les coordonnées du centre de la bouffée i émise au temps $t_i = i\Delta t_{\text{puff}}$ sont x_c^i, y_c^i et z_c^i. Les écarts types associés sont notés σ_x^i, σ_y^i, σ_z^i, et dépendent de la position de la bouffée par rapport à la source (ou de son âge). Ces coordonnées évoluent au cours du temps. Dans le cas d'un vent $\bar{u}$ constant et homogène dans la direction x, on a $x_c^i(t) = x_s + \bar{u}\ (t - t_i)$, $y_c = y_s$ et $z_c = z_s$.

Le passage du cas discret (équation 2.25) au cas du panache continu s'écrit en intégrant sur des bouffées émises à des pas de temps infinitésimaux dt'

$$\begin{aligned} C(x,y,z) \quad &= \int_{-\infty}^{\infty} \frac{Q_s \times dt'}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z} \exp\left(-\frac{(x-x_s-\bar{u}\ (t-t'))^2}{2\sigma_x^2}\right) \\ &\quad \exp\left(-\frac{(y-y_s)^2}{2\sigma_y^2}\right) \exp\left(-\frac{(z-z_s)^2}{2\sigma_z^2}\right). \end{aligned} \tag{2.26}$$

Reprenons à présent les hypothèses faites en partie 2.1.2 : (1) la concentration est donnée en moyennant sur un temps d'intégration T (car elle ne dépend pas du temps, et que la source émet en permanence), et (2) la dispersion turbulente dans la direction x est négligeable devant la distance à parcourir dans cette direction :

$$\sigma_x \ll x - x_s. \tag{2.27}$$

De plus, le panache est supposé dans un état stationnaire, ce qui revient à dire que la source a commencé à émettre très longtemps avant la mesure (à $t' = -\infty$). Les conditions météorologiques sont également stationnaires (hypothèse (3) dans la partie 2.1.2). Les valeurs de σ_x, σ_y et σ_z en un point donné (x,y,z) dépendent du temps de transport de chaque bouffée $t - t'$. Cependant, étant donné les conditions stationnaires, on considère que toutes les bouffées arrivent au point (x,y,z) avec le même temps de transport, et que les valeurs des écart types en un point donné sont donc constantes au cours du temps. L'équation 2.26 devient alors

$$\begin{aligned} C(x,y,z) = \frac{1}{T}\frac{Q_s}{2\pi\sigma_y\sigma_z} &\exp\left(-\frac{(y-y_s)^2}{2\sigma_y^2}\right) \exp(-\frac{(z-z_s)^2}{2\sigma_z^2}) \\ &\int_0^T \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(\bar{u}\ (t-t'))^2}{2\sigma_x^2}\right) dt'dt. \end{aligned} \tag{2.28}$$

En faisant le changement de variable $t'' = t - t'$, on calcule l'intégrale

$$\int_0^T \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(\bar{u}\ (t-t'))^2}{2\sigma_x^2}\right) dt'dt = \frac{T}{\bar{u}}, \tag{2.29}$$

ce qui permet donc de retrouver l'expression de la concentration du panache gaussien (équation 2.14)

$$C(y,z) = \frac{Q_s}{\bar{u}} G_y(y) G_z(z). \tag{2.30}$$

La démonstration précédente permet de comprendre les hypothèses et les limitations des deux modèles. Le modèle gaussien de panache fait des hypothèses plus contraignantes, notamment de météorologie constante et homogène sur tout le domaine, contrairement au modèle à bouffées.

Cette hypothèse n'est plus valide hors du champ proche, et peut poser problème en cas de cisaillement de vent important sur la verticale (ce qui est également le cas pour les bouffées). De plus, le fait de négliger la turbulence horizontale en regard du transport (hypothèse 2) n'est plus valide en cas de vent très faible.

Si l'on reste dans son domaine de validité, la formulation de panache gaussien est cependant plus appropriée, pour des sources émettant en continu, que la formulation instationnaire. En effet, le modèle à bouffée nécessite de discrétiser le panache de façon suffisamment fine, et de faire une moyenne sur un temps assez long pour retrouver la solution stationnaire. Si l'on considère qu'une bouffée a pour taille $2\sigma_x$ dans la direction x, deux bouffées successives émises à t_i et $t_{i+1} = t_i + \Delta t_{\text{puff}}$ se recouvrent au temps t lorsque

$$\sigma_x(t - t_i) + \sigma_x(t - t_{i+1}) \geq \frac{\bar{u}}{2}\Delta t_{\text{puff}}. \tag{2.31}$$

Il n'est pas nécessaire, pour bien modéliser un panache continu, de remplir cette condition à très faible distance de la source : il suffit qu'elle soit remplie aux points d'intérêt (par exemple, à un premier capteur de mesure). Si le vent est $\bar{u} = 5\,\mathrm{m\,s^{-1}}$, et que $\sigma_x(t) = 1000\,\mathrm{m}$ (taille atteinte au bout d'une heure environ dans une atmosphère neutre), on obtient $\Delta t_{\text{puff}} = 800\,\mathrm{s}$. Cette valeur minimale est adaptée dans une application régionale, où les capteurs sont éloignés de la source. Dans une application à échelle locale (quelques kilomètres), la valeur typique sera plutôt de l'ordre de 60 s ou moins.

2.1.4 Modèle de source linéique

Jusqu'à présent, les modèles de panache et de bouffées décrits ne modélisent que des sources ponctuelles. Cependant, ces formulations peuvent être adaptées pour représenter des sources linéiques telles que des tronçons routiers. Les équations décrites ici sont l'adaptation du modèle de panache, le modèle à bouffées pouvant être adapté de façon similaire (non encore implémentée dans le modèle à bouffées de Polyphemus). Pour une description plus complète de ce type de modélisation, on se référera par exemple à Yamartino [2008]. Une alternative, plus coûteuse en temps de calcul, est de discrétiser une source linéique en une série de sources ponctuelles.

Prenons le cas où la source est placée perpendiculairement à la direction du vent, suivant l'axe y. Les ordonnées des deux extrémités sont notées y_1 et y_2, et l'on suppose que l'abscisse de la source est $x = 0$. Il suffit donc d'intégrer la solution ponctuelle (équation 2.14) entre les deux extrémités

$$c(x, y, z, t) = \frac{Q}{2\pi u \sigma_y \sigma_z} \exp\left(-\frac{z^2}{2{\sigma_z}^2}\right) \int_{y_1}^{y_2} \exp\left(-\frac{y^2}{2{\sigma_y}^2}\right) \mathrm{d}y \tag{2.32}$$

et l'on obtient alors

$$c(x, y, z) = \frac{Q}{2\sqrt{2\pi} u \sigma_z} \left[\operatorname{erf}\left(\frac{y - y_1}{\sqrt{2}\sigma_{y_1}}\right) - \operatorname{erf}\left(\frac{y - y_2}{\sqrt{2}\sigma_{y_2}}\right)\right] \tag{2.33}$$

Si l'on considère à présent que le vent fait un angle θ par rapport à la source, il faut effectuer un changement de repère afin de se placer dans le repère du vent

$$x_{vent} = x\cos\theta + y\sin\theta \tag{2.34}$$

$$y_{vent} = -x\sin\theta + y\cos\theta \tag{2.35}$$

Les écarts types sont exprimés en fonction de x_{vent}, qui dépendent de x et de y. Il est nécessaire de faire une approximation, afin que σ_z ne dépende plus que de x, et ne soit pas

inclus dans l'intégrale sur y. Pour cela, on suppose que la distance prise pour calculer σ_z est la distance du point (x, y, z) à une droite passant par la source linéique [Venkatram et Horst, 2006]

$$d = \frac{x}{\cos\theta}. \tag{2.36}$$

L'équation que l'on doit intégrer est donc

$$c(x, y, z) = \frac{Q}{2\pi u \sigma_y \sigma_z} \exp\left(-\frac{z^2}{2{\sigma_z}^2}\right) \int_{y_1}^{y_2} \exp\left(-\frac{(y_{vent} - s)^2}{2\sigma_y(s)^2}\right) \mathrm{d}s. \tag{2.37}$$

Tous calculs faits, et changements de variables effectués, on obtient [Briant, 2009]

$$\begin{aligned} c(x, y, z) \quad = \quad & \frac{Q}{2\sqrt{2\pi} u \cos\theta \sigma_z} \exp\left(-\frac{z^2}{2{\sigma_z}^2}\right) \times \\ & \left[\mathrm{erf}\left(\frac{(y - y_1)\cos\theta - x\sin\theta}{\sqrt{2}\sigma_{y_1}}\right) - \mathrm{erf}\left(\frac{(y - y_2)\cos\theta - x\sin\theta}{\sqrt{2}\sigma_{y_2}}\right)\right]. \end{aligned} \tag{2.38}$$

Dans cette équation, σ_{y_1} et σ_{y_2} sont calculés à partir des distances du point à chacune des deux extrémités de la source dans la direction du vent. En revanche, σ_z est calculé à partir de la distance d donnée par l'équation 2.36.

2.2 Modélisation de la dispersion

Les formules présentées en partie 2.1 nécessitent d'évaluer les écarts types gaussiens, σ_x, σ_y et σ_z. Ces écarts types sont donnés par des formules empiriques, calibrées sur de nombreuses expériences permettant d'évaluer la dispersion pour différentes conditions météorologiques, ainsi que pour des types de sol (ruraux, urbains) variés. L'expression totale des écarts types gaussiens peut tenir compte, non seulement de l'extension du panache due à la turbulence ($\sigma_{y_{\mathrm{turb}}}$, $\sigma_{z_{\mathrm{turb}}}$), mais aussi du diamètre initial de la source d_s, et de l'élargissement ajouté par la surhauteur ($\sigma_{y_{\mathrm{pr}}}$, $\sigma_{z_{\mathrm{pr}}}$) décrit dans la partie 2.3. Les écarts types sont alors donnés par

$$\sigma_y^2 = \sigma_{y_{\mathrm{turb}}}^2 + \sigma_{y_{\mathrm{pr}}}^2 + \frac{d_s^2}{4}, \qquad \sigma_z^2 = \sigma_{z_{\mathrm{turb}}}^2 + \sigma_{z_{\mathrm{pr}}}^2 \tag{2.39}$$

Dans la suite, par simplicité, on notera les écarts types turbulents $\sigma_{y_{\mathrm{turb}}}$ et $\sigma_{z_{\mathrm{turb}}}$ simplement σ_y et σ_z. Ils sont paramétrés par les formules empiriques décrites ci-dessous.

2.2.1 Variables de la dispersion turbulente

L'équation eulérienne 2.1 modélise le transport des polluants dans l'atmosphère à travers l'advection, c'est-à-dire le déplacement horizontal par le vent, et la diffusion, qui correspond au brassage par la turbulence atmosphérique. Dans le cas eulérien, cette diffusion est représentée par le coefficient K ; dans le cas gaussien, elle est donnée par les écarts types gaussiens. Dans les deux cas, des formules empiriques sont utilisées pour donner une expression de cette diffusion turbulente. Ces formules sont basées sur quelques paramètres décrivant l'état de l'atmosphère, qui sont brièvement définis dans cette partie. Pour une description plus complète, on se référera à Hanna *et al.* [1982], Arya [1999] et Seinfeld et Pandis [1998].

En pratique, on peut distinguer deux phénomènes contribuant à la diffusion turbulente :

- la turbulence d'origine dynamique, qui correspond au cisaillement du vent dû au frottement sur le sol,
- la turbulence d'origine thermique, liée au gradient vertical de température potentielle. La température potentielle θ est la température qu'aurait un gaz ramené adiabatiquement de la température T et la pression P à la pression atmosphérique au sol $P_0 = 1000$ hPa.

Couche limite et cisaillement de vent

On définit la couche limite atmosphérique (ou, simplement « couche limite ») comme la partie de l'atmosphère qui est directement influencée par la surface terrestre, dans un temps caractéristique d'environ 1 heure. La description de la dispersion des polluants se fait à l'intérieur de la couche limite (aux échelles modélisées ici). La hauteur de couche limite h varie de quelques dizaines de mètres (la nuit) à quelques kilomètres. La partie de l'atmosphère située au-dessus de la couche limite se nomme « troposphère libre » et n'influence que peu les concentrations au sol, excepté par les nuages qui s'y développent. La couche limite se divise entre la « couche d'Ekman », où la direction du vent tourne avec la hauteur à cause de la force de Coriolis, et la « couche de surface », proche du sol. Dans cette couche, la direction du vent ainsi que les flux turbulents (turbulence mécanique et thermique) sont constants (ou varient de moins de 10%). On considère que la hauteur de la couche de surface est d'environ 10% celle de la couche limite. A l'intérieur de cette couche, il peut y avoir une sous-couche laminaire de quelques dizaines de centimètres. Elle dépend de la hauteur typique des obstacles générant de la turbulence (notée z_0, et appelée « hauteur de rugosité »). La figure 2.3 représente ces différents paramètres de la couche limite.

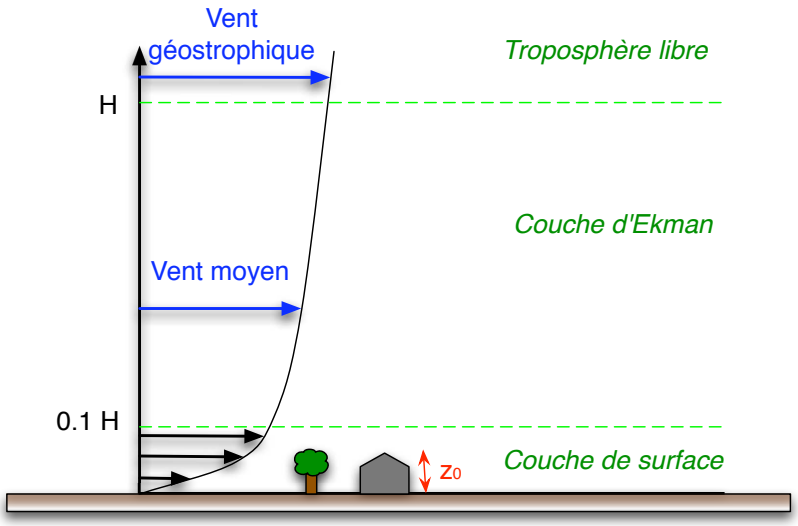


FIG. 2.3 – Représentation de la couche limite atmosphérique : couche limite de hauteur h, divisée entre la couche d'Ekman et la couche de surface.

Stabilité

Si les échanges horizontaux sont liés essentiellement au transport par le vent, le vent vertical est très faible et le transport vertical des polluants est principalement turbulent. Il est lié notamment à la turbulence d'origine thermique, due au réchauffement au sol par le rayonnement solaire. De plus, la convection nuageuse, lorsqu'elle se produit, est une source importante de transport vertical.

La turbulence atmosphérique est souvent associée à la notion de *stabilité*, qui dépend du profil vertical de température. Dans une atmosphère instable, une parcelle d'air chaud, en montant, se refroidit moins vite que le gradient vertical de température. Ainsi, la différence entre

la parcelle d'air chaud et l'air ambiant s'amplifie, accentuant le mouvement ascendant. Cela provoque de grands mouvements d'air convectifs, donc une dispersion importante des polluants. Notons d'ailleurs que dans ce cas, l'hypothèse d'une distribution gaussienne du panache sur la verticale n'est plus valable, car elle ne distingue pas les mouvements convectifs ascendants et descendants qui peuvent se produire. Certains modèles représentent alors la distribution verticale du panache par la somme de deux gaussiennes, l'une dans le mouvement ascendant, l'autre descendant. Inversement, dans une atmosphère stable, les mouvements convectifs d'origine thermique sont très faibles ou inexistants. Si une parcelle montante se refroidit plus vite que le gradient vertical de température, la différence de température avec l'air ambiant diminue et le mouvement ascendant est freiné. La limite entre les deux situations est appelée atmosphère neutre.

L'état le plus stable de l'atmosphère correspond au cas où le gradient de température s'inverse, et où l'air devient plus chaud avec l'altitude. On appelle ce cas l'« inversion de température », et la partie de l'atmosphère où il se produit est la « couche d'inversion ». La hauteur où le gradient vertical de température s'inverse est la hauteur z_i déjà introduite en partie 2.1.2. En pratique, il existe deux principaux cas d'inversion :

1. De jour, le gradient vertical de température est en général négatif. Toutefois, une inversion peut se produire en hauteur lorsqu'une masse d'air chaud est transportée au-dessus d'une masse d'air plus froid. Dans ce cas, la hauteur d'inversion est élevée (plusieurs centaines de mètres). L'atmosphère est donc neutre ou instable entre le sol et la hauteur d'inversion, et stable au-delà : les échanges sont bloqués à cette hauteur.
2. De nuit, lorsque le sol se refroidit par rayonnement infrarouge, il devient plus froid que l'air ambiant : il y a donc inversion de température près du sol, sur quelques dizaines de mètres. Les polluants émis dans la journée précédente sont bloqués au-dessus, dans ce que l'on appelle la « couche résiduelle ». Sa hauteur est celle de la couche limite du jour précédent.

La figure 2.4 illustre les deux cas d'inversion de température décrits ci-dessus. Notons que, lorsqu'il y a inversion, la hauteur de couche limite est la hauteur d'inversion puisqu'il n'y a pratiquement pas d'échanges entre la surface et la partie de l'atmosphère située au-dessus de la hauteur d'inversion.

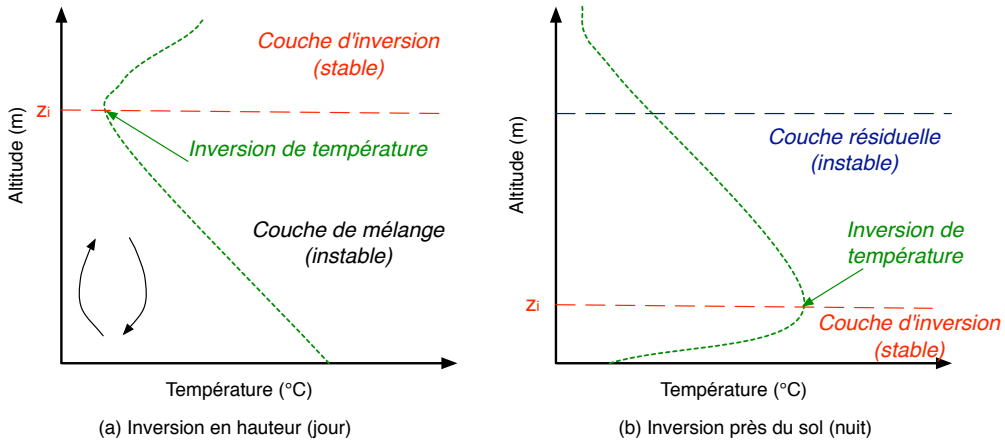


FIG. 2.4 – Représentation de l'inversion de température dans les deux cas les plus fréquents : (a) inversion en hauteur dans la journée (z_i de l'ordre de 100 m) et (b) inversion proche du sol la nuit (z_i de l'ordre de 10 m).

La stabilité de l'atmosphère dépend donc du gradient vertical de température, qui lui-même dépend de l'ensoleillement (de jour) et de la couverture nuageuse (de nuit). Elle dépend également du vent. Les classes de stabilité de Pasquill-Turner permettent de décrire la stabilité de l'atmosphère en fonction de ces quelques paramètres. Il y a six classes de stabilité, de A (très instable) à F (très stable). La façon de les déterminer d'après Turner [1969] est donnée par le tableau 2.1 pour le jour, et par le tableau 2.2 pour la nuit.

Vitesse du vent à 10 m $\mathrm{m\,s^{-1}}$	Fort rayonnement solaire $I > 700\ \mathrm{W\,m^{-2}}$	Rayonnement solaire modéré $350\ \mathrm{W\,m^{-2}} \leq I \leq 700\ \mathrm{W\,m^{-2}}$	Faible rayonnement solaire $I < 350\ \mathrm{W\,m^{-2}}$
≤ 2	A	A-B	B
2-3	A-B	B	C
3-5	B	B-C	C
5-6	C	C-D	D
≥ 6	C	D	D

Tab. 2.1 – Estimation des classes de stabilité de Pasquill, d'après Turner [1969] (journée).

Vitesse du vent à 10 m $\mathrm{m\,s^{-1}}$	Couverture nuageuse $\geq 1/2$	Couverture nuageuse $\leq 1/2$
≤ 2		
2-3	E	F
3-5	D	E
5-6	D	D
≥ 6	D	D

Tab. 2.2 – Estimation des classes de stabilité de Pasquill, d'après Turner [1969] (nuit).

De nombreuses paramétrisations des écarts types gaussiens reposent sur cette description discrète de l'état de l'atmosphère par des classes de stabilité, par exemple les formules de Briggs (partie 2.2.2). Plus simplement, les formules de Doury (partie 2.2.3) se basent sur deux classes de stabilité seulement, dépendant du gradient vertical de température.

Longueur de Monin-Obukhov

La caractérisation discrète de la stabilité par classes est grossière. Si davantage de données météorologiques sont disponibles, il est possible de calculer les écarts types en se basant sur d'autres paramètres décrivant la stabilité (partie 2.2.4). Un de ces paramètres est la longueur de Monin-Obukhov L, qui représente le rapport entre la production de turbulence d'origine mécanique et celle d'origine thermique. Si $L \to \pm\infty$, il n'y a pas de flux de chaleur turbulent vertical, et l'atmosphère est neutre. Si L est positive, l'atmosphère est stable (gradient vertical positif), et si L est négative, l'atmosphère est instable. Le tableau 2.3 donne les valeurs de L pour lesquelles l'atmosphère est considérée comme stable, instable, ou neutre, et la correspondance avec les classes de stabilité de Pasquill.

Autres variables

Pour compléter la description de l'état turbulent de l'atmosphère, d'autres variables sont définies par une analyse dimensionnelle :

- La vitesse de frottement u_* est une vitesse caractéristique de la turbulence provenant du cisaillement de vent. Elle est définie à partir du flux turbulent de surface (supposé

L : longueur de Monin-Obukhov (m)	Stabilité	Classe de Pasquill
$-100 \leq L < 0$	instable	B
$-10^5 \leq L < -100$	instable	C
$\lvert L \rvert > 10^5$	neutre	D
$10 \leq L < 10^5$	stable	E
$0 \leq L < 10$	stable	F

TAB. 2.3 – Correspondances entre la longueur de Monin-Obukhov et les classes de stabilité (Seinfeld et Pandis [1998]).

constant dans la couche limite de surface)

$$u_* = \left|\langle u'w' \rangle\right|_{z=0}. \tag{2.40}$$

– La vitesse convective w_* est caractéristique des mouvements verticaux convectifs produits par la turbulence d'origine thermique.

Ces variables sont reliées entre elles, à la longueur de Monin-Obukhov L et à la hauteur de couche limite h par la relation :

$$\frac{-L}{0.4h} = \frac{u_*^3}{w_*^3} \tag{2.41}$$

La turbulence engendrée par le frottement est naturellement liée à la hauteur de rugosité z_0. Dans le cas d'une surface très plane (glace par exemple), cette hauteur est très faible (pas d'aspérités engendrant de la turbulence) ; on prend alors $z_0 = 10^{-5}$ m. Inversement, dans le cas de zones urbanisées, il faut prendre en compte la turbulence engendrée par les bâtiments, et z_0 est comprise entre 2 m (quartier pavillonnaire), et 10 m (tours). En fonction de la viscosité de l'air ν, on peut donc définir un nombre de Reynolds caractéristique de l'écoulement dans la couche de surface :

$$Re_* = \frac{u_* z_0}{\nu}. \tag{2.42}$$

De façon expérimentale, on détermine que $Re_* < 0.13$ pour un écoulement laminaire et $Re_* > 2.5$ pour un écoulement turbulent.

Des relatioins empiriques peuvent être déterminées entre les paramètres z_0 et L, et la stabilité. En effet, la longueur de Monin-Obukhov mesure le rapport entre la turbulence d'origine mécanique (liée à z_0) et d'origine thermique (liée à la stabilité). Golder [1972] donne ainsi une relation entre L, la classe de stabilité de Pasquill, et z_0, sous la forme :

$$\frac{1}{L} = a + b \log z_0, \tag{2.43}$$

avec a et b des coefficients dépendant de la stabilité. Dans la suite, les paramètres utilisés, en particulier pour la théorie de la similitude sont u_*, w_*, h et L. Si l'un de ces quatre paramètres n'est pas disponibles dans les données météorologiques, il est possible de l'estimer en fonction des autres données, en particulier avec les équations 2.41 et 2.43.

2.2.2 Formules de Briggs

Les formules de Briggs font partie des paramétrisations « ancienne génération » qui dépendent des classes de stabilité de Pasquill. Parmi les autres paramétrisations de ce type couramment utilisées, on peut citer les écarts types de Pasquill-Guifford [Pasquill, 1961]. Les formules de

Briggs sont une tentative de synthétiser différentes paramétrisations existantes, en les interpolant pour des terrains ruraux ou urbains. Les écarts types de Briggs sont donnés sous la forme générale :

$$\sigma_y = \frac{\alpha\ x}{\sqrt{1+\beta\ x}}, \qquad \sigma_z = \alpha\ x\ (1+\beta\ x)^{\gamma}, \tag{2.44}$$

avec x la distance à la source dans la direction du vent, et α, β et γ des coefficients dépendant de la classe de stabilité. Le tableau 2.4 donne les formules pour le cas rural et le tableau 2.5 pour les terrains urbains.

Classe de Pasquill	σ_y	σ_z
A	$0.22x(1+0.0001x)^{-1/2}$	$0.20x$
B	$0.16x(1+0.0001x)^{-1/2}$	$0.12x$
C	$0.11x(1+0.0001x)^{-1/2}$	$0.08x(1+0.0002x)^{-1/2}$
D	$0.08x(1+0.0001x)^{-1/2}$	$0.06x(1+0.0015x)^{-1/2}$
E	$0.06x(1+0.0001x)^{-1/2}$	$0.03x(1+0.0003x)^{-1}$
F	$0.04x(1+0.0001x)^{-1/2}$	$0.016x(1+0.0003x)^{-1}$

Tab. 2.4 – Formules de Briggs pour terrain rural, en fonction de la classe de stabilité de Pasquill et de la distance x à la source, dans la direction du vent.

Classe de Pasquill	σ_y	σ_z
A-B	$0.32x(1+0.0004x)^{-1/2}$	$0.24x(1+0.001x)^{1/2}$
C	$0.22x(1+0.0004x)^{-1/2}$	$0.20x$
D	$0.16x(1+0.0004x)^{-1/2}$	$0.14x(1+0.0003x)^{-1/2}$
E-F	$0.11x(1+0.0004x)^{-1/2}$	$0.08x(1+0.0015x)^{-1}$

Tab. 2.5 – Formules de Briggs pour terrain urbain, en fonction de la classe de stabilité de Pasquill et de la distance x à la source, dans la direction du vent.

2.2.3 Formule de Doury

Une autre formulation des écarts types basée sur une représentation simplifiée de l'atmosphère est donnée dans Doury [1976]. Elle a été développée spécifiquement pour l'application aux radionucléides, et déterminée empiriquement, sur des expériences en champ plus lointain que celles utilisées en général (notamment par Briggs). Il n'y a que deux types de stabilité : la dispersion normale (variation de température inférieure à -0.5 °C sur 100 m), ou faible (variation supérieure à -0.5 °C sur 100 m). Si le gradient de température n'est pas connu, la classe de Pasquill peut être utilisée. A défaut, on considère que la dispersion normale a lieu de jour, ou de nuit avec un vent de plus de 3 $\mathrm{m\,s^{-1}}$, et la dispersion faible a lieu de nuit avec un vent faible. Cette représentation est cependant très simplifiée (une dispersion faible peut très bien avoir lieu de jour), mais présente l'avantage de ne nécessiter aucune donnée météorologique autre que le vent. Les écarts types sont sous la forme générale :

$$\sigma_y = (A_h t)^{B_h}, \qquad \sigma_z = (A_z t)^{B_z}, \tag{2.45}$$

où t est le temps de transfert, depuis l'émission de la bouffée. Dans le cas stationnaire, on a $t = x/\bar{u}$. Le tableau 2.6 donne les coefficients de calcul des écarts types pour la dispersion normale, et le tableau 2.7 pour les cas de faible dispersion.

Temps de trajet (s)	A_h	A_z	B_h	B_z
$0 < t \leq 240$	0.405	0.42	0.859	0.814
$240 < t \leq 3280$	0.135	1.00	1.130	0.685
$3280 < t \leq 97000$	0.135	20.0	1.130	0.50
$97000 < t \leq 508000$	0.463	20.0	1.00	0.50
$508000 < t \leq 1300000$	6.50	20.0	0.824	0.50
$t > 1300000$	2.10^5	20.0	0.50	0.50

TAB. 2.6 – Coefficients de Doury pour la diffusion normale.

Temps de trajet (s)	A_h	A_z	B_h	B_z
$0 < t \leq 240$	0.405	0.20	0.859	0.50
$240 < t \leq 97000$	0.135	0.20	1.130	0.50
$97000 < t \leq 508000$	0.463	0.20	1.00	0.50
$508000 < t \leq 1300000$	6.50	0.20	0.824	0.50
$t > 1300000$	2.10^5	0.20	0.50	0.50

TAB. 2.7 – Coefficients de Doury pour la diffusion faible.

2.2.4 Théorie de la similitude

La théorie de la similitude consiste, de façon générale, à déterminer des relations entre différentes échelles d'un écoulement, via une analyse dimensionnelle. Ici, il s'agit de relier les écarts types de la dispersion σ_y et σ_z aux variables liées à la petite échelle de la turbulence, et en particulier les écarts types des fluctuations du vent : $\sigma_v^2 = \langle v'^2 \rangle$ dans la direction horizontale, et $\sigma_w^2 = \langle w'^2 \rangle$ dans la direction verticale. Irwin [1979] donne la forme générale suivante pour cette relation :

$$\sigma_y = \sigma_v \; t \; F_y, \qquad \sigma_z = \sigma_w \; t \; F_z, \tag{2.46}$$

où F_y et F_z sont des fonctions déterminées de façon empirique. D'après la théorie de diffusion de Taylor [Taylor, 1921], la diffusion d'un panache de source ponctuelle a deux comportements distincts suivant le temps de transfert t : $\sigma \propto t$ lorsque $t \to 0$ et $\sigma \propto \sqrt{t}$ lorsque $t \to \infty$. La transition se fait lorsque les particules ont « oublié » leur vitesse initiale, et le temps caractéristique de cette transition s'appelle l'échelle de temps lagrangienne, notée τ_L. Les fonctions F_y et F_z sont déterminées en fonction des caractéristiques de la couche limite décrites en partie 2.2.1 : la longueur de Monin-Obukhov L, la vitesse de frottement u_* et la vitesse convective w_*, ainsi que la vitesse moyenne du vent u et la hauteur de couche limite notée h. De plus, les écarts types des fluctuations de vent σ_v et σ_w, ainsi que l'échelle de temps lagrangienne τ_L, sont également exprimées en fonction des grandeurs caractéristiques de l'écoulement.

Les formules utilisées dans Polyphemus sont détaillées ci-dessous. En plus de la description des paramètres dans la couche limite, des formulations alternatives pour les sources en hauteur (au-dessus de 100 m) sont proposées. Des paramétrisations pour la dispersion au-dessus de la couche limite sont également proposées.

Dans la couche limite

Écart type horizontal Déterminer l'écart type σ_y nécessite d'abord d'exprimer l'écart type horizontal du vent σ_v. Les formules utilisées pour cela sont issues de Hanna *et al.* [1982].

$$\sigma_v = u_* \left(12 - 0.5\frac{h}{L}\right)^{1/3} \quad \text{instable} \tag{2.47}$$

$$\sigma_v = 1.3\ u_*\ \exp\left(-2\ \frac{f\ z}{u_*}\right) \quad \text{neutre} \tag{2.48}$$

$$\sigma_v = \max(1.3\ u_*\ \left(1 - \frac{z}{h}\right), 0.2) \quad \text{stable} \tag{2.49}$$

avec h la hauteur de couche limite, L la longueur de Monin-Obukhov, f le paramètre de Coriolis et z la hauteur à laquelle les écarts types sont calculés.

La fonction F_y est donnée sous la forme

$$F_y = \left(1 + 0.5\ \frac{t}{\tau_L}\right)^{-1/2}, \tag{2.50}$$

avec τ_L l'échelle de temps lagrangienne, calculée d'après Hanna [1984]

$$\tau_L = 0.15\ \frac{h}{\sigma_v} \quad \text{instable} \tag{2.51}$$

$$\tau_L = 0.5\ \frac{z}{\sigma_v}\ /\ \left[1 + 15\ \frac{f\ z}{u_*}\right] \quad \text{neutre} \tag{2.52}$$

$$\tau_L = 0.07\ \frac{h}{\sigma_v}\sqrt{\frac{z}{h}} \quad \text{stable.} \tag{2.53}$$

Écart type vertical L'écart type vertical du vent σ_w est calculé d'après les formules de Venkatram *et al.* [1984] pour les cas stables et neutres, et de Weil [1988] pour les cas convectifs

$$\sigma_w = 0.6\ w_* \quad \text{instable} \tag{2.54}$$

$$\sigma_w = 1.3\ u_*\ \left(1 - \frac{z}{h}\right)^{3/4} \quad \text{stable/neutre} \tag{2.55}$$

avec w_* la vitesse convective, u_* la vitesse de frottement.

La fonction F_z est donnée par Irwin [1979] pour les cas stables et neutres, et par Weil [1988] pour les cas convectifs :

– Cas instables :

$$F_z = \left(1 + 0.5\ \frac{t}{\tau_{Lz}}\right)^{-1/2} \tag{2.56}$$

– Cas stables et neutres :

$$F_z = \begin{cases} \left[1 + 0.9\ (\frac{t}{50})\right]^{-1} & \text{si } z < 50 \\ \left[1 + 0.945\ (0.1\ t)^{0.806}\right]^{-1} & \text{si } z \geq 50 \end{cases} \tag{2.57}$$

L'échelle de temps Lagrangienne verticale τ_{Lz} est donnée dans Hanna *et al.* [1982] :

$$\tau_{Lz} = 0.15\ h\ \frac{1 - \exp\left(-5\ \frac{z}{h}\right)}{\sigma_w} \quad \text{instable}$$

$$\tau_{Lz} = 0.5\ \frac{z}{\sigma_w}\ /\ \left[1 + 15\ \frac{f\ z}{u_*}\right] \quad \text{neutre} \tag{2.58}$$

$$\tau_{Lz} = 0.10\ \frac{h}{\sigma_w}\left(\frac{z}{h}\right)^{0.8} \quad \text{stable} \tag{2.59}$$

Formules alternatives pour les sources en hauteur

Dans Hanna et Paine [1989], des formules alternatives sont proposées pour des sources situées à plus de 100 m de hauteur, arguant du fait que les formules précédentes ont été calibrées sur des expériences avec des rejets proches du sol. Ainsi, dans les cas de rejets en hauteur (expérience de Kincaid par exemple), Hanna et Paine [1989] estime que la dispersion donnée par les formules « classiques » est sous-estimée dans les cas stables et neutres. Les écarts types du vent sont donc calculés différemment :

– L'écart type vertical du vent est donné par

$$\sigma_w = 0.5\sqrt{1.2\ u_*^2 + 0.35\ w_*^2} \quad L \geq 100 \tag{2.60}$$

$$\sigma_w = 1.3\ u_* \quad 0 \leq L < 100 \tag{2.61}$$

– L'écart type horizontal du vent est donné par

$$\sigma_v = 0.7\sqrt{3.6\ u_*^2 + 0.35\ w_*^2} \quad L \geq 100 \tag{2.62}$$

$$\sigma_v = \max\left(1.5\sqrt{3.6\ u_*^2 + 0.35\ w_*^2}, 0.5\right) \quad 0 \leq L < 100 \tag{2.63}$$

Pour l'écart type vertical, l'échelle de temps lagrangienne est modifiée pour les cas stables et légèrement instables. Ici, le paramètre de stabilité statique de Briggs s_p est utilisé :

$$s_p = \frac{g\mathrm{d}\theta/\mathrm{d}z}{T}, \tag{2.64}$$

avec $g = 9.81\ \mathrm{m\,s^{-2}}$ l'accélération gravitationnelle, et $\mathrm{d}\theta/\mathrm{d}z$ le gradient de température potentielle.

– Dans le cas des situations stables/neutres ($L > 0$) :

$$\tau_{Lz} = z/\sigma_w \quad \text{si } z \leq L \tag{2.65}$$

$$\tau_{Lz} = 0.27\ \sqrt{s_p} \quad \text{si } L \leq 10 \tag{2.66}$$

$$\tau_{Lz} = (z/\sigma_w)\ (L-10)/(z-10) + 0.27\ \sqrt{s_p}\ (z-L)/(z-10) \quad \text{si } 10 < L < z \tag{2.67}$$

– Dans le cas de situations légèrement instables ($-100 \leq L \leq 0$) :

$$\tau_{Lz} = 0.27\ (z/\sigma_w)\ (0.55 - 0.38\ z/|L|) \quad \text{si } z \leq |L| \tag{2.68}$$

$$\tau_{Lz} = 0.3\ (h/\sigma_w)\ (1 - \exp(-5z/h) - 0.0003\exp(8z/h)) \quad \text{si } |L| < z < h \tag{2.69}$$

Au-dessus de la couche limite

Formules d'HPDM pour les écarts types du vent Les formules de dispersion données par la théorie de la similitude sont valides uniquement dans la couche limite. Les données disponibles pour la dispersion au-dessus de la couche limite étant rares, les formules utilisées dans ce cas sont très simples. Seuls les écarts types du vent sont modifiés, d'après Hanna et Paine [1989] :

$$\sigma_v = 0.1\sqrt{3.6\ u_*^2 + 0.35\ w_*^2} \tag{2.70}$$

$$\sigma_w = 0.1\sqrt{1.2\ u_*^2 + 0.35\ w_*^2} \tag{2.71}$$

Ces formules sont appliquées pour toutes les stabilités. Elles permettent de s'assurer que les paramètres de dispersion sont toujours positifs et qu'il n'y a pas de division par zéro, ou de racine carrée d'une valeur négative. Cependant, il faut noter aussi que la plupart des cas pratiques de dispersion ont lieu à l'intérieur de la couche limite, et que ces formules seront donc relativement peu utilisées.

Formule de Gillani pour l'écart type vertical Une formulation alternative pour σ_z au-dessus de la couche limite est donnée dans Gillani et Godowitch [1999]

$$\sigma_z^2 = \sigma_{z_{\mathrm{pr}}}^2 \ (1 + 2.3 \ \sqrt{t}) \tag{2.72}$$

avec t le temps de transfert du panache depuis l'émission (en prenant une émission à $t = 0$). Cette formule peut être appliquée au-dessus de la couche limite quelle que soit la paramétrisation utilisée par ailleurs. La valeur initiale de l'écart type due à la surhauteur, $\sigma_{z_{\mathrm{pr}}}$, est calculée d'une façon différente des formules données en partie 2.3

$$\sigma_{z_{\mathrm{pr}}} = \max\left(15 \ \exp(-117 \ \frac{\mathrm{d}T}{\mathrm{d}z}), 3\right) \tag{2.73}$$

avec $\mathrm{d}T/\mathrm{d}z$ le gradient vertical de température.

2.3 Modélisation de la surhauteur

Dans le cas de rejets beaucoup plus chauds que la température de l'air ambiant, ou émis avec une vitesse d'éjection importante, le panache atteint une certaine hauteur avant d'avoir un comportement identique à l'air ambiant et avant de satisfaire les équations de transport précédemment décrites. Cette différence entre la hauteur d'éjection et la hauteur finale atteinte par le panache s'appelle la surhauteur (« plume rise »), et peut être importante (de l'ordre de plusieurs centaines de mètres). Une modélisation exacte de la surhauteur supposerait de décrire précisément le comportement du fluide émis par un modèle de jet. Des formules empiriques plus simples fournissent une estimation de la surhauteur en fonction de la hauteur initiale, du diamètre de la source, de la température et de la vitesse d'éjection du fluide émis. Plusieurs de ces formules, fournies dans Polyphemus, sont décrites ici. Elles fournissent toutes une estimation de la hauteur finale du panache, en supposant qu'elle est atteinte instantanément à la sortie de la source, avant tout déplacement horizontal du panache (figure 2.5).

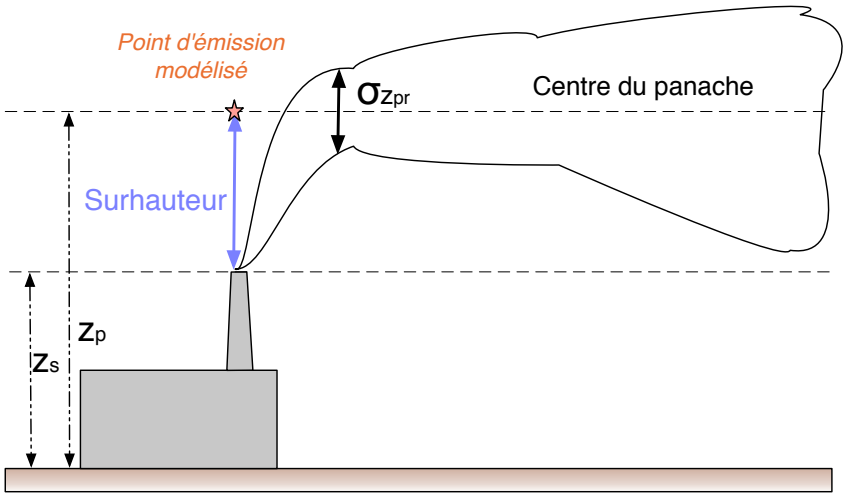


FIG. 2.5 – Représentation d'une source en hauteur : z_s est la hauteur de source, et z_p est la hauteur effective du panache, en tenant compte de la surhauteur. On modélise l'émission à la hauteur z_p, en négligeant le déplacement horizontal du panache avant d'atteindre cette hauteur. La largeur initiale du panache est $\sigma_{z_{\mathrm{pr}}}$.

Si l'on note la surhauteur Δh, la hauteur finale atteinte est donnée par $z_p = z_s + \Delta h$ and z_s. La hauteur de la source z_s est alors remplacée par z_p dans l'équation 2.14. De plus, l'élargissement du panache pendant cette phase d'élévation est représenté par un terme additionnel dans le calcul des écarts types (dans l'équation 2.39). L'estimation pour la direction y fournie ici provient de Irwin [1979], et provient pour la direction z de Hanna et Paine [1989]

$$\sigma_{y_{\mathrm{pr}}} = \frac{\Delta h}{3.5}, \qquad \sigma_{z_{\mathrm{pr}}} = \frac{\Delta h}{2}. \tag{2.74}$$

2.3.1 Surhauteur de Briggs et HPDM

Des estimations de la surhauteur sont données par les formules de Briggs [Briggs, 1969], et résumées dans Seinfeld et Pandis [1998]. Des estimations différentes sont données suivant que la surhauteur est principalement due à la vitesse d'éjection, ou à la flottabilité (c'est-à-dire à la différence de température avec l'air ambiant). Les formules décrites ici font la seconde hypothèse (équations 2.76 à 2.81). Dans ces formules, s_p est le paramètre de stabilité statique de Briggs (équation 2.64), et F_b est le paramètre de flottabilité initiale, défini par

$$F_b = g\ v_s\ d_s^2\ \frac{T_s - T}{T_s} \tag{2.75}$$

avec g l'accélération gravitationnelle, T la température ambiante, v_s la vitesse d'éjection du gaz émis en sortie de source, d_s le diamètre de la source et T_s la température du gaz émis.

Cas stables

Les formules de Briggs pour les cas stables donnent la surhauteur finale comme $\Delta h = \min(\Delta h_1, \Delta h_2)$, avec

$$\Delta h_1 = 2.6\ \left(\frac{F_b}{u\ s_p}\right)^{1/3}, \tag{2.76}$$

$$\Delta h_2 = 4\ F_b^{1/4} s_p^{-3/8}. \tag{2.77}$$

Cas instables et neutres

Les formules de Briggs dans les cas instables donnent

$$\Delta h = 21.4\ \frac{F_b^{3/4}}{u} \qquad \text{si } F_b < 55 \tag{2.78}$$

$$\Delta h = 38.71\ \frac{F_b^{3/5}}{u} \qquad \text{si } F_b \geq 55 \tag{2.79}$$

Les formules ci-dessus calculent la surhauteur en fonction des paramètres de l'émission, en tenant compte uniquement du vent et de la température ambiante pour décrire l'atmosphère extérieure. Cependant, dans les cas neutres et instables, la turbulence atmosphérique n'est pas négligeable, et peut stopper la trajectoire verticale du panache plus rapidement que l'estimation précédente ne le prévoit. Dans Briggs [1971], la hauteur d'arrêt du panache due à la turbulence extérieure (« plume breakup ») est définie comme la hauteur où les taux de dissipation d'énergie turbulente sont égaux à l'intérieur du panache et dans l'atmosphère. L'estimation de Briggs pour les cas neutres est rapportée dans Hanna [1984] et utilisée ici (équation 2.80). Pour les cas

instables, la formule 2.81 provient de celle utilisée dans le modèle HPDM et décrite dans Hanna et Paine [1989].

$$\Delta h = 4.3 \left(\frac{F_b}{u\ w_*^2} \right)^{3/5} z_i^{2/5} \quad \text{instable} \tag{2.80}$$

$$\Delta h = 1.54 \left(\frac{F_b}{u\ u_*^2} \right)^{2/3} z_s^{1/3} \quad \text{neutre} \tag{2.81}$$

Les deux formules ci-dessus donnent une estimation de la surhauteur maximale atteinte par le panache. La surhauteur finale est le minimum entre les différentes estimations possibles pour une stabilité donnée.

2.3.2 Formule de Concawe

Une étude comparative des différentes paramétrisations de surhauteur, y compris celle de Holland et Stümke (partie 2.3.3), est décrite dans Brummage [1968]. Parmi ces formules, toutes sont basées sur le flux de chaleur de la source Q_h et ne dépendent donc pas de la situation météorologique. Le but de cette étude était de trouver une formulation synthétique, représentative des meilleures paramétrisations, de la forme

$$\Delta h = K\ \frac{Q_h^\alpha}{u^\beta}, \tag{2.82}$$

avec K, α et β des constantes empiriques. Le résultat, appelé formule de Concawe, est donné par

$$\Delta h = 0.071\ \frac{Q_h^{0.55}}{u^{0.67}}. \tag{2.83}$$

Le flux de chaleur (par seconde) peut se calculer, à pression constante, par la formule

$$Q_h = n\ C_p\ \Delta T, \tag{2.84}$$

avec n le nombre de moles de gaz émis, C_p la capacité calorifique molaire du gaz émis (en $\mathrm{J\,mol^{-1}\,K^{-1}}$), et $\Delta T = T_s - T$ la différence de température entre le gaz émis et l'air ambiant. Par ailleurs, le volume de gaz éjecté par seconde est donné en fonction de la vitesse d'éjection v_s et du diamètre d_s

$$V_s = \pi\ v_s\ \frac{d_s^2}{4}. \tag{2.85}$$

Si on note le volume molaire des gaz parfaits $V_m = 24\ \mathrm{L\,mol^{-1}}$, le nombre de moles éjectées par seconde est alors donné par $n = V_s/V_m$. En prenant une capacité calorifique molaire égale à celle de l'air, $C_p = 29.5\ \mathrm{J\,mol^{-1}\,K^{-1}}$, on obtient alors l'expression du flux de chaleur

$$Q_h = 228.19\ v_s\ d_s^2\ \Delta T. \tag{2.86}$$

Dans le calcul précédent, le gaz émis est assimilé à de l'air chaud, en supposant que le traceur émis est en concentration suffisamment faible pour que cette approximation soit valide. Si le fluide émis était « pur », il serait plus simple de se baser sur le débit de la source Q_s

$$Q_h = Q_s\ c_p\ \Delta T \tag{2.87}$$

avec c_p la capacité calorifique massique du gaz émis. Cependant, il s'agit dans la plupart des cas d'un mélange entre un traceur à une concentration donnée, et de l'air chaud.

2.3.3 Formule de Holland et Stümke

Cette estimation de la surhauteur ne dépend pas de la stabilité atmosphérique, mais uniquement de la vitesse d'éjection, du diamètre, et de la différence de température avec l'air ambiant. La première estimation provient de Holland [1953], et la formule a été modifiée par Stümke [1963].

$$\Delta h = 1.5\ \frac{d_s\ v_s}{u} + 65\ \frac{d_s^{1.5}}{u}\ \left(\frac{T_s - T}{T_s}\right)^{1/4} \tag{2.88}$$

2.3.4 Pénétration partielle dans la couche d'inversion

Dans le cas d'une surhauteur importante, le panache peut pénétrer, entièrement ou partiellement, dans la couche d'inversion (au-dessus de la couche limite). Dans ce cas, la partie du panache qui est au-dessus de la couche limite est « piégée », et ne peut plus rentrer dans la couche limite (hypothèse, qui est faite dans les modèles gaussiens, de réflexion sur la couche d'inversion). Si l'on ne prend pas en compte cette pénétration, l'estimation des concentrations par le modèle peut être significativement biaisée

- Si l'on suppose que le panache est bloqué à l'intérieur de la couche limite, les concentrations seront surestimées,
- Si tout le panache est modélisé au-dessus de la couche limite, la concentration au sol sera estimée nulle.

Une façon simple de modéliser le fait que le panache puisse partiellement pénétrer dans la couche d'inversion est de supposer qu'il se sépare en deux panaches, modélisés séparément au-dessus et au-dessous de la couche limite (figure 2.6). Il faut donc déterminer à partir de quelle hauteur de panache on considère qu'il y a pénétration partielle, et quelle fraction du panache est supposée passer au-dessus de la couche limite. Les formules présentées ici sont tirées de Hanna et Paine [1989].

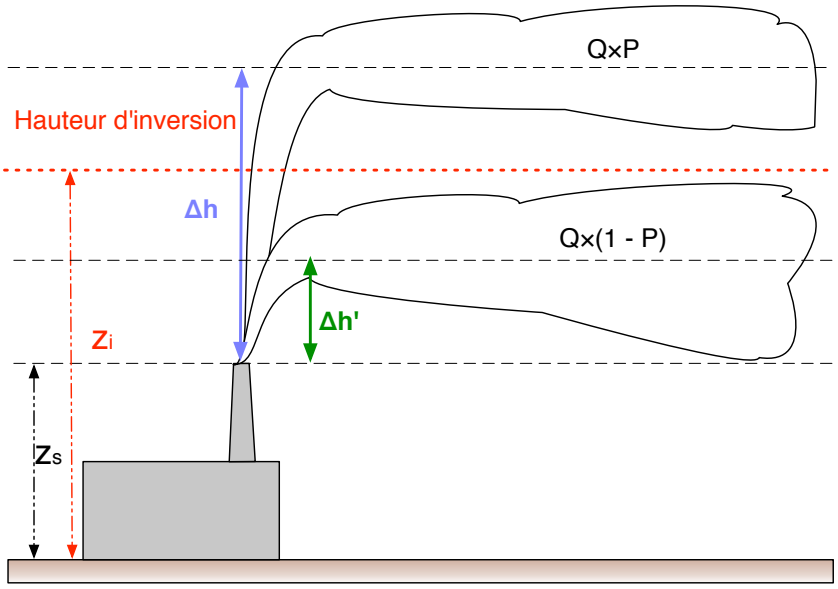


FIG. 2.6 – Pénétration partielle dans la couche d'inversion : z_s est la hauteur de la source et z_i est la hauteur d'inversion. Le facteur de pénétration est noté P. La surhauteur est Δh et la partie du panache restant dans la couche limite a pour surhauteur effective $\Delta h'$. Q est le taux d'émission total.

Si la surhauteur Δh est telle que

$$\Delta h \geq \frac{z_i - z_s}{1.5}, \tag{2.89}$$

alors le haut du panache touche la couche d'inversion. La fraction du panache qui pénètre dans la couche d'inversion est alors donnée par

$$P = \begin{cases} 1.5 - (z_i - z_s)/\Delta h & \text{si } 0.5 < (z_i - z_s)/\Delta h < 1.5 \\ 1 & \text{si } (z_i - z_s)/\Delta h \leq 0.5 \end{cases} \tag{2.90}$$

Dans le cas d'une pénétration partielle (équation 2.90), la partie du panache qui reste dans la couche limite a un nouveau taux d'émission

$$Q_s\ (1 - P) \tag{2.91}$$

et une nouvelle valeur de surhauteur (figure 2.6)

$$\Delta h' = (0.62 + 0.38P)(z_i - z_s). \tag{2.92}$$

Dans le cas d'une pénétration totale dans la couche d'inversion, il n'y a aucun panache restant dans la couche limite (nouveau taux d'émission égal à 0).

2.4 Modélisation des processus de perte

Jusqu'à présent, nous avons supposé la conservation de la masse dans les équations gaussiennes. En réalité, les processus de perte suivants peuvent être pris en compte dans les modèles :
- le dépôt sec au sol,
- le lessivage,
- la décroissance radioactive (cinétique chimique du premier ordre),
- la chimie gazeuse (par exemple, le mécanisme RACM [Stockwell *et al.*, 1997]).

La prise en compte de la chimie se fait uniquement dans le modèle à bouffées, et est décrite dans le chapitre 5. Nous considérons ici uniquement les processus simples qui conduisent à un appauvrissement du panache. Sauf indication contraire, les formulations utilisées ici sont celles du modèle de panache (partie 2.1.2). Les traceurs considérés ici peuvent être gazeux, ou sous forme de particules solides ou liquides (aérosols) de différents diamètres. Dans ce dernier cas, le dépôt sec et le lessivage sont calculés spécifiquement pour les différents diamètres. Les autres processus liés aux aérosols (coagulation, évaporation/condensation) ne sont pas pris en compte dans les modèles gaussiens de Polyphemus.

2.4.1 Décroissance radioactive

Si l'espèce considérée a un temps de demi-vie $t_{1/2}$, son coefficient de décroissance radioactive est $\lambda = \log(2)/t_{1/2}$. L'appauvrissement de la source à un temps t s'écrit alors

$$C_{rad}(x, y, z, t) = \exp(-\lambda t) \times C(x, y, z). \tag{2.93}$$

Avec t le temps de trajet du panache, qui est relié à la distance à la source x par $t = x/\bar{u}$.

2.4.2 Dépôt sec

Le dépôt sec consiste à prendre en compte le dépôt de l'espèce sur le sol, qu'il soit dû à la sédimentation gravitationnelle, ou à la capture par les végétaux, les bâtiments et le sol.

Vitesse de dépôt

Si le flux de dépôt au sol s'écrit F_{dry} (en unité de masse par mètre carré et par seconde), on définit alors la vitesse de dépôt v_d par

$$F_{dry}(x, y) = v_d \times C_{dry}(x, y, 0). \tag{2.94}$$

Pour une particule, on note v_g la vitesse de sédimentation gravitationnelle, et v'_d la partie dite « diffusive » de la vitesse de dépôt, correspondant aux autres processus de dépôt sec. La vitesse de sédimentation gravitationnelle, dépend du diamètre et est calculée comme la vitesse de Stokes (pour des diamètres inférieurs à 20 µm)

$$v_g = \frac{D^2(\rho_p - \rho_{air})gC_c}{18\mu_{air}}, \tag{2.95}$$

avec ρ_p (en $\mathrm{kg\,m^{-3}}$) la masse volumique de la particule, μ_{air} la viscosité dynamique de l'air (en $\mathrm{Pa\,s}$), g la constante gravitationnelle (en $\mathrm{m\,s^{-2}}$) et C_c le coefficient de Cunningham. Celui-ci est donné par

$$C_c = 1 + \frac{2\lambda_{air}}{D}\left[1.257 + 0.4\exp\left(-0.55\frac{D}{\lambda_{air}}\right)\right], \tag{2.96}$$

où λ_{air} (en m) est le libre parcours moyen de l'air

$$\lambda_{air} = \frac{2\mu_{air}}{P}\left(\frac{8}{\pi R_{air}T}\right)^{-1/2}, \tag{2.97}$$

avec R_{air} la constante molaire de l'air (en $\mathrm{J\,K^{-1}\,kg^{-1}}$), T la température (en K) et P la pression (en Pa).

La vitesse de dépôt totale dépend de la vitesse de sédimentation gravitationnelle ainsi que de la partie diffusive v'_d, suivant la relation

$$v_d = \frac{v_g}{1 - \exp\left(-\frac{v_g}{v'_d}\right)}. \tag{2.98}$$

Pour un gaz, v_g est négligée, et $v_d = v'_d$. En pratique, on considère (dans les modèles gaussiens utilisés ici) que les vitesses de dépôt sont constantes pour chaque espèce gazeuse, et qu'elles dépendent du diamètre pour les particules.

Il existe deux façons de prendre en compte le dépôt sec dans les modèles gaussiens :

1. en modélisant l'appauvrissement total du panache,
2. en utilisant une réflexion partielle au sol.

Appauvrissement de la source

Dans le premier cas, on considère que le débit de la source n'est plus constant, mais qu'il s'appauvrit en fonction de la distance à la source : $Q_s(x) = Q_s(0) \times q(x)$. Il suffit donc de calculer le facteur d'appauvrissement $q(x)$ [Hanna *et al.*, 1982].

En considérant la fraction de masse qui se dépose sur une portion $\mathrm{d}x$, on peut écrire

$$\frac{\partial Q_s}{\partial x} = \int_{-\infty}^{\infty} -F_{dry}(x, y)\mathrm{d}y = -v_d \int_{-\infty}^{\infty} C_{dry}(x, y, 0)\mathrm{d}y. \tag{2.99}$$

D'après l'équation 2.18, on a

$$C_{dry}(x,y,0) = \frac{Q_s}{\bar{u}}\ G_y(y-y_s)\ \left(G_z(-z_s)+G_z(z_s)\right). \tag{2.100}$$

Étant donné que l'intégrale sur tout l'espace (en y) de G_y vaut 1 (équation 2.16) , et que la fonction G_z est symétrique :

$$\int_{-\infty}^{\infty} C_{dry}(x,y,0)\mathrm{d}y = \frac{2Q_s}{\bar{u}}\ G_z(z_s), \tag{2.101}$$

et l'on peut finalement écrire

$$\frac{\partial Q_s}{\partial x} = Q_s \frac{-2v_d}{\bar{u}}\ G_z(z_s). \tag{2.102}$$

L'intégration sur x donne donc finalement le résultat

$$q(X) = \frac{Q_s(X)}{Q_s(0)} = \exp\left(\int_0^X \frac{-2v_d}{\bar{u}} G_z(z_s)\mathrm{d}x\right) \tag{2.103}$$

Le résultat n'est pas immédiat, étant donné que σ_z dépend de x, mais il est calculable analytiquement. On appelle parfois ce résultat l'intégrale de Chamberlain. Il présente l'inconvénient de modéliser un appauvrissement dans tout le panache, alors que le dépôt se fait essentiellement près du sol, mais donne cependant de bons résultats lorsque le panache est suffisamment loin de la source et bien mélangé sur la verticale. De plus, la formule 2.103 donne l'appauvrissement total de la source depuis l'émission. Dans le cas d'un modèle instationnaire à bouffées, l'appauvrissement « acquis » pendant $\Delta x = \bar{u}\Delta t$ est calculé à chaque pas de temps. Il n'y a donc pas besoin d'intégrer sur x, et le résultat s'écrit simplement :

$$Q_s(t+\Delta t) = \exp\left(-2v_d G_z(z_s)\Delta t\right) \times Q_s(t). \tag{2.104}$$

Réflexion partielle

Le dépôt sec est pris en compte dans le calcul des concentrations par la formule d'Overcamp (décrite par exemple dans Arya [1999]). Elle consiste à considérer que, lors de la réflexion du panache au sol, seule une partie du panache est réfléchie, l'autre partie correspondant à la quantité déposée. La concentration en un point s'écrit alors :

$$C_{dry}(x,y,z) = \frac{Q_s}{2\pi\sigma_y\sigma_z\bar{u}} \exp\left(-\frac{(y-y_s)^2}{2\sigma_y^2}\right) \times \tag{2.105}$$

$$\left[\exp\left(-\frac{(z-z_s)^2}{2\sigma_z^2}\right) + \alpha\ \exp\left(-\frac{(z+z_s)^2}{2\sigma_z^2}\right)\right]. \tag{2.106}$$

Le coefficient α permet de prendre en compte le dépôt : $\alpha = 1$ correspond à une réflexion totale au sol, sans dépôt, et $\alpha = -1$ correspondrait à une absorption totale par le sol. La fraction restante dans le panache est donnée par

$$1 - q(x) = (1-\alpha) \int_0^{z_i} G_z\left(z+z_s\right)\mathrm{d}z. \tag{2.107}$$

Le coefficient α est calculé de façon à faire correspondre la masse enlevée par cette méthode, avec la masse perdue par le flux de dépôt (équation 2.102). L'expression est donnée par :

$$\alpha(x) = 1 - \frac{2v_d}{v_d + \left(\bar{u}z_s\sigma_z^{-1}\right)\left(\mathrm{d}\sigma_z/\mathrm{d}x\right)}. \tag{2.108}$$

2.4.3 Lessivage

Le lessivage, ou dépôt humide, correspond à la perte de masse due aux nuages et à la pluie (dilution et entraînement dans les gouttes d'eau en suspension dans les nuages, et dans les gouttes de pluie). La capacité d'une espèce à être lessivée est modélisée par un coefficient de lessivage Λ.

Coefficient de lessivage

Les coefficients de lessivage des espèces peuvent être considérés comme constants, ou calculés suivant des paramétrisations simples. Pour les gaz, il s'agit de la formule de Belot [Belot *et al.*, 1988], qui donne :

$$\Lambda = A\, p_0^B, \tag{2.109}$$

où A et B sont des coefficients dépendant de l'espèce et p_0 est l'intensité de pluie en $\mathrm{mm\,hr^{-1}}$.

Les coefficients de lessivage pour les particules sont calculés avec la paramétrisation de Slinn (Underwood [2001], Sportisse et Du Bois [2002]), qui est adaptée au cas des particules et prend en compte les différents diamètres. Le coefficient de lessivage Λ est donné par :

$$\Lambda = \frac{3}{2}\frac{E\, p_0}{D_r} \tag{2.110}$$

avec E l'efficacité de lessivage et D_r le diamètre des gouttes. Le coefficient E dépend du diamètre D des aérosols et est donné par :

- $E = 0.1$ pour $D \leq D_1 = 1\,\mu m$;
- $E(D) = E(D_1) + (1 - E(D_1))\frac{D-D_1}{D_2-D_1}$ pour $D_1 \leq D \leq D_2 = 10\,\mu m$;
- $E(D) = 1$ pour $D \geq D_2$.

et le diamètre des gouttes est donné par :

$$D_r = 7.10^{-4}\, p_0^{0.25} \tag{2.111}$$

où p_0 est l'intensité de pluie en $\mathrm{mm\,hr^{-1}}$.

Flux de dépôt humide

La formule donnant la concentration est alors modifiée avec un facteur dépendant de Λ, qui représente la perte dans le panache due au dépôt humide :

$$C_{wet}(x, y, z) = \exp\left(-\Lambda t\right) \times C(x, y, z), \tag{2.112}$$

avec t le temps de trajet du panache, qui est relié à la distance à la source x par $t = x/\bar{u}$. Le flux de dépôt total est alors calculé en prenant l'intégrale sur toute la hauteur de la couche limite des pertes par lessivage (en supposant le lessivage homogène) [Hanna *et al.*, 1982] :

$$F_{wet}(x, y) = \int_0^{z_i} \Lambda \times C_{wet}(x, y, z)\mathrm{d}z. \tag{2.113}$$

Si l'on utilise la concentration calculée pour un mélange homogène sur la verticale (équation 2.23), étant donné que le lessivage est uniforme, on peut écrire :

$$F_{wet}(x, y) = \Lambda \frac{Q_s}{\sqrt{2\pi}\sigma_y \bar{u}} \exp\left(-\frac{(y - y_s)^2}{2\sigma_y^2}\right) \exp\left(-\Lambda t\right). \tag{2.114}$$

2.5 Bilan : modèles de Polyphemus

La théorie sur laquelle s'appuie la modélisation gaussienne des sources ponctuelles a été présentée dans cette partie. Les trois principaux processus qui nécessitent le recours à des paramétrisations empiriques sont la dispersion, la surhauteur, et le dépôt. Pour ces trois processus, seules les paramétrisations existantes dans Polyphemus ont été détaillées dans ce chapitre. On dresse ici un bilan des formules implémentées dans Polyphemus : le tableau 2.8 récapitule les possibilités pour calculer les données pré-traitées, spécifiques aux espèces modélisées, et utilisables ensuite par l'un des modèles gaussiens (« plume » ou « puff »), et le tableau 2.9 décrit les processus modélisés et les paramétrisations utilisées. Ces dernières sont communes aux deux modèles gaussiens, et à tous les types d'espèces (gaz ou particule), excepté pour la chimie.

Donnée d'espèce	Gaz	Particule
Temps de demi-vie $t_{1/2}$	Constant (par espèce)	Constant (par espèce)
Vitesse de dépôt v'_d	Constant (par espèce et météo.)	Constant (par diamètre et météo.)
Sédimentation gravitationnelle v_g	—	Calculée (équation 2.95)
Coefficient de lessivage Λ	Constant (par espèce et météo.)	Constant (par diamètre et météo.)
	ou Belot (équation 2.109)	*ou* Slinn (équation 2.110)

Tab. 2.8 – Récapitulatif des données spécifiques aux espèces, calculées en pré-traitement, dans Polyphemus.

Processus	Condition	Modélisation
Source linéique	Modèle de panache seulement	équation 2.38
Réflexion au sol	Si $\sigma_z > z_s$	équation 2.18
Réflexion sur z_i	Si journée et $z_s + \sigma_z > z_i$	équation 2.19
Champ lointain	Si $\sigma_z > 1.5\ z_i$	équation 2.23
Pénétration partielle	Si surhauteur et $\Delta h \geq \frac{z_i - z_s}{1.5}$	équation 2.90
Dispersion (calcul des σ)		Briggs (rural/urbain)
		ou Doury
		ou théorie de la similitude
Surhauteur	Si activé	Briggs-HPDM (avec ou sans « break-up »)
		ou Concawe
		ou Holland-Stümke
Dépôt sec	Si vitesses non nulles	Overcamp
		ou Chamberlain
Lessivage	Si coefficients non nuls	équation 2.93
Décroissance radioactive	Si coefficients non nuls	équation 2.112
Chimie	Gaz et modèle à bouffées seulement	RACM

Tab. 2.9 – Récapitulatif des processus modélisés dans les modèles gaussiens de Polyphemus. Les processus en gras sont modélisés par différentes paramétrisations empiriques, qui seront comparées dans le chapitre suivant.

Il existe également des processus non pris en compte, pour le moment, par les modèles présentés ici, et donc non détaillés dans ce chapitre. On peut notamment citer le fait de modéliser, dans les cas très convectifs, le panache par une superposition de deux gaussiennes sur la verticale [Cimorelli *et al.*, 2004], ou la prise en compte de l'effet des bâtiments. Enfin, pour le modèle à bouffées, il est possible de modéliser l'effet du cisaillement de vent par la scission des bouffées (Thykier-Nielsen *et al.* [1998], Sykes *et al.* [2000]). Les modèles décrits ici reprennent globalement l'état de l'art dans ce domaine, et implémentent les principales paramétrisations

utilisées aujourd'hui. Le chapitre 3 présente l'évaluation et la comparaison de ces différentes paramétrisations, comparées à d'autres modèles et aux observations, ainsi que les incertitudes associées.

Chapitre 3

Evaluation des modèles à l'échelle locale

Les modèles gaussiens pour la dispersion à l'échelle locale présentés dans le chapitre précédent sont évalués à l'aide de comparaison aux observations pour différentes campagnes de mesure, et les performances sont comparées à celles des modèles gaussiens les plus couramment utilisés. Il s'agit non seulement d'évaluer la performance des modèles gaussiens de Polyphemus, mais aussi de quantifier l'impact de la modélisation des principaux processus sur les résultats, en utilisant différentes paramétrisations pour la dispersion et la surhauteur. Cette approche permet d'évaluer l'incertitude inhérente aux modèles. En particulier, il s'agit de comparer la variabilité des performances obtenues avec les paramétrisations de Polyphemus, et celle issue de l'utilisation de différents modèles.
Pour avoir une première évaluation de la variabilité induite par les paramétrisations de la dispersion, les écarts types sont évalués pour différentes paramétrisations et situations météorologiques. Le modèle est comparé aux observations de la campagne Prairie Grass, qui correspond à une dispersion près du sol, sur terrain plat, sans surhauteur ni dépôt. Ensuite, la campagne de mesures de Kincaid est mise à profit pour évaluer les performances du modèle dans un cas convectif avec une source en hauteur, et une surhauteur importante. Enfin, un cas de dépôt en bord de route est traité, afin d'évaluer la modélisation des processus de perte (dépôt et lessivage) et des sources linéiques.

Sommaire

3.1 Critères d'évaluation **81**
3.1.1 Indicateurs statistiques 81
3.1.2 Critères de performance 81
3.2 Evaluation de la dispersion : Prairie Grass **81**
3.2.1 Présentation de l'expérience, et évaluation des écarts types 82
3.2.2 Comparaison avec d'autres modèles gaussiens 84
3.2.3 Statistiques par arc 85
3.3 Evaluation avec surhauteur : Kincaid **87**
3.3.1 Présentation de l'expérience, et évaluation de la surhauteur 87
3.3.2 Evaluation du modèle 90
3.3.3 Comparaison avec d'autres modèles gaussiens 95

3.4 Evaluation du dépôt : mesures de dépôt en bord de route 96
3.4.1 Configuration 98
3.4.2 Résultats moyens sur février 1997 98
3.4.3 Analyse des résultats sur certains jours 100
3.5 Bilan : évaluation des modèles gaussiens et variabilité 103
3.5.1 Performance des modèles 104
3.5.2 Variabilité des paramétrisations et des résultats 105

Les parties 3.1, 3.2 et 3.3 proviennent de Korsakissok et Mallet [2009]. La partie 3.4 est extraite d'une étude qui a été réalisée pour le SETRA, en collaboration avec le CETE de Lyon [Taghavi *et al.*, 2009].

Note : Toutes les simulations ont été effectuées avec le modèle gaussien stationnaire, car les observations correspondent à des concentrations moyennes en temps. Étant donné que les deux modèles stationnaire et instationnaire partagent les mêmes paramétrisations, le modèle à bouffées donne les mêmes résultats lorsqu'on considère la moyenne sur un temps suffisamment long.

3.1 Critères d'évaluation

La méthode d'évaluation du modèle décrite ici est utilisée dans la plupart des évaluations, et notamment pour les modèles avec lesquels les résultats de Polyphemus sont comparés. Les critères d'évaluation sont décrits dans Chang et Hanna [2004], et utilisés notamment dans le kit de validation de modèle [Olesen et Chang, 2005] qui a été développé et testé autour d'une série de conférences intitulée *Harmonization within Atmospheric Dispersion Modeling for Regulatory Purpose*. Dans cette approche, on considère des expériences où les capteurs sont placés en arcs, à différentes distances de la source. Il s'agit de comparer les maxima de concentration observés et simulés par arc. Le maximum sur un arc est déterminé sur des mesures moyennées sur une certaine durée (10 minutes pour Prairie Grass) afin de représenter un état stationnaire. Dans la suite, les comparaisons sont faites sur les valeurs appelées « centreline », qui correspondent aux maxima de concentration normalisés par le taux d'émission Q.

3.1.1 Indicateurs statistiques

Les résultats sont comparés au moyen de diagrammes de dispersion (scatter plots) donnant les valeurs simulées en fonction des valeurs observées. Les performances des modèles sont également évaluées par les indicateurs statistiques suivants, dont la définition est donnée en annexe A :

1. Le biais fractionnel (« fractional bias ») FB,
2. L'erreur quadratique normalisée (« normalized mean square error ») NMSE,
3. Le coefficient de corrélation Corr,
4. La proportion de valeurs simulées à moins d'un facteur 2 des valeurs observées FAC2.

Notons que, d'après les conventions utilisées (annexe A), les valeurs de FB sont négatives en cas de sur-estimation, et positives en cas de sous-estimation des valeurs.

3.1.2 Critères de performance

Un modèle parfait aurait une corrélation et un FAC2 égaux à 1, et des valeurs de FB et NMSE égales à 0. Étant donné qu'un modèle parfait n'existe pas, il est utile de se donner des critères donnant les performances acceptables. Ces critères ont été développés dans Chang et Hanna [2004] et Hanna *et al.* [2004] à partir de l'évaluation de nombreux modèles sur un grand nombre d'expériences. Un modèle est considéré comme acceptable si :

1. le biais moyen est égal à $\pm 30\%$ la valeur moyenne ($-0.3 \leq \mathrm{FB} \leq 0.3$) ;
2. la dispersion des valeurs est inférieure à 2 à 3 fois la moyenne ($\mathrm{NMSE} \leq 4$)
3. la proportion de valeurs prédites qui sont à moins d'un facteur 2 des valeurs observées est supérieure à 50% ($\mathrm{FAC2} \geq 0.5$).

Ces indicateurs et ces critères sont utilisés dans les parties 3.2 et 3.3, avec les expériences de Prairies Grass et Kincaid respectivement.

3.2 Evaluation de la dispersion : Prairie Grass

On évalue à présent les résultats du modèle gaussien « plume », comparé aux observations de Prairie Grass. Ce cas ne présente ni surhauteur, ni dépôt : il permet donc d'évaluer les paramétrisations des écarts types, et la variabilité associée en sortie de modèle. Dans un premier temps, les estimations typiques des écarts types données par les différentes paramétrisations sont comparées, afin de donner une idée de l'incertitude associée. Et puis, les résultats sur Prairie

Grass sont évalués suivant les critères donnés en partie 3.1, et les performances sont comparées à celles d'autres modèles.

3.2.1 Présentation de l'expérience, et évaluation des écarts types

Présentation de l'expérience de Prairie Grass

La campagne de mesures de Prairie Grass constitue une base de donnée sur laquelle de nombreux modèles ont été évalués, et qui a également servi à calibrer des paramétrisations (notamment les formules de Briggs). L'expérience a eu lieu à l'été 1956 à O'Neil, Nebraska. Le site est un terrain plat, herbeux. Une source ponctuelle émettait un panache continu de SO_2, sans surhauteur, près du sol ($z_s = 0.46$ m). Les capteurs de mesures étaient placés en 5 arcs, aux distances de 50, 100, 200, 400 et 800 m de la source (positions sur la figure 3.1). Sur les 68 expériences réalisées dans des conditions météorologiques variées, un ensemble de 43 expériences est en général utilisé pour l'évaluation de modèles. Pour l'évaluation des modèles de Polyphemus, les données proviennent de la base de données de l'EPA (Environmental Protection Agency, Etats-Unis) : `http://www.epa.gov/scram001/dispersion_prefrec.htm`. Les classes de stabilité proviennent de la base de données utilisée dans Hanna *et al.* [2004].

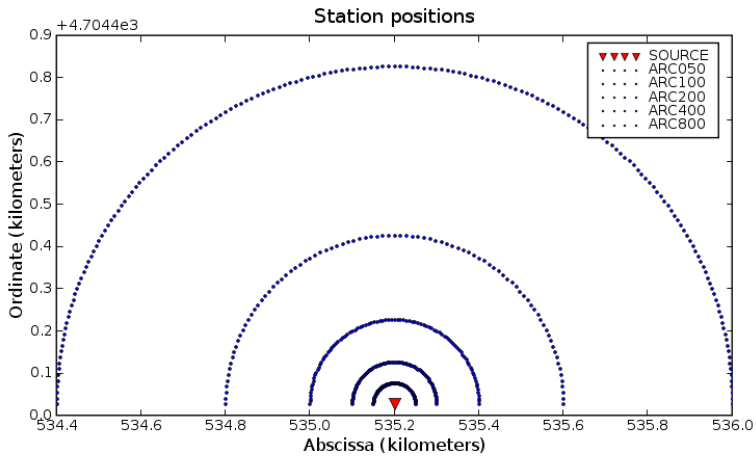
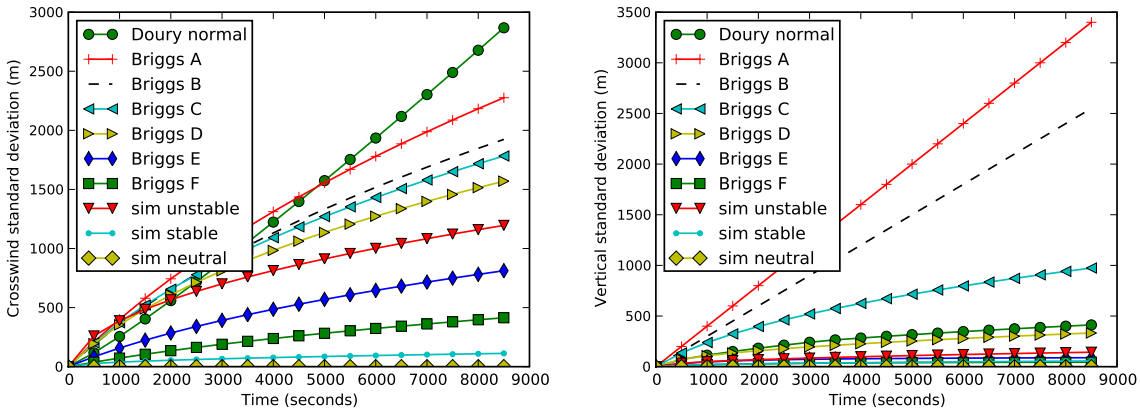


FIG. 3.1 – Position de la source et des capteurs pour l'expérience de Prairie Grass. Les capteurs sont disposés en arcs de rayons 50, 100, 200, 400 et 800 m.

Comparaison des écarts types

Il s'agit ici de comparer l'évolution des écarts types gaussiens, calculés avec les paramétrisations empiriques de Doury, Briggs, et la théorie de la similitude. On ne considère ici que les écarts types liés à la diffusion, sans l'ajout de termes éventuels dus au diamètre de la source ou à la surhauteur. Le but est d'avoir une première estimation de la variabilité due à ces écarts types, à comparer ensuite avec celle des concentrations obtenues avec les mêmes paramétrisations. On se place ici dans un cas similaire à celui de l'expérience de Prairie Grass, avec une source proche



(a) Ecart-type horizontal (b) Ecart-type vertical

FIG. 3.2 – Evolution des écarts types gaussiens σ_y et σ_z en fonction du temps de trajet depuis la source. Les écarts types sont calculés avec les formules de Briggs et les six classes de stabilité de Pasquill, avec les formules de Doury et une diffusion normale, et avec la théorie de la similitude.

du sol, sans surhauteur, et un terrain plat rural (longueur de rugosité prise à $z_0 = 0.01$ m). Les écarts types sont évalués pour différentes situations météorologiques. Pour les écarts types de Briggs, la classe de stabilité de Pasquill, ainsi que la vitesse du vent, sont nécessaires. On veut calculer les écarts types pour les six classes de stabilité. Les vitesses de vent associées sont prises de façon à être représentatives des différentes classes de stabilités (tableau 3.1).

Classe de Pasquill	Vitesse du vent ($\mathrm{m\,s^{-1}}$)
A	2
B	2.5
C	4
D	5.5
E	3
F	2

TAB. 3.1 – Vitesses de vent associées aux différentes classes de stabilité.

Pour calculer les écarts types avec la théorie de la similitude, il est nécessaire de disposer de données météorologiques supplémentaires. La vitesse de friction est prise à une valeur typique du cas de Prairie Grass : $u_* = 0.3\ \mathrm{m\,s^{-1}}$, et la hauteur de couche limite est $h = 500$ m. D'autre part, on estime une longueur de Monin-Obukhov correspondant à chaque classe de Pasquill, pour la longueur de rugosité donnée (Golder [1972] ; voir équation 2.43). Pour le cas instable (classe B), on obtient $L = -5.86$ m, et pour le cas stable (classe E), $L = 11.5$ m. Enfin, la vitesse convective w_* est calculée en fonction des paramètres précédents, d'après l'équation 2.41.

La figure 3.2(a) montre l'évolution de l'écart type horizontal, pendant trois heures. Les valeurs de σ_y, à $t = 9000$ s, vont de 100 m (théorie de la similitude) à 3000 m (Doury) dans les cas stables. Dans les cas instables, la variabilité est moins importante : σ_y varie dans l'intervalle $[1200\ \mathrm{m}, 3000\ \mathrm{m}]$. La théorie de la similitude donne l'écart type le plus faible dans les cas stables et neutres, et des valeurs comparables à celles données par les formules de Briggs dans les cas instables. La plus grande valeur de l'écart type horizontal, dans les cas stables, est donnée par la paramétrisation de Doury. Notons également que les écarts types de Briggs sont donnés en

fonction de la distance à la source x, alors que ceux de Doury et de la théorie de la similitude sont donnés en fonction du temps $t = x/u$ (avec u la vitesse du vent). Ainsi, dans des cas où le vent est faible (davantage que dans les cas moyens présentés ici), la distance x reste faible par rapport au temps de trajet t, donc les formules de Briggs donneront des valeurs relativement plus faibles que les deux autres paramétrisations.

L'écart type vertical a également une grande variabilité (figure 3.2(b)). Dans les cas stables, les valeurs de σ_z vont de 20 m (theorie de la similitude) à 400 m (Doury). Dans les cas instables, les valeurs sont plus dispersées, variant de 100 m (theorie de la similitude) à 3500 m (Briggs, classe A). Il s'agit donc d'un facteur 30 entre les valeurs. Cependant, seules les formules de Briggs pour les classes A, B et C donnent des valeurs de σ_z supérieures à 500 m : les autres estimations restent dans un intervalle beaucoup plus restreint ($100 - 500$ m). Les écarts types calculés par la théorie de la similitude sont les plus faibles, dans toutes les situations.

Ces premiers résultats montrent une large variabilité des écarts types, qui dépendent de la stabilité et des paramétrisations. On peut donc s'attendre à une grande variabilité, due à ces paramétrisations, dans les sorties des modèles. Ces résultats peuvent être en partie contrebalancés, par exemple du fait du diagnostic de stabilité qui est différent suivant les paramétrisations. Ainsi, dans le cas de Prairie Grass, le diagnostic de Pasquill-Turner utilisé pour les formules de Briggs donne 30% de cas instables, 20% de cas stables et 50% de cas neutres. A l'inverse, le diagnostic basé sur la longueur de Monin-Obukhov, utilisé pour la théorie de la similitude, donne 48% de cas instables et 52% de cas instables. Par conséquent, pour certaines situations, les formules de Briggs pour la classe D, et la théorie de la similitude pour les cas instables, seront utilisés. Ces deux formules ne donnent pas des estimations très différentes des écarts types.

3.2.2 Comparaison avec d'autres modèles gaussiens

De nombreux modèles gaussiens, ainsi que certains modèles de CFD, ont été évalués avec la campagne de Prairie Grass. Les résultats pour trois des modèles gaussiens les plus utilisés, repris de CERC [2007], sont repris ici : **ISCST3** (Industrial Source Complex Short Term, version 3) et **AERMOD** sont les modèles recommandés par l'EPA, et **ADMS4** (Atmospheric Dispersion Modeling System), développé en Grande-Bretagne, est très utilisé en Europe. ISCST3 est un modèle « ancienne génération », qui utilise une description discrète de l'atmosphère en classes de stabilité (Briggs pour les zones urbaines et Pasquill-Guifford pour les zones rurales). AERMOD et ADMS sont des modèles « nouvelle génération », qui utilisent la théorie de la similitude pour modéliser la dispersion, avec une description de l'atmosphère par la longueur de Monin-Obukhov. Les modèles de CFD **FLACS** (FLame Acceleration Simulator) [Hanna *et al.*, 2004], et **Mercure** [Demaël et Carissimo, 2008], ont également été évalués avec cette campagne de mesure.

Le tableau 3.2 donne les résultats sur Prairie Grass du modèle de Polyphemus avec les trois paramétrisations des écarts types, ainsi que pour les trois modèles gaussiens évalués dans CERC [2007]. Les statistiques sont calculées pour 43 expériences avec Polyphemus, et pour un sous-ensemble de ces 43 expériences pour les autres modèles (d'où les différences dans les valeurs moyennes observées). Les statistiques sont calculées sur les maxima de concentration normalisés par le taux d'émission, pour les cinq arcs.

La plupart des indicateurs présentés dans le tableau 3.2 remplissent les critères de performances présentés dans la partie 3.1. AERMOD, ISCST3 et Polyphemus avec la paramétrisation de Briggs donnent de bons résultats, ce qui n'est pas surprenant étant donné que l'expérience de Prairie Grass a servi à calibrer leurs paramétrisations de la dispersion. En revanche, les résultats d'ADMS4 et de Polyphemus avec les formules de Doury sont moins performants, et tous les indicateurs ne remplissent pas les conditions définies en 3.1. Ainsi, ADMS4 sous-estime les concentrations moyennes d'environ 30%, et la paramétrisation de Doury ne permet d'avoir

Model	Mean	FB	NMSE	Corr	FAC2
Observations	2.23	0.00	0.00	1.00	1.00
ADMS4	1.56	0.36	3.01	0.63	0.66
AERMOD	2.14	0.04	1.83	0.75	0.73
ISCST3	2.01	0.11	1.78	0.72	0.61
Observations	2.32	0.00	0.00	1.00	1.00
Briggs	2.33	0.00	1.83	0.78	0.74
Doury	1.74	0.29	2.58	0.67	0.29
Théorie de la similitude	2.34	-0.01	0.91	0.85	0.62

TAB. 3.2 – Statistiques pour différents modèles gaussiens, et pour le modèle « plume » de Polyphemus avec les trois paramétrisations pour les écarts types. Les statistiques sont calculées sur les maxima de concentration par arc, normalisés par le taux d'émission. Statistiques calculées sur 43 expériences de Prairie Grass pour les modèles de Polyphemus, et sur un sous-ensemble de ces expériences pour les autres modèles.

que 29% des valeurs prédites à moins d'un facteur 2 des valeurs observées. Enfin, la théorie de la similitude implémentée dans Polyphemus donne les meilleurs résultats : il n'y a pratiquement aucun biais, et la corrélation est la plus élevée des modèles présentés ici. Une comparaison plus complète des trois paramétrisations de Polyphemus, détaillée par arc, est présentée en partie 3.2.3.

3.2.3 Statistiques par arc

Les performances des trois paramétrisations des écarts types sont représentées, calculées séparément pour chacun des cinq arcs de mesures (figure 3.3). Les indicateurs statistiques représentés sont la corrélation, le biais fractionnel FB, la NMSE et le FAC2. Les limites pour qu'un modèle soit « acceptable » suivant les critères donnés en partie 3.1 sont représentées par les lignes noires. La plupart des indicateurs deviennent moins bons lorsque la distance à la source augmente, à l'exception de la corrélation. Les paramétrisations de Briggs et Doury ont tendance à sous-estimer les concentrations (FB est positif), tandis que la théorie de la similitude les sur-estime aux arcs les plus lointains (400 m and 800 m). Cela provient probablement des faibles valeurs données par cette paramétrisation pour les écarts types, et notamment σ_z, soulignées dans la partie 3.2.1 : les concentrations restent élevées sur de plus grandes distances. Enfin, ces résultats confirment les performances médiocres de la paramétrisation de Doury, notamment pour ce qui est du biais et du FAC2.

Ces résultats peuvent également être comparés avec les analyses par arc, effectuées pour les modèles de CFD FLACS et Mercure. Dans Hanna *et al.* [2004], les résultats pour le modèle FLACS sous-estiment les concentrations à faible distance de la source (jusqu'à 200 m), alors que la tendance est à une légère sur-estimation à plus grande distance. Les résultats pour Mercure, donnés dans Demaël et Carissimo [2008], montrent également cette tendance à la sous-estimation en champ très proche. En effet, ces deux modèles représentent la dispersion du panache en champ proche de la source par un comportement en $\sigma \propto \sqrt{t}$ (équation 2.15), ce qui n'est pas approprié à ces distances où l'on a plutôt $\sigma \propto t$. Les modèles gaussiens, à l'exception de la paramétrisation de Doury, donnent donc de meilleurs résultats.

La figure 3.4 montre les diagrammes de dispersion pour les trois paramétrisations des écarts types. Elle confirme les tendances montrées par l'analyse des indicateurs statistiques : Briggs prédit des valeurs proches des observations, la théorie de la similitude également, bien que la sur-estimation aux arcs éloignés soit clairement visible. Enfin, le diagramme de dispersion de Doury

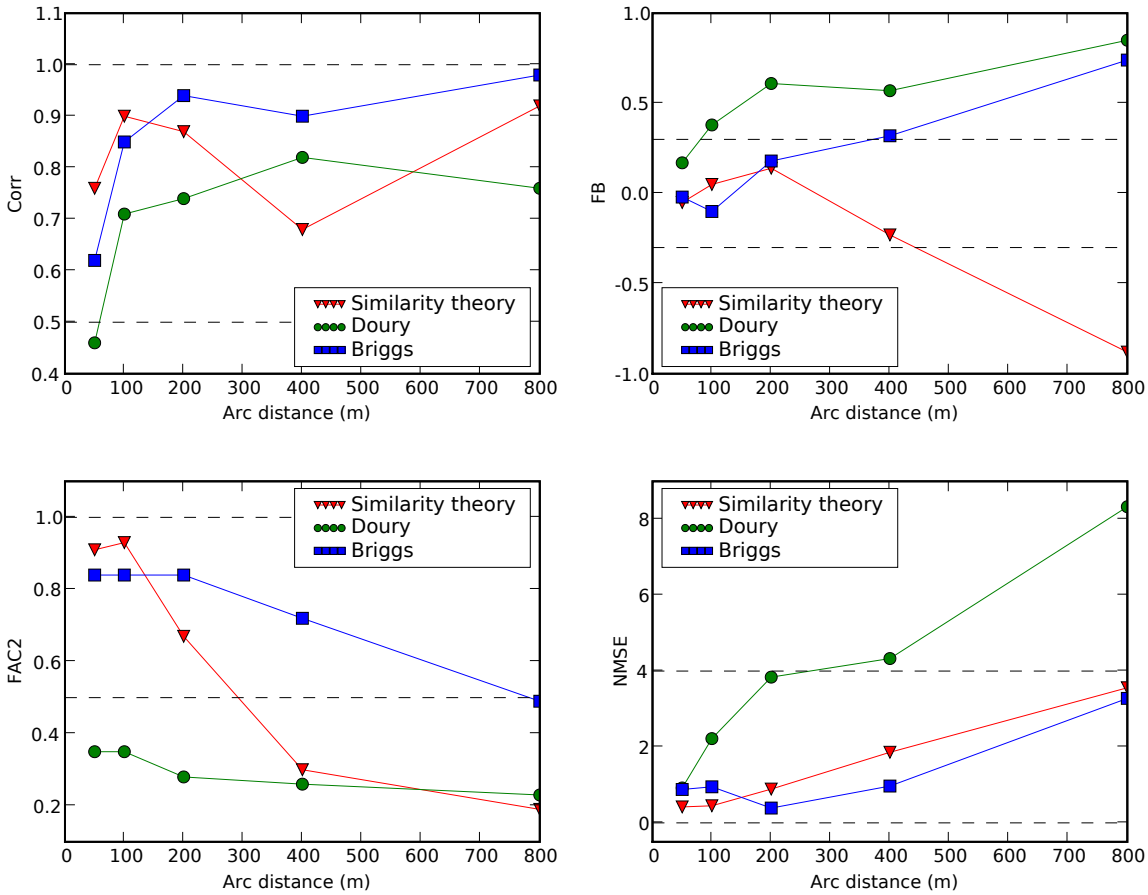
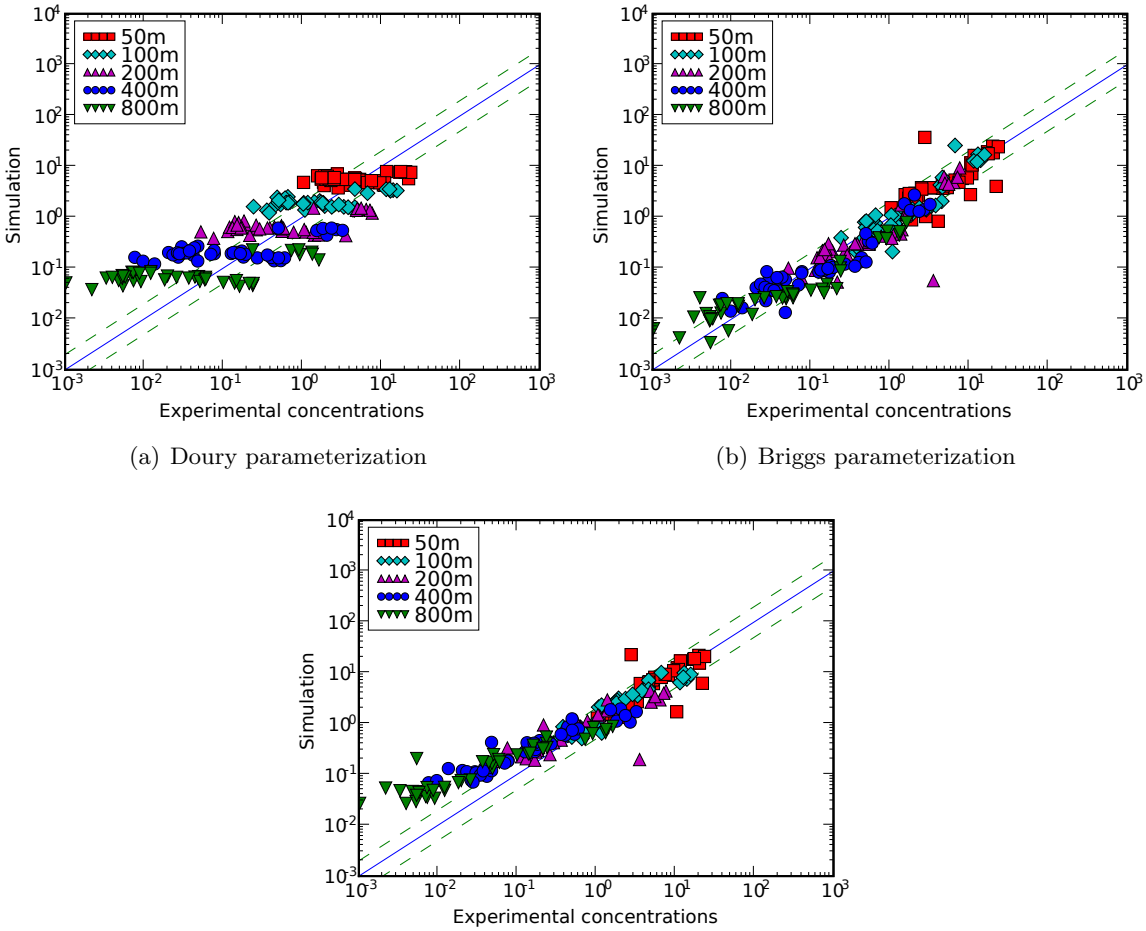


FIG. 3.3 – Evolution des statistiques par arc, en fonction de la distance à la source, pour les trois paramétrisations des écarts types utilisées dans Polyphemus. Les statistiques sont calculées aux cinq arcs, situés à 50 m, 100 m, 200 m, 400 m et 800 m de la source. Les limites pour une performance « acceptable » sont représentées (tirets).

met en évidence les mauvaises performances, notamment en terme de FAC2 : les valeurs, pour un arc donné, sont pratiquement constantes, ce qui signifie que les maxima de concentration simulés varient peu d'une expérience à l'autre. Cela peut provenir de la représentation très grossière de la stabilité de l'atmosphère (seulement deux cas, dispersion normale et faible). De plus, la paramétrisation de Doury donne également des écarts types horizontaux assez grands par rapport aux estimations fournies par les autres formules, notamment si le vent est faible, comme le montre la partie 3.2.1. Cela explique la sous-estimation globale des résultats.



(a) Doury parameterization

(b) Briggs parameterization

(c) Similarity theory

FIG. 3.4 – Diagrammes de dispersion pour les trois paramétrisations des écarts types de Polyphemus. Les maxima de concentration par arc sont utilisées, pour 43 expériences de Prairie Grass. L'unité est $\mathrm{mg\,m^{-3}/(g\,s^{-1})}$. Les résultats sont montrés pour les cinq arcs (50 m,100 m, 200 m, 400 m et 800 m).

3.3 Evaluation avec surhauteur : Kincaid

3.3.1 Présentation de l'expérience, et évaluation de la surhauteur

La campagne d'observations utilisée dans cette partie eut lieu à Kincaid, Illinois (USA), en 1980 et 1981. Il s'agissait de mesurer les concentrations au sol émises par une source en hauteur, une cheminée de 187 m de haut et de 9 m de diamètre, entourée de champs et de quelques lacs. Le panache de SF_6 gazeux, était émis par la cheminée avec une surhauteur importante. La position

des 200 capteurs était déterminée au jour le jour, en fonction de la situation météorologique (les conditions étant globalement neutres à convectives). La distance entre les arcs et la source varie ainsi de 0.5 à 50 km. La figure 3.5 donne un exemple de placement des capteurs, pour l'une des expériences, en fonction de la distance à laquelle le panache touche le sol. Les mesures consistent en des moyennes horaires du maximum de concentration pour chaque arc. La base de données ainsi obtenue contient en tout les résultats pour 171 expériences (ou mesures horaires), et est fréquemment utilisée pour tester des paramétrisations de surhauteur. Les données d'observation utilisées ici proviennent d'un kit de validation de modèles (Model Validation Kit, `http://www.harmo.org/kit/Download.asp`). Elles sont fournies avec un indicateur de qualité pour chacun des maxima de concentration mesurés. En général, seules les mesures les plus fiables, de qualité 2 et 3, sont utilisées pour la validation de modèle [Olesen, 1995]. Nous utilisons ici les données de qualité 2 et 3 pour 165 des 171 expériences.

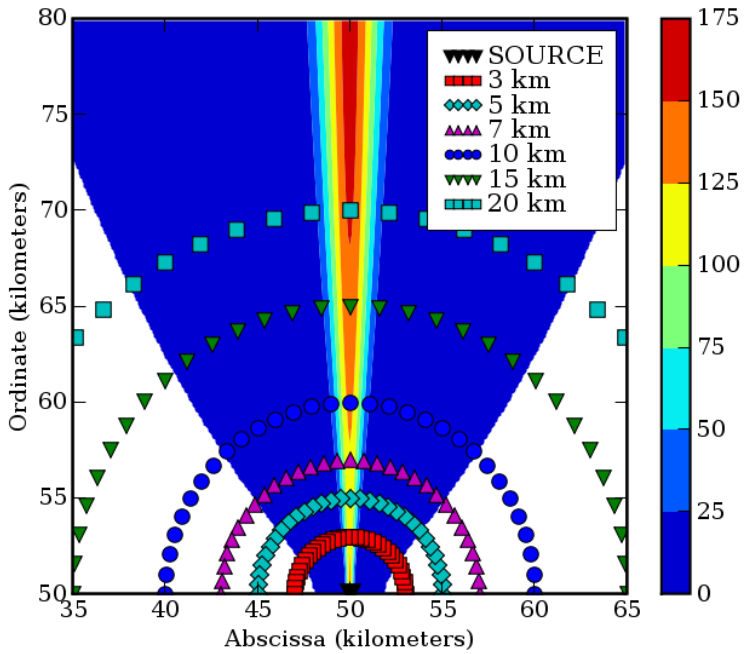


Fig. 3.5 – Position de la source et des capteurs pour l'expérience de Kincaid. Les concentrations au sol sont montrées pour l'une des expériences : le panache ne touche le sol qu'à une certaine distance de la source. Unité : $\mathrm{ng\,m^{-3}}$.

Comme pour l'expérience de Prairie Grass, les simulations effectuées avec les modèles gaussiens de Polyphemus sont comparées aux observations, mais aussi aux performances d'autres modèles tels qu'ADMS et AERMOD, notamment. A la différence de ces derniers, qui recalculent la hauteur d'inversion au cours de leur pré-traitement météorologique, les modèles de Polyphemus ont été utilisés en prenant en entrée la hauteur d'inversion observée, qui est fournie avec les autres données. Cette différence est mentionnée car la hauteur d'inversion peut fortement influencer les résultats obtenus, puisqu'elle détermine si le panache touche le sol, ou s'il est émis totalement ou partiellement au-dessus de la couche limite. Les simulations ont été réalisées avec les trois paramétrisations pour les écarts types (Briggs, Doury, théorie de la similitude) et les trois paramétrisations pour la surhauteur (Briggs-HPDM, Concawe, Holland) décrites dans le chapitre 2. Dans la suite, le diamètre de la cheminée ainsi que la largeur initiale due à la surhauteur sont pris en compte dans le calcul des écarts types (équation 2.39). Dans le cas où

la théorie de la similitude est utilisée, les formules alternatives pour les sources en hauteur sont choisies préférentiellement aux formules habituelles (équations de la section 2.2.4). Par rapport à la formulation classique, les résultats sont un peu améliorés en terme de corrélation, malgré une légère augmentation de la surestimation.

Surhauteur

Dans cette partie, il s'agit de faire une première analyse afin de quantifier la variabilité supplémentaire due à la surhauteur. Les valeurs de surhauteur calculées avec les différentes paramétrisations sont comparées avec les valeurs observées, pour les 165 expériences : la figure 3.6 montre les valeurs de surhauteur en fonction de la vitesse du vent. De façon prévisible, la surhauteur est d'autant plus faible que la vitesse du vent est élevée, et les différences entre les paramétrisations sont alors plus faibles. Lorsque le vent est faible, ces différences deviennent importantes. Le schéma de Briggs-HPDM donne des surhauteurs parfois très élevées, tandis que les formules d'Holland et de Concawe sont toutes deux plus faibles. Il est donc possible que Briggs-HPDM sur-estime la surhauteur dans certains cas, ce qui avait déjà été relevé pour les formules de Briggs, par exemple dans Canepa *et al.* [2000]. Cela peut, dans ce cas, conduire à une sous-estimation des concentrations, car le panache simulé est trop haut, et touche le sol plus tard, voire pas du tout s'il est placé au-dessus de la hauteur d'inversion. Le tableau 3.3 montre le pourcentage de cas (sur les 165 expériences) dans lesquels les paramétrisations prédisent une pénétration partielle ou totale du panache dans la couche d'inversion. Le nombre d'occurences de pénétration partielle ou totale est plus élevé pour le schéma de Briggs-HPDM (près de 60 % des cas), que pour les deux autres paramétrisations (40 à 47 %), ce qui confirme les observations précédentes. Les formules de Holland et de Concawe correspondent au même nombre d'occurence de pénétration totale, mais il y a moins de cas de pénétration partielle dans le cas d'Holland qu'avec la paramétrisation de Concawe.

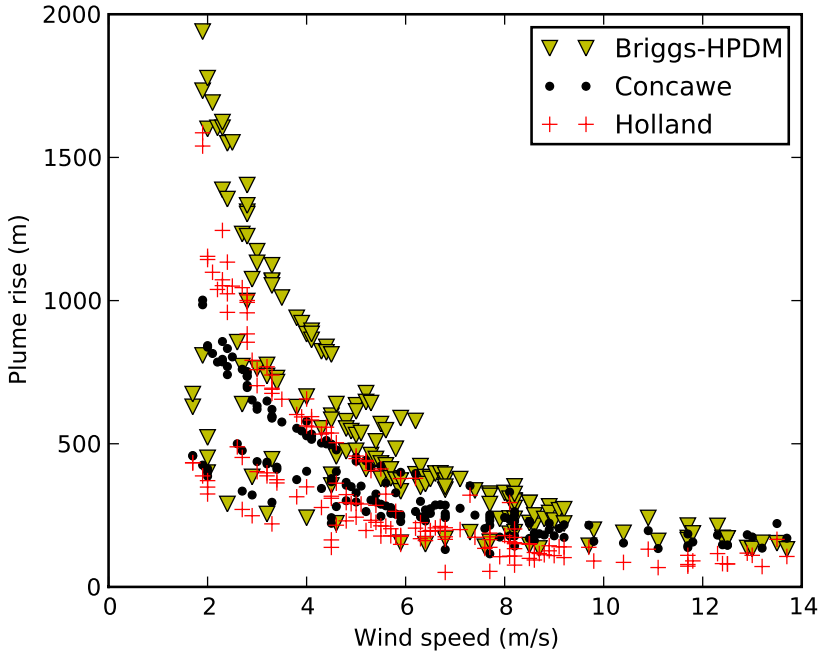


FIG. 3.6 – Surhauteur (en mètres), calculée pour les différentes paramétrisations de surhauteur, en fonction de la vitesse du vent. Résultats pour les 165 expériences de Kincaid.

Model	Total penetration	Partial penetration
Briggs-HPDM	29%	28%
Holland	23%	17%
Concawe	24%	23%

TAB. 3.3 – Pourcentage de cas de pénétration partielle ou totale du panache, pour les trois paramétrisations de surhauteur, sur les 165 expériences de Kincaid.

3.3.2 Evaluation du modèle

A présent, nous analysons les résultats pour les trois paramétrisations de surhauteur, successivement. Pour chacune d'elles, les résultats sont comparés aux observations (par des statistiques sur tous les arcs et toutes les expériences), pour les trois formules d'écarts types. La paramétrisation pour les écarts types qui donne le meilleur résultat est, à chaque fois, soulignée en gras.

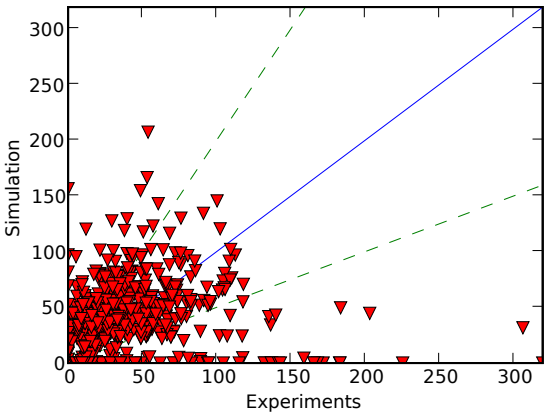
Résultats pour la paramétrisation de Briggs-HPDM

Le tableau 3.4 montre les résultats des simulations pour la surhauteur de Briggs-HPDM, et les trois paramétrisations pour les écarts types. On observe une sous-estimation des concentrations, d'un facteur deux pour la paramétrisation de Briggs, et d'un facteur trois pour celle de Doury. La théorie de la similitude donne des résultats quasiment sans biais, et globalement satisfaisants pour tous les indicateurs excepté la corrélation. Cette dernière est très faible pour toutes les paramétrisations.

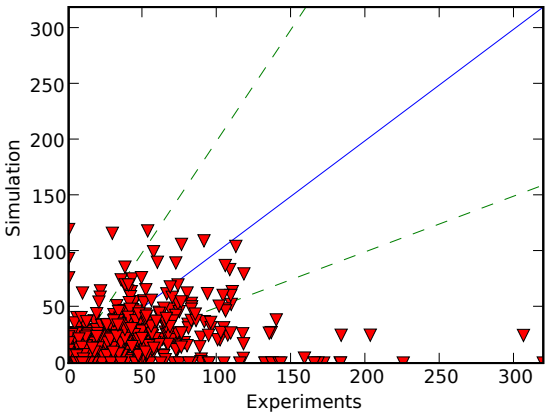
Model	Mean	std	FB	NMSE	Corr	FAC2
Obs	39.89	39.22	0.00	0.00	1.00	1.00
Briggs	21.33	21.10	0.61	2.37	0.19	0.43
Doury	12.77	14.09	1.03	4.70	0.07	0.22
Similitude	42.02	34.73	-0.05	1.33	0.19	0.45

TAB. 3.4 – Statistiques pour plusieurs paramétrisations des écarts types : comparaison des valeurs simulées et observées de maxima de concentrations normalisés par le taux d'émission C/Q, en ng $m^{-3}/(\mathrm{g\,s^{-1}})$. Données de qualité 2 et 3. Surhauteur de Briggs-HPDM.

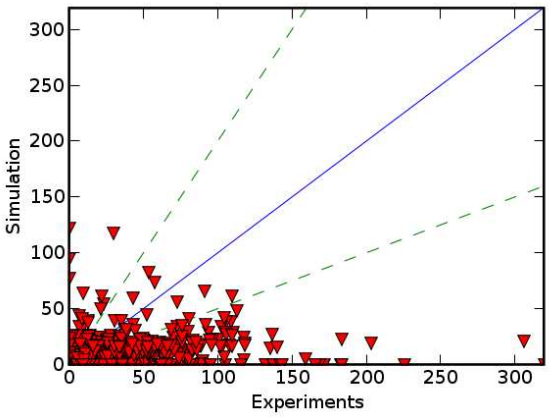
La figure 3.7 montre les diagrammes de dispersion pour les trois paramétrisations des écarts types, et la surhauteur de Briggs-HPDM. La tendance à la sous-estimation des concentrations au sol est tout à fait visible sur les diagrammes de Briggs et de Doury (figures 3.7(b) et 3.7(c)), tandis que le diagramme de dispersion pour la théorie de la similitude est beaucoup mieux centré (figure 3.7(a)). Dans tous les cas, certaines concentrations sont mesurées, mais pratiquement estimée à zéro par le modèle, ce qui se traduit par une série de points alignés sur l'axe des abscisses. Ces concentrations correspondent vraisemblablement à des cas où la surhauteur calculée est trop importante, ou la hauteur d'inversion utilisée trop faible. Si le panache simulé pénètre totalement dans la troposphère libre, les concentrations modélisées sont alors égales à zéro pour tous les arcs. Par ailleurs, une sur-estimation de la hauteur du panache sans qu'il y ait pénétration dans la couche d'inversion peut également produire des concentrations nulles simulées aux arcs les plus proches de la source – alors que les concentrations observées y sont souvent plus élevées qu'aux arc lointains. La faiblesse des corrélations peut s'expliquer en partie par ces erreurs de simulation.



(a) Similarity theory



(b) Briggs



(c) Doury

FIG. 3.7 – Diagrammes de dispersion : maxima de concentrations simulés et observés normalisés par le taux d'émission ($\mathrm{ng\,m^{-3}/(g\,s^{-1})}$). Expérience de Kincaid, 165 jours, données de qualité 3. Surhauteur calculée par les fomules de Briggs-HPDM.

Résultats pour la paramétrisation de Holland-Stümke

Le tableau 3.5 montre les mêmes résultats que précédemment, pour la surhauteur de Holland. Les concentrations modélisées sont globalement plus élevées qu'avec la surhauteur de Briggs-HPDM, car la surhauteur estimée est plus faible. La paramétrisation de Doury conduit toujours à une sous-estimation importante des concentrations, ce qui était déjà le cas pour l'expérience de Prairie Grass. Le biais et la NMSE pour les écarts types de Briggs sont bien meilleurs avec cette estimation de la surhauteur que dans le cas de Briggs-HPDM. A l'inverse, les écarts types calculés avec la théorie de la similitude sur-estiment davantage les concentrations au sol, et ont donc un biais et une NMSE plus élevés qu'avec la surhauteur de Briggs-HPDM.

Model	Mean	std	FB	NMSE	Corr	FAC2
Obs	39.89	39.22	0.00	0.00	1.00	1.00
Briggs	33.01	28.26	0.19	1.46	0.21	0.42
Doury	15.87	13.27	0.86	3.55	0.04	0.28
Similitude	61.97	47.52	-0.43	1.38	0.24	0.38

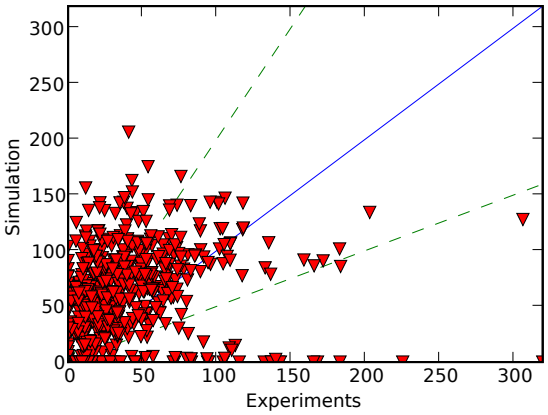
TAB. 3.5 – Statistiques pour plusieurs paramétrisations des écarts types : comparaison des valeurs simulées et observées de maxima de concentrations normalisés par le taux d'émission C/Q, en $\mathrm{ng\,m^{-3}/(g\,s^{-1})}$. Données de qualité 2 et 3. Surhauteur de Holland.

Les diagrammes de dispersion pour les trois paramétrisations des écarts types sont montrés figure 3.8. La tendance à la sur-estimation lorsque les écarts types sont calculés avec la théorie de la similitude est encore visible ici, tandis que le biais avec la paramétrisation de Briggs est plus faible qu'avec la surhauteur de Briggs-HPDM. Les écarts types de Doury produisent encore une importante sous-estimation des concentrations. On constate de nouveau des concentrations pratiquement nulles pour certaines observations élevées. Ce problème ne s'explique donc par une sur-estimation de la surhauteur due à une formulation plutôt qu'une autre. Il est possible que cela soit dû, au moins en partie, à une estimation erronée de la hauteur de la couche d'inversion.

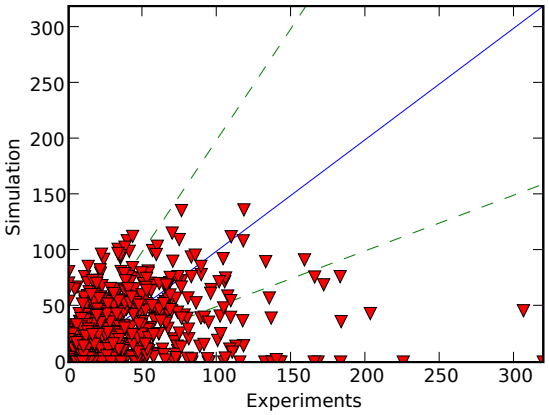
Résultats pour la paramétrisation de Concawe

Dans ce cas, les meilleurs résultats sont donnés par la paramétrisation de Briggs, comme dans le cas de la surhauteur de Holland (tableau 3.6). Pour la surhauteur de Briggs-HPDM, la meilleure performance était donnée par la théorie de la similitude. Les résultats obtenus avec la surhauteur de Concawe ne sont pas très différents de ceux donnés par la surhauteur de Holland, ce qui n'est pas surprenant, compte-tenu du fait que les deux estimations de la surhauteur sont très proches. Les concentrations moyennes ainsi que les biais des trois paramétrisations des écarts types, en particulier, sont similaires. Cependant, la corrélation ainsi que la proportion de valeurs simulées à moins d'un facteur deux des valeurs observées sont meilleures avec la formule de Concawe. Cette formulation de la surhauteur a été construite à partir d'un certain nombre de paramétrisations existantes, dont celle de Holland-Stümke, et ajustée sur des cas expérimentaux, en particulier neutres et convectifs comme dans le cas de Kincaid (Brummage [1968]). Il n'est donc pas surprenant qu'elle donne des résultats légèrement meilleurs que ceux de Holland.

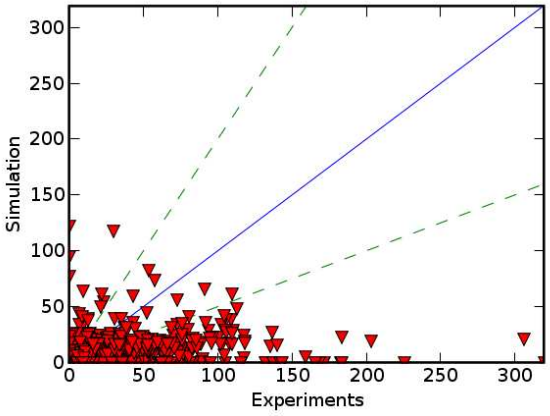
La figure 3.9 montre les diagrammes de dispersion pour les trois paramétrisations des écarts types, pour la surhauteur de Concawe. Ces diagrammes sont très similaires à ceux de la figure 3.8.



(a) Similarity theory



(b) Briggs

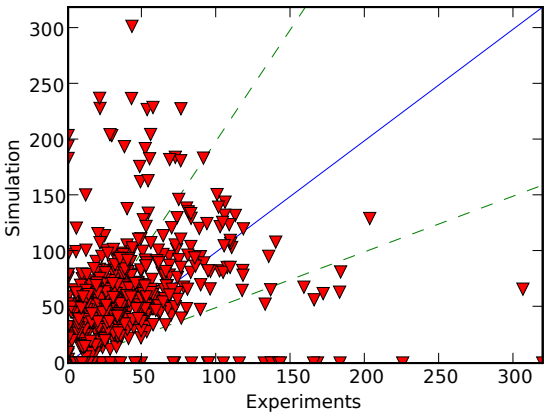


(c) Doury

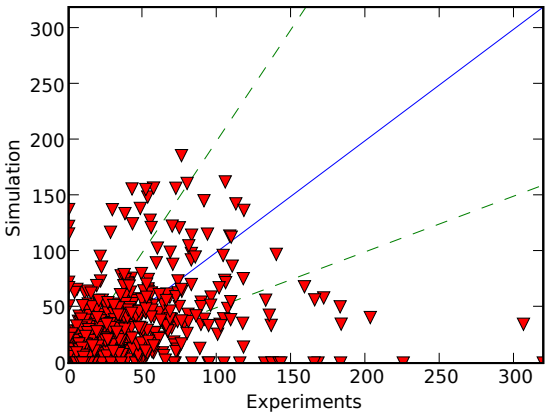
FIG. 3.8 – Diagrammes de dispersion : maxima de concentrations simulés et observés normalisés par le taux d'émission ($\mathrm{ng\,m^{-3}/(g\,s^{-1})}$). Expérience de Kincaid, 165 jours, données de qualité 3. Surhauteur calculée par les fomules de Holland.

Model	Mean	std	FB	NMSE	Corr	FAC2
Obs	39.89	39.22	0.00	0.00	1.00	1.00
Briggs	32.88	32.75	0.19	1.49	0.27	0.48
Doury	16.87	17.74	0.81	3.29	0.12	0.31
Similitude	61.82	53.79	-0.43	1.57	0.24	0.43

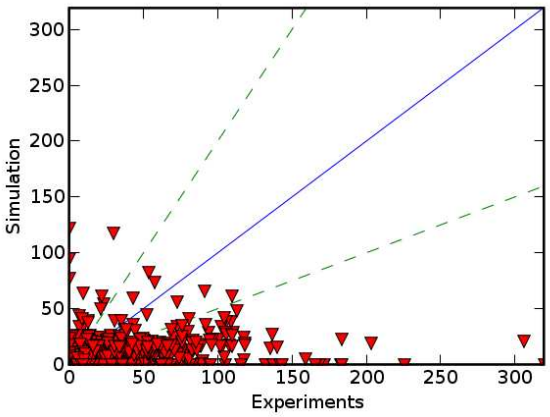
TAB. 3.6 – Statistiques pour plusieurs paramétrisations des écarts types : comparaison des valeurs simulées et observées de maxima de concentrations normalisés par le taux d'émission C/Q, en $\mathrm{ng\,m^{-3}/(g\,s^{-1})}$. Données de qualité 2 et 3. Surhauteur de Concawe.



(a) Similarity theory



(b) Briggs



(c) Doury

FIG. 3.9 – Diagrammes de dispersion : maxima de concentrations simulés et observés normalisés par le taux d'émission ($\mathrm{ng\,m^{-3}/(g\,s^{-1})}$). Expérience de Kincaid, 165 jours, données de qualité 3. Surhauteur calculée par les fomules de Concawe.

3.3.3 Comparaison avec d'autres modèles gaussiens

Il s'agit à présent de comparer les performances des modèles avec celles d'autres modèles gaussiens. Pour cela, on choisit les paramétrisations de surhauteur et d'écarts types donnant les meilleurs résultats parmi ceux montrés précédemment : les trois formulations d'écarts types avec la surhauteur de Concawe. On montre également les résultats donnés par la théorie de la similitude et la surhauteur de Briggs-HPDM, car la performance est plutôt bonne malgré la faible corrélation. Les résultats pour les autres modèles sont tirés de CERC [2007]. Les performances sont montrées dans le tableau 3.7, pour les données de qualité 2 et 3. Les performances pour les modèles gaussiens de Polyphemus ont été calculées sur 165 expériences et non 171, ce qui explique les légères différences entre les moyennes et écarts types des observations, par rapport aux données utilisées pour les autres modèles. Les performances des modèles de Polyphemus sont globalement situées entre celles d'AERMOD et d'ADMS4, excepté les corrélations qui sont plus faibles, et plus proches de celles d'ISCST3.

Model	Mean	std	FB	NMSE	Corr	FAC2
Observation	41.0	39.3	0.00	0.00	1.00	1.00
ADMS4	40.4	31.1	-0.05	0.80	0.50	0.58
AERMOD	20.3	24.1	0.68	2.30	0.35	0.33
ISCST3	23.1	53.3	0.56	3.83	0.26	0.26
Observation	39.89	39.22	0.00	0.00	1.00	1.00
Briggs/Concawe	32.83	32.80	0.19	1.50	0.27	0.47
Doury/Concawe	16.75	17.81	0.82	3.32	0.12	0.30
Similitude/Concawe	61.70	53.91	-0.43	1.58	0.25	0.43
Similitude/HPDM	41.99	34.76	-0.05	1.33	0.19	0.44

TAB. 3.7 – Statistiques pour plusieurs modèles gaussiens : comparaison des valeurs simulées et observées de maxima de concentrations normalisés par le taux d'émission C/Q, en $\mathrm{ng\,m^{-3}/(g\,s^{-1})}$. Données de qualité 2 et 3.

En utilisant des données de qualité 3 uniquement, on effectue les comparaisons avec d'autres modèles : HPDM (Hybrid Plume Dispersion Model, EU, Hanna et Paine [1989]), OML (Operationelle Meteorologiske Luftkvalitetsmodeller, Danemark), SAFE_AIR (Simulation of Air pollution From Emissions Above Inhomogeneous Regions, Italie, Canepa *et al.* [2000]), ainsi que le modèle lagrangien NAME (Numerical Atmospheric-dispersion Modelling Environment, GB, Webster et Thomson [2002]). Les statistiques pour les modèles HPDM et OML proviennent de Olesen [1995]. Le tableau 3.8 montre les résultats pour tous ces modèles. ISCST3, AERMOD et Polyphemus avec la paramétrisation de Doury sous-estiment fortement les valeurs moyennes, et ont un biais et une NMSE élevés. Tous les autres modèles montrent une légère sous-estimation des valeurs moyennes, à part SAFE_AIR et Polyphemus avec la théorie de la similitude et la surhauteur de Concawe, qui ont tendance à légèrement sur-estimer. Seules les corrélations vont à l'encontre de ces tendances : AERMOD a une corrélation de 40%, comparable aux meilleurs modèles, alors qu'il sur-estime fortement les concentrations. En revanche, des modèles ayant une performance globale plutôt bonne, tels que SAFE_AIR, OML, et Polyphemus avec la surhauteur de Concawe, ont des corrélations relativement faibles. La corrélation pour Polyphemus avec la théorie de la similitude et la surhauteur de Briggs-HPDM est très faible (5%), alors que cette configuration est parmi les meilleures pour tous les autres indicateurs. Cette faible corrélation est à rapprocher de celle donnée dans Hanna et Paine [1989] pour l'évaluation du modèle HPDM avec l'expérience de Kincaid, dans une configuration très proche de celle utilisée ici (surhauteur très proche, mêmes formules pour les écarts types des sources en hauteur). Des améliorations ont été faites par la suite dans le modèle HPDM, notamment pour prendre en

Model	Mean	std	FB	NMSE	Corr	FAC2
Observation	54.34	40.25	0.00	0.00	1.00	1.00
ADMS4	48.5	31.5	0.11	0.60	0.45	0.68
AERMOD	21.8	21.8	0.86	2.07	0.40	0.29
ISCST3	30.0	60.0	0.57	2.80	0.26	0.28
HPDM	44.84	38.55	0.19	0.75	0.44	0.56
SAFE_AIR	59.73	52.54	-0.09	1.10	0.20	0.50
OML	47.45	45.48	0.13	1.24	0.15	0.55
NAME	38.7	47.2	0.33	1.45	0.27	0.56
Observation	53.69	40.72	0.00	0.00	1.00	1.00
Briggs/Concawe	37.95	33.06	0.34	1.20	0.21	0.58
Doury/Concawe	17.36	14.21	1.02	3.19	0.18	0.30
Similitude/Concawe	68.91	51.25	-0.25	1.03	0.17	0.58
Similitude/HPDM	47.92	33.93	0.11	1.05	0.05	0.59

TAB. 3.8 – Statistiques pour plusieurs modèles gaussiens : comparaison des valeurs simulées et observées de maxima de concentrations normalisés par le taux d'émission C/Q, en $\mathrm{ng\,m^{-3}/(g\,s^{-1})}$. Données de qualité 3.

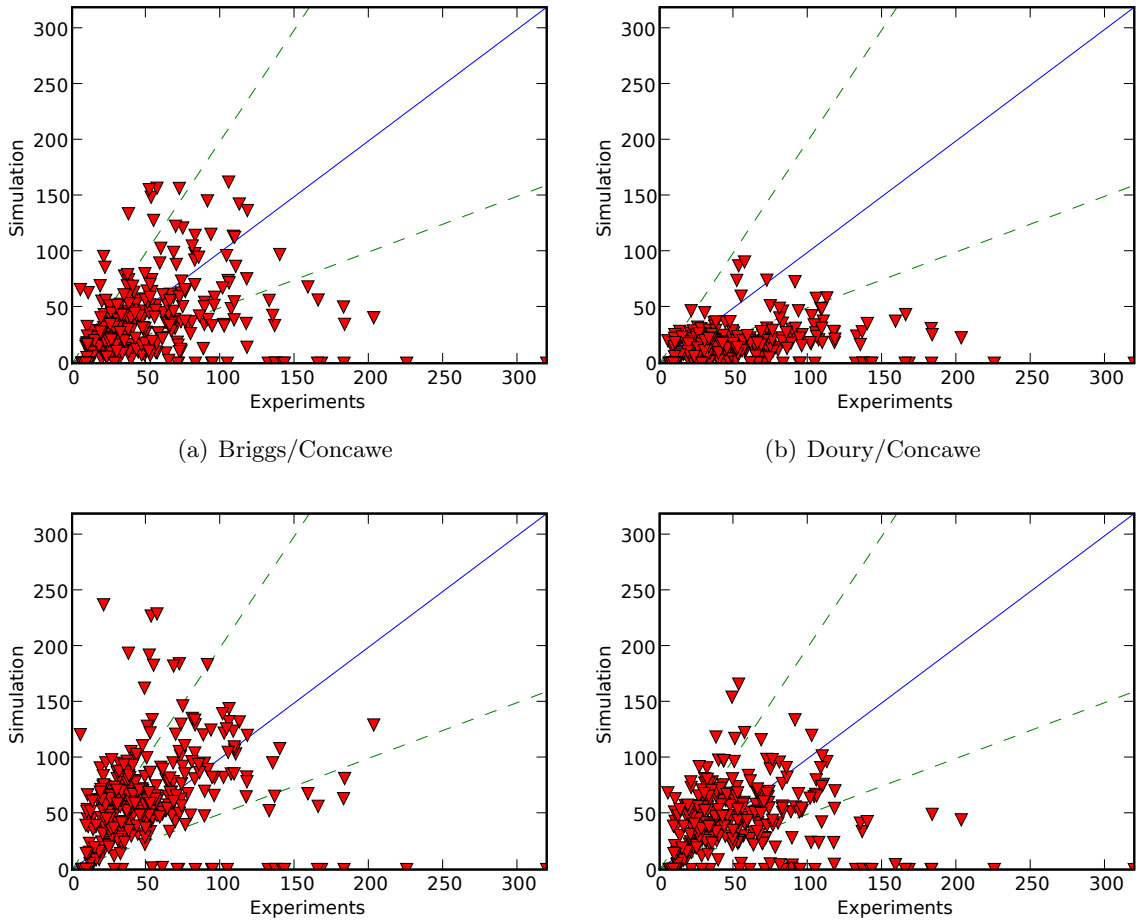
compte plus précisément les interactions du panache avec la couche limite.

Le tableau 3.8 montre une grande variabilité dans les résultats en fonction du modèle utilisé : les concentrations moyennes simulées sont comprises entre 20 $\mathrm{ng\,m^{-3}/(g\,s^{-1})}$ et 70 $\mathrm{ng\,m^{-3}/(g\,s^{-1})}$, ce qui correspond donc à un facteur 3 ou plus entre les modèles. En utilisant le modèle gaussien de Polyphemus et l'ensemble des paramétrisations de surhauteur et d'écarts types à notre disposition, il est possible de reproduire ces écarts. Une grande partie de la variabilité dans les résultats provient donc de l'incertitude sur ces paramétrisations. Seules les corrélations sont très faibles quelle que soit la paramétrisation utilisée, ce qui tend à prouver que cela provient de la modélisation, partagée par toutes les configurations, des interactions avec la couche d'inversion (pénétration partielle/totale, modélisation de la hauteur d'inversion, « lofting »).

La figure 3.10 montre les diagrammes de dispersion correspondant aux cas de qualité 3, pour les différentes paramétrisations. On constate de nouveau des concentrations simulées très proches de zéro pour certaines observations assez élevées. Les meilleures résultats sont donnés par la paramétrisation de Briggs avec la surhauteur de Concawe, et par la théorie de la similitude ave la surhauteur de Briggs-HPDM.

3.4 Evaluation du dépôt : mesures de dépôt en bord de route

Dans cette partie, le modèle gaussien à bouffées de Polyphemus est utilisé pour évaluer le dépôt de cadmium en bord de route. Il s'agit notamment de comparer les résultats à des observations, prises lors d'une campagne de mesure présentée dans Promeyrat-Qotbi [2001], réalisée en France dans la région de Metz. On se focalise sur les valeurs moyennes sur un mois (février 1997), pour un tronçon routier particulier correspondant à un terrain plat. Les simulations réalisées ici ont été faites dans le cadre d'un exercice d'inter-comparaison avec les modèles Mercure et ADMS, en partenariat avec le CETE de Lyon [Taghavi *et al.*, 2009]. Les données d'entrées utilisées pour les simulations sont identiques à celles d'ADMS, notamment concernant les vitesses de dépôt qui n'ont pas été recalculées, afin de comparer au mieux les modèles. Les résultats d'ADMS présentés ici proviennent de simulations réalisées par le CETE de Lyon.



(a) Briggs/Concawe

(b) Doury/Concawe

(c) Similarity theory/Concawe

(d) Similarity theory/HPDM

FIG. 3.10 – Diagrammes de dispersion : maxima de concentrations simulés et observés normalisés par le taux d'émission ($\mathrm{ng\,m^{-3}/(g\,s^{-1})}$). Expérience de Kincaid, 165 jours, données de qualité 3.

3.4.1 Configuration

Positionnement des émissions et des capteurs

L'emplacement de la route et les positions des capteurs ont été repris d'après la configuration d'ADMS. La route est discrétisée en sources ponctuelles de 20 mètres de diamètre, ce qui correspond à la largeur de la route. Il faut donc au minimum 50 sources ponctuelles pour le tronçon d'1 km. En pratique, le nombre de sources ponctuelles est pris à 200. Des simulations reprenant la même configuration et utilisant la modélisation des sources linéiques décrite dans le chapitre 2, ont également été réalisées [Briant, 2009]. La hauteur d'émission est prise à 1 m. Cela correspond à une hauteur de mélange de 2 m due à la turbulence induite par le trafic, en prenant la hauteur des sources au centre de cette couche mélangée.

Il y a 5 capteurs, placés à 5, 20, 40, 80, 160 et 320 mètres du bord de la route. Les positions de la route et des quatre premiers capteurs sont montrées figure 3.11. Ces capteurs étant fixes, une sélection a été effectuée sur les situations météorologiques du mois simulé, afin de ne retenir que les jours où la direction du vent permettait le transport de polluants vers les capteurs. Sur le mois de février 1997, 19 jours ont ainsi été retenus pour les simulations.

Données d'émission

La simulation a été effectuée avec du cadmium. Les diamètres des particules sont de 10 µm, 2.5 µm et 1 µm. La proportion de particules de ces différents diamètres est de 17%, 20% et 63% respectivement.

Le taux d'émission est de $1.16\,\mathrm{g\,km^{-1}s^{-1}}$: comme la longueur du tronçon est de 1 km, ce taux est divisé par le nombre de sources ponctuelles pour obtenir le taux d'émission de chacune des sources composant la route discrétisée.

Autres données

Les données météorologiques sont reprises des fichiers d'entrée ADMS. La stabilité, nécessaire pour calculer les écarts types gaussiens avec la formule de Briggs, est considérée comme neutre (classe D de Pasquill). Les écarts types σ_y et σ_z sont calculés avec les formules de Briggs pour les zones rurales.

Afin de reproduire les simulations ADMS, et de valider ainsi les autres aspects du modèle, les vitesses de dépôt totales ont été directement reprises du fichier de sortie ADMS. Elles ont été calculées pour une densité $\rho = 8647\,\mathrm{kg\,m^{-3}}$. Les coefficients de lessivage ont été, eux, calculés avec la paramétrisation de Slinn (équation 2.110) pour les aérosols (ADMS utilisant la paramétrisation de Belot, moins appropriée pour les particules).

Les simulations sont effectuées avec les formules d'Overcamp (équation 2.105) et de Chamberlain (équation 2.103) pour le dépôt sec.

3.4.2 Résultats moyens sur février 1997

Concentrations moyennes

La moyenne des concentrations de cadmium sur l'ensemble des jours simulés est représentée figure 3.11. Les quatre premiers capteurs sont visibles sur la figure, et sont visiblement atteints par les émissions de la route. Les deux capteurs les plus éloignés ne sont pas visibles, et sont apparemment situés dans une zone où les concentrations de fond prédominent, les émissions de la route étant négligeables.

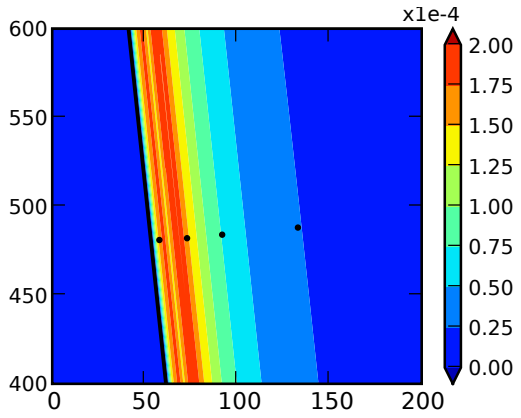


FIG. 3.11 – Concentrations moyennes au sol de cadmium ($\mu g\, m^{-3}$) sur 19 jours de février 1997. La ligne noire est le centre de la route, les points noirs sont les capteurs.

Comparaison des flux de dépôt Polyphemus et ADMS aux observations

Le flux de dépôt est calculé via la formule 2.94, auquel est rajouté le flux de dépôt humide calculé par l'équation 2.113. Les valeurs sont moyennées sur l'ensemble des jours simulés, et comparées aux observations tirées de Promeyrat-Qotbi [2001].

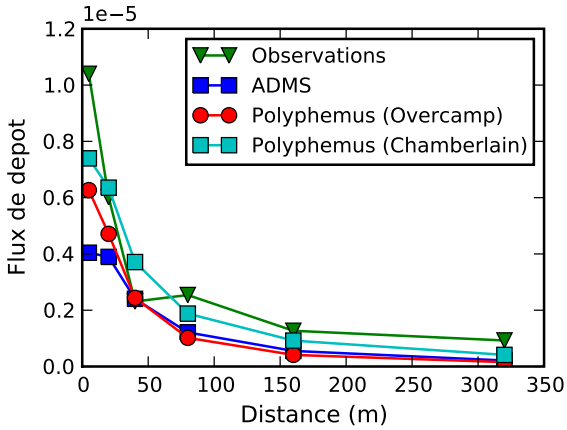


FIG. 3.12 – Flux de dépôt moyen de cadmium ($\mu g\, m^{-2}\, s^{-1}$) sur 19 jours de février 1997, simulé par Polyphemus et ADMS, et observé.

La figure 3.12 montre la comparaison entre les flux de dépôt simulés par Polyphemus et ADMS, et observés. Les simulations sont toutes les deux très proches des observations, avec une meilleure estimation donnée par Polyphemus au capteur le plus proche de la route, par rapport à ADMS. La modélisation du dépôt sec par la formulation de Chamberlain donne des valeurs un peu plus élevées que la formule d'Overcamp. En effet, dans le cas de Chamberlain, on suppose que tout le panache s'appauvrit sur la verticale et se dépose instantanément, alors qu'Overcamp utilise des réflexions partielles, pour simuler un dépôt proche du sol. En général, cette dernière paramétrisation donne de meilleurs résultats à faible distance. A titre de comparaison, ADMS modélise un appauvrissement dans tout le panache, avec un facteur de forme de façon à avoir un dépôt plus grand au sol qu'en hauteur. Les corrélations correspondantes (entre simulations

et observations) sont de 96% pour Polyphemus avec Overcamp, 93% avec Chamberlain, contre 88% pour ADMS.

Les deux modèles sous-estiment les valeurs aux capteurs éloignés de la route, étant donné que les simulations n'ont pas pris en compte les valeurs de fond, mais uniquement les émissions de la route. Le flux de dépôt mesuré par le capteur le plus éloigné de la route est égal à 9.25 $10^{-7}\mu g\,m^{-2}\,s^{-1}$. Si l'on considère que l'impact de la route est négligeable (figure 3.11) à cette distance, on peut prendre comme valeur de référence pour le fond une estimation de 9.10^{-7} $\mu g\,m^{-2}\,s^{-1}$. En ajoutant cette valeur aux flux de dépôt simulés par Polyphemus et ADMS, on obtient de très bons résultats (figure 3.13). Dans ce cas, les résultats modélisés avec le dépôt d'Overcamp sont plus proches des observations que les résultats donnés par Chamberlain. Cependant, les valeurs de concentrations de fond ont été prises à titre d'illustration et sont tout à fait incertaines. Elles sont probablement un peu trop élevées si l'influence de la route à cette distance a été sous-estimée.

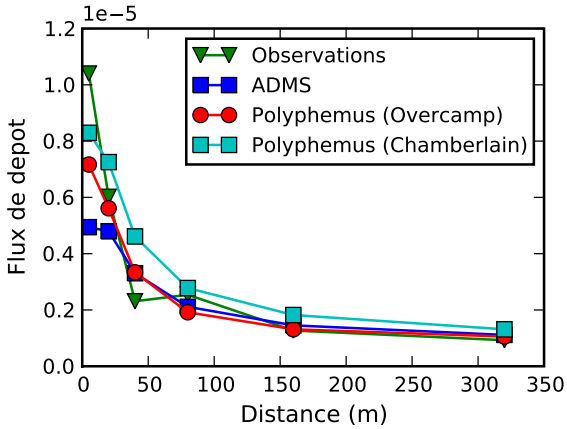


FIG. 3.13 – Flux de dépôt moyen de cadmium ($\mu g\,m^{-2}\,s^{-1}$) sur 19 jours de février 1997, simulé par Polyphemus (Overcamp et Chamberlain), ADMS, et observé. Les valeurs simulées sont corrigées avec une valeur de fond de 9.10^{-7} $\mu g\,m^{-2}\,s^{-1}$.

La figure 3.14 montre les mêmes résultats que la figure 3.13, mais en utilisant également la modélisation des sources linéiques. Celle-ci donne bien un résultat identique à la source discrétisée : cette modélisation permet essentiellement d'abaisser le temps de calcul, qui correspond à une seule source et non plus 200 sources. Sur certains jours, cependant, les résultats de ces deux modélisations diffèrent légèrement [Briant, 2009].

3.4.3 Analyse des résultats sur certains jours

Nous étudions maintenant les simulations ADMS et Polyphemus pour deux jours particuliers : le 3 et le 18 février. Le 3 février correspond à une situation de faible dispersion sans pluie, le 18 février à une dispersion normale avec pluie. Ces comparaisons ont été effectuées avec la formulation d'Overcamp pour le dépôt.

Résultats pour le 3 février

La figure 3.15 donne les résultats pour la journée du 3 février. Les résultats de Polyphemus et d'ADMS sont très proches, montrant tous les deux une valeur plus faible au premier capteur (le plus proche de la route), qu'au deuxième. Cela est dû au fait que la source est située à environ 1 mètre du sol, et, dans ce cas particulier de faible dispersion, le panache n'a pas encore

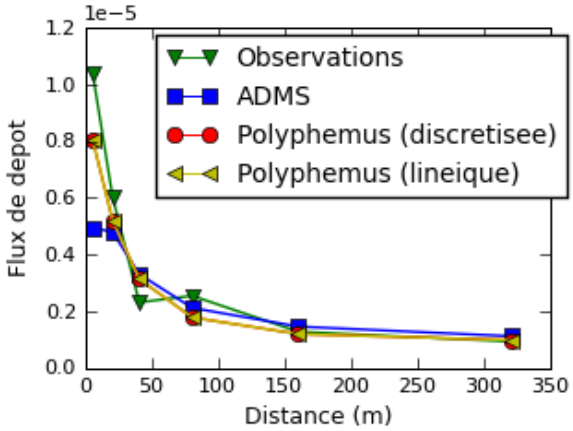


FIG. 3.14 – Flux de dépôt moyen de cadmium ($\mu g\, m^{-2}\, s^{-1}$) sur 19 jours de février 1997, simulé par Polyphemus (source discrétisée et source linéique), ADMS, et observé. Dépôt simulé par la formule d'Overcamp. Les valeurs simulées sont corrigées avec une valeur de fond de 9.10^{-7} $\mu g\, m^{-2}\, s^{-1}$. Crédit : Régis Briant.

complètement touché le sol au niveau du premier capteur. Le fait que les résultats des deux modèles présentent cette caractéristique confirme a posteriori le choix de la hauteur d'émission expliqué en partie 3.4.1 : la hauteur effective calculée par ADMS doit être très proche de 1 m. A titre de comparaison, la figure 3.16 montre les résultats dans la même configuration, mais avec une hauteur de source égale à 0.5 m. Dans ce cas, la valeur simulée au premier capteur est bien supérieure à la valeur du deuxième capteur, car le panache a déjà touché le sol.

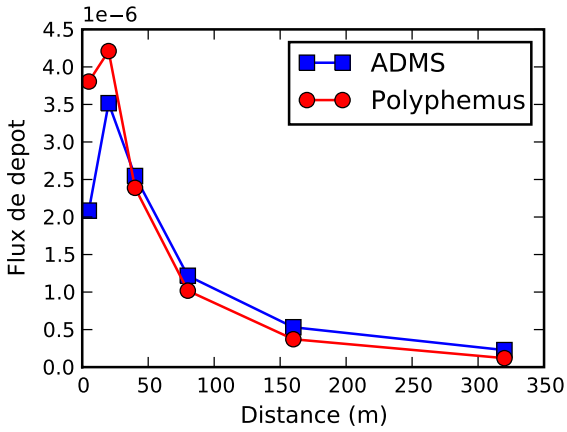


FIG. 3.15 – Flux de dépôt de cadmium ($\mu g\, m^{-2}\, s^{-1}$). Moyenne sur la journée, pour le 3 février 1997, simulé par Polyphemus et ADMS. La hauteur d'émission dans Polyphemus est prise à 1 m.

Ce résultat est important, car c'est au premier capteur que les différences entre les deux modèles, et avec les observations, sont notables. Ainsi, l'incertitude principale dans cette étude provient de la hauteur de la source, et non pas des paramétrisations.

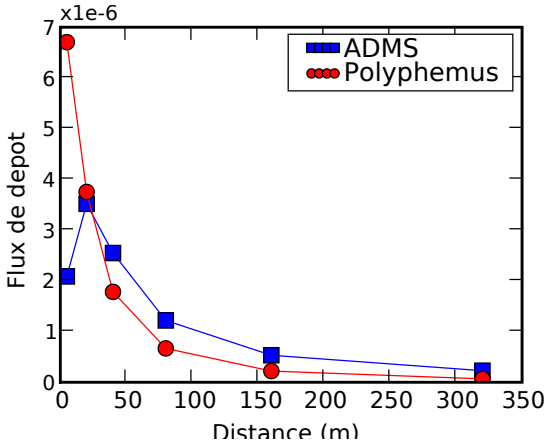


FIG. 3.16 – Flux de dépôt de cadmium ($\mu g\, m^{-2}\, s^{-1}$). Moyenne sur la journée, pour le 3 février 1997, simulé par Polyphemus et ADMS. La hauteur d'émission dans Polyphemus est prise à 0.5 m.

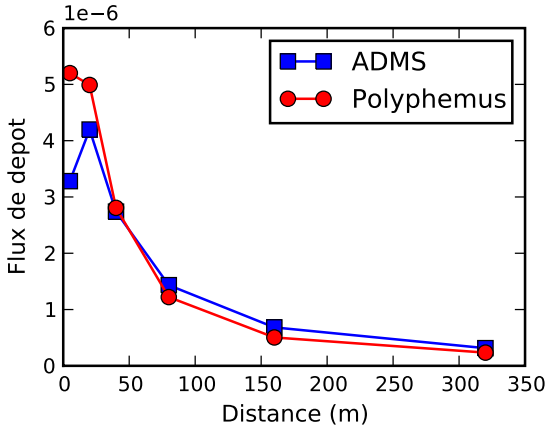


FIG. 3.17 – Flux de dépôt de cadmium ($\mu g\, m^{-2}\, s^{-1}$). Moyenne sur la journée, pour le 18 février 1997, simulé par Polyphemus et ADMS.

Résultats pour le 18 février

La figure 3.17 montre les flux de dépôt simulés pour la journée du 18 février. Il s'agit d'un cas de dispersion moyenne (vent de 7.3 $\mathrm{m\,s^{-1}}$) avec pluie (intensité de pluie $p_0 = 7.4\ \mathrm{mm\,hr^{-1}}$). Les résultats sont, là encore, très similaires pour les deux modèles.

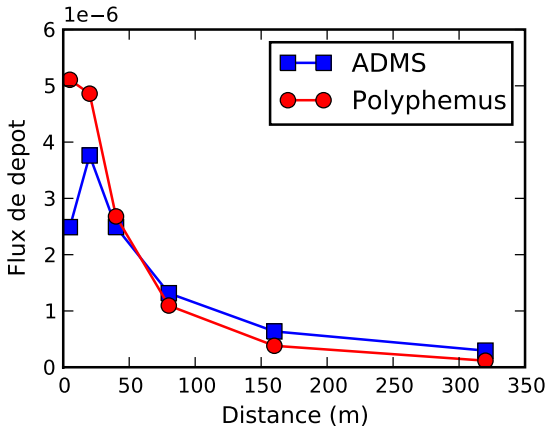


FIG. 3.18 – Flux de dépôt sec de cadmium ($\mu\mathrm{g\,m^{-2}\,s^{-1}}$). Moyenne sur la journée, pour le 18 février 1997, simulé par Polyphemus et ADMS.

Étant donné qu'il s'agit d'un jour avec pluie, on peut comparer la contribution des flux de dépôt sec et humide. La figure 3.18 donne les dépôts secs. Comme dans les cas sans pluie, Polyphemus donne une valeur de dépôt sec un peu plus élevée pour les capteurs les plus proches, probablement due en partie à l'utilisation de la formule d'Overcamp.

La figure 3.19 donne les flux de dépôt humides. Le flux de dépôt humide calculé par Polyphemus est plus faible que celui d'ADMS aux premiers capteurs, et quasiment constant égal à 1 $10^{-7}\mu\mathrm{g\,m^{-2}\,s^{-1}}$. En comparaison, le flux de dépôt humide moyen d'ADMS est égal à 2.7 $10^{-7}\mu\mathrm{g\,m^{-2}\,s^{-1}}$. Dans les deux cas, le flux de dépôt humide est inférieur d'environ un facteur 10 au flux de dépôt sec (de moyenne 2 $10^{-6}\mu\mathrm{g\,m^{-2}\,s^{-1}}$). Les coefficients de lessivage calculés par ADMS pour ce jour sont égaux à 0.36 $10^{-3}\mathrm{s^{-1}}$ pour tous les diamètres, alors que les coefficients calculés par la paramétrisation de Slinn sont égaux à 2.6 $10^{-3}\mathrm{s^{-1}}$, 0.74 $10^{-3}\mathrm{s^{-1}}$ et 0.26 $10^{-3}\mathrm{s^{-1}}$ respectivement pour les particules de diamètres 10 µm, 2.5 µm et 1 µm.

Les différences entre les deux simulations peuvent donc s'expliquer par une façon différente de calculer les coefficients de lessivage, et également peut-être dans le calcul du flux de dépôt total. Dans tous les cas, l'influence du dépôt humide sur le flux de dépôt total est faible, et négligeable si l'on considère les résultats moyens sur le mois.

L'évaluation de la modélisation du dépôt dans les modèles gaussiens de Polyphemus est satisfaisante. Cependant, étant donné le faible nombre d'observations utilisées, et le fait que les valeurs comparées soient des moyennes sur le mois, une étude détaillée de sensibilité aux paramétrisations, comme pour les écart types et la surhauteur, n'a pas été menée ici. L'influence des modèles de dépôt, ainsi que des paramétrisations, est faible dans le cas étudié, au regard de l'incertitude sur la hauteur de la source.

3.5 Bilan : évaluation des modèles gaussiens et variabilité

Dans ce chapitre, les modèles gaussiens de Polyphemus ont été évalués sur différentes expériences, et notamment les cas de Prairie Grass et de Kincaid. La variabilité dans les résultats

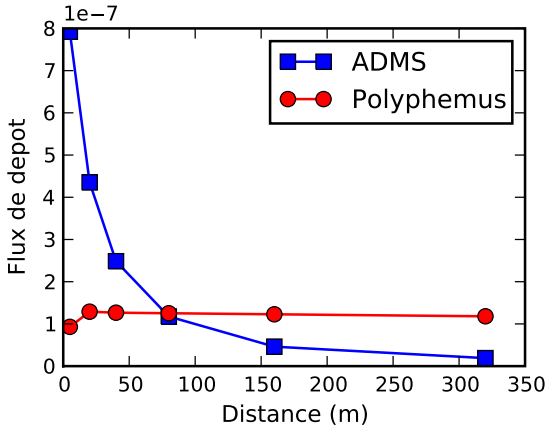


FIG. 3.19 – Flux de dépôt humide de cadmium ($\mu g\, m^{-2}\, s^{-1}$). Moyenne sur la journée, pour le 18 février 1997, simulé par Polyphemus et ADMS.

des modèles, due à l'incertitude sur les différentes paramétrisations empiriques, a été mise en évidence.

3.5.1 Performance des modèles

Dans l'ensemble, la comparaison modèle-mesures donne des résultats satisfaisants, en comparaison avec les performances d'autres modèles gaussiens connus et largement utilisés tels qu'ADMS et AERMOD. La meilleure paramétrisation des écarts types est la théorie de la similitude, en cohérence avec le fait qu'elle repose sur une description plus précise de l'atmosphère que les deux autres schémas. Les formules de Briggs donnent également de très bons résultats, non seulement pour Prairie Grass (expérience qui a servi dans l'élaboration des formules d'interpolation de Briggs) mais également pour Kincaid. Seule la paramétrisation de Doury donne des performances moins bonnes sur les deux cas présentés ici, avec une tendance à la sous-estimation des concentrations, probablement due à la sur-estimation de la dispersion du panache, notamment horizontale. Pour les cas étudiés ici, il n'y a pas de différence flagrante entre les modèles « ancienne génération », basés sur une représentation discrète de l'atmosphère dans la formulation des écarts types (Briggs et Doury pour Polyphemus, Pasquill-Guifford pour ISCST3), et les modèles « nouvelle génération », basés sur la théorie de la similitude (dont AERMOD et ADMS). Cette dernière approche a des performances légèrement meilleures. La comparaison devrait naturellement être étendue à davantage de cas, pour représenter davantage de situations météorologiques et de types de terrain.

Les résultats de l'expérience de Kincaid montrent l'importance de la modélisation de la surhauteur. En particulier, la meilleure performance n'est pas obtenue avec les mêmes formules d'écarts types suivant la paramétrisation utilisée pour calculer la surhauteur (théorie de la similitude avec la surhauteur de Briggs-HPDM, et écarts types de Briggs pour les surhauteurs de Holland et Concawe). Cela peut provenir, en partie, de compensations d'erreurs entre les formules utilisées pour la surhauteur et pour les écart types. Toutes les combinaisons entre les formules d'écarts types et de surhauteur ont des performances acceptables, comparables aux autres modèles, excepté celles utilisant les formules de Doury. Par ailleurs, ces résultats montrent également l'importance de la modélisation des interactions avec la couche d'inversion. Une meilleure modélisation du panache dans des conditions convectives, tenant compte de la

forme non gaussienne sur la verticale due aux mouvements ascendants et descendants de l'air, pourrait également améliorer les résultats sur ce genre d'expériences. L'utilisation d'un prétraitement météorologique recalculant la hauteur d'inversion, à l'instar de nombreux autres modèles, pourrait également être envisagée.

3.5.2 Variabilité des paramétrisations et des résultats

La partie 3.2.1 montre une grande variabilité des estimations des écarts types en fonction des paramétrisations choisies et des conditions météorologiques, jusqu'à un facteur 30 entre les valeurs obtenues. La variabilité de la surhauteur, estimée pour le cas de Kincaid en partie 3.3.2, n'est pas aussi grande, bien qu'il y ait des écarts importants pour de faibles vitesses de vent. L'utilisation de la surhauteur introduit une variabilité supplémentaire dans les résultats, en particulier dans les cas de source en hauteur où le panache peut pénétrer totalement ou partiellement dans la couche d'inversion.

Au final, l'utilisation de l'ensemble des paramétrisations disponibles pour les écarts types et la surhauteur, toutes choses étant par ailleurs égales dans la configuration du modèle, permet de retrouver une grande variabilité des résultats en sortie de modèle. En particulier, la variabilité obtenue avec les simulations de Polyphemus sur l'expérience de Kincaid est comparable à celle constatée avec l'utilisation d'un nombre important de modèles de la communauté. Cela souligne l'importance de l'incertitude sur les formules empiriques utilisées pour la modélisation des écarts types et de la surhauteur. De plus, cela suggère dans l'avenir l'utilisation d'une modélisation d'ensemble effectuée grâce aux différentes paramétrisations disponibles dans le système de modélisation.

Deuxième partie

Modélisation sous-maille des émissions ponctuelles : modèle de panache sous-maille

Chapitre 4

Modèle de panache sous-maille passif

Dans les modèles eulériens classiques, l'un des principaux processus sous-maille mal résolus par le modèle sont les émissions, et en particulier les sources ponctuelles. En effet, le panache émis par une source ponctuelle n'atteint pas la dimension horizontale d'une maille avant un certain temps, tandis que le modèle eulérien considère que tout émission est immédiatement diluée dans la maille. La méthode de panache sous-maille consiste à coupler un modèle gaussien à un modèle eulérien, afin de traiter en sous-maille les émissions ponctuelles. Au-delà d'un certain temps, lorsque le panache est suffisamment dispersé pour que l'erreur commise par le modèle eulérien ne soit pas trop importante, le traitement du panache est transféré au modèle eulérien. Cette méthode doit permettre de mieux représenter les émissions ponctuelles, sans augmenter le temps de calcul de façon significative.
Ce chapitre décrit le modèle de panache sous-maille couplant le modèle à bouffées présenté dans les chapitres précédents, et le modèle eulérien Polair3D, au sein de la plate-forme Polyphemus. Le modèle de panache sous-maille est appliqué à deux cas de rejet passif à l'échelle continentale, l'expérience ETEX-I et l'accident de Tchernobyl. Le but de cette étude est, d'une part, d'évaluer les performances du modèle de panache sous-maille par rapport à un modèle eulérien classique, mais également de réaliser une étude de sensibilité à différents paramètres. En particulier, l'influence des paramétrisations pour les écarts types gaussiens est estimée. Il s'agit enfin de donner un aperçu des échelles spatiales et temporelles en jeu dans l'utilisation d'un modèle de panache sous-maille, et des processus physiques impliqués.
Outre une réduction de la diffusion horizontale, l'utilisation d'un modèle de panache sous-maille pour une source proche du sol, émettant un traceur passif, dans des conditions météorologiques normales (stabilité neutre), a pour principale conséquence de mieux représenter la diffusion verticale aux alentours des sources ponctuelles. Le biais est ainsi amélioré de façon significative aux stations proches de la source, et jusqu'à quelques centaines de kilomètres. La sensibilité aux paramétrisations locales est importante, ce qui montre que la première phase de développement du panache a des répercussions sur de longues distances. Pour le cas d'ETEX-I, la paramétrisation de Doury et la méthode d'injection des bouffées sur une colonne eulérienne donnent les meilleurs résultats. Enfin, une étude de sensibilité sur la résolution du maillage, sans changer les données d'entrées, est également réalisée. On montre ainsi que l'utilisation du panache sous-maille est plus intéressante pour un maillage relativement fin, car elle compense la surestimation des concentrations induite par le faible volume d'émission dans le modèle eulérien.

Sommaire

4.1 Introduction **111**
4.1.1 Context 111
4.1.2 Issues 111
4.1.3 Methodology 112
4.1.4 Outline 112
4.2 Model description **112**
4.2.1 The Gaussian puff model 113
4.2.2 Model coupling 115
4.3 The ETEX case: impact of plume-in-grid at continental scale **117**
4.3.1 Modeling set-up 117
4.3.2 Convergence with the time step between two puffs 118
4.3.3 Comparison of results with/without the plume-in-grid approach 118
4.4 Sensitivity to the local-scale dispersion **122**
4.4.1 Local-scale dispersion parameters 123
4.4.2 Comparison between the Gaussian and Eulerian diffusion 123
4.4.3 Vertical profiles 125
4.5 Relevant timescales: a study on the injection time **125**
4.5.1 Impact on the concentrations 125
4.5.2 Impact on the arrival times 127
4.6 Influence of the grid resolution **127**
4.6.1 Impact of the initial volume of the emission 128
4.6.2 Use of plume-in-grid with several grid resolutions 130
4.7 Application to the Chernobyl disaster **130**
4.7.1 Modeling set-up 130
4.7.2 Results 133
4.8 Conclusions **134**
4.9 Appendix **136**
4.9.1 Coordinates 136
4.9.2 Formulae to compute the puff integral over a volume 136

Ce chapitre est constitué de Korsakissok et Mallet [2010b].

4.1 Introduction

4.1.1 Context

Classical air quality models at regional scale, based on Eulerian approaches, overlook subgrid-scale phenomena. In particular, emissions from major point sources are badly represented by Eulerian models since they are usually assumed to mix immediately within a grid cell, whereas a typical point-source plume (e.g., from a power plant) does not expand to the size of the grid cell for a substantial time period. This leads to an unrealistic near-source modeling, especially since the K-theory approach often used in Eulerian models does not properly represent the diffusion in the vicinity of the source [Maryon et Buckland, 1995]. Numerical problems can also be raised, such as oscillations caused by strong gradients due to point sources [Brandt *et al.*, 1996], if a non-monotonic advection scheme is used. On the other hand, Gaussian models provide a better representation of the near-source dispersion, but they rely on assumptions (e.g., no wind shear) that lead to increasing errors on the plume trajectory with distance [Stohl, 1998].

The plume-in-grid method is a multiscale modeling technique that couples a Gaussian (plume or puff) model with an Eulerian model in order to improve the treatment of point source emissions. This coupling method has been developed and evaluated both for photochemical applications (e.g., Seigneur *et al.* [1983], Morris *et al.* [1991], Kumar et Russell [1996], Byun et Schere [2006], Karamchandani *et al.* [2002]), and for the modeling of accidental releases with the DREAM model [Brandt, 1998].

4.1.2 Issues

This paper is aimed at investigating some issues of this multiscale modeling approach.

Physical processes The theoretical reasons why such a coupling would improve the near-source dispersion are well known, but we seek to have an insight on the key physical processes (e.g., horizontal or vertical diffusion) for which a subgrid-scale treatment is most relevant. One may also wonder whether the largest impact would show on the plume travel time, or on the concentrations.

Local-scale parameterizations Another question is the sensitivity of both the local-scale and the large-scale models to the input parameters and the dispersion schemes. While the sensitivity to the Eulerian parameterizations has already been investigated [Brandt *et al.*, 1998], the impact of the local-scale parameterizations on the plume-in-grid results was not, to the authors' knowledge, assessed. Here, the sensitivity to the Gaussian dispersion is highlighted, and compared to the impact of the Eulerian vertical diffusion schemes.

Relevant time and spatial scales It is also crucial to estimate the relevant spatial and time scales for both models. The "local scale" is the scale at which the use of a Gaussian model is relevant. It is the near-source area where the dispersion is not well represented by the Eulerian model, and the Gaussian model error due to trajectory uncertainties and wind shear is not too large. When the improvements in the dispersion are outweighed by the uncertainties in the plume trajectory, the plume must be transferred from the Gaussian model to the Eulerian model. This injection is decided according to some criterion on the puff age or size. For instance, in several photochemical plume-in-grid models, the transfer is done when the puff size is about the Eulerian-grid size at the transfer time. In DREAM, the Gaussian model is used on a nested subdomain, whose size is determined so that the puffs would be larger than the Eulerian-cell size at the interface—where the puffs are transferred into the Eulerian model. These conditions

are physically consistent with the primary goal of plume-in-grid models, that is, to minimize the artificial horizontal dilution. However, this does not prevent large errors in the plume trajectory in the case of very coarse grids. Alternatively, some models use a criterion on the plume travel time, usually 1 hour (e.g. in Kumar et Russell [1996]), but no sensitivity study is available on the subject. This issue is investigated here by means of a sensitivity study on the so-called "injection time" for a wide range of values. A comparison is also made between the time and size criteria, at several resolutions.

Relevant resolution Finally, we investigate whether the use of the plume-in-grid modeling technique is more relevant at a low or high resolution. In theory, a low resolution would imply more artificial dilution in the Eulerian model, hence more need for a subgrid-scale representation of the plume. In practice, using fine resolutions do not always improve simulation results, especially when no high-resolution input data is used. Thus, it is useful to have an insight on that subject, and to issue, if possible, general recommendations about the resolutions and corresponding injection times.

4.1.3 Methodology

These issues are investigated using the plume-in-grid model developed on the air quality modeling system Polyphemus [Mallet *et al.*, 2007]. This model was developed both for photochemistry and passive (chemically inert) tracers, at regional and continental scale. The aim is to provide an easy-to-use and modular model that is appropriate for a wide range of applications and modeling domains, with limited computational demands. The model can be used in the context of operational applications where detailed meteorological fields are not always available, and computational-time constraints do not allow high-resolution or adaptive grid simulations.

The model evaluation is carried out for passive-tracer releases with ETEX-1 (European Tracer EXperiment, first release) and Chernobyl cases. The ETEX campaign provides an appropriate case to carry out a detailed study, since it is well documented and the source parameters are well known, thus reducing the uncertainties. The plume-in-grid model, with the recommendations inferred from the ETEX sensitivity study, is then evaluated on the Chernobyl case.

4.1.4 Outline

Section 4.2 describes the model, with an emphasis on the local-scale model and the coupling methods. In Sections 4.3, 4.4, 4.5 and 4.6, the aforementioned issues (Section 4.1.2) are investigated on the ETEX case: Section 4.3 analyzes the results with/without plume-in-grid, Section 4.4 discusses the sensitivity to the local-scale dispersion parameters, compared to the Eulerian diffusion schemes, Section 4.5 details the model sensitivity to the puff injection time and Section 4.6 gives an insight on the influence of the grid resolution. Finally, Section 4.7 gives the overall model results on the Chernobyl case.

4.2 Model description

The plume-in-grid model presented here couples, on the Polyphemus platform, the Gaussian puff model [Korsakissok et Mallet, 2009] with the Eulerian model Polair3D [Boutahar *et al.*, 2004]. In Section 4.2.1, we briefly present the Gaussian puff model, and the necessary adjustments made to the Gaussian dispersion for the non-stationary case. Section 4.2.2 describes the method used to couple the two models.

4.2.1 The Gaussian puff model

Model presentation

The Gaussian puff model represents a continuous point emission as a series of Gaussian puffs. Each puff transports a given quantity of the emitted species; the speed and the direction of a puff are determined by the wind at its center. The puff size increases with turbulence, and it is determined by the Gaussian standard deviations in all three directions: σ_x (downwind), σ_y (crosswind) and σ_z (vertical). The concentration at one point (x, y, z) is then given by the sum of all the puffs' contributions:

$$C(x,y,z) = \sum_{i=1}^{N_{\text{puff}}} \frac{Q_s \times \Delta t_{\text{puff}}}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z} \exp\big(-\frac{(x-x_c^i)^2}{2\sigma_x^2}\big) \times \exp\big(-\frac{(y-y_c^i)^2}{2\sigma_y^2}\big) \exp\big(-\frac{(z-z_c^i)^2}{2\sigma_z^2}\big) \tag{4.1}$$

where N_{puff} is the number of puffs, Δt_{puff} is the time step between the release of two consecutive puffs, and (x_c^i, y_c^i, z_c^i) are the coordinates of the center of the i-th puff. Q_s is the source emission rate (in mass unit per second). Equation 4.1 can be modified to take into account the reflections on the ground and inversion layer if necessary. The continuous plume is well represented by the series of puffs if there is a sufficient overlap between two successively released puffs, that is, if Δt_{puff} is small enough. In practice, this condition is fulfilled at time t, for two successive puffs emitted at t_i and $t_{i+1} = t_i + \Delta t_{\text{puff}}$ respectively, if:

$$\sigma_x(t-t_i) + \sigma_x(t-t_{i+1}) \geq u\Delta t_{\text{puff}}, \tag{4.2}$$

where u is the wind speed at time t and at the puffs' locations, and the puffs' sizes ("radii") are assumed to be $\sigma_x(t-t_i)$ and $\sigma_x(t-t_{i+1})$ in the x direction. This criterion can be relaxed during the first time steps, when no measurement station has been reached: the overlap condition has to be fulfilled at the first station. Section 4.3.2 discusses this condition in the present case.

In Polyphemus, three empirical parameterizations may be used to compute puffs standard deviations: Briggs's, Doury's and similarity-theory. They are briefly described below, and more details can be found in Korsakissok et Mallet [2009], along with a validation against the Prairie Grass and Kincaid data sets.

Briggs's formulae The Briggs formulae are based on the Pasquill-Turner stability classes [Turner, 1969], and are different for rural and urban areas. The full formulae can be found in Arya [1999] for instance. The general form is given by:

$$\sigma_y = \frac{\alpha\ x}{\sqrt{1+\beta\ x}}, \qquad \sigma_z = \alpha\ x\ (1+\beta\ x)^{\gamma}, \tag{4.3}$$

with x the downwind distance from source, and α, β and γ coefficients depending on Pasquill stability class.

Doury's formulae An alternative parameterization, developed for the specific application of radionuclides dispersion, is described in Doury [1976]. The formulae use only two stability situations, corresponding to normal and low dispersion. The standard deviations are given in the general form:

$$\sigma_y = (A_h t)^{K_h}, \qquad \sigma_z = (A_z t)^{K_z}, \tag{4.4}$$

where t is the puff "age" (time since the puff release), and A_h, A_z, K_h and K_z are coefficients depending on the stability and the puff time.

Similarity theory If enough meteorological data is available, σ_y and σ_z can be estimated using the standard deviations of wind velocity fluctuations in horizontal direction σ_v and in vertical direction σ_w. Following Irwin [1979] dispersion coefficients are investigated in the form:

$$\sigma_y = \sigma_v \ t \ F_y, \qquad \sigma_z = \sigma_w \ t \ F_z, \tag{4.5}$$

where t is the time since the puff release, and F_y and F_z are functions of a set of parameters that characterize the atmospheric boundary layer. These functions are determined from experimental data. Here, the expressions of F_y and F_z, as well as σ_v and σ_w, come from Irwin [1979], Hanna *et al.* [1982], Hanna [1984] and Weil [1988].

Dealing with non-stationarity: treatment for dispersion parameters

The parameterizations to compute the standard deviations assume that the meteorological situation is stationary and homogeneous, which is not the case in the plume-in-grid model. Hence the need to adapt the usual Gaussian formulations to non-stationary situations. The Gaussian standard deviation in one direction can be written in the general form:

$$\sigma = f(t), \tag{4.6}$$

where t is the puff travel time, and f is a function depending on the meteorological situation. Equation 4.6 gives the puff evolution assuming that f has been constant since the puff emission time. Using this formula in non-stationary cases, where f is evaluated at each time step with the local meteorological situation, could lead to the unrealistic situation where the puff size would decrease. This would be the case for instance if the situation changed from unstable to stable (e.g., if night has fallen), as illustrated by Figure 4.1.

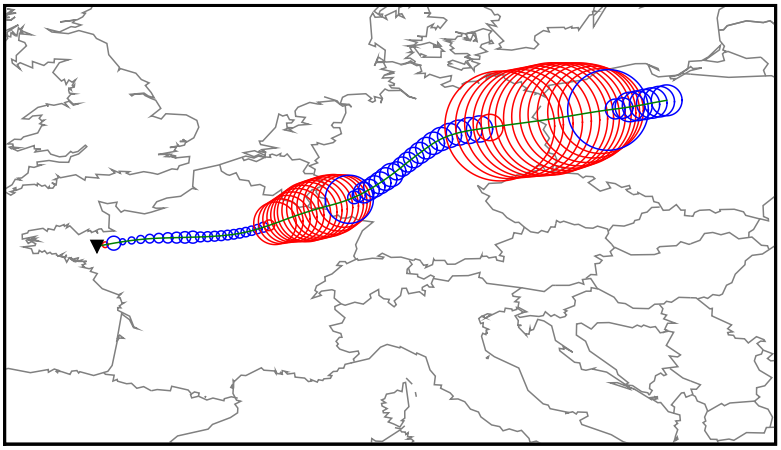


Figure 4.1: Unrealistic evolution of the size of one puff without correction on σ_y. The puff is represented at each time step, and the circle radius is proportional to σ_y. The puffs are drawn in red when it is daytime and in blue during nighttime. The green line is the puff center. The black triangle is the source location.

A physically consistent way to deal with this problem is to determine the puff growth time step per time step. The local growth of σ is perfectly determined by the value at the beginning of the time step (noted σ_1) and the knowledge of f (Equation 4.6) for the current situation.

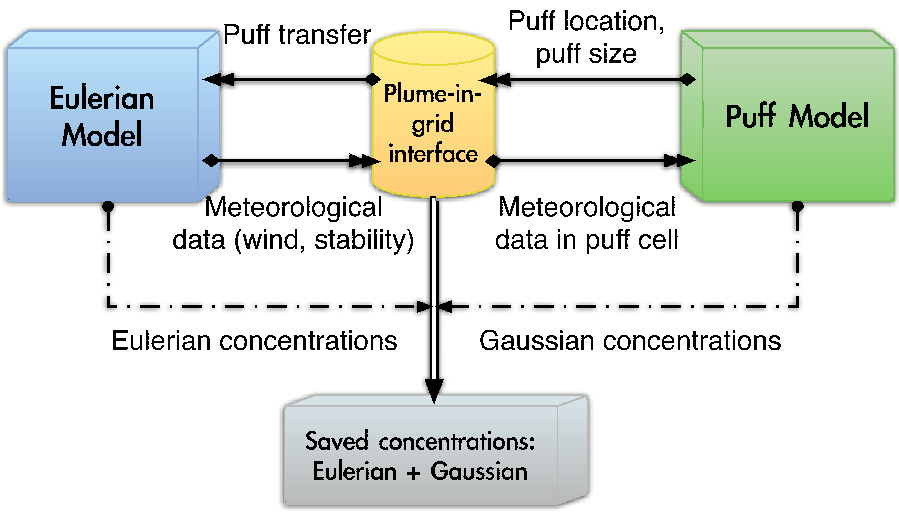


Figure 4.2: Description of the plume-in-grid coupling: information is exchanged between the Gaussian puff model and the Eulerian model. At each time step, the sum of the contributions from both models is saved.

Since f is an increasing function of time (for constant meteorological conditions), one may determine the virtual time t'_1 at which the standard deviation value is σ_1: $t'_1 = f^{-1}(\sigma_1)$. The standard deviation at the end of the time step can then be computed as $\sigma_2 = f(t'_1 + \Delta t)$. In some cases, however, the inverse function f^{-1} cannot be analytically computed; thus it is only ensured that the puff size cannot decrease during a time step ($\sigma_2 = \max(\sigma_1, f(t_1 + \Delta t))$).

4.2.2 Model coupling

In the plume-in-grid coupling, several point source emissions are treated by the Gaussian puff model while other sources, namely diffuse area emissions, are managed by Polair3D. When a puff has reached a given age or size, it is transferred into the Eulerian model. For an emission with a finite duration, the Gaussian puff model is not used anymore after the last puff transfer. During the time period when both models are used, they exchange some information at each time step (Figure 4.2):

Meteorological data They have been computed on the Eulerian grid and can be either interpolated at the puff center or taken at the center of the cell where the puff is located. They are then used by the Gaussian model, to carry out the puff advection and diffusion.

Puff data At each time step, the cell where the puff center resides is determined. For that purpose, the Cartesian coordinates used in the Gaussian puff model are converted into longitude/latitude coordinates used by Polair3D (Section 4.9.1). It is then checked whether the puff should be injected in the Eulerian model. If so, the puff is erased from the Gaussian puff model, and its mass is transferred into the Eulerian model (Section 4.2.2).

Puff injection criteria and method

The puff is transferred to the Eulerian model when one of the two following injection criteria is met:

1. The time period t_{puff} after the puff release (the puff age) has reached a given value, called the injection time and noted t_{inj}: the injection is applied if $t_{\text{puff}} \geq t_{\text{inj}}$,
2. The puff horizontal size has reached the cell size Δy: the puff is injected if

$$C_y \ \sigma_y \geq \Delta y, \tag{4.7}$$

where C_y is a coefficient set to 4.

The puff is then transferred, using either the column, or the integrated injection method.

Column injection The puff mass is equally distributed within $2\sigma_z$ of the puff center, within one column, bounded by the ground and by the boundary layer height. For a puff within the boundary layer, the puff center height is $z_c \leq z_i$, and the puff vertical extent is given by:

$$\min(2 \ \sigma_z, z_c) + \min(2 \ \sigma_z, z_i - z_c) \tag{4.8}$$

Using the column injection implies that the puff center is "shifted" to the cell center. It also adds some vertical dilution to the extent given in Equation 4.8: the total vertical extent actually is the sum of the heights of the vertical levels where the transfer is done (Figure 4.3).

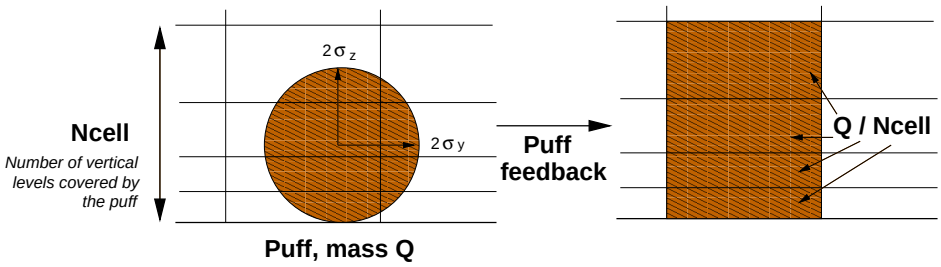


Figure 4.3: Column injection. The mass Q of the puff is equally distributed in the N_{cell} cells that it vertically covers. The puff vertical extent is 4 σ_z, unless it touched the ground or the inversion layer.

Integrated injection The puff mass is distributed over the cells within the puff horizontal or vertical extent (Figure 4.4). In each cell, the puff contribution is the integral of the Gaussian concentration over the cell volume, corrected to ensure the mass conservation (Section 4.9.2). While the column injection limits the artificial puff horizontal dilution, transferring a puff to several cells, as in the integrated-injection method, leads to a higher dilution volume.

Concentrations reconstructed at the measurement stations

Since the two models operate at the same time, the total concentrations are the sum of both models contributions (see Figure 4.2). Whereas the Eulerian model allows to estimate the mean concentrations in each grid cell, the Gaussian model gives an analytical formula (Equation 4.1) to compute the concentrations at the exact stations' locations. Therefore, for stations located in the local-scale domain—where the Gaussian model is used—the Gaussian contribution can either be analytically computed at the exact location, or averaged in the grid cells, then interpolated, as with the Eulerian model. Although the first way to reconstruct the concentrations relies on a more accurate representation of the Gaussian concentrations, it is more sensitive to the

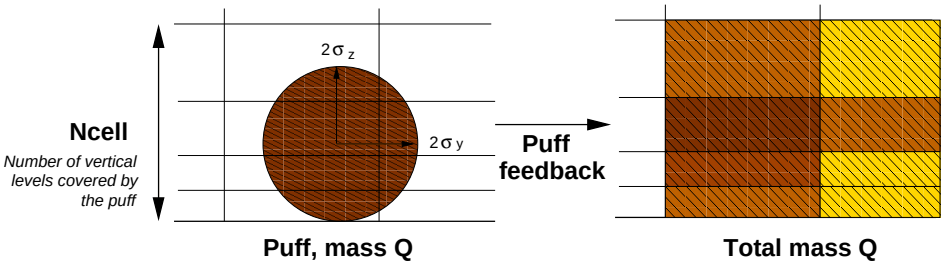


Figure 4.4: Integrated injection. The puff mass Q is added to each cell within the puff horizontal and vertical extent. The puff vertical extent is 4 σ_z (unless it touches the ground or the inversion layer), and the horizontal size is 4 σ_y.

uncertainty in the puff location. If the uncertainty is much larger than the puff horizontal size, the error at the station can be very large. A third option would be to average the concentrations on a volume whose extent would be related to the trajectory uncertainty and would be smaller than a grid cell. The results discussed hereafter are obtained by averaging the concentrations on the grid cells.

4.3 The ETEX case: impact of plume-in-grid at continental scale

The European Tracer Experiment (ETEX) campaign is a well-instrumented, dispersion experiment at continental scale (`http://rem.jrc.ec.europa.eu/etex/`). We model here the ETEX-I campaign, where 340 kg of a passive tracer (perfluoromethylcyclohexane, hereafter called PMCH) were continuously released during twelve hours. The source was located in west of France (48°03′N, 2°00′W, see Figure 4.1 for the source location). The emission started on October 23th, 1994, at 16:00 UTC. A total of 168 stations over 17 European countries measured the resulting plume, during more than one week after the release. This experiment was widely used to calibrate and compare atmospheric dispersion models (Nodop *et al.* [1998], Mosca *et al.* [1998]). Among the contributions on the subject, several were using Polyphemus with Polair3D: it was evaluated in Quélo *et al.* [2007], and used with inversion methods, e.g., in Bocquet [2005]. Lagrangian particle models were also evaluated on that case, for instance the NAME model [Ryall et Maryon, 1998], as well as Gaussian puff models [Sorensen, 1998] and the coupled model DREAM [Brandt *et al.*, 1998].

4.3.1 Modeling set-up

The configuration is similar to the simulation set-up used in Quélo *et al.* [2007]. The simulation grid covers all Europe and its cell width is 0.5625° in longitude and latitude. The source is located at the center of one grid cell. There are 12 vertical levels, up to 6090 m (the first level being at 30 m). The simulation starts at 00:00 UTC on October 23th, 1994, and the simulation time step is $\Delta t = 600$ s. The meteorological fields are the reanalyzed ECMWF fields ERA-40, with resolution 1.125° on the horizontal and with a 3-hour time step.

The release height was taken equal to 8 m. The observed plume rise was reported to be no more than 5 m, due to the effect of a significant buoyancy (the ejection temperature is 80 °C, the ejection velocity is 47.6 m s^{-1}) moderated by a small source diameter (2.5×10^{-2} m). The plume rise computed by the plume-in-grid model varies during the emission between 1 m and

4 m. In the Eulerian model (alone), the emission is naturally released at the center of the grid cell, whose height is 15 m. The emission rate is 7.98 $\mathrm{g\,s^{-1}}$.

The plume-in-grid simulations are carried out with the three Gaussian dispersion parameterizations described in Section 4.2.1, and the results are compared with those of Polair3D without the plume-in-grid treatment (hereafter called the "reference" results). Unless specified otherwise, the plume-in-grid results are shown with the column-injection method, which gave slightly better results than the integrated injection.

4.3.2 Convergence with the time step between two puffs

The plume-in-grid model discretizes the emitted plume into a series of puffs. As discussed in Section 4.2.1, the puffs overlap is important to ensure a physically consistent plume modeling. The puff standard deviation after 1 hour is about 1000 m (Section 4.4). The mean wind speed is about 5 $\mathrm{m\,s^{-1}}$. Thus, according to Equation 4.2, a reasonable time step between two puffs is about 400 s, to ensure an overlap at the first station. In comparison, in Sorensen [1998], the same condition is used to compute the puffs overlap, but with different parameterizations for σ_x. They advocated $\Delta t_{\mathrm{puff}} \leq 900$ s.

A sensitivity test was made on the time step between two puffs, to check that the impact of that time step on the simulation results is negligible (for small enough time steps). For that purpose, simulations are carried out with puff-time steps of 1, 5, 10, 30, 100, 200, 300 and 600 seconds. The simulation results are compared with the 1-second simulation, which ensures the best overlap. For all other time steps, the relative difference between the ground concentration and the 1-second simulation ground concentration, averaged in time, is computed in each cell (for concentrations above a threshold of 0.001 $\mathrm{ng\,m^{-3}}$). The maximum value of this indicator is computed for each time step (Figure 4.5). Below 200 s, the maximum difference is about 10% of the concentration value, and it is near-zero below 50 s. In the following, the time step between two puffs is taken equal to 5 s in order to ensure a sufficient puffs overlap with a reasonable computational time.

4.3.3 Comparison of results with/without the plume-in-grid approach

The model-to-data comparison is carried out with the following statistical indicators: fractional bias (FB), mean bias error (MBE), correlation (Corr) and proportion of values within a factor 2 and 5 of the observations (FAC2 and FAC5). The formulae are given in Appendix A. These statistics are computed for all stations with a minimum of 11 measurements, which amounts to 139 stations. The measured values are averaged over 3 hours. Measurements lower than a threshold of 0.05 $\mathrm{ng\,m^{-3}}$ are discarded, except when they occur within two time intervals (6 hours) of a value above the threshold.

Global results

Table 4.1 shows the statistics computed for the reference simulation and the plume-in-grid simulation (for the three local-scale configurations) with an injection time equal to 1 hour. The use of plume-in-grid improves the results no matter the dispersion parameterization used, reducing the overestimation of the concentrations (from a factor 3 to a factor 2). The correlation is also improved by the use of plume-in-grid with similarity-theory and Doury's parameterization, and decreases when using the Briggs formulae. The proportion of values within a factor of 2 and 5 of the observations are slightly improved.

Figure 4.6 shows how the use of plume-in-grid modifies the spatial distribution of the pollutant concentration. The difference between concentrations with and without plume-in-grid

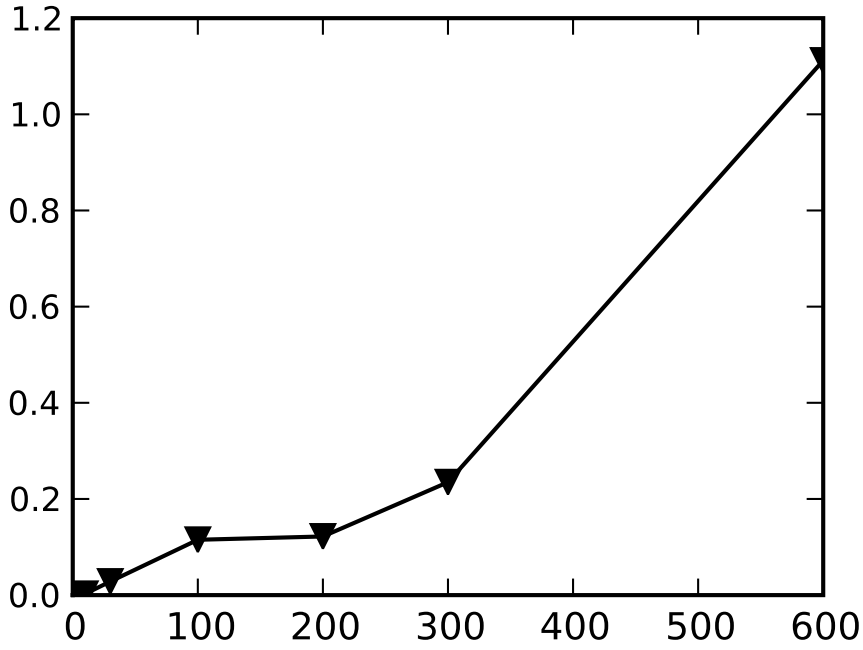


Figure 4.5: Convergence with the time step between two puffs. The reference is given by the simulation with the time step of 1 s. The indicator is the relative difference between the mean ground concentrations for one time step (between 600 s and 5 s), and the reference value (for 1 s).

Model	Mean	FB	MBE	Corr	FAC2	FAC5
Obs	0.21	0.00	0.00	1.00	1.00	1.00
Polair3D	0.68	1.07	0.48	0.60	0.19	0.36
Sim.th.	0.48	0.79	0.27	0.65	0.19	0.37
Doury	0.39	0.61	0.18	0.68	0.20	0.39
Briggs	0.43	0.70	0.22	0.51	0.22	0.39

Table 4.1: Statistics for the reference simulation (Polair3D), and plume-in-grid simulations with an injection time of 60 minutes, for all three parameterizations of the standard deviations. Column injection. "Sim.th." stands for similarity theory.

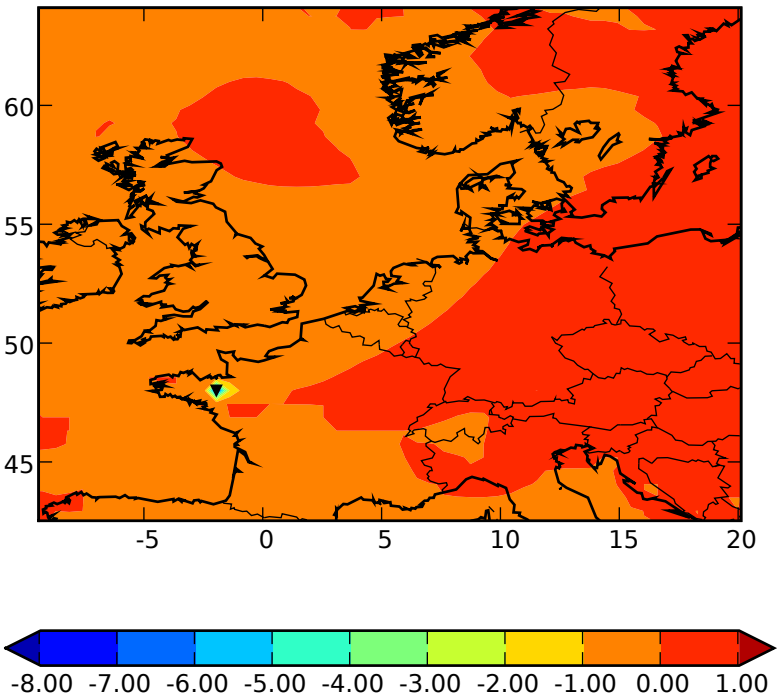
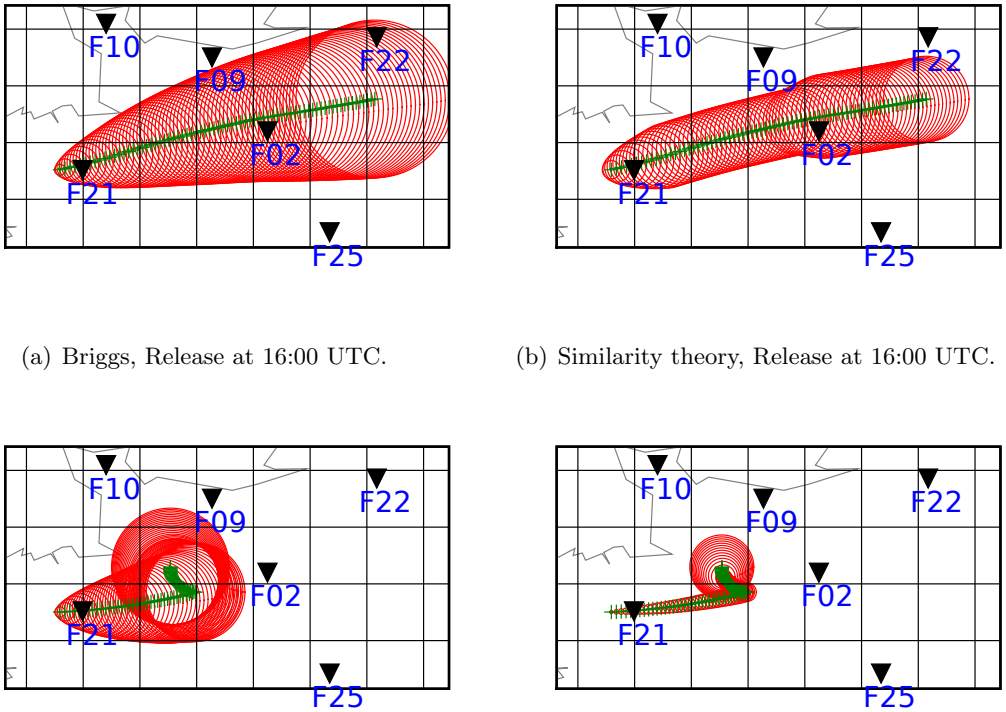


Figure 4.6: Difference between the mean ground concentrations (in $\mathrm{ng\,m^{-3}}$) averaged over the simulation period, between the plume-in-grid results and those of the reference simulation. The plume-in-grid simulation was carried out with similarity theory, an injection time of 1 hour and a column injection.



(a) Briggs, Release at 16:00 UTC.

(b) Similarity theory, Release at 16:00 UTC.

(c) Briggs, Release at 22:00 UTC.

(d) Similarity theory, Release at 22:00 UTC.

Figure 4.7: Puff trajectory during 12 hours. Green: puff center. Red: puff radius. The circle radius is $2 \times \sigma_y$, and one circle is drawn at each time step (ten minutes). Black triangles are the measurement stations. Top: puff emitted at 16:00 UTC. – Bottom: puff emitted at 22:00 UTC. – Left: Briggs parameterization – Right: similarity theory.

(with similarity theory, 1 hour injection time) is computed on ground concentrations, averaged over the whole simulation period. The difference is very high in the vicinity of the source, where the plume-in-grid concentrations are up to 80 % lower than the reference concentrations. At farther distance, the concentrations are much lower, and so are the differences. Thus, using the plume-in-grid model mostly modifies the concentrations at a few stations in the vicinity of the source, where the emission is treated by the Gaussian puff model. Reducing the bias at these stations—where the concentrations are the highest—has a significant impact on the overall bias. On the contrary, FAC2 and FAC5 show a smaller improvement, since they are less dependent on the first stations.

Results on local-scale stations

Figure 4.7 shows a puff's trajectory—and typical size—during 12 hours, for Briggs' and similarity-theory parameterizations, for a puff emitted at 16:00 UTC (figures 4.7(a) and 4.7(b)), and a puff emitted a 22:00 UTC (figures 4.7(c) and 4.7(d)). For all parameterizations, the puffs directly impact one station: Rennes (F21), which is reached between 40 and 60 minutes after the puff release. The earlier emitted puffs (between 16:00 UTC. and 20:00 UTC), as well as those emitted during the last hour of emission (not shown here) have a straightforward trajectory,

Station	MBE	NMSE	Corr	FMT	FAC5
F21	-23.5/-5.0	47.7/8.5	0.69/0.70	0.06/0.22	0/0.18
F02	-3.09/-1.26	12.61/4.25	0.83/0.78	0.16/0.31	0.27/0.33
F22	-1.23/-0.52	13.23/3.97	0.85/0.89	0.15/0.30	0.14/0.21
F19	-1.27/-0.75	5.11/2.34	0.92/0.92	0.29/0.41	0.37/0.50

Table 4.2: Statistics at stations located within 350 km from the source. Statistics are given in this order: reference value / plume-in-grid value. Plume-in-grid values are given for the Doury parameterization and a 60-minute injection time.

and the plume also impact Alençon (F02). However, only the station F21 can be considered as a "local-scale" station, reached before the puffs are injected into the Eulerian model.

The statistics at particular stations give an insight on the scope of the plume-in-grid impact on statistics. We focus here on the "near-source" stations, located within 350 km of the release location, where a high enough number of above-threshold measurements allows to compute the statistics. Most of these stations can be seen Figure 4.7 (except station F19 which is south-east from F22, within the main plume trajectory). Table 4.2 shows the statistics for these stations: the correlation, the MBE, the normalized mean square error (NMSE), the figure of merit in time (FMT), and the FAC5 (see Appendix A). The plume-in-grid results are shown for the Doury parameterization and a 1-hour injection time. As expected, the higher differences are seen at Rennes, where concentrations are highest. All other stations show a significant improvement, especially in terms of NMSE, FMT and FAC5. Thus, the use of plume-in-grid modeling is relevant at continental scale, even when few local-scale stations are available: the impact is carried for a few hundreds of kilometers, even though the Gaussian model is used on a much smaller domain. The impact of plume-in-grid at farther stations with lower concentrations is much less significant (not shown).

In Section 4.2.2, it was pointed out that the concentrations at local-scale stations (here, Rennes) could also be computed with the analytical Gaussian formula, but this leads to highly overestimated values. As explained, this may come from trajectory errors (due to the wind direction modeling, especially time and space interpolation). If the error in the wind direction is 15% of the traveled distance [Stohl et Koffi, 1998], it represents about 3 km when the station is reached, while the puff horizontal "radius" is about 1 km. Moreover, a part of the plume in the upper vertical levels might have been carried away from the main trajectory, while the Gaussian model does not account for wind shear.

4.4 Sensitivity to the local-scale dispersion

Section 4.3 showed that the near-source concentrations are lower with the plume-in-grid model than with the Eulerian model, especially with the Doury parameterization. This could be surprising, since the use of plume-in-grid is supposed to prevent artificial dilution of the plume. However, this dilution is mostly horizontal: the puff vertical size reaches the top of the first vertical level (30 m) after only a few time steps. Besides, the Gaussian puff model allows to better represent the vertical diffusion, which the Eulerian model underestimates in the vicinity of the source. This mechanism is highlighted in this section, along with the sensitivity to the local-scale dispersion parameterizations.

4.4.1 Local-scale dispersion parameters

Figure 4.8 shows the evolution of σ_y and σ_z for the three parameterizations, for a puff emitted during daytime (16:00 UTC, Figures 4.8(a) and 4.8(b)) and nighttime (22:00 UTC, Figures 4.8(c) and 4.8(d)), during three hours ($\sigma_y \simeq \sigma_x$). The tendency is similar over the next nine hours. The Pasquill stability class was D (neutral) during most of the emission. Thus, there is little difference between the standard deviations during nighttime and daytime, when computed with the Briggs and Doury parameterizations. Similarity theory, which relies on a more accurate description of the boundary layer (boundary height, Monin-Obukhov length) rather than on stability classes, shows more variability. The Doury parameterization, which shows the highest standard deviations, was fitted on tracer experiments conducted on wider fields than the usual experiments used to derive the Briggs formulae and the similarity-theory formulae. Thus, the Doury scheme is better at European scale, while the other two parameterizations are better at local scale [Korsakissok et Mallet, 2009].

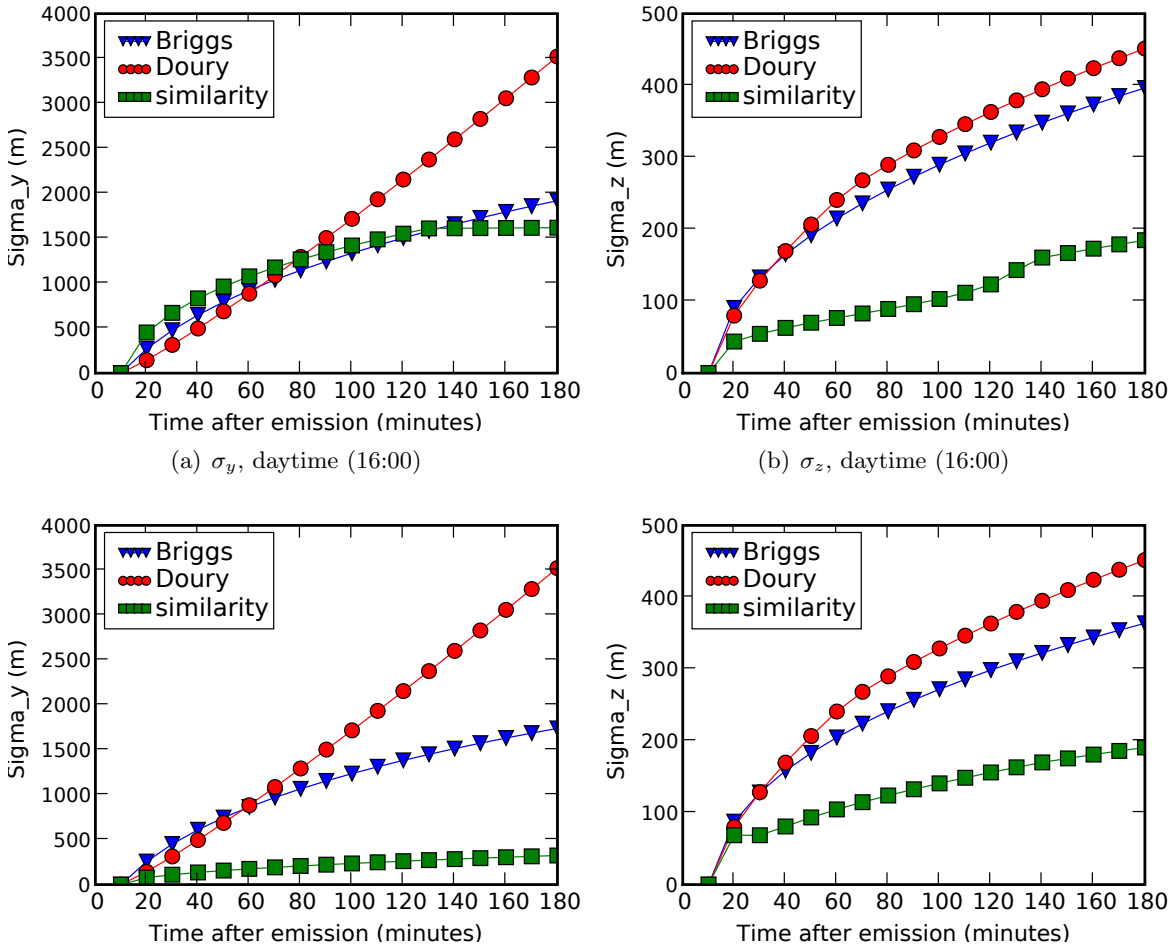
4.4.2 Comparison between the Gaussian and Eulerian diffusion

The Eulerian model represents the turbulence through the vertical diffusion coefficient K_z, which is parameterized and depends on the meteorological conditions. On the other hand, the Gaussian equation derives from the same dispersion equation as the Eulerian model, but simplified, especially by assuming a constant and homogeneous wind. Under these assumptions, the standard deviation σ_z can be related to the vertical diffusion coefficient K_z [Seinfeld et Pandis, 1998]:

$$\sigma_z = \sqrt{2\ K_z\ t} \tag{4.9}$$

Thus, the plume evolution computed with the Eulerian model has a behavior in $\sigma_z \propto \sqrt{t}$, which represents well the plume dispersion at long range, but underestimates the dispersion at short range, where it is theoretically closer to a behavior in $\sigma_z \propto t$ [Taylor, 1921]. Gaussian models, fitted on short-range experiments, better represent this behavior by adapting the dispersion to the puff distance from source (or "age", t). This was shown for instance in Demaël et Carissimo [2008], where a CFD code and a Gaussian model are compared on short-range experiments. Figure 4.9(a) shows the comparison between the evolution of σ_z given by the Gaussian models, and by Equation 4.9—taking the value of K_z in the puff cell. Figure 4.9(b) shows the puff vertical extent, that is given in practice by the sum of the heights of the vertical levels covered by the puff (see Figure 4.3). The figures also show K_z as computed by two possible parameterizations: Louis's [Louis, 1979] and Troen & Mahrt's [Troen et Mahrt, 1986]. The dispersions corresponding to the two parameterizations for K_z are lower than those estimated with Doury's and Briggs's formulae, but they are close to that of the similarity-theory estimation without the additional dilution due to the injection.

In this study, the Louis parameterization is used. Although the Troen-Mahrt parameterization ensures slightly more vertical diffusion (as seen on Figure 4.9), it did not give quite as good results. Thus, changing the Eulerian parameterization to ensure more diffusion does not allow to accurately represent point emissions near the source. The plume-in-grid model can be considered as a subgrid-scale diffusion scheme, which adapts to the puff distance from the source, and proves to be more efficient than the usual diffusion parameterizations. However, the model results are still sensitive to the Eulerian diffusion, as shown in Brandt *et al.* [1998] using the DREAM model. This study also highlighted the insufficient vertical diffusion given by Eulerian schemes, since an homogeneous mixing of the pollutants within the boundary layer was found to be more efficient than the two Eulerian parameterizations evaluated (the Louis and similarity-theory schemes).



(a) σ_y, daytime (16:00)

(b) σ_z, daytime (16:00)

(c) σ_y, nighttime (22:00)

(d) σ_z, nighttime (22:00)

Figure 4.8: Evolution of the horizontal standard deviation σ_y and vertical standard deviation σ_z, during 3 hours, for the three parameterizations and two puff release times. Figures 4.8(a) and 4.8(b) are for a puff emitted at the beginning of emission (16:00 UTC). Figures 4.8(c) and 4.8(d) correspond to puff emitted during the night (22:00 UTC). The stability class is D (neutral).

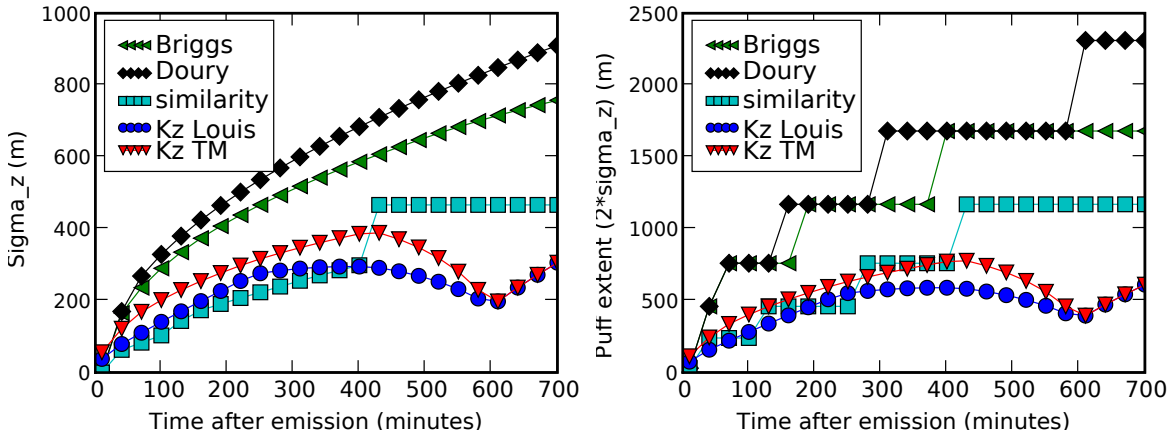


Figure 4.9: Evolution of the puff vertical extent during 12 hours for the three Gaussian parameterizations, and for the two Eulerian parameterizations for K_z. Left: the puff extent is the Gaussian standard deviation σ_z—Right: the puff extent is the sum of the heights of the vertical levels covered by the puff. The puff extent is the same in both figures for the Eulerian parameterizations.

4.4.3 Vertical profiles

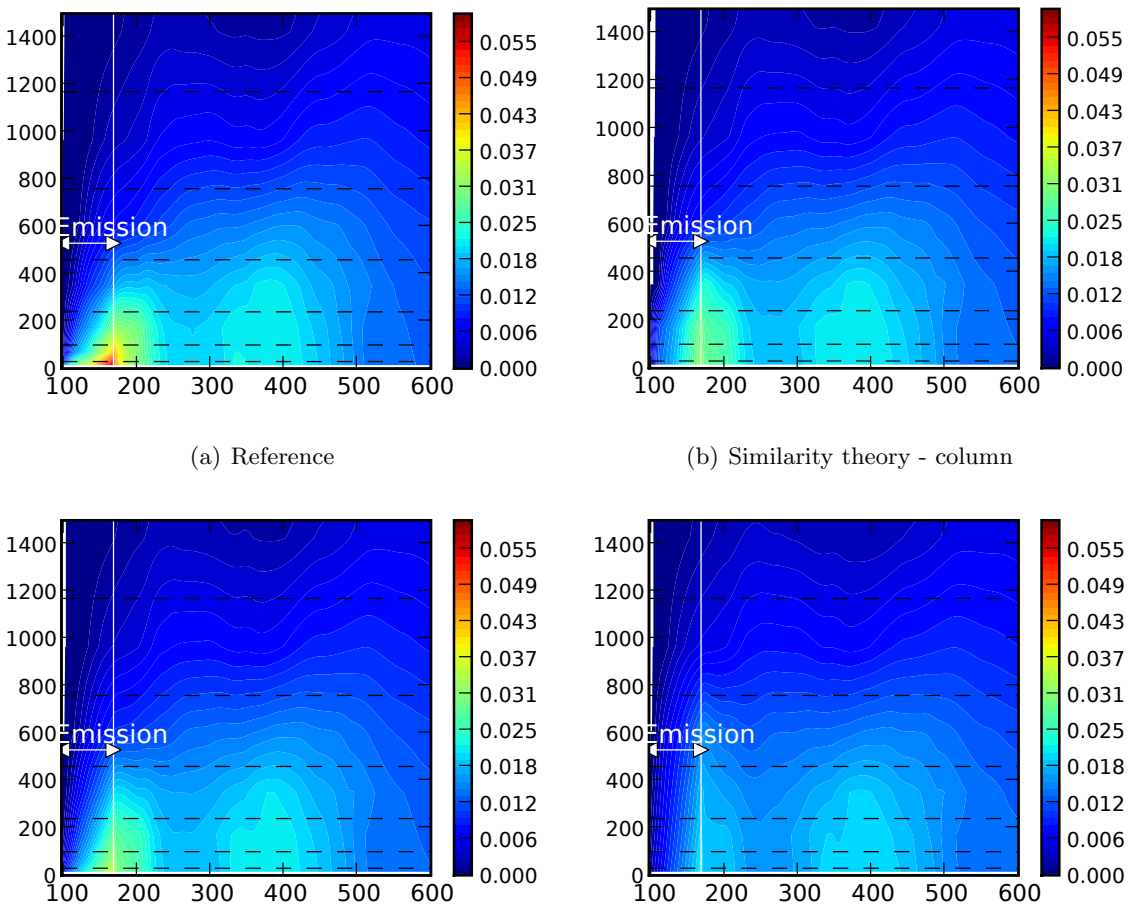
Coming back to the simulation results, the influence of the vertical diffusion schemes can be observed on Figure 4.10, which shows the vertical profiles of the mean concentration. The concentrations are averaged over the whole horizontal domain, and the vertical profile is plotted against the simulation time step. Figure 4.10(a) shows the profile for the reference simulation. The emission begins at time step 96, and the source emits during twelve hours within the first vertical level. Hence, the concentration increases with time within this level during the emission period, and slowly diffuses on the vertical. The three other figures show the same vertical profiles for the plume-in-grid model in several configurations. There is indeed more vertical diffusion, especially during the emission time. This tendency is stronger with the column injection than with the integrated injection. The Doury parameterization also shows a higher vertical diffusion than similarity-theory, which is consistent with the previous results (Figure 4.9).

4.5 Relevant timescales: a study on the injection time

The injection time determines how long the emission is handled by the Gaussian model. The sensitivity to this parameter therefore gives an insight on the relevant timescale for the use of the local-scale model, before the errors made in the Gaussian modeling approach—due to trajectory uncertainties as well as wind shear—become too large. For that purpose, plume-in-grid simulations were carried out with various injection times, ranging from $t_{\text{inj}} = 0$ to $t_{\text{inj}} = 3$ hours with a discretization step of 10 minutes, and ranging from $t_{\text{inj}} = 0$ to $t_{\text{inj}} = 12$ hours with a discretization step 1 hour. The puffs are injected at the end of the time step (600 s) when the injection criterion is met.

4.5.1 Impact on the concentrations

Figure 4.11 shows the evolution of two statistical indicators presented Section 4.3 (correlation and fractional bias) against the injection time. The trends are the same for the three param-



(a) Reference

(b) Similarity theory - column

(c) Doury - integrated

(d) Doury - column

Figure 4.10: Vertical profile of the concentrations (in $\mu g\,m^{-3}$), averaged over the whole horizontal domain. Ordinate: vertical levels (meters), abscissa: simulation time step ($\Delta t = 600$ s). The plume-in-grid injection time is 1 hour. The horizontal lines represent the interfaces of the model vertical levels.

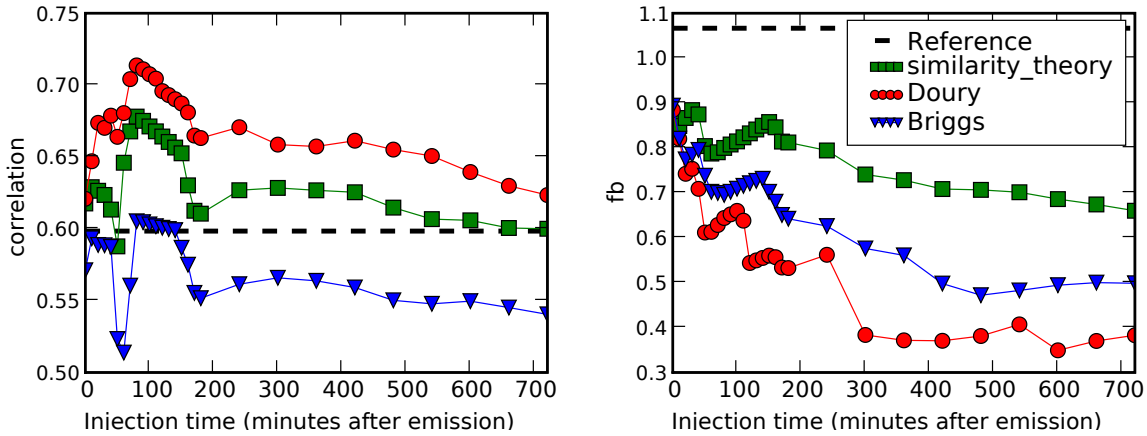


Figure 4.11: Evolution of the correlation and fractional bias against the puff injection time, ranging from zero to twelve hours. Statistics are computed for all concentrations and all stations. The simulations are carried out with the column injection, for all three Gaussian parameterizations.

eterizations. Similarity theory does not give as good a bias as the two others but still much better than the reference simulation (dashed line). The Briggs parameterization gives a lower correlation than the reference simulation for almost all injection times. The correlation is more sensitive to the parameterization than other indicators (not reported here), especially between 0 and 3 hours, while the bias is always good. As shown in Section 4.3, the main impact of plume-in-grid on concentrations is localized in the vicinity of the source, which explains why even small injection times have an impact on the global results.

4.5.2 Impact on the arrival times

Figure 4.12 shows the same indicators computed for the arrival times at the stations. The statistics are computed on values averaged over 3 hours, and the arrival time is the first time period of 3 hours when the mean value at the station is above the threshold value. This explains why the statistics are so good: the simulation has essentially to predict the plume arrival within three hours of the observation to be considered as "perfect". These statistics are sensitive to the use of a plume-in-grid treatment: since the Eulerian model diffuses the plume more rapidly on the horizontal, it shortens the plume arrival time at stations. Using the plume-in-grid model delays this arrival, and the higher the injection time is, the later the plume arrives at the stations. Between 0 and 180 minutes, the correlation increases and the bias is almost constant. Between 3 and 12 hours, however, the correlation and bias become worse than the reference simulation. Thus, an injection time of 3 hours seems to be the upper limit for acceptable results, both for the arrival time and the concentrations statistics.

4.6 Influence of the grid resolution

The spatial impact of the plume-in-grid treatment is now assessed from the viewpoint of the sensitivity to the grid resolution. The aim is to determine at what grid resolution the subgrid-scale treatment is more relevant, and to investigate the possible relationship between the grid resolution and the injection time. The grid resolution used in the previous parts of this study is henceforth called the "original" grid resolution; it corresponds to a cell size equal to 0.5625° (in

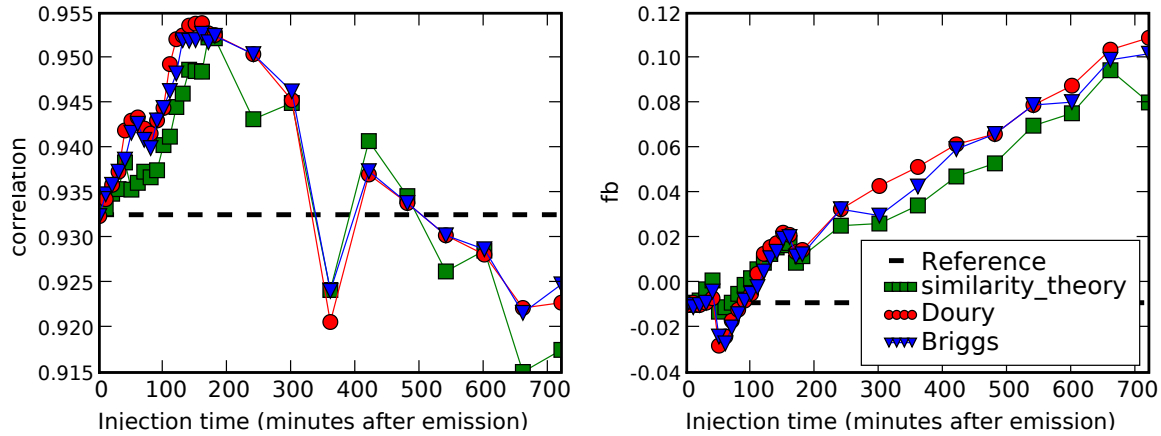


Figure 4.12: Evolution of the correlation and fractional bias for arrival times at stations, against the puff injection time. The simulations are carried out with the column injection, for all three Gaussian parameterizations.

longitude and latitude). Simulations were carried out for finer grids (÷8, ÷4, ÷2) and coarser grids (×2,×4,×6 and ×8). All the input data are the same as with the original resolution, except that they are interpolated (from the raw ECMWF data) at the new resolution. Hence, a finer resolution does not enhance the description of parameters such as the meteorological situation, and only the numerical impact of the grid resolution is assessed. This is consistent with our main goal, which is to improve the modeling of emissions without depending too much on the input data accuracy. Besides, the aim of such a model is to be used in an operational case, when no high-quality data are available, and the computational time is limited.

4.6.1 Impact of the initial volume of the emission

Before comparing the results with and without plume-in-grid, it is useful to assess the impact of the resolution on the Eulerian model alone. In particular, since the simulations are carried out with the same input parameters for all resolutions, the main changes may come from the initial volume of the emission: in the Eulerian model, the point source is immediately diluted within the volume of the cell where the source is located. Since the model has a tendency to over-predict the concentrations, a smaller emission volume will clearly increase this deficiency. To verify this hypothesis, the Eulerian model Polair3D was run on the finest grids (÷8, ÷4, ÷2), for two configurations. In the "reference" configuration, the source is emitted in one cell, as usual. The "diluted emission" consists in distributing the source in several cells, so that the total emission volume is equal to the cell volume of the original grid resolution. The result is shown in Figure 4.13. With the emission released in a single cell, the model statistics are worse with finer resolutions than with the original resolution. The FAC2 and FAC5 are less sensitive to emission changes—as already observed in Section 4.3. On the contrary, the change in the emission volume is responsible for most of the decrease observed in the correlation and bias, when refining the resolution: with an unchanged emission volume, the correlation and bias are almost constant. This highlights the importance of the initial volume where the emission is released. In the Eulerian model, this volume depends on the grid resolution, and it would be difficult to find an "ideal" volume for any emission source, hence the need for a subgrid-scale treatment that limits the dependence.

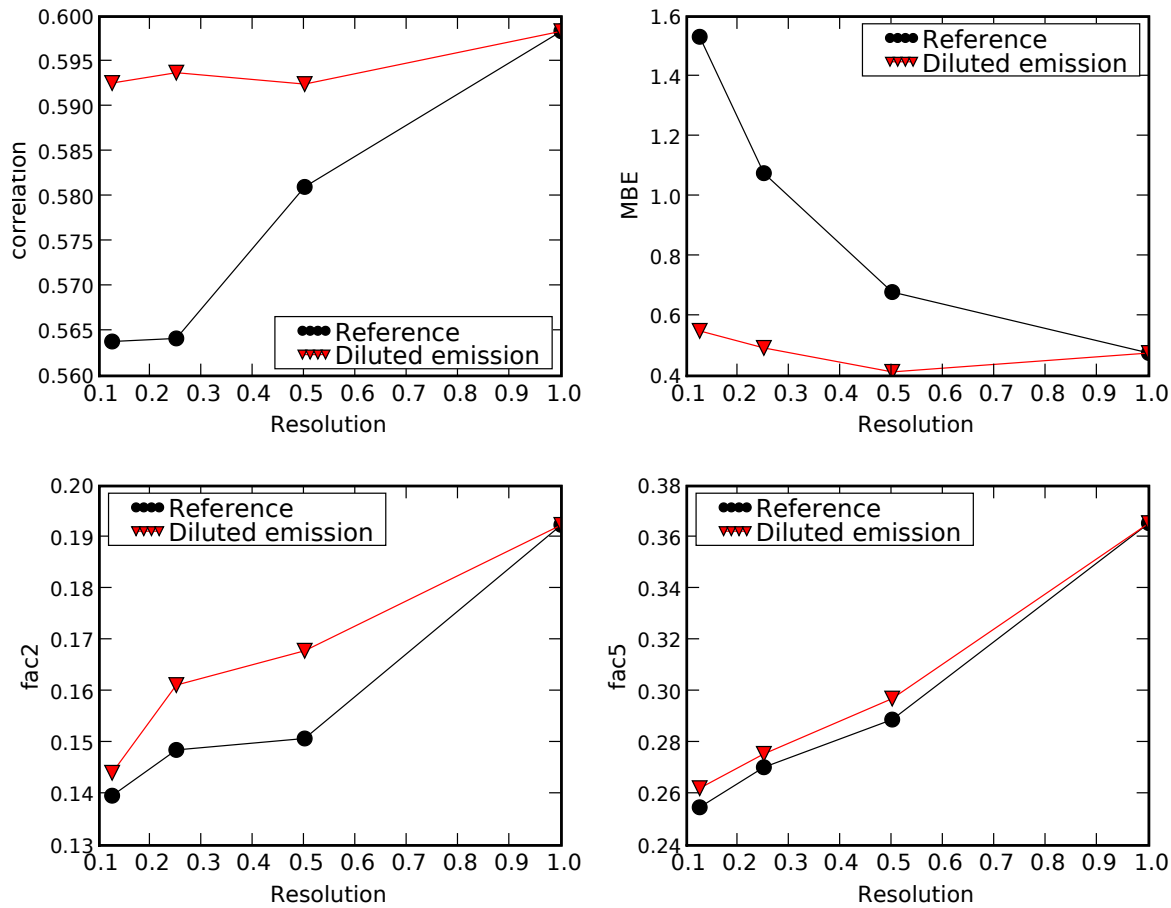


Figure 4.13: Evolution of the correlation, bias, FAC2 and FAC5 for the Eulerian model, for several factors applied to the original grid resolution (÷8, ÷4, ÷2 and ×1). The label "Reference" is the simulation with the emission in one grid cell, and "diluted emission" refers to the simulations made with a constant emission volume.

4.6.2 Use of plume-in-grid with several grid resolutions

Figure 4.14 shows the impact of the grid resolution for Polair3D, as well as for the plume-in-grid model in several configurations. Several statistical indicators (correlation, mean bias error MBE, FAC2 and FAC5) are plotted against the grid resolution. In the sequel, the resolutions will be named after their abscissa in the figure: grid-0 is the finest ($\div 8$), grid-3 is the original resolution, and grid-7 is the coarsest ($\times 8$). The reference simulation with Polair3D (dashed line) shows a decrease in performance for finer resolutions, as detailed in Section 4.6.1. The best results of the reference simulation are given for grid-3 (original), grid-4 and grid-5.

The plume-in-grid results are also shown for the three parameterizations and the two injection methods, with a 1-hour injection time. The spread of the statistics for these six simulations is wider for the finer resolutions. The use of plume-in-grid tends to improve all the statistics for the fine resolutions (grid-0 to grid-3), in almost all cases. For coarser resolutions (grid-4 to grid-7), the results are globally close to the reference, or slightly worse (especially the correlations). The small spread in the output statistics for the coarse resolutions comes from the dilution of the puff concentrations within the cell, that smooths the differences between the parameterizations. The "best" injection time would depend on the grid resolution, as well as the parameterization, which makes it difficult to determine.

The other injection criterion, based on the puff size—as defined in Equation 4.7—is also evaluated. It has the advantage of adapting to the grid resolution. The results with the size injection criterion are shown in Figure 4.15, with $C_y = 8$—the puff is injected when it reaches half the cell size. They are quite good for fine resolutions: they are better, especially in terms of bias, than the 1-hour results (Figure 4.14). However, the results with the original resolution (grid-3) and coarser resolutions are bad, especially for the correlation: the puffs are not injected before 6 hours, or more, which leads to large errors in their trajectories. Thus, it is advised to use an injection criterion based on the puff size only for grid cells smaller than 25 km, and to ensure the puff injection time is no larger than 3 hours.

4.7 Application to the Chernobyl disaster

An explosive accident took place at the Chernobyl nuclear power plant in Ukraine (51°23′N, 30°06′W), on April 25th, 1986 at 21:23 UTC. This disaster triggered a dispersion of radionuclides at continental scale. Radioactivity was measured in many European countries after the accident. The measured data used here come from the REM-database (Joint Research Center, Ispra, Italy) and consist in 88 stations, scattered mainly through central and western European countries. While the ETEX case provided a well-documented case to carry out a comprehensive sensitivity study, Chernobyl is a means to evaluate the model in a real-case study with many uncertainties, especially concerning the source characteristics.

4.7.1 Modeling set-up

The simulation was carried out with the ERA-40 fields from ECMWF, of 1.125° resolution. The simulation grid also has a 1.125° resolution in both horizontal directions. The simulation time step is 900 s and there are 12 levels, up to 5000 m. The plume-in-grid model is used with the column injection and the Doury parameterization, and a 3-hours injection time (chosen because of the very coarse resolution).

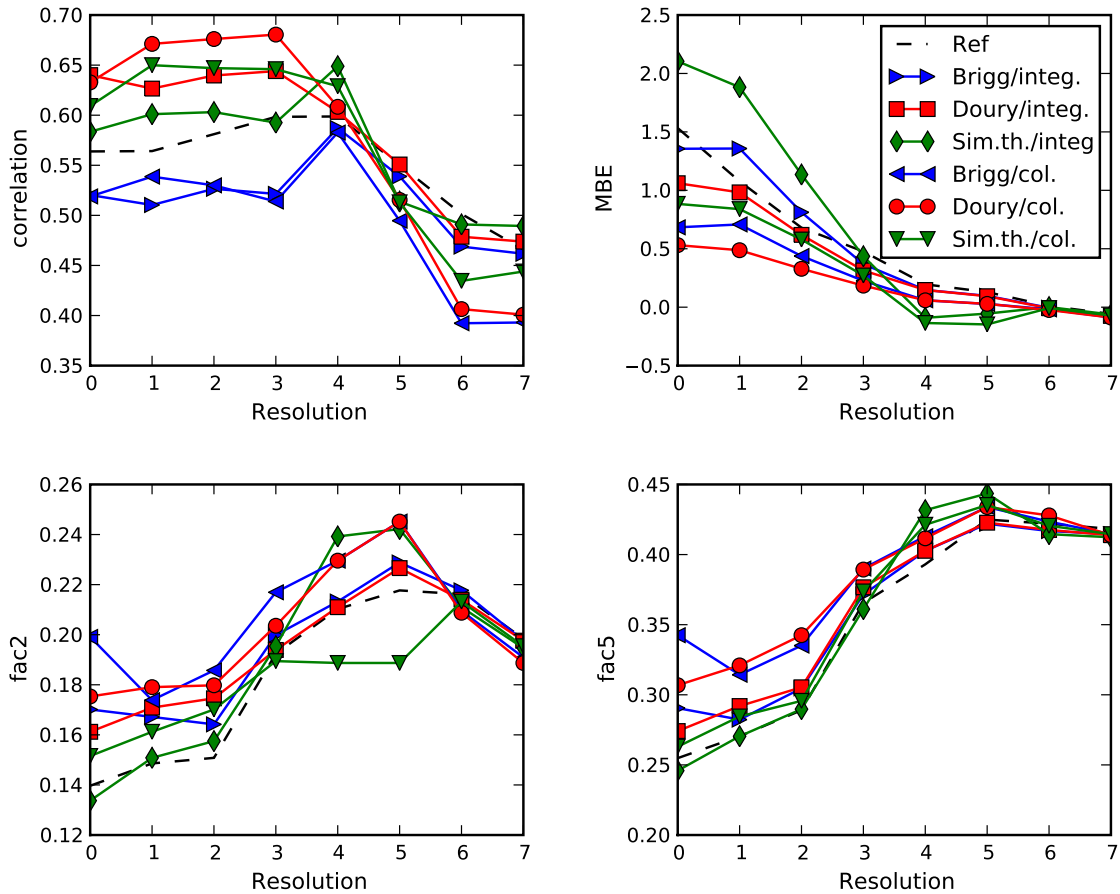


Figure 4.14: Evolution of the correlation, bias, FAC2 and FAC5 for reference and plume-in-grid models, for several factors applied to the original grid resolution (÷8, ÷4, ÷2, ×1, ×2, ×4, ×6 and ×8). Statistics on concentrations, 139 stations. The injection time is one hour.

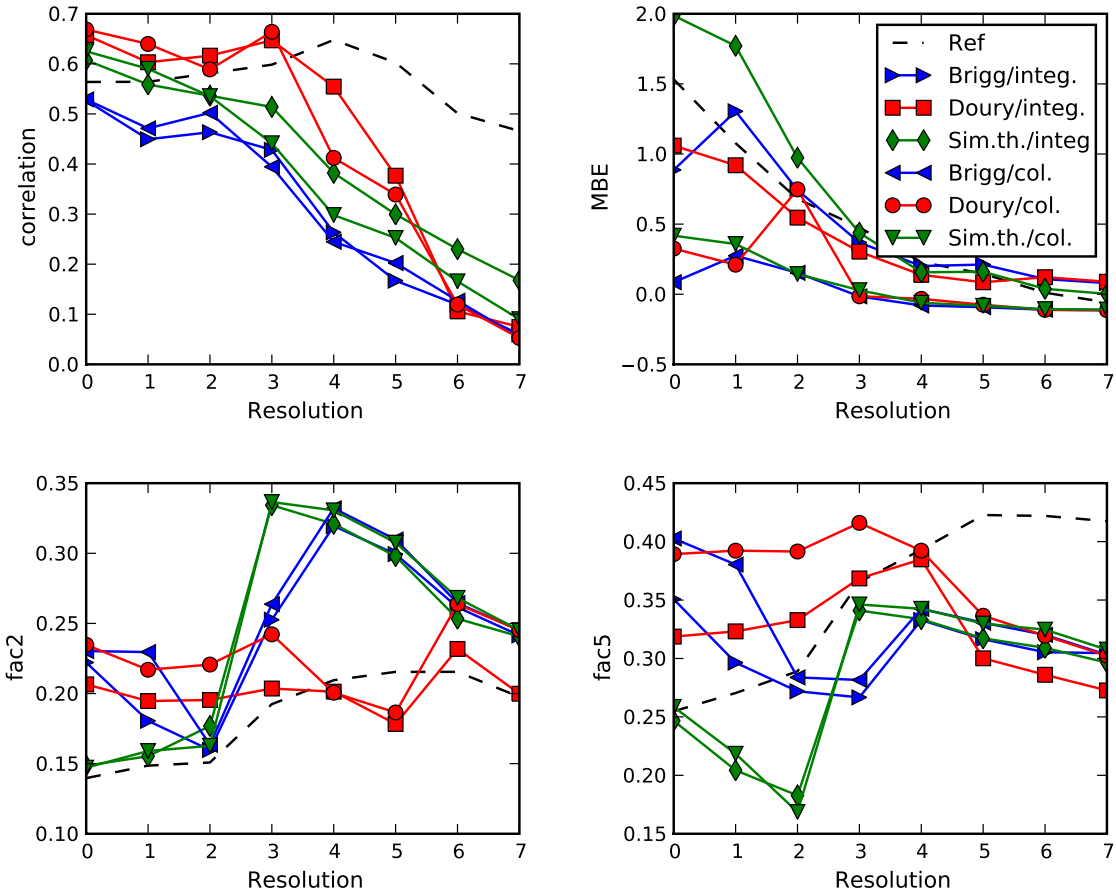


Figure 4.15: Evolution of the correlation, bias, FAC2 and FAC5 for reference and plume-in-grid models, for several factors applied to the original grid resolution (÷8, ÷4, ÷2, ×1, ×2, ×4, ×6 and ×8). Statistics on concentrations, 139 stations. Plume-in-grid simulations use an injection criterion on the puff size.

Factors	26/04	27/04	28/04	29/04	30/04	01/05	02/05	03/05	04/05	05/05
UNSCEAR	0.40	0.116	0.085	0.058	0.039	0.035	0.058	0.061	0.074	0.074
Inv. Mod.	0.45	0.42	0.05	0.001	0.015	0	0.015	0.01	0.01	0.032

Table 4.3: Temporal factors for the Chernobyl emission, for the UNSCEAR source term and the source term coming from inverse modeling (labeled "Inv. Mod"). Factors are given for each day, and applied to the total emission rate.

Species

The released radionuclides were both particulate matter and gaseous species. Here, we model only iodine (I^{131}) and caesium (Cs^{137} and Cs^{134}) in a gaseous form. For this model evaluation, we have chosen to use a simple modeling of the species dry deposition and scavenging coefficients, similarly to Quélo *et al.* [2007]. The deposition velocity is constant, equal to $0.2\,\mathrm{cm\,s^{-1}}$ for caesium and $0.5\,\mathrm{cm\,s^{-1}}$ for iodine. The scavenging coefficient is parameterized with the Belot formula [Belot *et al.*, 1988], as $\Lambda = A\ p_0^B$, where p_0 is the rain intensity (in $\mathrm{mm\,hr^{-1}}$), $A = 8 \times 10^{-5}$ and $B = 0.8$. For a more comprehensive study of these parameterizations, we refer to Brandt *et al.* [2002]. The radioactive decay of the three species is taken into account, with half-life values of 8.04 days for I^{131}, 742 days for Cs^{134} and 11000 days for Cs^{137}.

Source term

The main uncertainties come from the estimation of the source. The total released activity and temporal distribution, as well as the vertical distribution, are highly uncertain: the high core temperature lead to very high effective release heights during the initial explosion, while the emission height was lower during the two following weeks. Here, we use the vertical distribution of Brandt *et al.* [2002]. Simulations are carried out with two estimations of the total activity and time evolution: one comes from UNSCEAR [2000] and was used in Quélo *et al.* [2007], while the other was constructed by inverse modeling, using Polair3D [Davoine et Bocquet, 2007]. The Polair3D configuration used for inverse modeling was the same as used here (same meteorological fields, simulation grid, species parameterizations). The temporal factors for the two sources are given in Table 4.3. The total estimated activity is:

- I^{131}: 1.76×10^{18}Bq (UNSCEAR), 1.62×10^{18}Bq (reconstructed)
- Cs^{137}: 8.5×10^{16}Bq (UNSCEAR), 13.6×10^{16}Bq (reconstructed)
- Cs^{134}: 5.4×10^{16}Bq (UNSCEAR), 3.5×10^{16}Bq (reconstructed).

4.7.2 Results

In this case, there are no stations in the vicinity of the source: the smallest arrival time is 3 days after the explosion. Hence, even though using a plume-in-grid model still has an impact on the global statistics, the model results are quite insensitive to the local-scale configuration. Table 4.4 shows the models statistics with the UNSCEAR source term as well as the source term reconstructed by inverse modeling, for the three species. The source term specification is, as expected, the most important parameter, and has much more influence on the results than the model in use. The comparison with and without plume-in-grid with the UNSCEAR source term does not show a significant trend. The plume-in-grid results tend to be better in terms of FAC2 but slightly more biased. Using the plume-in-grid model with the reconstructed

Model	Mean obs	Mean sim	MBE	NMSE	Corr	FAC2	FMT
UNSCEAR							
Cs^{137}	0.98	0.49/0.48	-0.49/-0.50	5.62/5.73	0.46/0.47	0.23/0.26	0.31/0.31
Cs^{134}	0.57	0.34/0.33	-0.23/-0.23	4.77/4.88	0.46/0.47	0.26/0.28	0.33/0.33
I^{131}	3.83	3.45/3.42	-0.36/-0.40	6.08/6.37	0.43/0.39	0.34/0.35	0.36/0.35
Inverse modeling							
Cs^{137}	0.98	1.3/1.2	0.32/0.23	3.01/2.52	0.53/0.56	0.39/0.41	0.42/0.45
Cs^{134}	0.57	0.36/0.34	-0.20/-0.23	4.06/4.26	0.54/0.57	0.33/0.34	0.39/0.39
I^{131}	3.83	5.59/5.18	1.76/1.35	6.06/5.43	0.47/0.47	0.38/0.39	0.36/0.39

Table 4.4: Chernobyl statistics for the UNSCEAR source term, and the source term reconstructed with inverse modeling. Statistics are given in the following order: reference / plume-in-grid. The plume-in-grid simulations are run with a 3 hour injection time, the Doury parameterization and the column injection.

source term shows, however, a clear tendency to improve all indicators, even when looking at the performance at stations (not shown). Since the inverse modeling was performed with Polair3D (without plume-in-grid), this configuration is optimized for the reference model, but the plume-in-grid model performs better nonetheless. However, these results confirm that the use of a plume-in-grid model is mostly efficient at "local-scale" stations, and that no significant improvement can be expected when stations are located so far from the source.

4.8 Conclusions

Physical processes The plume-in-grid treatment of the point source emissions implemented within Polyphemus is described in this paper, with an emphasis on the available parameterizations for the local-scale dispersion and the model coupling. The associated study, based on the ETEX passive tracer experiment, shows that the main role of the plume-in-grid model is to better represent the vertical diffusion in the vicinity of the source. This theoretical result, and its practical implications, were analyzed. The subsequent improvement can be observed on the statistics on concentrations, especially at the closest stations to the source, and up to a few hundreds of kilometers from the release location. Thus, the scale of the plume-in-grid impact is larger than the mere local-scale domain, where the Gaussian model is used. There is also an impact on the arrival times at all stations, that are more accurate with the delay implied by the plume-in-grid.

Sensitivity to the local-scale parameterizations While the results with plume-in-grid are almost always better than with the Eulerian model alone, there is a substantial variability in the outputs when changing the parameterizations and injection method. The Doury parameterization is found to be the best in this case, ensuring more vertical diffusion. The column injection gives slightly better results than the integrated injection for the same reason. The sensitivity to the local-scale parameterizations is mostly seen at the few stations located in the vicinity of the source. As confirmed with the Chernobyl case, this sensitivity is much less important at farthest stations.

Spatial and time scales The upper time limit, for the use of the local-scale model, was found to be about 3 hours. Afterward, the global statistics on the concentrations and the arrival times tend to worsen. This can be observed in the ETEX case and in the Chernobyl case

as well (not shown here). The use of a plume-in-grid coupling tends to be more efficient at finer resolutions: it allows to depend less on the volume of the Eulerian cells, and to compensate for the subsequent over-prediction. The key point when using the plume-in-grid model is to determine a good injection time. It is proposed here to use a criterion on the puff size when the resolution is finer than 25 km. For coarser resolutions, this methods leads to very high injection times, and errors in the puff trajectory tend to decrease the model's performance. Using a fixed injection time (the default 1 hour value seems appropriate in most cases) is therefore recommended.

Perspectives The results of this application and sensitivity study on the ETEX case are representative of near-ground releases during neutral conditions. However, the conclusions may be different for elevated releases—for instance, buoyant plumes from power plants may be more than 200 m high. In these cases, the plume is maintained aloft for a longer time with the plume-in-grid model, especially during nighttime and very stable situations [Karamchandani *et al.*, 2002]. Thus, the next step would be to extend the plume-in-grid model to reactive cases, to be evaluated with elevated releases. Another step would be to use the plume-in-grid model for inverse modeling and network design studies.

4.9 Appendix

4.9.1 Coordinates

Since Polair3D uses longitude and latitude coordinates, and the Gaussian model uses Cartesian coordinates, it is necessary to carry out a coordinate transformation. The Cartesian coordinates are the coordinates in the tangent plane at the point source. The transformation formulae are:

$$\begin{pmatrix} \mathrm{d}x \\ \mathrm{d}y \end{pmatrix} = \begin{pmatrix} R\ \cos\phi\ \mathrm{d}\lambda \\ R\ \mathrm{d}\phi \end{pmatrix} \tag{4.10}$$

Where x, y, z are the point Cartesian coordinates, λ is the longitude and ϕ is the latitude, in radians, and R is the Earth radius in meters.

Note that we assume that the distance between the source and the puff center is not too great, otherwise the error in assimilating the puff trajectory to that in the tangent plane could not be neglected anymore.

4.9.2 Formulae to compute the puff integral over a volume

Let a puff α contain a given quantity of species A. The concentration of A in α is denoted c_A^α. The integral over space of a variable v is denoted with brackets $< v >$.

The total quantity of A in the puff is given by:

$$Q_A^\alpha =< c_A^\alpha >= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c_A^\alpha \tag{4.11}$$

If we assume that the puff has a Gaussian shape, then the concentration can be written as:

$$c_A^\alpha(x,y,z) = Q_A^\alpha \times G_\alpha^x(x)\ G_\alpha^y(y)\ G_\alpha^z(z) \tag{4.12}$$

where $G_\alpha^x(x)$ is the Gaussian distribution in the x direction:

$$G_\alpha^x(x) = \frac{1}{\sqrt{2\ \pi}\ \sigma_x}\ \exp\left(-\frac{(x-x_\alpha)^2}{2\ \sigma_x^2}\right) \tag{4.13}$$

with x_α the puff center coordinate and σ_x the puff standard deviation in the x direction.

We want to compute the integral of the puff within a given volume, noted Q_A^{vol}. We note x_1 and x_2 the lower and upper boundaries in the x direction, and with a similar notation for the y and z directions, we have:

$$Q_A^{vol} = \int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} c_A^\alpha(x,y,z)\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z \tag{4.14}$$

Combining this with Equation 4.12 leads to write:

$$Q_A^{vol} = Q_A^\alpha \times H_\alpha^x(x_1,x_2)\ H_\alpha^y(y_1,y_2)\ H_\alpha^z(z_1,z_2) \tag{4.15}$$

with

$$\begin{aligned} H_\alpha^x(x_1,x_2) &= \int_{x_1}^{x_2} G_\alpha^x(x)\ \mathrm{d}x \\ &= \underbrace{\int_{-\infty}^{x_2} G_\alpha^x(x)\ \mathrm{dx}}_{F_\alpha^x(x_2)} - \underbrace{\int_{-\infty}^{x_1} G_\alpha^x(x)\ \mathrm{dx}}_{F_\alpha^x(x_1)} \end{aligned} \tag{4.16}$$

To compute the integral over the volume Q_A^{vol}, we rely on the cumulative distribution function $F_\alpha^x(u)$ for a normal distribution (in the x direction):

$$F_\alpha^x(u) = \frac{1}{\sqrt{2\ \pi}\ \sigma_x} \int_{-\infty}^{u} \exp\left(-\frac{(x - x_\alpha)^2}{2\ \sigma_x^2}\right)\ \mathrm{d}x\ . \tag{4.17}$$

It can be expressed with the function $\mathrm{erfc}(u)$ defined as

$$\mathrm{erfc}(u) = \frac{2}{\sqrt{\pi}} \int_{u}^{\infty} \exp(-x^2)\ \mathrm{d}x\ , \tag{4.18}$$

which satisfies $\mathrm{erfc}(-u) = 1 + \mathrm{erf}(u)$. A simple change of variable in Equation 4.17 shows that

$$F_\alpha^x(u) = \frac{1}{2}\ \left[1 + \mathrm{erf}\left(\frac{u - x_\alpha}{\sqrt{2}\ \sigma_x}\right)\right]\ . \tag{4.19}$$

This allows to compute the integral over a given volume for a Gaussian-shaped puff. To take into account the reflections on the ground and capping inversion, additional reflection terms are needed, and their integral is computed following the same procedure.

The cell concentration is corrected to ensure mass conservation: if Q_A^i is the mass of species A within the cell i (whose volume is V_i), then the puff contribution in this cell is given by:

$$c_A^i = \frac{Q_A}{\sum_{i=1}^{N_{\mathrm{grid}}} Q_A^i} \times \frac{Q_A^i}{V_i} \tag{4.20}$$

where Q_A is the total mass of the puff for the species A, and N_{grid} is the number of cells into which the puff is transferred.

Chapitre 5

Modèle de panache sous-maille réactif

Le modèle de panache sous-maille, décrit et évalué pour des traceurs passifs à l'échelle continentale dans le chapitre 4, est utilisé dans ce chapitre pour modéliser des panaches réactifs. Pour cela, il faut tout d'abord introduire un modèle de chimie dans les bouffées gaussiennes. Cela suppose de prendre en compte non seulement les réactions entre les espèces au sein d'une bouffée, mais également les interactions entre deux bouffées superposées, ainsi que les réactions entre les espèces de la bouffée et les espèces présentes dans l'atmosphère en-dehors du panache (le « fond »). Pour cela, certaines hypothèses simplificatrices sont faites sur la géométrie des bouffées. De plus, les réactions chimiques sont calculées sur les concentrations moyennes au sein des bouffées. Le modèle de chimie est décrit dans la partie 5.1.

Dans un panache dont la composition diffère de celle de l'atmosphère environnante, les réactions prépondérantes ne sont pas les mêmes. Ainsi, non seulement l'utilisation du panache sous-maille a un effet sur la dispersion, comme pour les espèces passives, mais aussi sur les réactions chimiques en champ proche. Cet aspect est décrit de façon théorique, et testé ensuite sur un cas simplifié, dans la partie 5.2. Il y est montré, en particulier, que la titration de l'ozone est la réaction principale dans un panache de NOx, tandis que la formation de radicaux OH est retardée jusqu'à une phase où le panache est suffisamment développé.

Enfin, le modèle de panache sous-maille réactif est appliqué au cas de la photochimie en Île-de-France, sur six mois pour l'été 2001 (partie 5.3). Cela permet d'évaluer l'impact du traitement sous-maille des émissions ponctuelles sur un cas d'application réel et sur les performances du modèle en comparaison aux observations. L'utilisation du traitement sous-maille améliore les statistiques globales, en particulier pour les polluants primaires (SO_2 et NO). La différence est plus faible pour les espèces secondaires NO_2 et O_3. L'utilisation du modèle de panache sous-maille a pour principale conséquence de modifier les concentrations localement autour des sources. L'impact est plus grand lorsque l'on retient certains jours particuliers, notamment aux stations situées en aval de sources importantes. Il est enfin important de remarquer que, si le temps d'injection est un paramètre très important pour les espèces passives, il affecte beaucoup moins les espèces réactives.

Sommaire

5.1 Description du modèle 141
5.1.1 Modèle de chimie pour les bouffées réactives 141
5.1.2 Interaction entre deux bouffées superposées 144
5.2 Application du panache sous-maille à la chimie de l'ozone 149
5.2.1 Chimie de l'ozone 149
5.2.2 Impact du panache sous-maille réactif 152
5.2.3 Influence du nombre de sources 157
5.3 Application à la qualité de l'air en Île-de-France 157
5.3.1 Configuration 157
5.3.2 Émissions 158
5.3.3 Configuration de panache sous-maille 160
5.3.4 Simulations 160
5.4 Évaluation du modèle de panache sous-maille réactif 160
5.4.1 Critères d'évaluation 160
5.4.2 Impact du panache sous-maille sur les statistiques 161
5.4.3 Répartition spatiale de l'impact 163
5.5 Résultats pour certains jours 165
5.5.1 Impact sur les polluants primaires 166
5.5.2 Impact sur les polluants secondaires 167
5.6 Analyse de sensibilité 170
5.6.1 Influence de la diffusion verticale 170
5.6.2 Influence de la modélisation locale 173
5.7 Conclusions 173
5.7.1 Chimie dans un panache 173
5.7.2 Impact sur la photochimie en région parisienne 175
5.7.3 Bilan et perspectives 176

Les résultats présentés dans la partie 5.3 et les suivantes sont tirés de Korsakissok et Mallet [2010a].

5.1 Description du modèle

5.1.1 Modèle de chimie pour les bouffées réactives

Le modèle de chimie décrit ici est largement inspiré du modèle SCICHEM, décrit dans Karamchandani *et al.* [2000], notamment en ce qui concerne le traitement des interactions entre les bouffées superposées, et entre les bouffées et le fond (pour ce qui est de la « méthode 1 » décrite en partie 5.1.1).

Principe et notations

Chaque bouffée transporte toutes les espèces utilisées par le modèle. Les bouffées transportent une quantité pour chaque espèce, éventuellement zéro si l'espèce n'est pas présente dans la bouffée. La chimie des bouffées a les caractéristiques suivantes :

- Les espèces d'une bouffée réagissent entre elles,
- Les espèces de deux bouffées superposées réagissent entre elles,
- Les espèces d'une bouffée réagissent avec les espèces de fond, c'est-à-dire celles présentent dans l'atmosphère.

Les quantités transportées par les bouffées sont, en réalité, des *perturbations* des concentrations de fond : si une espèce présente dans l'atmosphère réagit avec la bouffée, les concentrations de fond diminuent localement, ce qui se traduit par une perturbation négative. Cette perturbation est ajoutée aux concentrations de fond, pour former une concentration totale, représentative de l'état de l'atmosphère dans le volume de la bouffée.

La chimie est calculée avec les concentrations moyennes de chaque bouffée. Les concentrations de fond sont également considérées comme homogènes à l'intérieur de la bouffée, ce qui est cohérent avec le fait que le volume des bouffées est inférieur au volume des mailles du modèle eulérien, où sont prises les concentrations de fond. Ces concentrations sont prises dans la maille correspondant au centre de la bouffée. Dans la suite, l'exposant α (resp. β) correspond à la variable de la bouffée α (resp. β), et l'indice A (resp. B) correspond à l'espèce A (resp. B). La concentration de l'espèce A dans la bouffée α sera donc notée c_A^α. L'exposant b correspond aux concentrations de fond. La moyenne en temps d'une variable v est notée $\overline{v}$, et l'intégrale sur tout l'espace est notée $\langle v \rangle$.

Moments d'ordre 1 et 2 d'une bouffée gaussienne

La quantité totale de l'espèce 1 dans une bouffée α est donnée par

$$Q_A^\alpha = \langle c_A^\alpha \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_A^\alpha \, \mathrm{d}V. \tag{5.1}$$

Pour une bouffée de forme gaussienne, la concentration correspondante s'écrit

$$c_A^\alpha(x, y, z, t) = Q_A^\alpha \times G_\alpha(x) \; G_\alpha(y) \; G_\alpha(z), \tag{5.2}$$

avec $G_\alpha(x)$ la distribution gaussienne dans la direction x

$$G_\alpha(x) = \frac{1}{\sqrt{2\,\pi}\;\sigma_x} \; \exp\left(-\frac{(x - x_\alpha)^2}{2\;\sigma_x^2}\right), \tag{5.3}$$

avec x_α la coordonnée du centre de la bouffée, σ_x l'écart type gaussien dans la direction x. On a donc $\langle G_\alpha(x) \rangle = 1$, de même pour les directions y et z, et $Q_A^\alpha = \langle c_A^\alpha \rangle$ c'est le moment de

premier ordre de la distribution des concentrations.

Le moment de second ordre de la distribution des concentrations est donné par

$$\langle {c_A^\alpha}^2 \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[Q_A^\alpha G_\alpha(x)\ G_\alpha(y)\ G_\alpha(z)\right]^2 \mathrm{d}x\mathrm{d}y\mathrm{d}z = (Q_A^\alpha)^2\ I_\alpha, \tag{5.4}$$

avec

$$I_\alpha = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[G_\alpha(x)\ G_\alpha(y)\ G_\alpha(z)\right]^2 \mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_{-\infty}^{\infty} G_\alpha^2(x)\mathrm{d}x \int_{-\infty}^{\infty} G_\alpha^2(y)\mathrm{d}y \int_{-\infty}^{\infty} G_\alpha^2(z)\mathrm{d}z. \tag{5.5}$$

L'intégrale dans les trois directions est donc donnée par

$$\int_{-\infty}^{\infty} G_\alpha^2(x)\ \mathrm{d}x = \frac{1}{2\sqrt{\pi}\sigma_x}, \tag{5.6}$$

ce qui signifie que l'équation 5.5 devient

$$I_\alpha = \frac{1}{2^3\pi^{3/2}\sigma_x\sigma_y\sigma_z}. \tag{5.7}$$

L'expression du moment de second ordre se déduit alors de l'équation 5.4 et de l'équation 5.7

$$\langle {c_A^\alpha}^2 \rangle = \langle c_A^\alpha \rangle^2\ I_\alpha = \frac{\langle c_A^\alpha \rangle^2}{2^3\pi^{3/2}\sigma_x\sigma_y\sigma_z}. \tag{5.8}$$

Bouffée avec une concentration uniforme

On fait à présent l'hypothèse que les concentrations sont uniformes à l'intérieur de la bouffée. On suppose également que le volume de la bouffée est fini, et noté V_α. Il s'agit d'exprimer ce volume V_α en fonction des moments de premier et second ordre de la distribution de concentration. La concentration de l'espèce A, supposée à présent homogène dans la bouffée, est égale à C_A^α à l'intérieur du volume V_α, et zéro ailleurs.

Le premier moment, ou la quantité totale de l'espèce A dans la bouffée α est à présent $\langle c_A^\alpha \rangle = Q_A^\alpha = C_A^\alpha \times V_\alpha$.

Le moment de second ordre est donné par

$$\langle {c_A^\alpha}^2 \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {C_A^\alpha}^2\ dV = {C_A^\alpha}^2 \times V_\alpha = \frac{\langle c_A^\alpha \rangle^2}{V_\alpha}. \tag{5.9}$$

Ainsi, le volume de la bouffée α peut se déduire des moments de premier et second ordre par la relation

$$V_\alpha = \frac{\langle c_A^\alpha \rangle^2}{\langle {c_A^\alpha}^2 \rangle}. \tag{5.10}$$

Si l'on étend cette définition au cas des bouffées gaussiennes, d'après l'équation 5.8, on a

$$V_\alpha = \frac{1}{I_\alpha} = 2^3\pi^{3/2}\sigma_x\sigma_y\sigma_z. \tag{5.11}$$

Chimie dans une bouffée

Note : dans la suite, les concentrations utilisées sont prises à l'intérieur des bouffées. Donc, c_A^α est égale à la concentration uniforme C_A^α utilisée dans l'équation 5.9 : $c_A^\alpha = C_A^\alpha$.

On considère deux espèces A et B, transportées par la bouffée α et réagissant entre elles pour produire une troisième espèce P, avec un taux de réaction k :

$$\mathrm{A} + \mathrm{B} \rightarrow P \tag{R 5.4}$$

Si l'on ne considère que les réactions entre les espèces présentes dans la bouffée, sans interaction extérieure, et que l'on suppose les concentrations uniformes dans la bouffée, la vitesse de disparition de A est donnée par :

$$\frac{\mathrm{d}c_A^\alpha}{\mathrm{d}t} = -k\ c_A^\alpha c_B^\alpha. \tag{5.12}$$

Si la chimie se produit durant un pas de temps Δt, la concentration de l'espèce A dans la bouffée à la fin du pas de temps est alors donnée au premier ordre (en temps) par

$$c_A^\alpha(t + \Delta t) = c_A^\alpha(t) + \Delta c_A^\alpha, \tag{5.13}$$

avec

$$\Delta c_A^\alpha = -k\ c_A^\alpha(t)\ c_B^\alpha(t)\ \Delta t. \tag{5.14}$$

La quantité correspondante, transportée par la bouffée, est alors donnée par

$$Q_A^\alpha(t + \Delta t) = c_A^\alpha(t + \Delta t) \times V_\alpha, \tag{5.15}$$

et la quantité de A dans la bouffée a donc diminué, durant le pas de temps, de

$$\Delta Q_A^\alpha = \Delta c_A^\alpha \times V_\alpha. \tag{5.16}$$

Prise en compte des interactions avec les concentrations de fond (méthode 1)

On suppose à présent que les espèces A et B sont présentes à la fois dans la bouffée, et dans les concentrations de fond, c'est-à-dire dans l'atmosphère à l'endroit où se situe la bouffée. Les concentrations de fond pour l'espèce A sont notées c_A^b. Pour prendre en compte les réactions entre l'ensemble des espèces présentes dans le volume de la bouffée, on utilise la concentration totale (bouffée et fond), notée $\widetilde{c_A^\alpha} = c_A^\alpha + c_A^b$. La nouvelle vitesse de disparition de A est donc

$$\frac{\mathrm{d}(c_A^\alpha + c_A^b)}{\mathrm{d}t} = -k\ (c_A^\alpha + c_A^b)\ (c_B^\alpha + c_B^b). \tag{5.17}$$

Cependant, les réactions chimiques entre les espèces de fond sont également prises en compte indépendamment de la modélisation des bouffées, dans le modèle eulérien (si l'on se situe dans le cadre d'un couplage de modèles de type panache sous-maille). Dans ce cas, le taux de disparition de l'espèce A pris en compte par le modèle eulérien est donné par

$$\frac{\mathrm{d}c_A^b}{\mathrm{d}t} = -k\ c_A^b c_B^b. \tag{5.18}$$

Cependant, la réaction entre les concentrations de fond des espèces A et B est prise en compte une deuxième fois dans la chimie des bouffées telle que décrite dans l'équation 5.17 ; celle-ci peut en effet se décomposer de la façon suivante

$$\frac{\mathrm{d}(c_A^\alpha + c_A^b)}{\mathrm{d}t} = -k(\underbrace{c_A^\alpha\ c_B^\alpha}_{(1)} + \underbrace{c_A^b\ c_B^b}_{(2)} + \underbrace{c_A^\alpha\ c_B^b + c_B^\alpha\ c_A^b}_{(3)}) \tag{5.19}$$

(1) représente les réactions entre les espèces de la bouffée seulement, (2) représente les réactions entre les espèces de fond seulement, et (3) est un terme supplémentaire représentant les interactions entre le fond et la bouffée, qui intervient car la chimie décrite ici est non-linéaire. Le terme (2) est bien le même que le terme de l'équation 5.18, déjà pris en compte dans le modèle eulérien. Pour prendre en compte les interactions non linéaires, tout en éliminant ce terme, il faut donc calculer la chimie d'après l'équation 5.17, puis soustraire la contribution du terme (2). La bouffée transporte ainsi l'ensemble des perturbations (chimie dans la bouffée et interactions avec le fond).

Pour calculer les réactions se produisant durant un pas de temps Δt, et en déduire les concentrations dans la bouffée à la fin du pas de temps, $c_A^\alpha(t+\Delta t)$, connaissant la concentration initiale $c_A^\alpha(t)$, on utilise la procédure suivante :

1. Ajouter la concentration de fond à la concentration de la bouffée : $\widetilde{c_A}(t) = c_A^\alpha(t) + c_A^b(t)$
2. Calculer la variation de concentration totale durant le pas de temps Δt due à la chimie (équation 5.17) : $\Delta\widetilde{c_A}$, et la nouvelle concentration totale $\widetilde{c_A}(t+\Delta t) = \widetilde{c_A}(t) + \Delta\widetilde{c_A}$
3. Calculer la variation de concentration de fond pendant le pas de temps Δt due à la chimie entre espèces de fond (équation 5.18) : Δc_A^b, et la nouvelle concentration de fond $c_A^b(t+\Delta t) = c_A^b(t) + \Delta c_A^b$
4. En déduire la nouvelle concentration de la bouffée :

$$c_A^\alpha(t+\Delta t) = \widetilde{c_A}(t+\Delta t) - c_A^b(t+\Delta t) = \underbrace{\widetilde{c_A}(t) - c_A^b(t)}_{c_A^\alpha(t)} + \underbrace{\Delta\widetilde{c_A}}_{(1)+(2)+(3)} - \underbrace{\Delta c_A^b}_{(2)}$$

De cette façon, l'interaction entre le fond et la bouffée est entièrement prise en compte dans la bouffée, qui est donc bien une perturbation des concentrations de fond : si une espèce A de fond, non présente initialement dans la bouffée ($c_A^\alpha(t) = 0$), disparaît à cause des réactions chimiques avec la bouffée, sa concentration dans la bouffée avec prise en compte de la chimie et des interactions de fond sera négative : $c_A^\alpha(t+\Delta t) < 0$. La concentration totale (prise en compte en sortie de modèle, au moment où les concentrations sont sauvegardées) est toujours positive, et tient compte de la diminution locale de concentration : $0 < \widetilde{c_A}(t+\Delta t) < c_A^b(t+\Delta t)$. Il faut toutefois remarquer que la concentration de fond utilisée est celle où se trouve le centre de la bouffée, alors que la bouffée peut couvrir plusieurs mailles (horizontalement et/ou verticalement).

5.1.2 Interaction entre deux bouffées superposées

A présent, on ne tient à nouveau plus compte du fond, et l'on considère juste deux bouffées α and β, contenant toutes les deux les espèces A et B. Quand les bouffées sont suffisamment proches, elles sont partiellement superposées. Elle interagissent alors entre elles, de la même façon que les espèces de la bouffée et du fond, mais dans un volume plus petit (le volume d'interaction, ou de superposition, $V_{\alpha\beta}$).

Moment d'ordre 2 pour deux bouffées gaussiennes

Le moment d'ordre deux de la distribution de concentration pour les deux bouffées superposées peut se définir ainsi :

$$\langle c_A^\alpha c_A^\beta \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[Q_A^\alpha G_\alpha(x)\ G_\alpha(y)\ G_\alpha(z)\right] \times \left[Q_A^\beta G_\beta(x)\ G_\beta(y)\ G_\beta(z)\right] \mathrm{d}x\mathrm{d}y\mathrm{d}z.$$

Il peut donc se calculer de façon similaire à celle décrite dans la partie 5.1.1 :

$$\langle c_A^\alpha c_A^\beta \rangle = (Q_A^\alpha \; Q_A^\beta) \times I_{\alpha\beta}, \tag{5.20}$$

avec

$$I_{\alpha\beta} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[G_\alpha(x)\; G_\alpha(y)\; G_\alpha(z)\right] \times \left[G_\beta(x)\; G_\beta(y)\; G_\beta(z)\right] \mathrm{d}x\mathrm{d}y\mathrm{d}z. \tag{5.21}$$

Le coefficient $I_{\alpha\beta}$ peut se calculer analytiquement. On appellera ce terme le *coefficient d'interaction* entre les deux bouffées α et β, car il représente le degré d'interaction entre les deux bouffées : $I_{\alpha\beta} = 0$ si les deux bouffées n'interagissent pas du tout, et $I_{\alpha\beta} = I_\alpha = 1/V_\alpha$ si $\beta = \alpha$.

Moment d'ordre 2 pour deux bouffées à concentration uniforme

On suppose à présent que les concentrations des deux bouffées sont uniformes, comme dans la partie 5.1.1. Le volume d'interaction entre les deux bouffées $V_{\alpha\beta}$ est représenté figure 5.1.

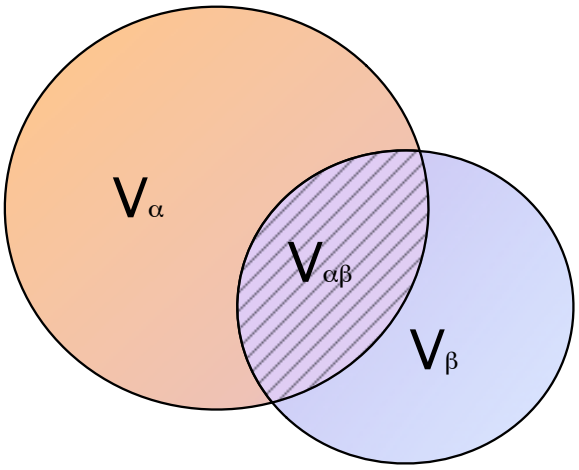


FIG. 5.1 – Superposition de deux bouffées α et β. Le volume de superposition est $V_{\alpha\beta}$.

Le moment d'ordre deux est donné par

$$\langle c_A^\alpha c_A^\beta \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C_A^\alpha C_A^\beta \; dV = C_A^\alpha C_A^\beta \times V_{\alpha\beta} = \frac{Q_A^\alpha Q_A^\beta}{V_\alpha V_\beta} \times V_{\alpha\beta}. \tag{5.22}$$

L'équation 5.22 provient du fait qu'en dehors du volume d'interaction $V_{\alpha\beta}$, les concentrations de l'espèce A sont nulles, soit pour la bouffée α, soit pour la bouffée β, d'après l'hypothèse de concentration uniforme dans un volume fini.

Des équations 5.20 et 5.22, on peut déduire une relation entre le coefficient d'interaction $I_{\alpha\beta}$, les volumes des bouffées V_α et V_β, et le volume d'interaction $V_{\alpha\beta}$

$$I_{\alpha\beta} = \frac{V_{\alpha\beta}}{V_\alpha V_\beta}. \tag{5.23}$$

Réactions entre deux bouffées superposées

On considère à présent la réaction décrite dans la partie 5.1.1. Pour cette réaction, le taux de disparition de l'espèce A en un point (x, y, z) est donné par

$$\frac{\mathrm{d}c_A}{\mathrm{d}t}(x, y, z, t) = -k \; c_A(x, y, z, t) \; c_B(x, y, z, t) \tag{5.24}$$

avec la concentration en ce point donnée par la somme des contributions de toutes les bouffées (sur tout l'espace) :

$$c_A(x,y,z,t) = \sum_\alpha c_A^\alpha(x,y,z,t). \tag{5.25}$$

La quantité exacte de l'espèce A disparaissant dans tout l'espace durant Δt à cause de la réaction chimique est donc

$$\Delta Q_A(t) = -k\ \Delta t \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sum_{\alpha,\beta} c_A^\alpha(x,y,z,t)\ c_B^\beta(x,y,z,t)\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z \tag{5.26}$$

On suppose à présent que l'on peut permuter somme et intégrale, ce qui est possible car il s'agit de fonctions gaussiennes dérivables dans tout l'espace

$$\Delta Q_A(t) = -k\ \Delta t \sum_{\alpha,\beta} \langle c_A^\alpha c_B^\beta \rangle. \tag{5.27}$$

D'après l'équation 5.20, on a alors

$$\Delta Q_A(t) = -k\ \Delta t \sum_{\alpha,\beta} Q_A^\alpha Q_B^\beta\ I_{\alpha\beta}. \tag{5.28}$$

L'équation 5.28 donne la quantité totale de l'espèce A disparaissant dans tout l'espace, à cause de la chimie ayant lieu dans toutes les bouffées, en prenant en compte les interactions non-linéaires entre les bouffées. Cette quantité doit alors être distribuée entre les différentes bouffées. Une répartition évidente (mais non unique) est

$$\Delta Q_A^\alpha = -k\ \Delta t\ Q_A^\alpha \sum_\beta Q_B^\beta\ I_{\alpha\beta}, \tag{5.29}$$

avec ΔQ_A^α la quantité de l'espèce A disparaissant dans la bouffée α durant Δt, et $\Delta Q_A = \sum_\alpha \Delta Q_A^\alpha$. L'équation 5.28 donne l'exacte formulation du taux de disparition de l'espèce A pour la réaction étudiée ici, sans l'approximation des concentrations uniformes dans les bouffées.

Comme le modèle utilise des concentrations et non des quantités, pour résoudre le système d'équations décrivant les réactions chimiques, il est nécessaire d'avoir des formules de transformation permettant d'exprimer l'équation 5.29 en terme de concentrations.

On remarque que la somme $\sum_\beta Q_B^\beta\ I_{\alpha\beta}$ est homogène à une concentration. On définit alors une *concentration de superposition* pour la bouffée α et l'espèce A, donnée par

$$\widehat{c_A^\alpha}(t) = \sum_\beta Q_A^\beta(t)\ I_{\alpha\beta}(t). \tag{5.30}$$

L'équation 5.29 s'écrit alors

$$\Delta Q_A^\alpha = -k\ \Delta t\ Q_A^\alpha\ \widehat{c_A^\alpha}(t). \tag{5.31}$$

Pour mieux comprendre ce que signifie la concentration de superposition, on fait l'hypothèse des concentrations uniformes, et des bouffées ayant un volume fini. L'équation 5.30 peut s'écrire

$$\widehat{c_A^\alpha} = \sum_\beta Q_A^\beta\ \frac{V_{\alpha\beta}}{V_\alpha V_\beta} = \sum_\beta c_A^\beta\ \frac{V_{\alpha\beta}}{V_\alpha}. \tag{5.32}$$

L'équation 5.32 montre que la concentration de superposition correspond à la concentration obtenue si toutes les quantités contenues dans les volumes de superposition de la bouffée α avec

les autres bouffées étaient diluées dans le volume V_α. S'il n'y a que deux bouffées superposées, et que $\beta = \alpha$, alors $V_{\alpha\beta} = V_\alpha$ et $\widehat{c_A^\alpha} = c_A^\alpha$. Si les deux bouffées ne sont que partiellement superposées, $V_{\alpha\beta} \leq V_\alpha$. La quantité de l'espèce A provenant de la bouffée β dans le volume de superposition est $V_{\alpha\beta} c_A^\beta$. Dans l'équation 5.32, cette quantité est diluée dans le volume de la bouffée α.

On repasse à présent au cas général. Si l'on calcule le taux de disparition de l'espèce A dans la bouffée α, en prenant en compte les concentrations de superposition, on obtient

$$\Delta \widehat{c_A^\alpha} = -k \ \Delta t \ \widehat{c_A^\alpha}(t) \ \widehat{c_B^\alpha}(t). \tag{5.33}$$

Une fois les réactions prises en compte avec les concentrations de superposition, on cherche à repasser des concentrations aux quantités transportées par les bouffées, et donc à retrouver l'équation 5.33 :

$$\Delta Q_A^\alpha = \Delta \widehat{c_A^\alpha} \times \frac{Q_A^\alpha(t)}{\widehat{c_A^\alpha}(t)}. \tag{5.34}$$

Le facteur multiplicatif est homogène à un volume, on le nomme donc *volume de superposition*, noté $\widehat{V_A^\alpha}$. Il dépend de l'espèce et de la bouffée

$$\widehat{V_A^\alpha} = \frac{Q_A^\alpha}{\widehat{c_A^\alpha}}. \tag{5.35}$$

Pour une espèce ayant une concentration de superposition nulle au début du pas de temps ($\widehat{c_A^\alpha}(t) = 0$), le volume de superposition est calculé avec la formule obtenue d'après l'équation 5.35 en considérant que toutes les concentrations des bouffées sont uniformes et identiques dans toutes les bouffées, d'une valeur notée C :

$$\widehat{V_A^\alpha} = \frac{C \ V_\alpha}{\sum_\beta C \ V_\beta \ I_{\alpha\beta}} = \frac{V_\alpha^2}{\sum_\beta \ V_{\alpha\beta}}. \tag{5.36}$$

Ou encore, pour exprimer le volume de superposition en fonction des intégrales I_α, I_β et $I_{\alpha\beta}$:

$$\widehat{V_A^\alpha} = \frac{1}{\sum_\beta \ I_{\alpha\beta} \ I_\alpha / I_\beta}. \tag{5.37}$$

Prise en compte des interactions avec les concentrations de fond (méthode 2)

Si l'on ajoute les concentrations de fond aux concentrations de superposition, pour prendre en compte à la fois les interactions entre les bouffées et avec le fond, l'équation 5.19 est modifiée de la façon suivante :

$$\frac{\mathrm{d}(\widehat{c_A^\alpha} + c_A^b)}{\mathrm{d}t} = -k(\underbrace{\widehat{c_A^\alpha} \ \widehat{c_B^\alpha}}_{(1)} + \underbrace{c_A^b \ c_B^b}_{(2)} + \underbrace{\widehat{c_A^\alpha} \ c_B^b + \widehat{c_B^\alpha} \ c_A^b}_{(3)}). \tag{5.38}$$

Ici, le terme (3) prend en compte l'interaction entre le fond et toutes les bouffées superposées, et non seulement l'interaction entre le fond et la bouffée α. Si l'on soustrait la contribution du terme (2), de la même façon qu'en partie 5.1.1, il reste donc

$$\Delta \widehat{c_A^\alpha} = -k \ \Delta t \ \left(\widehat{c_A^\alpha}(t) \ \widehat{c_B^\alpha}(t) + \widehat{c_A^\alpha} \ c_B^b + \widehat{c_B^\alpha} \ c_A^b \right). \tag{5.39}$$

Le terme additionnel pour les interactions avec le fond prend donc en compte les concentrations de superposition. Il serait également possible de ne calculer les interactions qu'entre les bouffées et le fond pour chacune des bouffées (sans interactions entre les bouffées). Dans tous les cas, il est possible d'avoir une concentration négative au total (même en ajoutant les concentrations de fond), dans le cas où la concentration de fond d'une espèce serait très faible. En effet, dans ce cas, chaque bouffée présente dans la maille interagit avec l'espèce de fond comme si l'ensemble de l'espèce de fond était consommable par la bouffée. En théorie, le cadre d'application du modèle de panache sous-maille garantit que ce cas ne se produit pas, car le volume des bouffées est beaucoup plus petit que le volume des mailles. On suppose donc qu'il reste suffisamment de quantités de fond dans la maille pour être consommées par toutes les bouffées présentes. Cette hypothèse peut cependant être invalidée si le nombre de bouffées dans une même maille est trop grand.

Une autre façon de faire est de traiter les mailles où se situent les bouffées de la même façon qu'une autre bouffée, dont on connaîtrait la quantité totale (d'après la concentration et le volume de la maille). Le volume d'interaction entre une maille et la bouffée α serait donc égal à zéro si le centre de la bouffée n'est pas dans la maille, et V_α sinon. Ainsi, le coefficient d'interaction entre la maille et la bouffée est

$$I_{\alpha b} = 1/V_b, \tag{5.40}$$

et la concentration de superposition (dans le cas où il n'y a qu'une seule bouffée dans la maille) pour la bouffée α est

$$\widehat{c_A^\alpha}(t) = c_A^\alpha + c_A^b. \tag{5.41}$$

Cela correspond exactement à la concentration totale utilisée dans la partie 5.1.1 et notée $\widetilde{c_A^\alpha}(t)$. La concentration de superposition calculée dans la maille est alors

$$\widehat{c_A^b}(t) = Q_A^\alpha/V_b + c_A^b. \tag{5.42}$$

Après avoir pris en compte la chimie pour les concentrations de superposition, on peut recalculer les quantités respectives de la bouffée et du fond, grâce aux volumes de superposition (équation 5.34) :

$$\Delta Q_A^\alpha = -k\Delta t\ Q_A^\alpha \left(c_B^\alpha + c_B^b\right) \tag{5.43}$$

$$\Delta Q_A^b = -k\Delta t\ Q_A^b \left(c_B^\alpha\ V_\alpha/V_b + c_B^b\right), \tag{5.44}$$

ou encore

$$\Delta Q_A^\alpha = -k\Delta t\ V_\alpha \left(c_A^\alpha\ c_B^\alpha + c_A^\alpha\ c_B^b\right) \tag{5.45}$$

$$\Delta Q_A^b = -k\Delta t\ V_b \left(c_A^b\ c_B^\alpha\ V_\alpha/V_b + c_A^b\ c_B^b\right). \tag{5.46}$$

Dans la précédente méthode pour prendre en compte les interactions avec le fond, le terme (3) de l'équation 5.19 donnait la quantité de l'espèce A qui était modifiée par les interactions avec le fond, dans le volume V_α. Cette quantité s'écrit

$$\Delta Q_A = \underbrace{-k\Delta t\ V_\alpha\ c_A^\alpha\ c_B^b}_{(a)}\underbrace{-k\Delta t\ V_\alpha\ c_B^\alpha\ c_A^b}_{(b)}. \tag{5.47}$$

Ici, la partie (a) de cette quantité est distribuée à la bouffée α, et la partie (b) est ajoutée à la quantité de A dans la maille. Ainsi, si l'espèce A est une espèce de fond, non présente dans

la bouffée (typiquement le cas de l'ozone), (a) est égal à zéro et la perturbation est ajoutée aux concentrations de fond. De la sorte, on perd l'information sur la perturbation locale de la concentration de A dans le panache, puisque cette perturbation est diluée dans la maille.

L'avantage de cette méthode par rapport à la précédente est de gagner en temps de calcul. En effet, la chimie dans la maille de fond est encore calculée, mais seulement une seule fois par maille contenant des bouffées. Dans la méthode précédente, la chimie de la maille de fond était calculée pour chaque bouffée présente dans cette maille. Or, de nombreuses bouffées provenant de la même source ponctuelle peuvent se situer dans la même maille.

Remarque : Dans tous les cas, on a supposé qu'une bouffée était majoritairement contenue dans une seule maille, où l'on prenait les valeurs des concentrations de fond (supposées uniformes dans la bouffée). Cependant, les bouffées peuvent rapidement couvrir plusieurs niveaux verticaux. Dans la méthode 1, la bouffée transporte la perturbation un certain temps, avant d'être injectée dans le modèle eulérien. A ce moment-là, les quantités contenues dans la bouffée (y compris négatives dans le cas d'une perturbation négative d'une espèce de fond) sont réparties dans les différents niveaux verticaux couverts par la bouffée. L'ajout de cette perturbation dans d'autres niveaux que celui du centre de la bouffée est donc une approximation supplémentaire, et les concentrations de fond sont mises à zéro si la perturbation négative était trop grande par rapport à la concentration de fond d'une espèce dans l'un des niveaux. Dans le cas de la méthode 2, une partie de cette perturbation est injectée dans la maille du centre de la bouffée à chaque pas de temps. En revanche, les formules de cette méthode ne sont plus valables si le volume de la bouffée est plus grand que celui de la maille, ce qui peut se produire si la bouffée a une extension verticale beaucoup plus grande que la maille où se trouve son centre. Dans ce cas, le volume d'interaction entre la bouffée et la maille est pris égal à V_b. Il est cependant préférable d'injecter la bouffée dans le modèle eulérien avant que cela ne se produise.

5.2 Application du panache sous-maille à la chimie de l'ozone

5.2.1 Chimie de l'ozone

Dans la suite, le modèle de panache sous-maille réactif est appliquée à la modélisation de la chimie troposphérique et en particulier de l'ozone. Il est donc utile de décrire brièvement les principales réactions en jeu. Pour une description plus détaillée, on se référera à Jacob [1999], Seinfeld et Pandis [1998] et Sportisse [2007a].

Réaction avec les NOx

La réaction de production de l'ozone fait intervenir le dioxygène O_2 ainsi que l'oxygène atomique O, suivant

$$O_2 + O + M \longrightarrow O_3 + M \tag{R 5.5}$$

où M est un tiers corps qui est en général N_2 ou O_2.

L'oxygène atomique O peut provenir de la photolyse de O_2

$$O_2 + h\nu \longrightarrow O + O \tag{R 5.6}$$

avec $h\nu$ correspondant à un rayonnement de longueur d'onde $\lambda \leq 242$ nm. Cette longueur d'onde est principalement absorbée dans la stratosphère, où ce type de réaction peut donc avoir

lieu. Dans la troposphère, la source d'oxygène moléculaire est la dissociation photolytique de NO_2

$$NO_2 + h\nu \xrightarrow{J_{NO_2}} NO + O \tag{R 5.7}$$

avec $h\nu$ correspondant à un rayonnement de longueur d'onde 300 nm $\leq \lambda \leq$ 400 nm. La réaction de production de l'ozone dépend donc directement de la dissociation photolytique du NO_2. D'autre part, la réaction de destruction de l'ozone par oxydation de NO, appelée « titration » à cause de la rapidité de sa cinétique, s'écrit

$$O_3 + NO \xrightarrow{k} NO_2 + O_2 \tag{R 5.8}$$

Si l'on suppose que NO est à l'équilibre (les réactions de production et de consommation ont lieu à la même vitesse), on obtient d'après les réactions R 5.7 et R 5.8

$$c_{O_3} \simeq \left(\frac{J_{NO_2}}{k}\right)\left(\frac{c_{NO_2}}{c_{NO}}\right). \tag{5.48}$$

En pratique, cependant, cet équilibre est perturbé par de nombreux facteurs tels que les émissions, ainsi que par une oxydation accrue de NO en NO_2 par d'autres oxydants que l'ozone.

L'ozone est aussi détruit par photolyse :

$$O_3 + h\nu \longrightarrow O_2 + O^{1D} \tag{R 5.9}$$

où O^{1D} est un atome d'oxygène excité.

Réaction avec les COV

La production d'ozone repose également sur l'oxydation des COV, c'est-à-dire des molécules organiques telles que les alcanes, les alcènes ou les aldéhydes. Si l'on prend un hydrocarbure générique noté RH, celui-ci est oxydé par OH pour produire des espèces telles que des peroxy-radicaux RO_2 ou des radicaux carbonyles/aldéhydes R'CHO (avec R' une chaîne contenant moins d'atomes de carbone que la chaîne initiale). En présence de NO, ces radicaux produisent du NO_2. Le bilan de cette chaîne de réaction s'écrit

$$RH + 2\,O_2 + 2\,NO \longrightarrow 2\,NO_2 + H_2O + R'CHO \tag{R 5.10}$$

Le NO_2 ainsi produit contribue à la formation d'ozone via les réactions R 5.7 et R 5.5. Une façon d'écrire le bilan global de production d'ozone par les COV est de considérer que la production est catalysée par les NOx

$$RH \xrightarrow{NOx} R'CHO + 2\,O_3 \tag{R 5.11}$$

Pour que la réaction R 5.11 ait lieu, il faut la présence du radical hydroxyle OH afin d'initier la chaîne. Ce radical a différentes sources, dont les plus importantes sont, de jour, la réaction

$$O^{1D} + H_2O \longrightarrow 2\,OH \tag{R 5.12}$$

avec O^{1D} provenant de la photolyse de l'ozone (réaction R 5.9), et la photolyse de l'acide nitreux

$$HONO + h\nu \longrightarrow OH + NO \tag{R 5.13}$$

De nuit, ces réactions n'ont pas lieu par faute de photolyse, et les concentrations de radicaux OH sont négligeables

Les COV contribuent donc à la production de l'ozone. Des réactions concurrentes peuvent diminuer l'efficacité de ces réactions, notamment la formation de HNO_3 qui diminue les concentrations de NO_2 et OH via

$$NO_2 + OH + M \longrightarrow HNO_3 + M \tag{R 5.14}$$

et la formation de peroxydes (peroxydes d'hydrogène ou peroxydes organiques) telle que

$$HO_2 + HO_2 \longrightarrow H_2O_2 + O_2 \cdot \tag{R 5.15}$$

Régimes chimiques

La présence de NOx peut donc conduire soit à la diminution de l'ozone, via la réaction de titration R 5.8, soit à la production d'ozone en catalysant la réaction R 5.11. Dans un régime pauvre en NOx (ou NOx-limité), la production d'ozone est limitée par les concentrations de NOx. Une diminution des COV aura donc peu d'effet sur l'ozone, alors qu'une diminution des NOx sera favorable. Dans un régime COV-limité, une augmentation des COV conduit à une augmentation de l'ozone, de même qu'une diminution des NOx (via la diminution de la titration). Dans ce dernier cas, une stratégie de réduction des émissions de NOx est donc défavorable car cela conduit à une augmentation de O_3. Il arrive fréquemment que des agglomérations soient en régime COV-limité. Le rapport entre les concentrations de NOx et COV fournit une indication du régime.

Chimie dans un panache

Dans un panache, les réactions chimiques prépondérantes ne sont pas les mêmes que dans le reste de l'atmosphère. D'après Karamchandani *et al.* [1998], on peut distinguer trois phases dans l'évolution d'un panache, illustrées par la figure 5.2 :

1. Quand le panache vient juste d'être émis, en champ très proche de la cheminée, il est essentiellement composé de NOx et est très peu mélangé au reste de l'atmosphère. On peut alors négliger le radical OH dans la composition du panache, ce qui inhibe la formation d'ozone par les COV qui débute par une oxydation (réaction R 5.11). La chimie dans cette phase de développement du panache est donc réduite à l'équilibre photostationnaire du système $NO/NO_2/O_3$ (réactions R 5.5 à R 5.8). Il faut y ajouter l'oxydation de NO qui a lieu en champ très proche

$$2\,NO + O_2 \longrightarrow 2\,NO_2 \tag{R 5.16}$$

   Cette réaction devient très rapidement négligeable, car son taux est proportionnel au carré des concentrations de NO, qui diminuent avec la dispersion du panache.

2. Dans l'étape suivante du développement du panache, les concentrations de NOx sont toujours plus grandes que celles des COV, et les concentrations d'ozone sont donc toujours plus faibles dans le panache que dans le milieu environnant. Dans cette phase, les concentrations de OH et HO_2 ne sont plus négligeables (produits notamment via les réactions de photolyse), et conduisent à l'oxydation de NO, NO_2 et SO_2 en HNO_2, HNO_3 et H_2SO_4 respectivement. On note en particulier l'oxydation de NO par HO_2

$$NO + HO_2 \longrightarrow NO_2 + OH \tag{R 5.17}$$

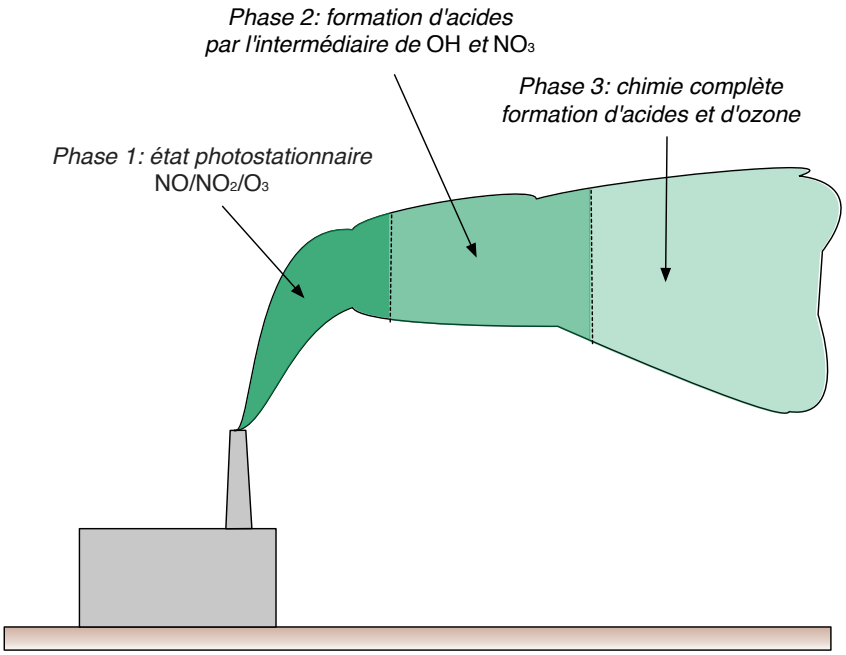


FIG. 5.2 – Illustration des trois phases d'évolution de la chimie dans un panache émis par une cheminée. D'après Karamchandani *et al.* [1998].

qui permet la formation de NO_2 sans consommer d'O_3, bien que la réaction de titration R 5.8 reste prépondérante à ce stade. De nuit, les réactions de photolyse n'ont plus lieu et la chimie impliquant NO_3 est prépondérante.

3. Enfin, lorsque le panache est suffisamment mélangé, les réactions impliquant la production d'ozone par les COV ne sont plus négligeables, et la chimie dans le panache est bien représentée par un mécanisme chimique complet.

Dans Karamchandani *et al.* [1998], des critères de transition entre les phases sont proposés et évalués pour différentes heures de la journée. Les résultats obtenus correspondent à une durée de 20 minutes à une demi-heure pour la première phase, et de trois à quatre heures pour la deuxième phase. Ces phases durent plus longtemps de nuit.

5.2.2 Impact du panache sous-maille réactif

Configuration

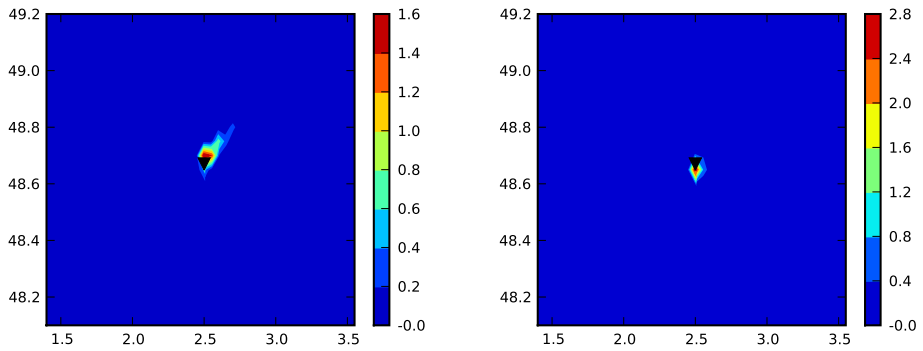
Les simulations présentées ici sont un test très simple. On utilise le domaine de simulation qui servira dans la partie 5.3. La résolution est de $\Delta x = \Delta y = 0.05°$ (environ 5 km), $N_y = 23$ et $N_x = 44$, y étant la latitude et x la longitude. Les données météorologiques proviennent d'ECMWF, pour la date du 2 avril 2001. La simulation est réalisée durant la journée. Le pas de temps est de $\Delta t = 60$ s, pour les modèles eulérien et gaussien ainsi que pour la discrétisation du panache en bouffées. La simulation dure une heure.

Afin d'observer ce qui se passe lorsque quelques espèces réagissent entre elles, seules les concentrations de fond pour l'ozone sont non nulles, avec une concentration homogène de 40 $\mu g\,m^{-3}$. On modélise une source ponctuelle au centre du domaine, qui émet des NOx avec les taux d'émission suivants : 10^7 $\mu g\,s^{-1}$ pour NO_2 et $2.1 \times 1\,0^7$ $\mu g\,s^{-1}$ pour NO. Ces conditions

ne sont pas un cas réel, mais permettent d'observer la titration de O_3 par NO. Il n'y a pas d'autres émissions, ni de conditions aux limites, de façon à ce que le panache soit clairement visible. La source est située à une altitude de 30 m, et le premier niveau vertical du modèle eulérien a pour hauteur 50 m. La simulation est réalisée avec le modèle de panache sous-maille ainsi qu'avec le modèle eulérien Polair3D seul (appelé « référence »). Le modèle de panache sous-maille est utilisé ici avec la théorie de la similitude, un temps d'injection d'une heure, et la méthode d'injection sur une colonne. Les concentrations de fond sont prises en compte de la façon décrite en partie 5.1.1.

Impact spatial du traitement sous-maille

La figure 5.3 montre les concentrations au sol (moyenne sur une heure) avec une seule source ponctuelle, pour NO. Les résultats sont montrés pour le panache sous-maille (figure 5.3(a)) ainsi que pour le modèle eulérien Polair3D sans traitement sous-maille (figure 5.3(b)). L'utilisation du modèle de panache sous-maille permet clairement de visualiser la forme du panache jusqu'à une certaine distance en aval de la source, avant d'être dilué, ce qui n'est pas le cas du modèle eulérien pour lequel le panache est réduit à la maille de la source. De plus, le panache obtenu par modèle eulérien est étalé autour de la source, sans direction privilégiée, y compris en amont de la source.

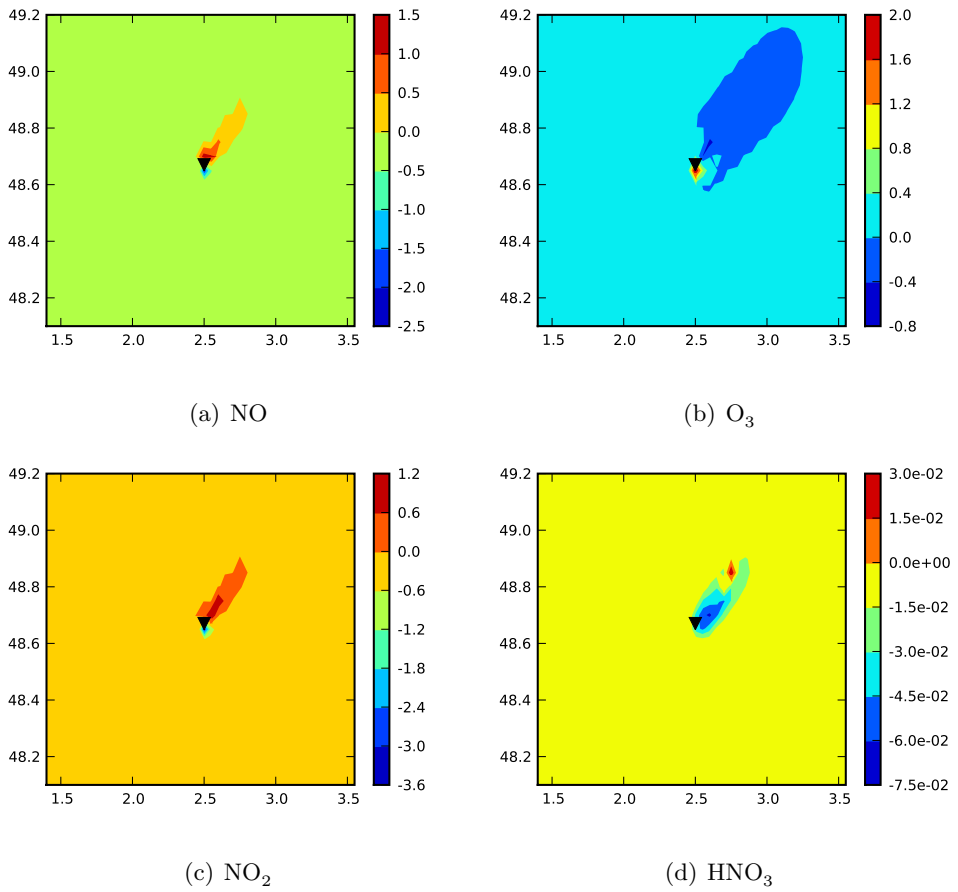


(a) NO – Panache sous-maille (b) NO – Eulérien

FIG. 5.3 – Concentrations au sol (moyenne sur une heure), en $\mu g\,m^{-3}$, de NO avec et sans traitement sous-maille des émissions. Simulation avec une seule source ponctuelle de NOx dans un fond initialement uniforme de O_3 à 40 $\mu g\,m^{-3}$. L'emplacement de la source est marqué par un triangle.

La figure 5.4 montre les différences entre les concentrations au sol du modèle de panache sous-maille et du modèle de référence, pour NO, NO_2, O_3 et HNO_3. La figure 5.4(a) confirme que les concentrations simulées par le panache sous-maille sont plus faibles que celles de référence à l'emplacement de la source, puis deviennent plus grandes (lorsque le panache touche le sol). Comme la réaction de titration R 5.8 est prépondérante dans le panache, les concentrations de O_3 données par le panache sous-maille sont plus grandes en champ proche (pas de titration au sol car le panache est en hauteur), puis plus faibles en aval (figure 5.4(b)). Le NO_2 étant un produit de cette titration, il est donc plus faible en champ proche et plus élevé en aval (figure 5.4(c)). Enfin, la production d'acide nitrique est beaucoup plus importante dans le modèle eulérien que dans le modèle de panache sous-maille (figure 5.4(d)), où elle ne débute qu'à une certaine distance (environ 30 km) en aval de la source. Comme le modèle eulérien dilue

prématurément le panache, la formation de radicaux OH et donc d'acides est en effet quasiment immédiate, alors qu'elle ne se produit théoriquement qu'au bout d'un certain temps (phase 2 de la chimie dans un panache décrite en partie 5.2.1).



(a) NO (b) O_3

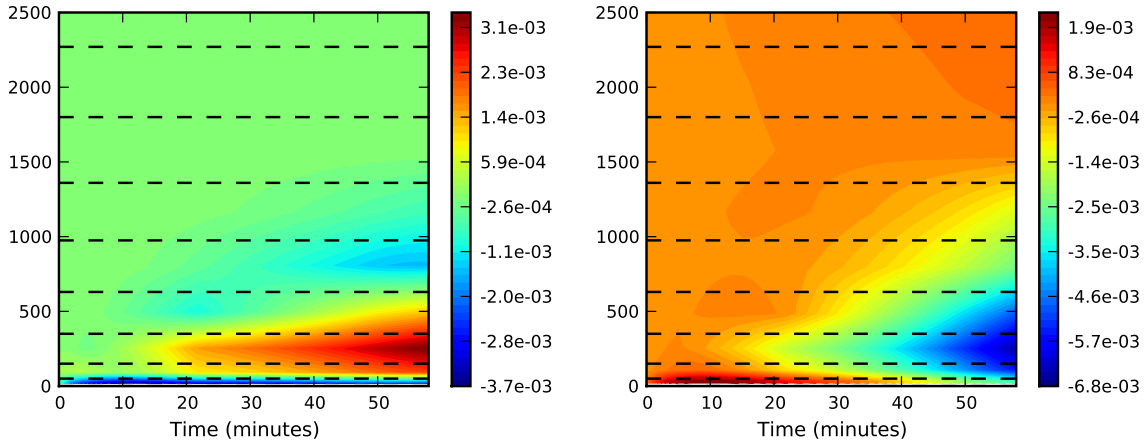
(c) NO_2 (d) HNO_3

FIG. 5.4 – Différences entre les concentrations au sol (moyenne sur une heure) en $\mu g\,m^{-3}$, avec et sans traitement sous-maille des émissions (panache sous-maille moins la référence). Simulation avec une seule source ponctuelle de NOx dans un fond initialement uniforme de O_3 à 40 $\mu g\,m^{-3}$. L'emplacement de la source est marqué par un triangle. Les résultats sont montrés pour NO, NO_2, O_3 et HNO_3.

Enfin, la figure 5.5 confirme le fait que la titration de O_3 pour former NO_2 est plus importante dans le modèle de panache sous-maille, mais maintenue en hauteur. Cette figure montre en effet les profils verticaux, en moyenne sur tout le domaine, en fonction du temps. Au début ($t = 0$), la source commence à émettre, et le profil correspond à l'impact du modèle de panache sous-maille en champ très proche. Ensuite, il s'agit de la moyenne, à chaque pas de temps, de l'impact sur tout le domaine. Les profils montrent l'impact cumulatif du modèle de panache sous-maille, en champ proche et en aval de la source.

Masse totale

La figure 5.6 montre l'évolution de la masse totale du panache pour NO, NO_2 et O_3. La masse totale du panache est donnée par la somme des quantités transportées par toutes les bouffées



(a) NO_2 (b) O_3

FIG. 5.5 – Différences entre les profils verticaux en fonction du temps (moyenne sur tout le domaine), avec et sans traitement sous-maille des émissions (panache sous-maille moins la référence). Les concentrations sont en $\mu g\,m^{-3}$, le temps (abscisse) est en minutes. Les interfaces des niveaux verticaux sont indiquées (lignes en pointillés). Simulation avec une seule source ponctuelle de NOx dans un fond initialement uniforme de O_3 à 40 $\mu g\,m^{-3}$. Les résultats sont montrés pour NO_2 et O_3.

non encore injectées dans le modèle eulérien. Étant donné que les bouffées transportent des perturbations négatives des concentrations de fond, la diminution des concentrations d'ozone est représentée par une perturbation négative dans les bouffées. La masse totale d'O_3 attribuée panache est donc négative, et représente la quantité totale d'ozone qui a été titrée. Bien que le taux d'émission de NO soit supérieur au double du taux d'émission de NO_2, la masse de NO_2 dans le panache dépasse celle de NO après dix minutes environ, à cause de la production de NO_2 par la titration de l'ozone. La source émet en continu (une bouffée par minute) et, au bout d'une heure, les premières bouffées commencent à être injectées dans le modèle eulérien. Elles ne sont donc plus prises en compte dans la masse totale du panache, ce qui explique que celle-ci devient pratiquement constante : à chaque pas de temps, il y a autant de bouffées émises que de bouffées ayant atteint le temps d'injection transférées dans le modèle eulérien.

L'utilisation du modèle de panache sous-maille permet de mieux représenter la chimie qui a lieu dans le panache, et modifie la répartition spatiale des concentrations. Il est intéressant de vérifier si la masse totale des polluants (dans tout le domaine) en est modifiée de façon significative. La figure 5.7 montre cette masse totale, avec et sans représentation sous-maille du panache. L'effet du panache sous-maille est principalement de modifier la répartition spatiale des polluants, la masse totale n'étant pas beaucoup modifiée par rapport à celle obtenue avec le modèle eulérien seul. Au début de la simulation, il y a moins de NO_2 créé dans le modèle de panache sous-maille car les bouffées sont petites et non encore superposées. Après un certain temps, les bouffées sont assez grandes pour que la quantité d'ozone contenue dans leur volume soit importante, et les concentrations de NOx dans les bouffées sont toujours relativement élevées. Ainsi, il y a une production légèrement plus grande de NO_2 avec le traitement sous-maille du panache.

Il s'agit ici de tests simples destinés à visualiser clairement l'impact du modèle de panache sous-maille sur un panache. En pratique, le milieu ambient contient également des COV, qui

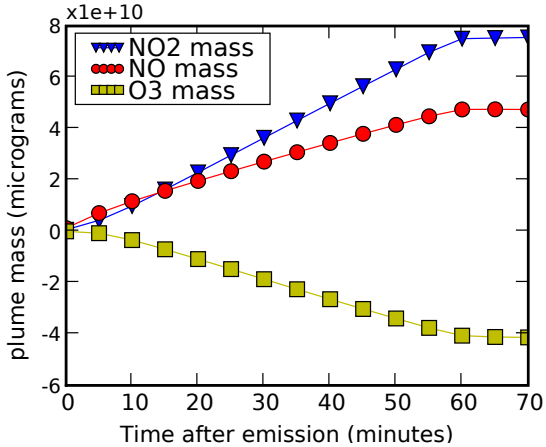


FIG. 5.6 – Masse totale du panache, en µg, pour NO_2, NO et O_3 avec le modèle de panache sous-maille et une seule source ponctuelle.

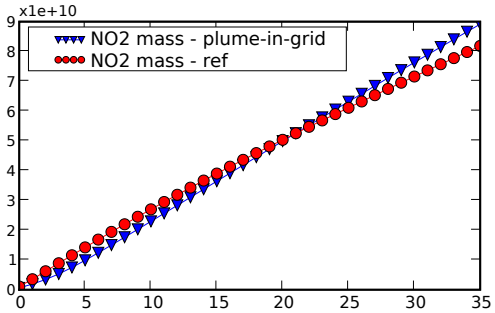


FIG. 5.7 – Masse totale, en µg, dans tout le domaine, pour NO_2 avec le modèle de panache sous-maille et Polair3D seul (« ref »). Simulation avec une seule source ponctuelle.

peuvent intervenir dans la dernière phase d'évolution du panache (phase 3 décrite dans la partie 5.2.1).

5.2.3 Influence du nombre de sources

On réalise une simulation identique à la précédente, mais en prenant 40 sources ponctuelles situées exactement au même point, avec un taux d'émission égal au 1/40e de la source utilisée précédemment. Le nombre élevé de sources ponctuelles a été choisi afin de tester le modèle lorsque de nombreuses bouffées interagissent entre elles dans la même maille. Cette simulation doit évidemment donner les mêmes résultats que celle avec une seule source. Il s'agit, en particulier, de vérifier que la prise en compte des concentrations de fond avec les bouffées superposées ne conduit pas à une sur-estimation de la titration. La figure 5.8 montre la masse totale du panache pour une source et pour 40 sources. Les masses des différentes espèces sont les mêmes pour les deux simulations. La différence maximale entre les masses de NO_2 pour une ou pour 40 source est inférieure à 1 g. De plus, dans la simulation avec 40 sources, il peut y avoir jusqu'à 1200 bouffées dans la même maille, sans faire apparaître de problème numérique. La prise en compte des interactions avec les concentrations de fond décrite en partie 5.1.1 est donc valide, dans ce contexte. Des problèmes numériques dans la résolution de la chimie pourraient cependant se produire avec un très grand nombre de bouffées comme dans le cas présent, en particulier, si une espèce de fond ayant une concentration très faible réagit avec toutes les bouffées. Ce genre de problème se résout en prenant un pas de temps plus faible pour la chimie : dans les simulations suivantes, un pas de temps adaptatif pour la chimie est utilisé afin d'éviter ce type de problème.

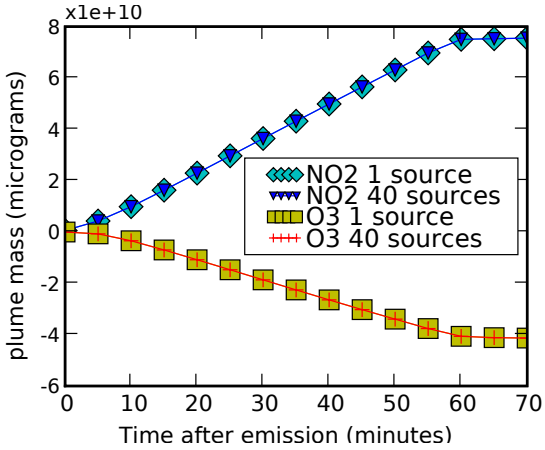


FIG. 5.8 – Masse totale du panache, en µg, pour NO_2, NO et O_3 avec le modèle de panache sous-maille. Résultats pour 1 seule source ponctuelle, et 40 sources superposées donnant le même taux d'émission au total.

5.3 Application à la qualité de l'air en Île-de-France

5.3.1 Configuration

Le modèle de panache sous-maille est utilisé pour la qualité de l'air en Île-de-France, pour l'été 2001 (du 1er avril au 27 septembre). Le domaine de simulation couvre l' Île-de-France, de 1.40° E à 3.55° E (44 mailles) et de 48.10° N à 49.20° N (23 mailles) (Figure 5.9). La résolution en longitude et latitude est de 0.05°, ce qui correspond environ à 5 km. Cette configuration

correspond à celle utilisée dans Tombette et Sportisse [2007], pour la modélisation des particules sur la région parisienne. Le pas de temps de la simulation eulérienne Polair3D est de 100 s, et est utilisé également pour l'advection et la diffusion des bouffées dans le modèle de panache sous-maille.

Il y a neuf niveaux verticaux, de taille croissante en partant du sol jusqu'à 2730 m, le premier niveau vertical ayant une hauteur de 50 m. Les champs météorologiques sont interpolés depuis des données ECMWF. Les conditions aux limites sont prises d'après une simulation à l'échelle continentale, à une résolution de 0.5° ($\simeq$ 50 km).

Le coefficient de diffusion verticale est calculé avec la paramétrisation de Troen-Mahrt [Troen et Mahrt, 1986] dans la couche limite, en imposant une valeur minimale de $K_z = 0.5\ \mathrm{m^2\,s^{-1}}$ pour les zones urbaines et $K_z = 0.2\ \mathrm{m^2\,s^{-1}}$ pour les zones rurales. La valeur plus élevée en zone urbaine permet de prendre en compte l'augmentation de la diffusion verticale due en particulier au phénomène d'îlot de chaleur urbain. L'impact de ce changement sur les zones urbaines est évalué dans la partie 5.6.1. La paramétrisation de Louis [Louis, 1979] est utilisée au-dessus de la couche limite.

Seule la chimie gazeuse est prise en compte, avec le mécanisme RACM (Regional Atmospheric Chemistry Mechanism, Stockwell *et al.* [1997]). Il est utilisé à la fois dans le modèle eulérien Polair3D et pour calculer la chimie dans les bouffées.

5.3.2 Émissions

Les émissions proviennent de l'inventaire Airparif pour l'année 2000. Les émissions surfaciques et volumiques sont interpolées sur le maillage de la simulation, tandis que les émissions provenant des sources ponctuelles importantes sont traitées séparément. Il y a au total 295 sources ponctuelles, parmi lesquelles on sélectionne les sources ayant un taux d'émission de NO_x ou de SO_2 supérieur à $Q_{\min} = 10^6\ \mu\mathrm{g\,s^{-1}}$. On obtient ainsi 89 sources ponctuelles qui sont traitées en sous-maille, les autres étant traitées directement par le modèle eulérien. Les sources sélectionnées pour le traitement sous-maille représentent 94% de la masse de NO_x et 98% de la masse de SO_2 emises par les sources ponctuelles. Cela ne correspond pas à la majorité des émissions totales, puisque le secteur industriel, qui contient en grande partie les sources ponctuelles, correspond à 20% des émissions de NO_x et 55% de celles de SO_2. On peut donc s'attendre à un impact plus important du traitement sous-maille pour le SO_2 que pour les autres espèces. L'emplacement des sources et des stations de mesure est montré figure 5.9 pour le SO_2 et figure 5.10 pour le NO. On se focalise ici sur les NO_x et le SO_2, mais le traitement sous-maille modélise l'ensemble des espèces émises par les sources, y compris les COV.

L'inventaire d'émissions fournit les coordonnées des sources ainsi que les taux d'émissions par espèce, mais ne contient pas les informations nécessaires au calcul de la surhauteur (température de l'émission, vitesse d'éjection et diamètre de la source). La surhauteur est donc calculée, dans les simulations, avec des valeurs estimées de $12\ \mathrm{m\,s^{-1}}$ pour la vitesse d'éjection, 100 °C pour la température d'émission, et 5 $\mathrm{m^2}$ pour la section de la cheminée. La surhauteur calculée avec ces paramètres est d'environ 30 m, allant jusqu'à 100 m dans certains cas. Les formules utilisées pour le calcul de la surhauteur sont les formules de Briggs, données en partie 2.3.1. L'utilisation de ces valeurs est une source supplémentaire d'incertitudes. Cependant, cela permet de prendre en compte les variations de surhauteur en fonction des conditions météorologiques. De plus, une approximation sur la hauteur de la source est également faite pour les sources ponctuelles traitées par le modèle eulérien, puisqu'elles sont placées au centre d'un niveau vertical. Ainsi, plusieurs sources importantes sont situées entre 60 m et 80 m de hauteur, ce qui correspond dans le modèle eulérien à un niveau vertical dont le centre est à 100 m.

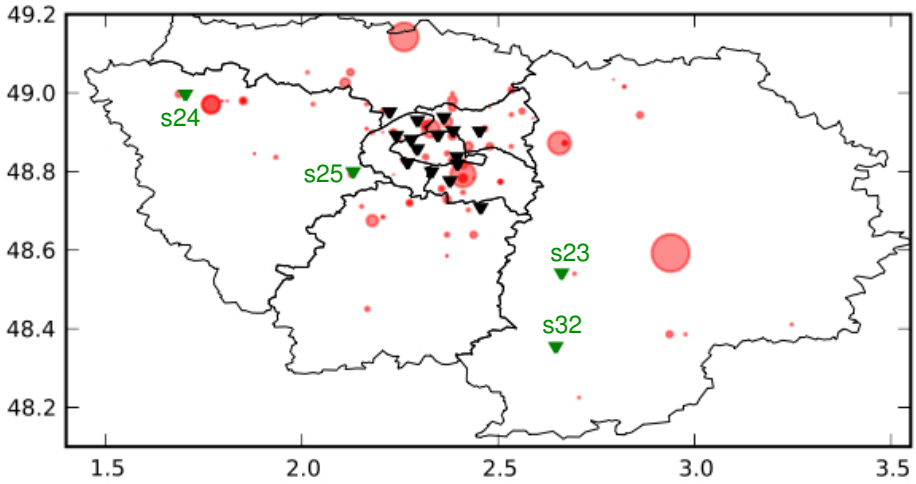


FIG. 5.9 – Position des stations de mesure (triangles) et des sources ponctuelles importantes, pour le SO_2 ; la surface des cercles est proportionnelle au taux d'émission de SO_2. Les stations de mesure sont représentées par des triangles noirs, et les stations périurbaines et rurales par des triangles verts.

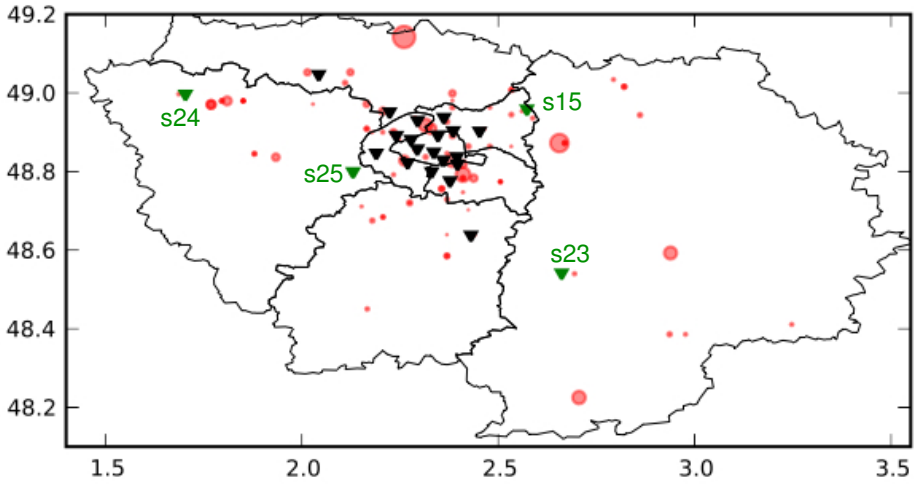


FIG. 5.10 – Position des stations de mesure (triangles) et des sources ponctuelles importantes, pour le NO ; la surface des cercles est proportionnelle au taux d'émission de NO. Les stations de mesure sont représentées par des triangles noirs, et les stations périurbaines et rurales par des triangles verts.

5.3.3 Configuration de panache sous-maille

Dans chaque bouffée, l'ensemble des espèces transportées subissent des transformations chimiques, calculées avec le mécanisme RACM. Cela augmente le temps de calcul de façon conséquente, en fonction du nombre de bouffées. Ce nombre dépend de deux paramètres, à nombre de sources fixé : le pas de temps entre les bouffées, et le temps d'injection. Ces deux paramètres sont donc choisis avec soin, de façon à limiter le temps de calcul. Le pas de temps entre deux bouffées est $\Delta t_{\text{puff}} = 100$ s, afin de correspondre au pas de temps de la simulation. Afin de garder un nombre constant de bouffées dans le modèle, on prend un critère d'injection sur l'âge de la bouffée, qui est fixé à 20 minutes. Cela correspond approximativement à la durée de la phase 1 de la chimie dans le panache (définie en partie 5.2.1), où les réactions prépondérantes sont différentes de celles qui ont lieu dans le milieu ambiant. Des simulations supplémentaires ont également été réalisées avec un temps d'injection de 40 minutes afin d'évaluer l'impact de ce paramètre (partie 5.6.2). Le modèle gaussien à bouffées est utilisé avec les trois paramétrisations d'écarts types : Briggs, Doury et la théorie de la similitude. La méthode d'injection utilisée est l'injection sur une colonne, car les simulations réalisées avec la méthode d'injection par intégration (la bouffée est injectée sur les mailles adjacentes horizontales aussi bien que verticales) donnent systématiquement des résultats légèrement moins bons, en raison probablement d'une trop grande dilution spatiale.

5.3.4 Simulations

Plusieurs simulations sont réalisées avec la configuration décrite en partie 5.3.1.

1. Une *simulation de référence*, avec le modèle eulérien Polair3D, où toutes les sources ponctuelles sont traitées par le modèle eulérien sans traitement sous-maille,
2. Trois simulations en utilisant un traitement sous-maille pour les 89 sources ponctuelles sélectionnées, pour les trois paramétrisations des écarts types, appelées *simulations avec panache sous-maille*,

Les résultats obtenus avec ces différentes simulations sont détaillés en partie 5.4. D'autres simulations ont été réalisées en complément afin de réaliser une étude de sensibilité, qui est décrite en partie 5.6.

5.4 Évaluation du modèle de panache sous-maille réactif

5.4.1 Critères d'évaluation

L'évaluation du modèle est effectuée sur les concentrations horaires, observées aux stations de mesure Airparif. Les indicateurs utilisés ici sont la RMSE (Root Mean Square Error), la corrélation, le biais fractionnel (MFBE) et l'erreur fractionnelle (MFGE). Pour l'ozone, des indicateurs supplémentaires sont utilisés conformément aux recommandations de l'EPA [EPA, 2005] : MNGE, MNBE et UPA. Les définitions de ces indicateurs sont données en annexe A.

Les critères de performance pour qu'un modèle soit « acceptable » sont un biais fractionnel de ±30% [Chang et Hanna, 2004], et une erreur fractionnelle plus faible que 50%. Les recommandations de l'EPA pour l'ozone sont [EPA, 1991] MNGE $\leq$ 35%, MNBE dans l'intervalle ±15% et UPA dans l'intervalle ±20%.

Les statistiques sont calculées pour les stations de mesure où au moins 60% des observations sont fournies pour l'ensemble de la période de simulation. Cela correspond à 19 stations pour le SO_2 (Figure 5.9), 21 stations pour O_3, 24 pour NO (Figure 5.10) et 26 pour NO_2, sur un total de 48 stations. Le calcul des indicateurs spécifiques à l'ozone se fait sur les valeurs supérieures à un certain seuil, suivant là encore les recommandations de l'EPA. Le seuil est pris ici à 30 $\mu g\,m^{-3}$.

5.4.2 Impact du panache sous-maille sur les statistiques

Statistiques globales

Le tableau 5.1 montre les statistiques sur toute la période et l'ensemble des stations, pour les concentrations horaires de SO_2, NO, O_3 et NO_2. Dans la plupart des cas, les critères de performance décrits en partie 5.4.1 sont remplis pour O_3 and NO_2. En revanche, le modèle surestime fortement les concentrations de SO_2 et NO.

Les résultats sont donnés pour la simulation de référence ainsi que pour les simulations avec le panache sous-maille avec les trois écarts types gaussiens. La théorie de la similitude donne les meilleurs résultats, suivie par la paramétrisation de Briggs. Celle de Doury est la moins bonne, mais sa performance reste en général supérieure au modèle de référence malgré tout. Alors que les paramétrisations de Briggs et de la théorie de la similitude donnaient de bons résultats à l'échelle locale (chapitre 3), celle de Doury était en revanche la meilleure sur le cas continental (chapitre 4). Ici, ce sont à nouveau les deux premières formules qui obtiennent une meilleure performance car l'échelle est beaucoup plus petite que celle de l'expérience d'ETEX par exemple.

L'utilisation d'un traitement sous-maille ne change pas de façon significative les statistiques globales. Ce n'est pas surprenant, si l'on considère d'une part la proportion relativement faible des émissions ponctuelles dans l'ensemble des émissions, et d'autre part le fait que les statistiques sont calculées sur des concentrations de fond, à des stations qui ne sont pas situées, en général, près des sources importantes. Cependant, l'amélioration est notable pour le NO et le SO_2. Le traitement sous-maille permet de réduire la RMSE de 9% dans le cas du SO_2 et de 4.5% pour le NO. Les résultats pour NO_2 et O_3 changent par contre assez peu. Il semble donc que le panache sous-maille ait un impact principalement sur les espèces primaires, plutôt que sur les polluants secondaires. De plus, les sources ponctuelles constituent une part plus importante des émissions pour le SO_2, alors que les NOx et l'ozone dépendent davantage des émissions dues au trafic routier. Enfin, le SO_2 est ici une espèce quasiment passive (on ne considère que la chimie gazeuse, qui est lente pour l'oxydation du SO_2), et l'impact sur ses concentrations peut donc être transporté sur de plus longues distances (partie 5.4.3).

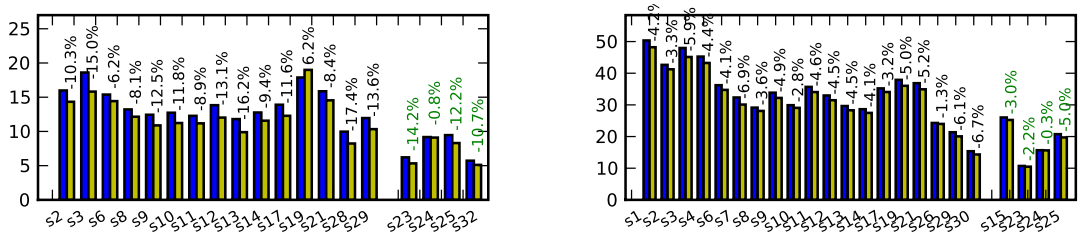
Résultats aux stations

L'impact sur les statistiques globales n'est pas très important, mais il varie en fonction des stations. Certaines stations, plus proches des sources ponctuelles, sont plus fortement impactées par le traitement sous-maille. La figure 5.11 montre la réduction de RMSE station par station, entre les statistiques du modèle de référence et celles du modèle de panache sous-maille. La paramétrisation utilisée ici est la théorie de la similitude, mais celle de Briggs donne des résultats similaires. La baisse de RMSE varie entre -6.2% et -17.4% suivant les stations (SO_2, figure 5.11(a)), et entre -1.3% et -6.9% (NO, figure 5.11(b)), aux stations urbaines. Les résultats aux stations péri-urbaines et rurales sont montrés séparément. Ces stations sont moins influencées par les émissions dues au trafic, et l'impact des sources ponctuelles y est donc, comparativement, plus élevé. Cependant, elles sont également plus éloignées des sources ponctuelles. Les statistiques globales y sont donc comparables à celles des stations urbaines, pour les moyennes sur six mois. Ces stations peuvent cependant montrer un impact important dans certaines conditions, lorsque le vent est tel que les stations sont situées en aval de certaines sources, dans le panache (partie 5.5).

Les résultats pour NO_2 et O_3 ne sont pas montrés aux stations. L'impact est en effet plus faible, et peu de différences sont observées entre les stations : la RMSE diminue de 0.2% à 2% pour l'ozone, et augmente de façon similaire pour le NO_2.

Indicateur	Obs	Référence	Doury	Briggs	Similitude
SO_2					
Moyenne ($\mu g\, m^{-3}$)	6.20	13.76	12.59	12.24	**11.98**
RMSE		13.08	12.16	12.06	**11.88**
Corrélation		**0.35**	0.34	0.33	0.31
MFBE		0.66	0.59	0.55	**0.55**
MFGE		0.81	0.77	0.76	**0.75**
NO					
Moyenne ($\mu g\, m^{-3}$)	10.42	20.93	20.26	19.94	**19.64**
RMSE		33.08	32.44	32.37	**31.62**
Corrélation		**0.47**	**0.47**	**0.47**	0.46
MFBE		0.41	0.39	0.36	**0.36**
MFGE		0.94	0.93	0.92	**0.92**
O_3					
Moyenne ($\mu g\, m^{-3}$)	56.87	40.24	40.58	40.94	**41.05**
RMSE		30.58	30.41	30.22	**30.18**
Corrélation		**0.68**	**0.68**	**0.68**	**0.68**
MFBE		-0.46	-0.45	**-0.44**	**-0.44**
MFGE		0.47	0.46	0.46	**0.45**
MNBE		-0.34	**-0.33**	**-0.33**	**-0.33**
MNGE		0.35	0.35	**0.34**	**0.34**
UPA		-0.21	-0.21	-0.21	**-0.20**
NO_2					
Moyenne ($\mu g\, m^{-3}$)	34.64	35.84	35.47	35.28	**35.23**
RMSE		**20.57**	20.64	20.74	20.81
Corrélation		**0.58**	0.57	0.57	0.57
MFBE		0.06	0.05	**0.04**	**0.04**
MFGE		**0.47**	**0.47**	**0.47**	**0.47**

TAB. 5.1 – Statistiques horaires pour le modèle de référence Polair3D ("Référence"), et les trois simulations sous-mailles correspondant aux trois paramétrisations des écarts types. Les statistiques sont calculées sur la période de simulation, du 01-04-2001 au 27-09-2001. Les meilleures statistiques sont en gras.



(a) SO_2 – RMSE aux stations (b) NO – RMSE aux stations

FIG. 5.11 – RMSE aux stations pour la période de simulation (avril à septembre), pour le modèle de référence (bleu) et le panache sous-maille avec la théorie de la similitude (jaune). La différence entre les deux RMSE, divisée par la RMSE du modèle de référence, est indiquée au-dessus des barres (en pourcentage). Le premier groupe de barres correspond aux stations urbaines (différence indiquée en noir), le second aux stations péri-urbaines et rurales (différence indiquée en vert). Figure 5.11(a) : RMSE pour le SO_2. Figure 5.11(b) : RMSE pour le NO.

5.4.3 Répartition spatiale de l'impact

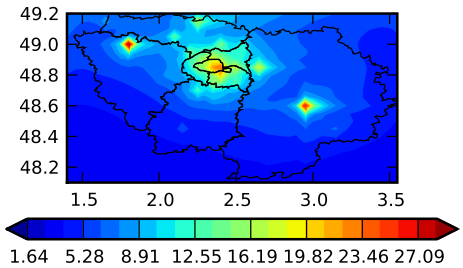
On étudie à présent l'impact spatial du traitement sous-maille des émissions sur les concentrations au sol. Les résultats sont à présent toujours montrés pour le modèle de panache sous-maille avec la théorie de la similitude, puisque c'est avec cette paramétrisation que sont obtenus les meilleurs résultats (partie 5.4.2).

Impact sur les concentrations au sol de SO_2

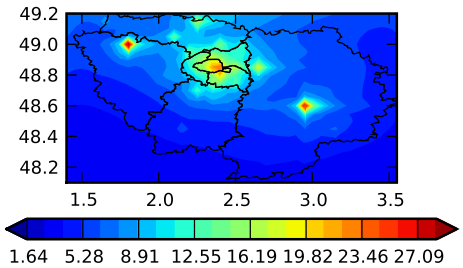
La figure 5.12 montre les concentrations au sol de SO_2, en moyenne sur la période de simulation. Les deux premières figures correspondent aux concentrations au sol pour la simulation de référence (5.12(a)) et la simulation de panache sous-maille (5.12(b)). Les deux figures suivantes correspondent à la différence entre les concentrations de référence et les concentrations de fond (5.12(c)), et à la différence entre les concentrations de référence et les concentrations de panache sous-maille (5.12(d)). Cela permet ainsi de mesurer l'impact du modèle panache sous-maille, à comparer avec l'impact total des sources ponctuelles sur les concentrations moyennes.

L'utilisation du modèle de panache sous-maille a pour effet principal de diminuer la concentration des espèces primaires à l'endroit des sources ponctuelles. Dans la plupart des cas, les sources sélectionnées ici sont au deuxième niveau vertical dans le modèle eulérien (entre 50 m et 150 m), parfois au troisième (au-dessus de 150 m). De la sorte, utiliser un traitement sous-maille permet de les maintenir plus longtemps en hauteur, alors que le modèle eulérien simule un panache touchant le sol rapidement. Les concentrations dans le panache, en altitude, sont plus importantes avec le modèle de panache sous-maille pour les espèces primaires. En aval des sources, lorsque le panache simulé en sous-maille touche le sol, il est possible que les concentrations au sol soient alors plus élevées que pour le modèle de référence. Cependant, cela n'est pas visible sur les cartes qui représentent une moyenne sur six mois, car il n'y a pas de direction de vent privilégiée. De plus, pour les sources proches du sol, la diffusion modélisée par le modèle de panache sous-maille peut être plus élevée, à proximité des sources, que la diffusion eulérienne (mis en évidence sur ETEX, chapitre 4). Cela peut donc également contribuer à cette baisse des concentrations. L'impact des sources ponctuelles est également plus localisé avec le modèle de panache sous-maille, car la dilution horizontale du panache due au modèle eulérien disparaît.

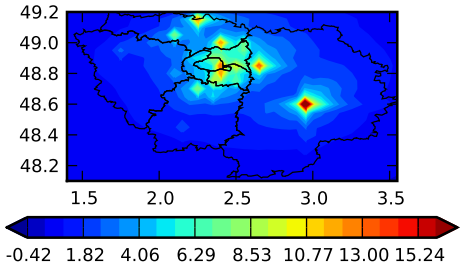
Sur la figure 5.12(a), on observe deux sources principales de SO_2, hors du centre urbain : au nord-ouest et au sud-ouest de la région. La première n'est pas une source ponctuelle (elle



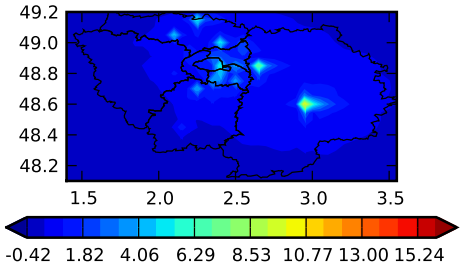
(a) SO_2 – Référence



(b) SO_2 – panache sous-maille



(c) SO_2 – Référence − fond



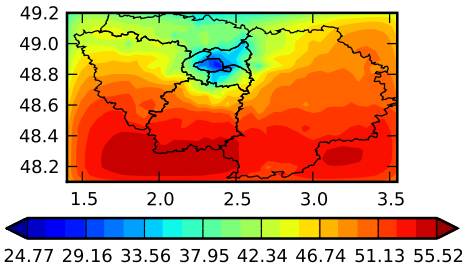
(d) SO_2 – Référence − panache sous-maille

FIG. 5.12 – Concentrations au sol de SO_2 sur la région parisienne, en moyenne sur les six mois de la simulation. Les figures 5.12(a) et 5.12(b) montrent les concentrations pour la simulation de référence et celle de panache sous-maille avec la théorie de la similitude, respectivement, en $\mu g\,m^{-3}$. La figure 5.12(c) montre la différence entre les concentrations de référence avec et sans les 89 sources ponctuelles. La figure 5.12(d) montre la différence entre les concentrations avec et sans le traitement sous-maille (référence moins les résultats avec panache sous-maille).

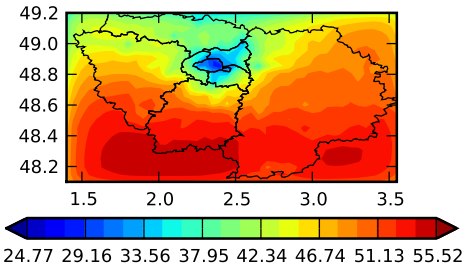
apparaît également sur la simulation qui exclut les sources ponctuelles, figure 5.12(c)), alors que la seconde est l'une des sources principales (visible sur la figure 5.9). C'est à cet endroit que l'impact du traitement sous-maille est le plus important (figure 5.12(d)).

Impact sur les concentrations au sol de O_3

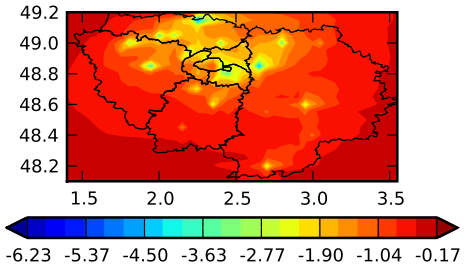
L'utilisation d'un modèle de panache sous-maille a également pour effet de diminuer les concentrations au sol de NOx, que l'on ne montre pas ici. Les concentrations d'ozone sont donc plus élevées, car la titration par les NOx est moins importante (figure 5.13(d)). Un effet similaire est constaté en comparant les simulations de référence avec et sans les sources ponctuelles (figure 5.13(c)). Dans cette région, le régime chimique est en général limité par les COV, et une diminution des concentrations de NOx a pour effet de diminuer la titration, donc d'augmenter les concentrations de O_3. Le régime COV-limité sur la région parisienne et dans les panaches a déjà été mis en évidence [Deguillaume *et al.*, 2008; Kim *et al.*, 2009]. Il dépend cependant de la météorologie, et des occurences de régimes NOx-limité peuvent également se produire [Sillman *et al.*, 2003; Honore *et al.*, 2000].



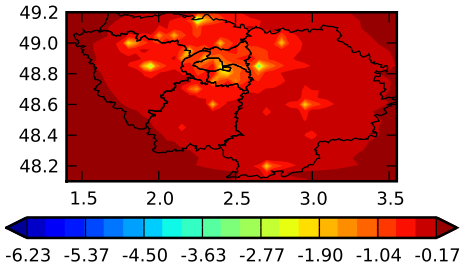
(a) O_3 – Référence



(b) O_3 – panache sous-maille



(c) O_3 – Référence – fond



(d) O_3 – Référence – panache sous-maille

FIG. 5.13 – Concentrations au sol de O_3 sur la région parisienne, en moyenne sur les six mois de la simulation. Les figures 5.13(a) et 5.13(b) montrent les concentrations pour la simulation de référence et celle de panache sous-maille avec la théorie de la similitude, respectivement, en $\mu g\,m^{-3}$. La figure 5.13(c) montre la différence entre les concentrations de référence avec et sans les 89 sources ponctuelles. La figure 5.13(d) montre la différence entre les concentrations avec et sans le traitement sous-maille (référence moins les résultats avec panache sous-maille).

5.5 Résultats pour certains jours

La partie précédente était consacrée à l'étude des résultats moyens sur l'ensemble des six mois simulés. Si l'utilisation du panache sous-maille permet d'améliorer les performances globales dans une certaine mesure, l'impact de ce traitement peut varier d'un jour à l'autre en fonction des conditions météorologiques. L'étude de certains jours bien choisis peut mettre en évidence les situations météorologiques où l'utilisation du panache sous-maille est particulièrement intéressante. Si l'on s'intéresse aux performances aux stations, les jours d'intérêt sont ceux où le vent transporte des panaches de sources ponctuelles dans la direction des stations de mesure. De plus, les situations de faible dispersion sont particulièrement intéressantes, car la surestimation de la dilution horizontale, et éventuellement verticale, du panache par le modèle eulérien est alors importante.

5.5.1 Impact sur les polluants primaires

On sélectionne deux jours consécutifs de faible dispersion, où la direction du vent est telle que certaines stations sont directement impactées par des panaches. Il s'agit du 23 et 24 août 2001. La vitesse du vent durant ces deux jours varie entre $0.4\,\mathrm{m\,s^{-1}}$ et $2.6\,\mathrm{m\,s^{-1}}$, avec une hauteur moyenne de couche limite de 600 m. Ces valeurs sont typiques de situations de faible dispersion, où la concentration des polluants primaires est particulièrement élevée.

Concentrations de SO_2

La figure 5.14 montre l'évolution de la différence entre les concentrations au sol de SO_2 avec la simulation de référence et la simulation de panache sous-maille, entre 3 heure du matin et 14 heure (heure locale), pour chaque heure, le 23 août. La direction du vent tourne durant cette période, et la vitesse du vent décroît. Les sept premières cartes montrent clairement le panache dû à la source située au sud-est de la région, dont le traitement sous-maille diminue les concentrations (différence positive). Le panache est transporté vers le sud/sud-ouest, et passe directement sur la station MELUN ("s23" sur la figure 5.9). On observe également une deuxième zone où l'impact du panache sous-maille est important, avec des concentrations beaucoup plus faibles que celles de la simulation de référence. Il s'agit du sud de Paris (12e arrondissement) et de la petite couronne, vers les stations VITRY-sur-SEINE et IVRY-sur-SEINE notamment.

Sur les cinq dernières cartes de la figure 5.14, la situation est très stable, avec un vent très faible (de l'ordre de $0.5\,\mathrm{m\,s^{-1}}$), et les concentrations sont très élevées. Dans le modèle de panache sous-maille, les bouffées ne sont pas transportées sur une grande distance avant d'être transférées au modèle eulérien (600 m pour un temps d'injection de 20 minutes). Les concentrations des bouffées sont particulièrement élevées, étant donné que la diffusion est très faible. Ainsi, lorsque les bouffées sont injectées dans le modèle eulérien et touchent le sol, les concentrations restent plus importantes que pour la simulation de référence. Cela se traduit par une différence importante (négative) en aval de certaines sources.

La figure 5.15 montre les profils de SO_2 pour six stations durant les deux jours étudiés. Les simulations de référence et de panache sous-maille sont comparées aux observations. Ici encore, les concentrations sont globalement surestimées. La première station, MELUN, est rurale, et montre l'impact du panache situé au sud-est au début de la période. En dehors de cet impact, les concentrations y sont globalement faibles. L'utilisation d'un traitement sous-maille ne modifie pas le temps d'arrivée du panache. Celui-ci arrive légèrement plus tard, car la diffusion horizontale est plus faible dans le modèle de panache sous-maille, mais cette différence est négligeable sur les moyennes horaires de concentrations utilisées ici. Les trois autres profils sont aux stations VITRY-SUR-SEINE, IVRY-SUR-SEINE and PARIS12eme, dans la zone où

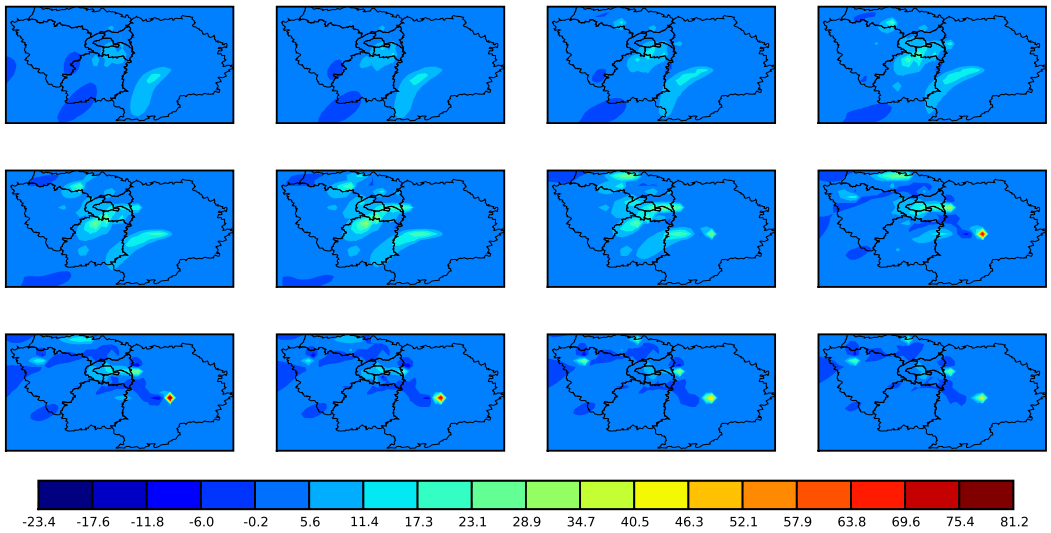


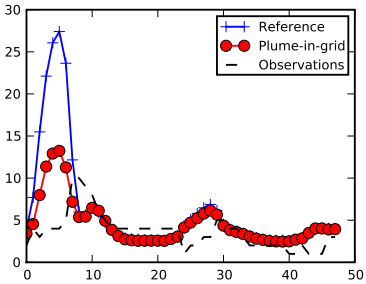
FIG. 5.14 – Évolution de la différence entre la simulation de référence et celle de panache sous-maille pour les concentrations au sol de SO_2 durant 12 heures, le 23 août de 3 heures du matin à 14 heures. Les concentrations sont données en $\mu g\,m^{-3}$. Les concentrations de panache sous-maille sont soustraites aux concentrations de référence.

l'impact du panache sous-maille est également important (Figure 5.14). Les profils du modèle de panache sous-maille sont en effet beaucoup plus proches des observations que ceux du modèle de référence. Enfin, on donne également les profils pour deux stations situées au nord de Paris, en amont de la plupart des sources, afin de fournir un élément de comparaison. Ici, l'impact du modèle de panache sous-maille est beaucoup plus faible, mais permet malgré tout de diminuer la sur-estimation.

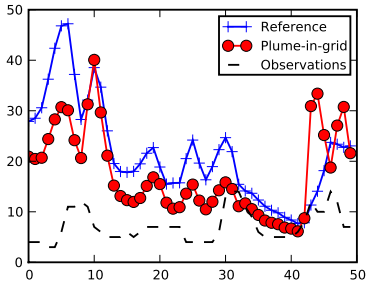
Concentrations de NO

Les émissions de NO sont beaucoup plus concentrées dans la zone urbaine de Paris, au centre du domaine de simulation. La baisse des concentrations de NO due à l'utilisation du modèle de panache sous-maille est également plus forte sur Paris (Figure 5.16). On ne peut pas distinguer clairement de zones où cette différence serait positive, à cause de l'impact au sol du panache, comme pour le SO_2. De plus, l'impact du panache sous-maille est beaucoup plus localisé autour des sources, car le NO émis est entièrement transformé par les réactions chimiques (notamment la titration) en aval des sources. Les différences entre les concentrations de référence et de panache sous-maille sont localement grandes, mais restent relativement faibles en comparaison des valeurs très élevées obtenues pour les concentrations moyennes. En effet, les émissions de NO sont principalement dues au trafic, et le traitement sous-maille des sources ponctuelles ne modifie qu'environ 20% des émissions.

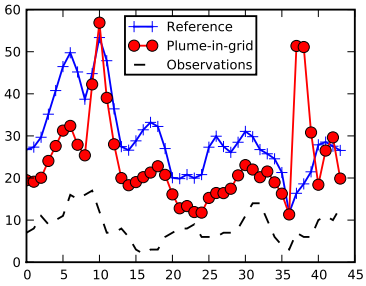
Pour le NO aussi bien que pour le SO_2, les différences sont plus grandes durant la matinée (deuxième rangée de cartes sur les figures 5.14 et 5.16) que plus tard dans la journée. En effet, les concentrations sont plus grandes le matin car les émissions sont élevées, et la couche limite n'a pas encore atteint sa hauteur maximale. Le mélange vertical s'effectue donc sur une hauteur plus faible que durant le reste de la journée. Cette observation est valable pour tous les jours et n'est pas propre aux journées particulières étudiées dans cette partie.



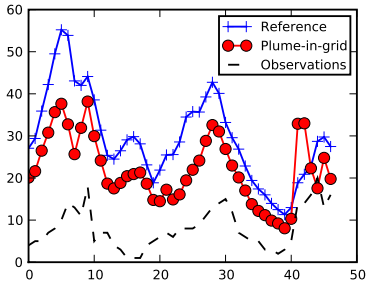
(a) SO_2–MELUN



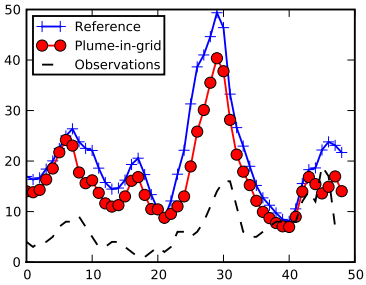
(b) SO_2–VITRY-SUR-SEINE



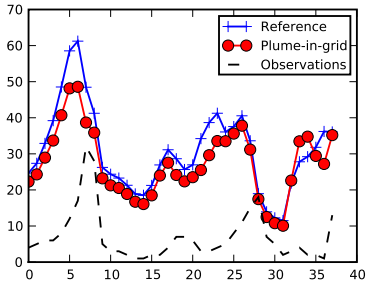
(c) SO_2–IVRY-SUR-SEINE



(d) SO_2–PARIS12eme



(e) SO_2–AUBERVILLIERS



(f) SO_2–LADEFENSE

FIG. 5.15 – Profils de concentration de SO_2 durant deux jours, entre le 23 août 2001 à 3 heures du matin et le 25 août 2001 à la même heure. Les profils de concentration observés sont comparés aux profils simulés par le modèle de référence et le modèle de panache sous-maille avec la théorie de la similitude.

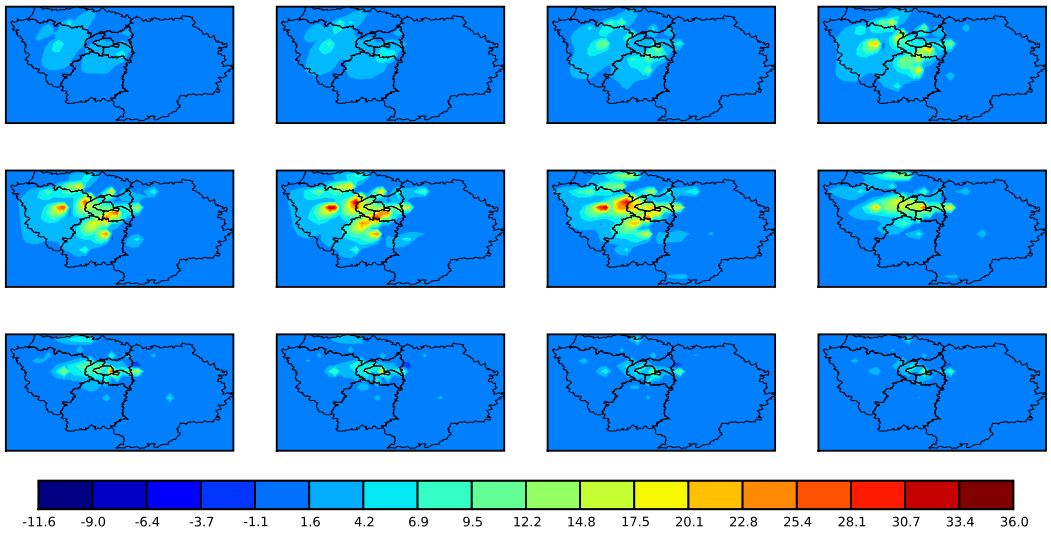


FIG. 5.16 – Évolution de la différence entre la simulation de référence et celle de panache sous-maille pour les concentrations au sol de NO durant 12 heures, le 23 août de 3 heures du matin à 14 heures. Les concentrations sont données en $\mu g\,m^{-3}$. Les concentrations de panache sous-maille sont soustraites aux concentrations de référence.

5.5.2 Impact sur les polluants secondaires

L'étude de NO_2 et O_3 est effectuée ici pour la journée du 20 août. Ce jour correspond à une dispersion faible à moyenne, et une vitesse de vent allant de 1 à 2 $m\,s^{-1}$. Cette journée a été retenue car l'impact spatial du panache sous-maille est plus étendu que pour d'autres jours, notamment ceux retenus dans la partie 5.5.1, pour les polluants secondaires.

Concentrations au sol de NO_2

L'utilisation du panache sous-maille a pour effet de diminuer les concentrations de NO_2, principalement car il y a moins de titration et donc moins NO_2 produit. La figure 5.17 montre les différences entre les concentrations au sol de NO_2 données par le modèle de référence, et par le modèle de panache sous-maille. Ces différences sont principalement positives.

Concentrations au sol d'O_3

La figure 5.18 montre l'évolution des différences pour les concentrations au sol d'ozone pour cette même journée du 20 août. La répartition spatiale des différences est assez semblable à celle montrée en figure 5.17 pour le NO_2, quoique plus localisée en général. L'utilisation du modèle de panache sous-maille a pour effet d'augmenter les concentrations de O_3 car la titration diminue, comme expliqué en partie 5.4.3. Les concentrations d'ozone sont cependant diminuées par endroits (différences positives sur la figure 5.18). Cela peut s'expliquer par l'arrivée au sol d'un panache maintenu en hauteur par le modèle de panache sous-maille, dans lequel aurait eu lieu la titration. Lorsque ce panache est ramené au sol, cela produit donc une baisse locale des concentrations dans le modèle de panache sous-maille. Cette différence peut également être due à des écarts de direction du vent entre les deux premiers niveaux verticaux. Plus en aval (en dehors du domaine de simulation), si le régime chimique devenait limité par les NOx, on aurait alors à nouveau une différence négative due à la production d'ozone, plus importante dans le

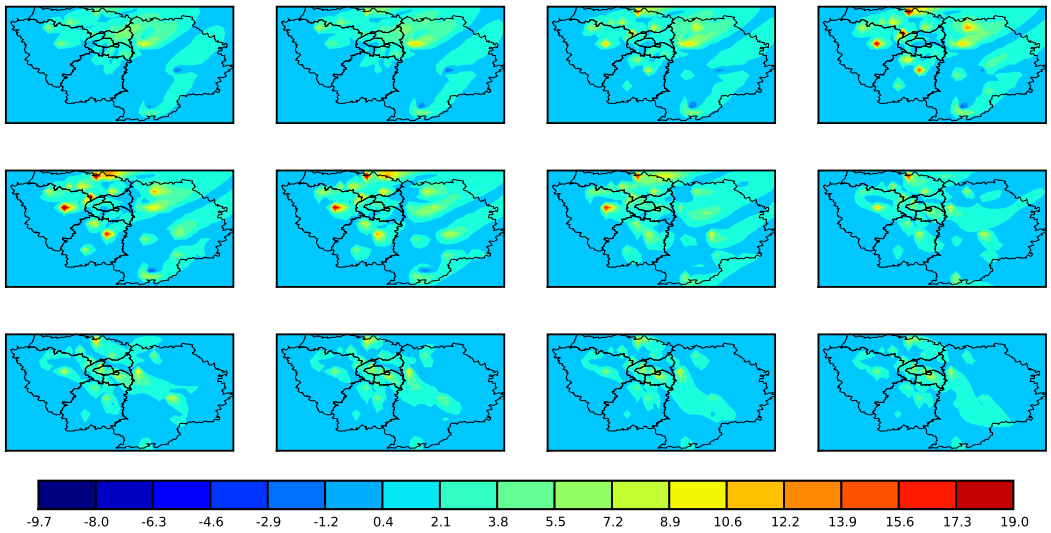


FIG. 5.17 – Évolution de la différence entre la simulation de référence et celle de panache sous-maille pour les concentrations au sol de NO_2 durant 12 heures, le 20 août de 3 heures du matin à 14 heures. Les concentrations sont données en $\mu g\,m^{-3}$. Les concentrations de panache sous-maille sont soustraites aux concentrations de référence.

modèle de panache sous-maille. Ce phénomène est mis en évidence dans Vijayaraghavan *et al.* [2006].

L'étude de ce jour particulier permet de mettre en évidence que l'impact du traitement sous-maille sur les concentrations d'ozone peut, dans certains cas, être transporté sur une distance importante en aval (voir notamment les deux dernières cartes de la figure 5.18). Cependant, cet impact reste souvent localisé près des sources (notamment en zone urbaine), ce qui ne permet donc pas de l'observer aux stations d'observation.

5.6 Analyse de sensibilité

5.6.1 Influence de la diffusion verticale

Diffusion verticale urbaine

Le coefficient de diffusion verticale K_z utilisé par le modèle eulérien Polair3D est donné par la paramétrisation de Troen–Mahrt. Comme présenté en partie 5.3.1, on utilise une valeur minimale imposée partout, afin de s'assurer d'un minimum de diffusion verticale sur tout le domaine. Cette valeur minimale est prise à $K_{z\min} = 0.2\ m^2\,s^{-1}$ pour les zones rurales, et $K_{z\min} = 0.5\ m^2\,s^{-1}$ dans les zones urbaines. La valeur plus grande pour les zones urbaines est utilisée afin de prendre en compte la diffusion verticale accrue, due au phénomène d'îlot de chaleur urbain. La valeur minimale de K_z est utilisée principalement dans les situations très stables, notamment de nuit. Dans cette partie, il s'agit de quantifier l'impact du changement de cette valeur minimale dans les zones urbaines sur les concentrations.

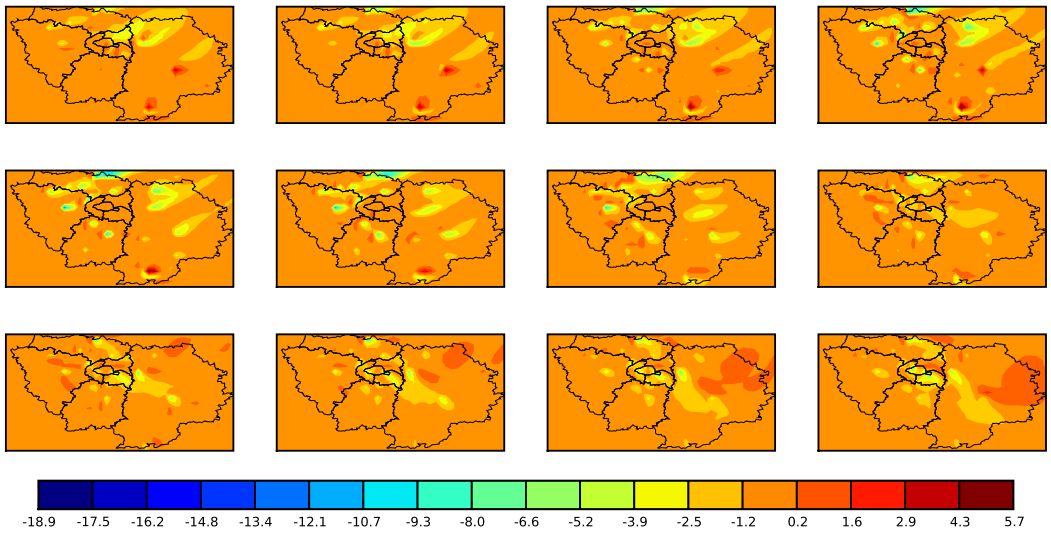


FIG. 5.18 – Évolution de la différence entre la simulation de référence et celle de panache sous-maille pour les concentrations au sol de O_3 durant 12 heures, le 20 août de 3 heures du matin à 14 heures. Les concentrations sont données en $\mu g\,m^{-3}$. Les concentrations de panache sous-maille sont soustraites aux concentrations de référence.

Diffusion verticale pour le modèle eulérien et le panache sous-maille

L'utilisation du modèle de panache sous-maille modifie la modélisation de la diffusion verticale autour des sources [Korsakissok et Mallet, 2010b]. La diffusion du modèle eulérien correspond à une taille du panache qui évolue en $\sigma_z \propto \sqrt{t}$, ce qui est représentatif du comportement à longue distance. Le modèle gaussien à bouffées utilise, en revanche, des paramétrisations empiriques qui correspondent à un comportement en $\sigma_z \propto t$, plus proche de la diffusion en champ proche. Il est donc intéressant de comparer l'apport du panache sous-maille avec celui d'un simple changement dans la verticale diffusion eulérienne tel que décrit dans cette partie. Naturellement, cette comparaison a des limites, car le changement dans la diffusion eulérienne a un impact sur toutes les émissions, mais limité à la zone urbaine et aux situations stables où la valeur minimale est utilisée, tandis que l'utilisation du panache sous-maille n'affecte que les émissions ponctuelles, sur l'ensemble du domaine.

Résultats

Les résultats sont analysés pour trois simulations : (1) une simulation eulérienne avec $K_{z\min} = 0.2\ m^2\,s^{-1}$ utilisé comme valeur minimale sur tout le domaine (notée "K_z"), (2) une simulation eulérienne avec $K_{z\min} = 0.5\ m^2\,s^{-1}$ pour les zones urbaines (notée "K_z urbain") et (3) une simulation avec le traitement sous-maille des sources ponctuelles, et $K_{z\min} = 0.5\ m^2\,s^{-1}$ sur les zones urbaines (notée "panache sous-maille"). Les simulations (2) et (3) sont les mêmes que les simulations de référence, et de panache sous-maille avec la théorie de la similitude, analysées dans les parties 5.4 et 5.5.

Le tableau 5.2 montre les mêmes indicateurs statistiques que ceux utilisés dans le tableau 5.1, pour les quatre espèces étudiées et les six mois de simulation. Augmenter la diffusion verticale dans les zones urbaines apporte une légère amélioration de ces statistiques, en réduisant la surestimation des concentrations de polluants primaires. De même que pour le panache sous-

Indicateur	Obs	K_z	K_z urbain	panache sous-maille
SO_2				
Moyenne ($\mu g\,m^{-3}$)	6.20	13.95	13.76	**11.98**
RMSE		13.61	13.08	**11.88**
Corrélation		**0.35**	0.34	0.31
MFBE		0.66	0.65	**0.55**
MFGE		0.81	0.81	**0.75**
NO				
Moyenne ($\mu g\,m^{-3}$)	10.42	21.72	20.93	**19.64**
RMSE		35.34	33.08	**31.62**
Corrélation		**0.48**	0.47	0.46
MFBE		0.42	0.41	**0.36**
MFGE		0.94	0.94	**0.92**
O_3				
Moyenne ($\mu g\,m^{-3}$)	56.87	40.18	40.24	**41.05**
RMSE		30.58	30.58	**30.18**
Corrélation		**0.68**	**0.68**	**0.68**
MFBE		-0.46	-0.46	**-0.44**
MFGE		0.47	0.47	**0.45**
MNBE		-0.34	-0.34	**-0.33**
MNGE		0.35	0.35	**0.34**
UPA		-0.21	-0.21	**-0.20**
NO_2				
Moyenne ($\mu g\,m^{-3}$)	34.64	36.04	35.84	**35.23**
RMSE		20.65	**20.57**	20.81
Corrélation		**0.58**	**0.58**	0.57
MFBE		0.07	0.06	**0.04**
MFGE		**0.47**	**0.47**	**0.47**

TAB. 5.2 – Statistiques horaires pour (1) Polair3D avec une valeur de $K_{z\min}$ uniforme sur tout le domaine (« K_z ») (2) Polair3D avec une valeur de $K_{z\min}$ spécifique pour les zones urbaines (« K_z urbain ») et (3) le modèle de panache sous-maille avec la théorie de la similitude. Les statistiques sont calculées sur la période de simulation, du 01-04-2001 au 27-09-2001. Les meilleures statistiques sont en gras.

maille, les espèces les plus sensibles à ce changement sont les espèces primaires SO_2 et NO. En revanche, les statistiques pour O_3 ne sont pratiquement pas modifiées.

Les statistiques observées aux stations (non détaillées ici) montrent les mêmes tendances que les statistiques globales. NO est l'espèce la plus sensible au changement de diffusion verticale, avec une réduction de la RMSE lorsque l'on augmente la diffusion urbaine comprise entre 0.5% et 7.9% aux stations urbaines. Pour le SO_2, cette différence est comprise entre -2.4% et -6.7%. En comparaison avec les résultats obtenus pour le panache sous-maille (détaillés en partie 5.4.2), l'amélioration est moins grande. De façon prévisible, les résultats aux stations péri-urbaines et rurales sont inchangés lorsque l'on modifie K_z en zone urbaine.

5.6.2 Influence de la modélisation locale

Le modèle local est le modèle gaussien à bouffées utilisé près des sources. L'échelle « locale » est définie comme la zone autour de la source où le modèle local est utilisé, et dont l'étendue

Indicateur	Obs	Base	$\Delta t_{\text{puff}} = 600$ s	$t_{\text{inj}} = 40$ min
SO_2				
Moyenne ($\mu g\, m^{-3}$)	6.20	11.98	11.98	**11.86**
RMSE		11.88	11.87	**11.57**
Corrélation		0.31	0.31	**0.32**
MFBE		0.55	0.55	**0.54**
MFGE		**0.75**	**0.75**	**0.75**
NO				
Moyenne ($\mu g\, m^{-3}$)	10.42	19.64	19.59	**19.56**
RMSE		31.62	31.60	**31.52**
Corrélation		**0.46**	**0.46**	**0.46**
MFBE		0.36	**0.34**	0.36
MFGE		**0.92**	0.93	**0.92**
O_3				
Moyenne ($\mu g\, m^{-3}$)	56.87	41.05	40.87	**41.09**
RMSE		30.18	30.25	**30.17**
Corrélation		**0.68**	**0.68**	**0.68**
MFBE		**-0.44**	-0.45	**-0.44**
MFGE		**0.45**	0.46	**0.45**
MNBE		**-0.33**	**-0.33**	**-0.33**
MNGE		**0.34**	**0.34**	**0.34**
UPA		**-0.20**	**-0.20**	**-0.20**
NO_2				
Moyenne ($\mu g\, m^{-3}$)	34.64	35.23	35.32	**35.22**
RMSE		20.81	20.81	**20.80**
Corrélation		**0.57**	**0.57**	**0.57**
MFBE		**0.04**	**0.04**	**0.04**
MFGE		**0.47**	**0.47**	**0.47**

TAB. 5.3 – Statistiques horaires pour (1) la simulation de base avec $t_{\text{inj}} = 20$ min et $\Delta t_{\text{puff}} = 100$ s, (2) $\Delta t_{\text{puff}} = 600$ s, (3) $t_{\text{inj}} = 40$ min. Dans tous les cas, il s'agit du modèle de panache sous-maille avec la théorie de la similitude. Les statistiques sont calculées sur la période de simulation, du 01-04-2001 au 27-09-2001. Les meilleures statistiques sont en gras.

est déterminée par le critère d'injection choisi. La sensibilité des résultats au choix des formules des écarts types a été évaluée en partie 5.4. On analyse ici l'influence de deux paramètres importants du modèle de panache sous-maille : le temps d'injection t_{inj} et le pas de temps entre l'émission de deux bouffées Δt_{puff}. Dans cette partie, la simulation de « base » correspond à la simulation de panache sous-maille avec la théorie de la similitude déjà étudiée. Pour cette simulation, $t_{\text{inj}} = 20$ min et $\Delta t_{\text{puff}} = 100$ s. Le tableau 5.3 donne les statistiques globales pour la simulation de base, ainsi que pour les simulations avec (1) un pas de temps plus grand entre les bouffées $\Delta t_{\text{puff}} = 600$ s, et (2) un temps d'injection plus long $t_{\text{inj}} = 40$ min.

Les résultats ne sont pas très sensibles au pas de temps entre les bouffées, notamment en ce qui concerne le SO_2. Pour les espèces plus réactives, les résultats diffèrent légèrement du cas de base, avec une amélioration pour le NO et une légère dégradation des statistiques pour NO_2 et O_3. Le pas de temps entre deux bouffées détermine le volume de superposition entre ces bouffées. Pour un traceur passif, il suffit de s'assurer que les bouffées se recouvrent afin d'avoir une simulation réaliste du panache (voir chapitre 4). Pour des espèces réactives, en revanche,

cela influe également sur les réactions qui ont lieu dans les bouffées. Si le pas de temps est grand, le volume de superposition entre deux bouffées consécutives est faible (peu de réactions entre les espèces des deux bouffées), mais les concentrations dans chaque bouffée sont plus grandes. La quantité transportée par une bouffée est en effet $Q = Q_s \times \Delta t_{\text{puff}}$, avec Q_s le taux d'émission de la source en unité de masse par seconde, pour une espèce donnée. Dans le cas d'un panache contenant des NOx émis dans un environnement contenant de l'ozone, une augmentation locale de la concentration de NOx dans la bouffée conduit à davantage de titration (en régime COV-limité). Cela augmente localement la formation de NO_2 et diminue les concentrations de NO et O_3, ce que l'on observe sur les concentrations moyennes données dans le tableau 5.3. Ce phénomène est très local et n'a pas un impact très important sur les statistiques globales.

Le temps d'injection a une influence beaucoup plus grande sur les résultats pour le SO_2 que pour les autres espèces. En comparaison avec la simulation de base (avec $t_{\text{inj}} = 20$ min), la RMSE pour un temps d'injection de 40 minutes est améliorée de 0.4% à 8%, et de 2% environ pour la plupart des stations. Cette amélioration s'ajoute à celle apportée par l'utilisation du panache sous-maille, montrée sur la figure 5.11. Comme le SO_2 est considéré ici comme une espèce passive, l'augmentation du temps d'injection permet d'élargir l'impact spatial du traitement sous-maille. En revanche, le temps d'injection plus long n'a quasiment pas d'effet sur les autres espèces. En effet, pour les espèces réactives, utiliser le traitement sous-maille pour un panache n'a plus d'effet lorsque la composition du panache est proche de celle de l'atmosphère. Lorsque la chimie dans le panache atteint la phase 2, puis la phase 3 décrites en partie 5.2.1, l'impact du traitement sous-maille est moindre. Au-delà d'une vingtaine de minutes, le panache est probablement passé dans la phase 2, et l'augmentation du temps d'injection a donc peu d'impact sur les résultats.

5.7 Conclusions

5.7.1 Chimie dans un panache

Le modèle de panache sous-maille est étendu dans ce chapitre au cas des espèces réactives, avec une application à la photochimie. L'effet du traitement sous-maille sur la chimie dans un panache a été tout d'abord détaillé en partie 5.2 sur un cas simple avec une seule source. Cela a permis de vérifier l'impact théorique du traitement sous-maille sur les réactions chimiques, discuté en partie 5.2.1. L'effet principal du modèle de panache sous-maille est d'augmenter la titration de l'ozone à l'intérieur du panache, tout en le maintenant en hauteur, ce qui diminue (resp. augmente) les concentrations au sol de NOx (resp. O_3) en champ proche. De plus, la formation d'acides, notamment HNO_3, est retardée et ne débute que quelques dizaines de kilomètres en aval de la source, contrairement à ce qui se produit dans le modèle eulérien.

5.7.2 Impact sur la photochimie en région parisienne

L'application de ce modèle à la photochimie à l'échelle de la région parisienne a permis d'évaluer l'impact du traitement sous-maille des émissions ponctuelles sur les performances du modèle, par comparaison avec les observations. Les performances sont évaluées pour une période de six mois, durant l'été 2001. Les sources les plus importantes de NOx et SO_2 ont été sélectionnées, ce qui correspond à 89 sources ponctuelles.

Impact global

Les statistiques globales sont améliorées par le traitement sous-maille, particulièrement en utilisant les écarts types calculés avec la théorie de la similitude. Cela permet de diminuer la RMSE aux stations, jusqu'à -17% pour le SO_2 et -6% pour le NO. Pour NO_2 et O_3, l'impact

est plus faible. L'impact du traitement sous-maille des émissions ponctuelles est relativement faible si l'on regarde les statistiques globales, pour différentes raisons :

1. les sources ponctuelles ne constituent qu'une partie des émissions totales, plus faible que le trafic dans le cas des NOx,
2. l'impact est surtout local, alors que les stations de mesure ne sont pas situées à proximité immédiate des sources,
3. les résultats sont pour une moyenne de six mois, ce qui « lisse » l'impact que l'on peut voir en aval du panache pour une direction du vent donnée.

Considérer ces facteurs permet donc de relativiser les résultats obtenus. Cela met également en évidence l'intérêt d'étudier les résultats pour des jours particuliers, en complément des statistiques globales. Les jours choisis ici sont représentatifs de situations de faible dispersion, où la surestimation de la dilution du panache par le modèle eulérien est particulièrement importante. Durant ces jours, l'impact du traitement sous-maille est très grand, particulièrement autour des sources mais aussi à certaines stations situées en aval.

Impact de la diffusion urbaine

Par ailleurs, dans les simulations, la valeur minimale du coefficient de diffusion verticale K_z utilisé dans le modèle eulérien est augmenté en milieu urbain (par rapport à la valeur par défaut prise en milieu rural), pour tenir compte de la diffusion accrue due au phénomène d'îlot de chaleur urbain. L'impact de ce changement est évalué, et comparé à celui du panache sous-maille. Dans les deux cas, les polluants primaires SO_2 et NO sont les plus affectés par les modifications.

Impact du temps d'injection et du pas de temps entre deux bouffées

Deux paramètres importants du modèle de panache sous-maille ont également été évalués. Il s'agit du pas de temps entre l'émission de deux bouffées consécutives, ainsi que du temps d'injection. Ces deux paramètres déterminent le nombre total de bouffées gérées par le modèle en même temps (pour un nombre de sources donné), donc le temps de calcul. Le pas de temps entre deux bouffées influence principalement les espèces réactives, car les taux de réaction en champ très proches sont légèrement surestimées. Pour une espèce passive, ce pas de temps a moins d'influence (ce qui a déjà été discuté au chapitre 4). Le chapitre 4 a également mis en évidence l'importance du temps d'injection pour la dispersion des traceurs passifs. Ici, cette influence est visible pour le SO_2, mais devient négligeable pour les espèces réactives. En effet, lorsque la composition du panache devient proche de celle de l'air ambiant, l'utilisation d'un traitement sous-maille a peu d'effet sur les résultats. Au bout d'une vingtaine de minutes environ, la chimie du panache devient proche de celle de l'atmosphère, et prolonger le traitement sous-maille a donc peu d'effet.

5.7.3 Bilan et perspectives

Les perspectives d'évolution pour le modèle de panache sous-maille réactif consistent, d'une part, à inclure la chimie des particules, et d'autre part à étendre ce modèle à la simulation des routes. Une modélisation performante des sources linéiques a déjà été introduite dans les modèles gaussiens et validée (chapitre 2). Elle peut donc à présent être étendue au modèle de panache sous-maille. Utiliser un traitement sous-maille pour les routes est en effet pertinent, étant donné la forte variabilité locale près des grands axes routiers (chapitre 1), et l'impact prépondérant du trafic sur les émissions de NOx.

Par ailleurs, le modèle développé ici utilise un mécanisme chimique complet pour les réactions à l'intérieur des bouffées. Or, la chimie dans le panache ne fait intervenir que quelques réactions prépondérantes en champ proche (décrites en partie 5.2.1). Introduire une modélisation simplifiée de la chimie dans les bouffées, en ne prenant en compte que les réactions prépondérantes, permettrait de réduire le temps de calcul de façon significative [Karamchandani *et al.*, 1998].

Troisième partie

Méthodes statistiques pour la réduction d'échelle et application à la qualité de l'air

Chapitre 6

Méthodes statistiques pour la réduction d'échelle

Ce chapitre est une présentation générale des méthodes statistiques utilisées pour la réduction d'échelle, qui ont été principalement appliquées à modélisation du climat. Il décrit l'état de l'art dans le domaine, et en particulier les points communs et les différences avec une application à la modélisation de la qualité de l'air. Il s'agit, d'une part, de décrire le cadre dans lequel les méthodes statistiques de réduction d'échelle sont pertinentes. En s'appuyant sur la modélisation du climat, différentes contraintes sont identifiées : (1) le lien physique entre le prédicteur (variable à grande échelle, connue) et le prédictand (variable à petite échelle que l'on souhaite reconstituer), (2) la stationnarité de la relation (pour l'application en prévision), (3) la bonne connaissance du prédicteur, c'est-à-dire en pratique une bonne estimation fournie par le modèle utilisé. Pour ce dernier point, l'utilisation d'un ensemble de modèles peut être envisagée. D'autre part, on identifie dans ce chapitre les différentes méthodes possibles pour la réduction d'échelle, ainsi que leurs points forts et leurs inconvénients. Des méthodes telles que la régression linéaire multiple, la décomposition en valeurs singulières et l'analyse de corrélation canonique sont décrites. Celles-ci peuvent être appliquées sur les prédicteurs lui-même, ou après un pré-traitement destiné à les rendre indépendants. L'ensemble de ces méthodes est décrit dans le cadre très général de l'application au climat. Dans la dernière partie, on définit le cadre dans lequel la réduction d'échelle sera effectuée pour la qualité de l'air. Les méthodes choisies sont adaptées à l'application particulière visée ici et détaillées. Les résultats seront présentés dans le chapitre 7.

Sommaire

6.1 Introduction 181
- 6.1.1 Historique 181
- 6.1.2 Réduction d'échelle statistique et dynamique 181
- 6.1.3 Une question d'échelles... 181

6.2 Principes de la réduction d'échelle statistique 182
- 6.2.1 Principe général 182
- 6.2.2 Stationnarité 183
- 6.2.3 Prédicteur et prédictand 183
- 6.2.4 Représentativité des modèles 184

6.2.5 Le problème de la variance 184
6.3 Méthodes statistiques 185
6.3.1 Pré-traitement des données 185
6.3.2 Méthodes de régression 187
6.3.3 Comparaison des méthodes 190
6.4 Utilisation des méthodes statistiques pour la qualité de l'air 190
6.4.1 Régression sur un ensemble de modèles 191
6.4.2 Décomposition en composantes principales (PCA) 192
6.4.3 Décomposition en composantes principales ajustées (PFC) 193
6.5 Conclusion 194
6.5.1 Méthodes de réduction d'échelle 194
6.5.2 Questions 194

6.1 Introduction

6.1.1 Historique

Les méthodes de réduction d'échelle (« downscaling ») sont très utilisées en science du climat, où la prédiction du climat futur implique de déduire des caractéristiques régionales à partir des résultats de modèles globaux. Il existe donc une littérature fournie sur le sujet, notamment depuis ces 15 dernières années, et dont l'état de l'art est décrit par exemple par Wilby *et al.* [2004] et Benestad *et al.* [2008]. Le but de ce chapitre est de dégager les aspects importants de ces techniques, et la façon dont elles peuvent être adaptées à la modélisation de la qualité de l'air. L'utilisation de ces techniques dans le domaine de la qualité de l'air est en effet récente (Guillas *et al.* [2006] et Guillas *et al.* [2008]).

6.1.2 Réduction d'échelle statistique et dynamique

La réduction d'échelle peut se définir comme l'ensemble des méthodes qui permettent de faire le lien entre l'état d'une variable à une grande échelle, et l'état d'une autre variable représentant une échelle beaucoup plus petite [Benestad *et al.*, 2008]. Il existe principalement deux types de méthodes utilisées à cette fin :

- les méthodes de « nesting », consistant à utiliser un modèle de climat à l'échelle régionale, dont les conditions aux limites sont forcées par le modèle global, et parfois appelées en climatologie « dynamical downscaling »,
- les méthodes statistiques, appelées « statistical downscaling » ou « empirical downscaling », ou encore ESD pour « empirical-statistical downscaling ».

Ces deux méthodes ont été comparées dans le contexte du climat (e.g. Kidson et Thompson [1998], Murphy [1999], Spak *et al.* [2007]...), et donnent des résultats globalement similaires, les méthodes statistiques ayant l'avantage d'un moindre coût de calcul.

6.1.3 Une question d'échelles...

Échelle minimale et échelle bien représentée. Le principe de l'ESD est de déterminer une relation statistique entre une variable locale et une variable globale. Il reste à préciser la signification des termes « local » et « global ». Une interprétation en est faite dans von Storch *et al.* [1993], qui, se référant à Grotch et MacCracken [1991], définit la notion d' « échelle bien représentée » (« skillful scale ») comme étant l'échelle des phénomènes qui sont correctement simulés par un modèle donné. En définissant l' « échelle minimale » comme la distance entre deux centres de maille (donc, la résolution spatiale du modèle), l'échelle bien représentée est alors supérieure ou égale à N fois l'échelle minimale (dans les références citées, $N = 8$). Dans ce contexte, la variable « locale » est une variable dont l'échelle caractéristique est inférieure à cette distance, et l'ESD est alors définie comme l'ensemble des techniques statistiques permettant de passer d'une échelle bien représentée à une échelle plus fine. Comme souligné dans Benestad *et al.* [2008], la définition de « skillful scale » reste floue, les estimations évoquées précédemment n'ont pas été faites sur des modèles récents, et cette notion dépend également de la façon dont la performance est estimée (« bonne » représentation de la moyenne, de la variance, autocorrélation...).

Réduction d'échelle et assimilation. Benestad *et al.* [2008] s'appuie sur cette définition de l'ESD pour avancer l'argument qu'appliquer une méthode statistique sur une seule maille d'un modèle (l'échelle minimale), pour reconstituer le comportement à une station à l'intérieur de cette maille, n'est pas de la réduction d'échelle, mais plutôt de l' « ajustement » ou de l' « assimilation ». En effet, les méthodes statistiques de réduction d'échelle peuvent être considérées

comme une façon parmi d'autres de corriger les sorties d'un modèle en fonction des observations. En d'autres termes, il s'agirait d'une forme alternative à l'assimilation de données. En assimilation également, les données simulées sont corrigées par apport d'information fournie par les observations. Cependant, le but est de corriger l'erreur modèle : l'erreur de représentativité est prise en compte, mais n'est pas corrigée. En ESD, on utilise une méthode statistique pour obtenir une relation reflétant le lien « systématique » entre les deux variables. On suppose que la relation statistique obtenue reflétera alors principalement l'erreur de représentativité, donc liée au changement d'échelle, ce qui justifie alors bien l'utilisation du terme « ESD ». De plus, un apport supplémentaire de l'ESD est de prendre en compte la variabilité locale, qui n'est jamais déterminée complètement par les variables à plus grande échelle (partie 6.2.5).

Représentativité des stations. L'erreur de représentativité, évoquée plus haut, vient de la mauvaise représentation des processus à une échelle inférieure à celle de la maille (et même, inférieure à l'échelle « bien représentée »), tels que les émissions, l'occupation des sols, ou les précipitations. Une station proche d'une route ne sera pas nécessairement représentative des concentrations moyennes à plus grande échelle (on pourrait parler d'échelle « bien représentée » par la station). On peut également dire qu'une observation à une station donnée n'est qu'une réalisation particulière d'une variable aléatoire (représentable par une densité de probabilité). L'estimation du modèle étant une autre réalisation, elle pourrait différer des observations même si le modèle était parfait.

6.2 Principes de la réduction d'échelle statistique

Notations

Dans la suite, pour une variable x représentant une donnée physique (concentration, température...), on notera

1. x un scalaire,
2. $\boldsymbol{x}$ un vecteur, par exemple une série d'observations à m stations, à un instant t

$$\boldsymbol{x}(t) = (x_1(t), ..., x_m(t))^T,$$

3. X une matrice. Par exemple, X pourra avoir pour colonne les observations pour un ensemble de stations, à un instant donné. Alors, pour n échéances temporelles :

$$X = [\boldsymbol{x_1} \dots \boldsymbol{x_j} \dots \boldsymbol{x_n}]$$

où l'élément x_{ij} de la matrice X est donc la valeur au pas de temps j pour la i^{e} station.

Avec ces notations, $\boldsymbol{y} = \Psi\boldsymbol{x}$ est une régression linéaire simple, $\boldsymbol{y} = \Psi X$ est une régression linéaire multiple, et $Y = \Psi X$ est une régression multidimensionnelle [Benestad *et al.*, 2008].

La moyenne de $\boldsymbol{x}$, c'est-à-dire $\frac{1}{n}\sum_{i=1}^{n} x_i$, sera notée $\bar{\boldsymbol{x}}$ (moyenne temporelle) si $\boldsymbol{x}$ est une série temporelle, et elle sera notée $\langle \boldsymbol{x} \rangle$ (moyenne spatiale) si $\boldsymbol{x}$ est une série de valeurs en différents points de l'espace. La variance sera notée $\text{var}(\boldsymbol{x}) = \sigma_x^2$.

6.2.1 Principe général

Les méthodes statistiques de réduction d'échelle consistent à décrire une variable locale $\boldsymbol{y}$ (le « prédictand »), comme une fonction d'une ou plusieurs caractéristiques globales $\boldsymbol{x_j}$ (les P « prédicteurs »). Cela peut s'écrire sous la forme générique de l'équation 6.1

$$\hat{\boldsymbol{y}} = f(\boldsymbol{x_1}, ..., \boldsymbol{x_P}) \tag{6.1}$$

La fonction f est une relation empirique entre les variables, calibrée sur un ensemble de données d'observation (période d'apprentissage). Les méthodes statistiques utilisées pour calculer f sont détaillées dans la partie 6.3. Nous nous intéressons ici aux implications, physiques et statistiques, de l'utilisation de telles méthodes, et aux conditions dans lesquelles elles sont pertinentes. Benestad *et al.* [2008] donne plusieurs conditions indispensables à une bonne utilisation de l'ESD :

- que la relation f entre prédicteurs et prédictand soit stationnaire, pour qu'elle puisse être calibrée sur le passé et utilisée en prédiction de façon pertinente,
- qu'il y ait une relation physique forte entre le ou les prédicteur(s), et le prédictand,
- que le modèle représente correctement les variables à grande échelles utilisées comme prédicteurs,

Les trois conditions ci-dessus sont des critères très importants pour bien utiliser les techniques d'ESD, qui seront développés dans la suite de cette partie. On y rajoutera le problème de la représentation de la variabilité à l'échelle locale (partie 6.2.5), qui est au cœur de la problématique de réduction d'échelle.

6.2.2 Stationnarité

L'idée sous-jacente est que la relation 6.1 est constante dans le temps, et pourra donc être utilisée pour la prévision. Cette hypothèse de stationnarité est l'un des points cruciaux dans le cadre du changement climatique car elle est difficilement vérifiable à long terme. En qualité de l'air, pour des prévisions à beaucoup plus court terme, cette hypothèse semble plus réaliste. Les méthodes de réduction d'échelle nécessitent tout de même de se ramener à un processus stationnaire, ce qui n'est le cas ni en climatologie ni en qualité de l'air. Pour les variables climatiques, la fonction f est donc calculée pour chaque mois, ou chaque saison. Pour la qualité de l'air, il est nécessaire de la calculer pour chaque heure de la journée, afin de prendre en compte le cycle journalier des concentrations. La distinction été/hiver est également nécessaire. Enfin, la réduction d'échelle est faite en général sur des valeurs centrées sur la moyenne, afin de ne chercher à relier que les variations autour de la moyenne (Wilby *et al.* [2003], Benestad *et al.* [2008]) (voir partie 6.3.1).

6.2.3 Prédicteur et prédictand

Pour justifier l'emploi de cette méthode statistique, il faut s'assurer que les variables prédicteur et prédictand aient un lien physique entre elles, faute de quoi une relation statistique pourra peut-être être calculée mais n'aura aucun sens. La variable à grande échelle (prédicteur) est par exemple en climatologie la pression au niveau de la mer (SLP), ou la circulation à l'échelle synoptique, et la variable locale (prédictand), peut être la précipitation, le vent, la température, en une ou plusieurs stations. L'un des problèmes rencontrés dans l'application au climat est donc de choisir correctement les prédicteurs : il faut qu'il y ait un lien physique avec le prédictand, et que les prédicteurs « expliquent » une grande partie de la variable locale. De même, il faut que les prédicteurs soient pris sur un domaine spatial suffisamment grand pour contenir toute l'information nécessaire, sans englober des variables non corrélées à la variable locale. La comparaison des prédicteurs et le choix des domaines sont traités par exemple dans Huth [1999] et Benestad [2001].

Il peut s'agir en particulier de prédire le comportement à l'échelle locale de la même variable connue à l'échelle synoptique (la température par exemple). C'est, a priori, ce qui sera fait en qualité de l'air : les concentrations d'ozone, par exemple, seront reconstruites à une station à partir des concentrations fournie à l'échelle du maillage d'un modèle. Dans ce cas, le lien physique entre le prédicteur et le prédictand est évident. Il reste en revanche la question du domaine

spatial où prendre le prédicteur (une seule maille ou différentes mailles autour d'une station, par exemple). Il peut alors être intéressant de calculer les corrélations entre les concentrations modélisées dans différentes mailles avec une station donnée, pour avoir une idée de la taille représentative du domaine où les concentrations sont corrélées avec celles de la station. D'autre part, on peut également tenter de trouver un lien avec d'autres variables explicatives. Ainsi, dans Guillas *et al.* [2008], une régression est d'abord effectuée entre les concentrations d'ozone mesurées et modélisées, puis le résidu est à son tour soumis à une régression avec des paramètres tels que le vent et les précipitations. Dans ce cas, le problème de la pertinence physique du choix des prédicteurs se pose donc également.

6.2.4 Représentativité des modèles

Les prédicteurs utilisés en ESD pour le climat proviennent de simulations de modèles type GCM (du moins pour les prévisions – pour la phase de calibration, il est courant d'utiliser des réanalyses ECMWF ou NCEP). La bonne estimation du prédicteur dépend donc de la performance du modèle. Il en est de même en qualité de l'air, où un CTM calcule les concentrations qui servent ensuite de prédicteurs. L'erreur due au modèle contribue donc aux incertitudes de l'ESD. Pour réduire cette erreur, et estimer son impact, un ensemble de modèles peut être utilisé pour la réduction d'échelle. C'est le cas dans Benestad [2002], Pryor *et al.* [2005] et Najac *et al.* [2009]. En général, la réduction d'échelle est faite sur les différents modèles, et l'incertitude sur les résultats due aux modèles peut alors être estimée. Dans Benestad [2002], le pré-traitement (détaillé partie 6.3.1) permet de trouver les modes dominants par une décomposition en composantes principales ou EOF (Empirical Orthogonal Functions) qui est effectuée sur la moyenne des modèles. De plus, Benestad [2002] utilise différents prédicteurs issus d'un ensemble de modèles (les prédicteurs étant pris sur des domaines de différentes tailles), afin de comparer l'incertitude due aux erreurs de modélisation et celle due au choix des prédicteurs (la première étant en général la plus grande).

6.2.5 Le problème de la variance

Les méthodes de réduction d'échelle fournissent une estimation du prédictand $\boldsymbol{y}$ en fonction des variables explicatives à grande échelle. Or, même si l'on inclut tous les prédicteurs ayant un lien avec le comportement de la variable $\boldsymbol{y}$, celui-ci n'est jamais entièrement déterminé par les variables à grande échelle. Il restera toujours une part de variabilité dans $\boldsymbol{y}$ qui ne peut être expliquée que par des phénomènes locaux : la topographie, par exemple, le type et l'occupation du sol, ou plus généralement tout phénomène dont l'échelle caractéristique est très petite devant les échelles résolues par le modèle utilisé. Négliger purement et simplement cette part revient paradoxalement à ne pas tenir compte du problème de changement d'échelles, qui a motivé au départ l'utilisation de techniques de réduction d'échelle. C'est pourquoi les modèles de régression classiques peuvent donner de bons résultats en terme d'erreur quadratique moyenne (RMSE) mais sous-estiment la variance de façon significative (par exemple dans Wilby *et al.* [2003] et Najac *et al.* [2009]).

Si l'on représente la variance inexpliquée comme un « bruit » noté ϵ, la variable locale $\boldsymbol{y}$ est alors la somme de la part déterministe (signal) donnée par l'équation 6.1, notée $\hat{\boldsymbol{y}}$, et du bruit :

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \epsilon. \tag{6.2}$$

Si ϵ est un bruit blanc, de moyenne nulle et de variance $\mathrm{var}(\epsilon) = \sigma_\epsilon^2$, non corrélé au signal, il s'ensuit que la variance de la variable $\boldsymbol{y}$ s'écrit

$$\mathrm{var}(\boldsymbol{y}) = \mathrm{var}(\hat{\boldsymbol{y}}) + \sigma_\epsilon^2. \tag{6.3}$$

La variance var($\hat{\boldsymbol{y}}$) est celle obtenue par les modèles de régression déterministes, et l'équation 6.3 montre bien qu'elle est inférieure à la variance totale var($\boldsymbol{y}$). Le problème se pose donc de reconstituer la part supplémentaire de la variance introduite par ϵ. Une approche très utilisée (e.g. Murphy [1999] et Huth [1999]), appelée « inflation », consiste à appliquer un facteur de correction à $\hat{\boldsymbol{y}}$, calculé pour que la formule finale ait exactement la variance var($\boldsymbol{y}$) (connue sur la période d'apprentissage). On obtient alors simplement la formule suivante

$$\tilde{\boldsymbol{y}} = \sqrt{\frac{\mathrm{var}(\boldsymbol{y})}{\mathrm{var}(\hat{\boldsymbol{y}})}}\ \hat{\boldsymbol{y}}. \tag{6.4}$$

Or, si l'équation 6.4 permet effectivement de reconstituer la variance var($\boldsymbol{y}$), elle perd également une bonne partie du bénéfice apporté par la régression dans le calcul de $\hat{\boldsymbol{y}}$, en terme de biais et de RMSE. La démonstration en est faite dans von Storch [1999] où l'utilisation de l'inflation est critiquée.

La bonne façon de faire est donc de revenir à l'équation 6.2, en trouvant une manière adéquate de spécifier ϵ. En général, on prend une distribution gaussienne $\epsilon \sim N(0, \sigma_\epsilon^2)$, avec σ_ϵ estimé sur la période d'apprentissage (Wilby *et al.* [2003], Hessami *et al.* [2008], Guillas *et al.* [2008]).

6.3 Méthodes statistiques

La procédure de réduction d'échelle peut se décomposer en trois étapes, une fois les prédicteurs choisis.

1. Pré-traitement des données : découpage des séries temporelles en fonction des saisons et heures du jour à séparer, éventuellement transformation pour enlever la tendance et réduire la matrice (partie 6.3.1) X, afin d'obtenir une relation stationnaire comme décrit dans la partie 4.5. Si nécessaire, décomposition en EOF pour réduire l'espace des prédicteurs et les rendre orthogonaux (partie 6.3.1).
2. La réduction d'échelle proprement dite : reconstituer la fonction f de l'équation 6.1 par des méthodes statistiques, en particulier de régression (partie 6.3.2).
3. Reconstitution du prédictand : application de f en prévision, ajout de la tendance si nécessaire et/ou de la variabilité locale (déjà évoquée partie 6.2.5).

Parmi les méthodes statistiques utilisées en réduction d'échelle pour le climat, certaines sont basées sur la classification par type de météorologie (« weather classification »), qui permettent de procéder par analogie, ou des méthodes stochastiques (« weather generators »). Dans la suite, on s'intéressera à une troisième famille de méthodes, les modèles de régression, et en particulier de régression linéaire. Différentes méthodes comme la régression linéaire multiple (MVR, Murphy [1999]), l'analyse de corrélation canonique (CCA, von Storch *et al.* [1993]), la décomposition en valeurs singulières (SVD), sont décrites partie 6.3.2. Des méthodes de régression non linéaires, comme les réseaux de neurones, ont également été utilisées pour le climat [Crane et Hewitson, 1998], mais ne seront pas évoquées ici. Des méthodes statistiques basées sur les séries temporelles et les réseaux de neurones, ainsi que leurs applications en qualité de l'air, sont décrites de façon très complète dans Guillas *et al.* [2000].

6.3.1 Pré-traitement des données

le pré-traitement s'effectue sur la matrice X contenant des valeurs (simulées ou observées) du prédicteur, à m stations et n pas de temps, de taille $m \times n$. On a $X = [\boldsymbol{x_1} \dots \boldsymbol{x_n}]$, chaque colonne étant un vecteur représentant une série d'observations au temps t_j : $\boldsymbol{x_j} = (x_{1j}, ..., x_{mj})^T$.

Matrice réduite et centrée

La matrice X est éventuellement centrée sur la moyenne temporelle suivant l'équation 6.5, voire réduite (équation 6.6).

$$x_{ij} = x_{ij} - \bar{\boldsymbol{x}}_i \tag{6.5}$$

$$x_{ij} = \frac{x_{ij} - \bar{\boldsymbol{x}}_i}{\sigma_{x_i}}. \tag{6.6}$$

La transformation peut se faire, de façon similaire, à partir de la moyenne spatiale $\langle \boldsymbol{x} \rangle$ et de l'écart type spatial.

Décomposition en composantes principales

La décomposition d'une variable en composantes principales permet de réduire l'espace sur lequel elle est décrite, en éliminant les redondances d'informations (dues à une corrélation spatiale, ou temporelle, non nulle), et en gardant les « modes » dominants. Les modes spatiaux, ou EOF (« empirical orthogonal functions ») spatiales, sont les structures spatiales qui expliquent le maximum de variance temporelle. De même, les modes temporels sont les trajectoires qui expliquent le maximum de variance spatiale. Mathématiquement, il s'agit des vecteurs propres de la matrice de variance-covariance, temporelle dans le premier cas, spatiale dans le second. Une analyse basée sur la décomposition en EOF spatiale est appelée « S-mode », et celle basée sur la décomposition en EOF temporelles est alors une analyse « T-mode ».

Les EOF spatiales sont les vecteurs propres, notés $\boldsymbol{p_k}$, de la matrice de variance-covariance $C_{XX} = XX^T$, de taille $m \times m$. Si l'on note P la matrice dont les colonnes sont les EOF $P = (\boldsymbol{p_1}, ..., \boldsymbol{p_m})$ (on a supposé que la matrice C_{XX} est de rang m), de taille $m \times m$, on a

$$C_{XX} = P \Lambda P^T, \tag{6.7}$$

avec Λ la matrice diagonale contenant les valeurs propres λ_k. Les vecteurs propres $\boldsymbol{p_k}$ forment une base orthogonale sur laquelle on peut projeter tout vecteur $\boldsymbol{x}(t)$ de taille m :

$$\boldsymbol{x}(t) = \sum_{k=1}^{m} \alpha_k(t) \boldsymbol{p_k}. \tag{6.8}$$

Les coefficients $\alpha_k(t)$ sont les poids, à l'instant t, représentant l'importance des différents modes à cet instant. Le vecteur $\boldsymbol{\alpha_k} = (\alpha_{1k}, ..., \alpha_{nk})^T$ est le coefficient d'expansion correspondant au mode k, c'est-à-dire le poids de ce mode au cours du temps. Si A est la matrice (de taille $n \times m$) dont les colonnes sont les coefficients d'expansion $A = [\boldsymbol{\alpha_1} \dots \boldsymbol{\alpha_m}]$, on a

$$X = PA^T \qquad \text{ou} \qquad A^T = P^T X \tag{6.9}$$

car P est orthogonale, de même que A. On note par ailleurs que les coefficients d'expansion ont la même variance temporelle que X

$$AA^T = X^T P P^T X = X^T X. \tag{6.10}$$

On a également :

$$A^T A = P^T X X^T P = \Lambda. \tag{6.11}$$

On peut également appeler les EOF spatiales, ou modes, les « directions principales », et les coefficients d'expansion associés « composantes principales », pour revenir à un vocabulaire classique associé à l'analyse en composantes principales. On trouve, de façon similaire, les EOF temporelles, qui sont les vecteurs propres de la matrice de variance-covariance $C'_{XX} = X^T X$.

On peut ensuite tronquer le résultat à un nombre limité de modes (correspondant aux valeurs propres les plus élevées), qui expliquent une grande partie de la variance, afin de réduire l'espace des prédicteurs. Par exemple, on peut tronquer à 20 modes, ou aux m' premiers modes qui expliquent 90% de la variance, avec $m' \ll m$.

Décomposition en valeurs singulières

En pratique, il est souvent plus rapide (numériquement) de décomposer la matrice rectangle X en valeurs singulières plutôt que de diagonaliser la matrice de variance-covariance. X s'écrit alors sous la forme

$$X = UMV^T, \tag{6.12}$$

avec M une matrice diagonale contenant des valeurs singulières σ_k, et U et V des matrices orthogonales. Dans ce cas, la matrice de covariance temporelle C_{XX} peut s'écrire sous la forme

$$C_{XX} = XX^T = UM^2U^T, \tag{6.13}$$

et la matrice de covariance spatiale s'écrit

$$C'_{XX} = X^T X = VM^2V^T. \tag{6.14}$$

Ainsi, les EOF spatiales sont les colonnes de U ($P = U$), les EOF temporelles sont les colonnes de V, et les valeurs propres correspondantes sont $\lambda_k = \sigma_k^2$ dans les deux cas. On montre [Bretherton *et al.*, 1992] que la variance « expliquée » par le premier mode est σ_1^2, donc le pourcentage expliqué par ce mode est

$$\sigma_1^2 / \sum_k \sigma_k^2. \tag{6.15}$$

L'analyse en composantes principales décrite précédemment peut donc se déduire de la décomposition en valeurs singulières. Rechercher les valeurs propres et vecteurs propre de la matrice de covariance, ou la décomposition en valeurs singulières de X, est équivalent. Dans ce cas, les coefficients d'expansion pour la décomposition en EOF spatiales, définis par l'équation 6.9, sont donnés par :

$$A = X^T U = VM. \tag{6.16}$$

Cette méthode porte le même nom que la méthode de régression présentée en partie 6.3.2, car toutes les deux reposent sur la pseudo-diagonalisation d'une matrice rectangle. Ici, il s'agit de décomposer la matrice X, pour en déduire ses modes principaux (spatiaux ou temporels). Dans le cas de la partie 6.3.2, on décompose la matrice de covariance de deux variables différentes (le prédicteur et le prédictand), afin de trouver les modes correspondant au maximum de covariance entre les deux variables.

6.3.2 Méthodes de régression

Trois méthodes sont principalement utilisées : la régression linéaire multiple, l'analyse de corrélation canonique et la décomposition en valeurs singulières. La première consiste à minimiser la RMSE sur chacune des stations, alors que les deux autres méthodes consistent à trouver les modes spatiaux donnant la plus forte relation entre les deux variable (prédicteur et prédictand) : on recherche le maximum de corrélation dans le cas de la CCA, et de covariance pour la SVD. Les méthodes pour trouver des couplages entre deux variables (CCA et SVD), et leurs applications au climat, sont décrites dans Bretherton *et al.* [1992]. Des comparaisons entre les différentes méthodes de régression, linéaires ou non, ont été effectuées, notamment dans Wilby *et al.* [1998] et Huth [1999].

Régression linéaire multiple (MVR)

La régression linéaire multiple a pour objectif de minimiser la RMSE des séries temporelles aux stations. Elle peut être appliquée au prédicteur X sans pré-traitement, ou bien uniquement aux EOFs dominantes de X, décrites en partie 6.3.1. Dans le second cas, les coefficients de régression correspondent à la contribution de chaque EOF à la prédiction de Y. On se place ici dans le cas général où Y est une matrice correspondant par exemple à plusieurs stations (régression linéaire multiple multivariée). Dans l'application faite dans la suite, on utilisera le cas particulier où le prédictand est un vecteur (régression station par station).

Le modèle de régression s'écrit

$$Y^T = X^T \Psi, \tag{6.17}$$

et la matrice Ψ contenant les coefficients de régression peut s'exprimer par [Benestad *et al.*, 2008]

$$\Psi = (XX^T)^{-1}(XY^T) = C_{XX}^{-1} C_{XY}, \tag{6.18}$$

avec $C_{XX} = XX^T$ et $C_{XY} = XY^T$. Ainsi, on détermine Ψ si la matrice C_{XX} est inversible, donc si les vecteurs lignes de X (les séries temporelles aux stations) sont indépendants. Un pré-traitement pour décomposer X en EOFs peut s'imposer dans le cas contraire. Dans le cas d'une analyse S-mode, on peut ainsi faire la projection sur l'espace engendré par les EOF spatiales (vecteurs colonne de U), ou sur un sous-espace avec les k premiers vecteurs. D'après l'équation 6.12, l'équation 6.17 s'écrit :

$$Y^T = VMU^T \Psi. \tag{6.19}$$

On cherche à exprimer cette équation en fonction des EOF spatiales (matrice U) ou des coefficients d'expansion associés A (équation 6.16)

$$Y^T = AU^T \Psi. \tag{6.20}$$

L'équation 6.20 montre que l'on peut donc faire la régression en prenant les coefficients d'expansion comme prédicteurs. L'avantage est que les prédicteurs sont alors orthogonaux. Les coefficients sont alors donnés par $U^T\Psi$. La régression peut se faire sur les k premiers coefficients d'expansion, correspondant aux k premiers modes (expliquant le plus de variance).

Décomposition en valeurs singulières (SVD)

Le principe est de décomposer la matrice $C_{XY} = XY^T$ de covariance temporelle entre deux variables X (prédicteur) et Y (prédictand) en valeur singulières. On peut la décomposer de la façon suivante :

$$C_{XY} = LMR^T, \tag{6.21}$$

avec L et R des matrices orthogonales, et M une matrice diagonale contenant des valeurs singulières. On montre alors que les vecteurs colonne de L et R constituent les bases permettant d'expliquer le maximum de covariance entre les deux variables [Bretherton *et al.*, 1992]. Pour cela, il faut projeter X sur L, et Y sur R.

$$\boldsymbol{x}(t) = \sum_k \alpha_k(t)\boldsymbol{l_k}, \tag{6.22}$$

$$\boldsymbol{y}(t) = \sum_k \beta_k(t)\boldsymbol{r_k}. \tag{6.23}$$

Les vecteurs $\boldsymbol{\alpha_k}$ et $\boldsymbol{\beta_k}$ sont les coefficients d'expansion correspondant au mode k, contenant les valeurs discrètes à des pas de temps donnés des fonctions $\alpha_k(t)$ et $\beta_k(t)$. Si l'on définit, comme dans la partie 6.3.1, les matrices A et B dont les colonnes sont respectivement $\boldsymbol{\alpha_k}$ et $\boldsymbol{\beta_k}$, on a donc :

$$LA^T = X \tag{6.24}$$

$$RB^T = Y \tag{6.25}$$

Notons que les coefficients d'expansion A et B ne sont pas orthogonaux (AA^T et BB^T ne sont pas égaux à I), mais vérifient la relation

$$A^T B = M. \tag{6.26}$$

De plus, les coefficients d'expansion d'une variable (X ou Y) ont la même covariance spatiale que celle-ci (équation 6.10) : $AA^T = X^T X$ et $BB^T = Y^T Y$.

Le problème de la réduction d'échelle est donc d'exprimer le prédictand Y en fonction du prédicteur X, des modes L et R, et de M, supposés stationnaires. Pour cela, il est utile de calculer la covariance entre le prédictand Y et les coefficients d'expansion A (d'après les équations 6.25 et 6.26) :

$$C_{YA} = YA = RB^T A = RM. \tag{6.27}$$

Les cartes de covariance (ou de corrélation) entre une variable et les coefficients d'expansion de l'autre variable (appelées carte « hétérogènes ») forment une représentation utile de la répartition spatiale des contributions des différents coefficients. De l'équation 6.27, on tire une expression de Y :

$$Y = RMA^T C_{XX}'^{-1} \tag{6.28}$$

car les covariances spatiales de A et X sont identiques (équation 6.10). En remplaçant A^T dans l'équation 6.28 par l'expression donnée dans l'équation 6.24, on a :

$$Y = RML^T X C_{XX}'^{-1}, \tag{6.29}$$

et l'on obtient après commutation, étant donné que $XC_{XX}'^{-1} = C_{XX}^{-1}X$:

$$Y = \Psi X \quad \text{avec } \Psi = C_{YX}C_{XX}^{-1}. \tag{6.30}$$

Analyse de corrélation canonique (CCA)

Il s'agit, de même que pour la SVD, d'écrire X et Y sous la forme générique :

$$HV^T = X \tag{6.31}$$

$$GU^T = Y \tag{6.32}$$

Dans le cas de la SVD, on a $H = L$ et $G = R$, donc G et H sont orthogonales. Ici, on suppose que U et V sont orthogonales, mais pas nécessairement G et H.

Le principe est de déterminer les directions expliquant le maximum de corrélation entre les variables X et Y. La matrice de corrélation s'écrit :

$$C = C_{XX}^{-1/2} C_{XY} C_{YY}^{-1/2}. \tag{6.33}$$

On effectue une décomposition de C en valeurs singulières : $C = LMR^T$. Les matrices L et R donnent les combinaisons expliquant le maximum de corrélation entre les deux variables. De plus, on a :

$$C_{XY} = HV^T UG^T, \tag{6.34}$$

donc on peut écrire :

$$C_{XX}^{-1/2} HV^T UG^T C_{YY}^{-1/2} = LMR^T. \tag{6.35}$$

Les relations $H = C_{XX}^{1/2} L$ et $G = C_{YY}^{1/2} R$ permettent d'expliquer le maximum de corrélation. On a alors :

$$V^T U = M, \tag{6.36}$$

de façon similaire à l'équation 6.26.

Il s'agit donc, à nouveau, d'exprimer Y en fonction de X, G, H, et M. Comme U et V sont orthogonales, les équations 6.32 et 6.36 donnent :

$$Y = GMV^T. \tag{6.37}$$

En écrivant V^T en fonction de X d'après l'équation 6.31, on obtient alors :

$$Y = \Psi X \quad \text{avec } \Psi = GM(H^T H)^{-1} H^T. \tag{6.38}$$

La CCA décrite ici est la CCA « simple » si elle est appliquée sur X sans pré-traitement. En pratique, elle est souvent appliquée sur les EOF de X.

6.3.3 Comparaison des méthodes

Dans Huth [1999], ces différentes méthodes sont comparées pour la climatologie. En particulier, la MVR et la CCA sont appliquées à des prédicteurs pré-traités, en intégrant un plus ou moins grand nombre d'EOF dans la régression. En général, on considère qu'au-delà d'un certain nombre d'EOF, on va intégrer des composantes qui correspondent à du « bruit » instationnaire, et qui vont dégrader les résultats de la régression (« overfitting »). Cependant, dans le cas de Huth [1999], ce phénomène n'est pas constaté.

Pour la CCA et la SVD, la sensibilité au nombre de modes de la régression est étudiée (le nombre de vecteurs propres de L et R inclus dans le résultat). Pour la SVD, augmenter le nombre de modes dégrade le résultat (en terme de RMSE) à partir du 1e mode, alors que pour la CCA, il semble qu'il y ait un optimum à 4 modes.

Au final, la régression linéaire aux stations, sans pré-traitement, donne le meilleur résultat en terme de RMSE, ce qui n'est pas étonnant puisque cette méthode, contrairement aux deux autres, vise explicitement à réduire la RMSE. Il semble également que le pré-traitement en EOFs ne soit pas indispensable, sauf à devoir rendre les vecteurs de X indépendants. Enfin, il faut noter que l'inflation est utilisée dans ce cas, ce qui dégrade donc les résultats en terme de RMSE (partie 6.2.5).

6.4 Utilisation des méthodes statistiques pour la qualité de l'air

Il s'agit d'utiliser les méthodes statistiques de réduction d'échelle (downscaling) afin de mieux prévoir les concentrations aux stations, en réduisant l'erreur de représentativité. On utilise pour cela un seul modèle, ou un ensemble de modèles. Les prédicteurs sont les sorties des modèles (interpolées aux stations), et le prédictand est la série temporelle des concentrations aux stations.

Pour l'application au chapitre 7, on utilisera la régression linéaire décrite en partie 6.3.2, avec ou sans pré-traitement, en faisant varier le nombre de prédicteurs. Dans cette partie, le cadre général de la régression linéaire est adapté au cas d'application : régression station par station sur un ensemble de modèles.

6.4.1 Régression sur un ensemble de modèles

Calcul des coefficients de régression

On suppose que la régression se fait sur une seule station (le prédictand $\boldsymbol{y}$ est un vecteur, contenant une série de N_t observations à la station). X est la matrice de prédicteurs, contenant les concentrations simulées par l'ensemble des modèles (interpolées à la station). X est donc de taille $N_s \times N_t$, avec N_s le nombre de modèles inclus dans l'ensemble et N_t le nombre d'observations à la station. Chaque ligne contient donc la série temporelle à la station donnée par l'un des modèles de l'ensemble. La régression s'écrit alors :

$$\boldsymbol{y} = X^T \boldsymbol{b} + \boldsymbol{\epsilon}, \tag{6.39}$$

avec $\boldsymbol{b}$ le vecteur contenant les coefficients de régression (de taille N_s). $\boldsymbol{\epsilon}$ est un vecteur contenant les composantes de bruit blanc, de variance σ_ϵ^2.

Dans le cas où la régression est effectuée avec une composante constante (« intercept »), il suffit d'augmenter la matrice X^T d'une colonne contenant des 1. Le vecteur $\boldsymbol{b}$ est alors de taille $N_s + 1$ car il contient la constante en plus des autres coefficients. En pratique, ce sera toujours le cas ici (on incluera toujours une composante constante dans la régression).

On a une estimation des coefficients de régression $\widehat{\boldsymbol{b}}$, qui est calculée en fonction de $\boldsymbol{y}$:

$$\widehat{\boldsymbol{b}} = D\ \boldsymbol{y} \quad \text{avec } D = (XX^T)^{-1}X. \tag{6.40}$$

L'estimation de $\boldsymbol{y}$ est alors donnée par :

$$\widehat{\boldsymbol{y}} = X^T \widehat{\boldsymbol{b}}. \tag{6.41}$$

Notons que l'on a $DX^T = I$ (avec I la matrice identité) : on appelle D le pseudo-inverse de X^T (égal à $X^{T^{-1}}$ si X est inversible). La relation entre l'estimation $\widehat{\boldsymbol{b}}$ et la « vraie » valeur $\boldsymbol{b}$ est donnée par :

$$\begin{aligned} \widehat{\boldsymbol{b}} &= D\boldsymbol{y} \\ &= DX^T\boldsymbol{b} + D\boldsymbol{\epsilon} \\ &= \boldsymbol{b} + D\boldsymbol{\epsilon}. \end{aligned} \tag{6.42}$$

Variance des coefficients de régression

Tout d'abord, $\widehat{\boldsymbol{b}}$ est une variable aléatoire, contrairement à $\boldsymbol{b}$. On a donc $E(\boldsymbol{b}) = \boldsymbol{b}$, ce qui conduit à :

$$E\left(\widehat{\boldsymbol{b}}\right) = E(\boldsymbol{b}) = \boldsymbol{b}. \tag{6.43}$$

On peut également calculer la variance des coefficients de régression :

$$\begin{aligned}
\mathrm{var}(\widehat{\boldsymbol{b}}) &= E\left(\left(\widehat{\boldsymbol{b}} - E(\widehat{\boldsymbol{b}})\right)\left(\widehat{\boldsymbol{b}} - E(\widehat{\boldsymbol{b}})\right)^T\right) \\
&= E\left(\left(\widehat{\boldsymbol{b}} - \boldsymbol{b}\right)\left(\widehat{\boldsymbol{b}} - \boldsymbol{b}\right)^T\right) \\
&= \mathrm{var}\left(\widehat{\boldsymbol{b}} - \boldsymbol{b}\right) \\
&= \mathrm{var}\left(D\boldsymbol{\epsilon}\right) \qquad (6.44)
\end{aligned}$$

On peut donc écrire la variance de l'équation 6.44 de la façon suivante :

$$\mathrm{var}\left(\boldsymbol{b}\right) = E\left(D\left(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right)D^T\right) \qquad (6.45)$$

A priori, les variables D et $\boldsymbol{\epsilon}$ sont décorrélées. De plus, $\boldsymbol{\epsilon}$ est un vecteur contenant des composantes de bruit blanc de variance σ_ϵ^2. On a donc

$$\mathrm{var}\left(\epsilon\right) = \sigma_\epsilon^2 I. \qquad (6.46)$$

L'équation 6.45 s'écrit donc finalement :

$$\mathrm{var}\left(\boldsymbol{b}\right) = DE\left(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right)D^T = \sigma_\epsilon^2 DD^T = \sigma_\epsilon^2\ \left(XX^T\right)^{-1}. \qquad (6.47)$$

Cette équation 6.47 montre que la variance des coefficients de régression peut devenir très grande s'il existe une forte colinéarité entre les prédicteurs. Ici, les modèles représentent tous une même évolution aux stations : ils sont donc probablement très colinéaires. Ce phénomène est appelé « inflation de la variance », à ne pas confondre avec l'inflation brièvement évoquée en partie 6.2.5 : il s'agit de la variance des coefficients de régression, et non du prédictand lui-même.

Notons que le résultat 6.47 est obtenu en supposant que X n'est pas une variable aléatoire. On ne modélise donc pas l'incertitude des modèles.

Application en prévision

En pratique, il s'agit de prévoir les concentrations à une station au temps $n+1$, connaissant les observations jusqu'au temps n (notées $\boldsymbol{y}_n$). La matrice X tronquée au temps n est notée X_n. La régression est effectuée sur la période d'apprentissage, jusqu'au pas de temps n, puis appliquée au pas de temps $n+1$. Les étapes sont les suivantes :

1. Calcul des coefficients $\widehat{\boldsymbol{b}}_n$ par la régression de $\boldsymbol{y}_n$ sur X_n (équation 6.40),
2. Constitution de la matrice X_{n+1} avec les prévisions pour l'échéance $n+1$,
3. Calcul des estimations pour les N jours suivants : $\widehat{\boldsymbol{y}}_{n+i} = X_{n+i}^T\widehat{\boldsymbol{b}}_n$ (équation 6.41), avec $i \in \{1, \ldots, N\}$.

Le paramètre N détermine la période sur laquelle les coefficients de régression peuvent être appliqués sans être recalculés (stationnarité).

6.4.2 Décomposition en composantes principales (PCA)

La décomposition en composantes principales permet d'éliminer la colinéarité entre les régresseurs. Cette méthode a déjà été décrite dans la partie 6.3.1, on ne la rappelle donc que brièvement ici.

Principe

X est toujours matrice contenant les régresseurs, de taille $N_s \times N_t$ (où N_s est le nombre de modèles, et N_t le nombre de pas de temps), à une station donnée. La réduction en composantes principales consiste à trouver les vecteurs propres de la matrice de variance-covariance $C_{XX} = XX^T$, notés P, qui sont les « directions principales » de X. Les valeurs propres associées donnent la part de variance de X expliquée par ces directions. Les composantes principales A, ou coefficients d'expansion, définis par l'équation 6.8, se retrouvent par la projection de X sur la base formée par les directions principales. La régression peut alors s'effectuer sur ces composantes principales. La matrice P est de taille $N_s \times N_s$, et A est de taille $N_t \times N_s$.

Application en prévision

Les directions principales sont supposées stationnaires : en prévision, elles sont calculées au temps n et sont utilisées au temps $n+1$ pour calculer les nouvelles composantes principales. L'algorithme est le suivant :

1. Calcul des directions principales au temps n (notées P_n), d'après la matrice X_n (par une SVD de X_n par exemple, équation 6.12),
2. Calcul des composantes principales au temps n (équation 6.9) : $A_n = X_n^T P_n$,
3. Régression de $\boldsymbol{y}_n$ sur A_n^T : obtention des coefficients de régression $\widehat{\boldsymbol{b}}_n$ (équation 6.40),
4. Calcul des composantes principales au temps $n+1$: $A_{n+1} = X_{n+1}^T P_n$,
5. Calcul de l'estimation au temps $n+1$ (équation 6.41) : $\widehat{\boldsymbol{y}}_{n+1} = A_{n+1}\widehat{\boldsymbol{b}}_n$.

6.4.3 Décomposition en composantes principales ajustées (PFC)

Principe

La réduction en composantes principales permet de réduire le nombre de régresseurs et de les rendre orthogonaux. Cependant, les directions choisies sont celles qui expliquent le plus de variance de X, indépendamment de toute information sur $\boldsymbol{y}$. Rien n'indique a priori que ce seront les directions les plus pertinentes pour la régression : une composante expliquant très peu de variance de X peut apporter de l'information sur la relation avec la variable $\boldsymbol{y}$. Ainsi, dans le choix des directions, il est utile d'inclure ce que l'on sait a priori des liens entre X et $\boldsymbol{y}$. C'est le but de la réduction en composantes principales ajustées (« Principal Fitted Components ») décrite dans Cook [2007]. Notons que les deux autres méthodes d'analyse que la régression, la SVD et la CCA, permettent aussi un tel couplage dans le choix des directions. Cependant, on souhaite rester ici dans le cadre de la régression linéaire, qui permet de réduire la RMSE (les autres méthodes étant a priori moins efficaces sur ce point).

Application en prévision

Le principe est d'effectuer la décomposition, puis la régression, décrites en partie 6.4.2 sur une matrice $\widehat{X}$ qui a été « ajustée » sur $\boldsymbol{y}$ (comme précédemment, les variables ont été centrées). Pour cela, on procède de la façon suivante :

1. On commence par estimer une matrice F_y dépendant de $\boldsymbol{y}$, qui décrit la relation a priori entre X et Y. Si ces variables sont a priori linéairement dépendantes, F_y pourra être simplement $\boldsymbol{y}^T$ (de taille $1 \times N_t$). Si, par exemple, on a une relation polynomiale, on pourra prendre $F_y = \left(\boldsymbol{y}, \boldsymbol{y}^2, \boldsymbol{y}^3\right)^T$. Ainsi, F_y est de taille $N_r \times N_t$, avec N_r en général faible, donnant le nombre de composantes de la relation entre X et $\boldsymbol{y}$. F_y doit être centré.

2. Une fois F_y déterminée, on effectue une régression de X sur F_y : $\widehat{X}$ est alors l'estimation de X donnée par cette régression. En réalité, il s'agit d'effectuer une projection de X sur un espace réduit engendré par les lignes de F_y. On a en effet de façon similaire à l'équation 6.17,
$$\widehat{X}^T = P_F X^T \quad \text{avec} \quad P_F = F_y^T \left(F_y F_y^T\right)^{-1} F_y, \tag{6.48}$$
3. On recherche les directions principales ajustées au temps n, $\widehat{P}_n$ en effectuant une SVD de la matrice $\widehat{X}_n$ (Le nombre de directions principales ajustées est $N_r' \leq N_r$),
4. On calcule les composantes principales ajustées, qui sont données par $\widehat{A}_n = X_n^T \widehat{P}_n$ (projection de X, et non $\widehat{X}$, sur les directions principales ajustées)
5. Régression de $\boldsymbol{y}_n$ sur A_n^T : obtention des coefficients de régression $\widehat{\boldsymbol{b}}_n$ de façon similaire à l'équation 6.40,
6. On calcule les composantes principales ajustées au temps $n+1$: $\widehat{A}_{n+1} = X_{n+1}^T \widehat{P}_n$,
7. Calcul de l'estimation au temps $n+1$ (équation 6.41) : $\widehat{\boldsymbol{y}}_{n+1} = \widehat{A}_{n+1} \widehat{\boldsymbol{b}}_n$.

6.5 Conclusion

6.5.1 Méthodes de réduction d'échelle

Les méthodes statistiques pour la réduction d'échelle présentées dans ce chapitre permettent de corriger les estimations fournies par les modèles en tenant compte de l'erreur de représentativité des stations. Différentes conditions sont nécessaires afin que l'utilisation de méthodes statistiques soit pertinente : (1) le lien physique entre le prédicteur et le prédictand, (2) la stationnarité de la relation pour l'application en prévision, (3) la bonne estimation du prédicteur par les modèles utilisés. Dans le cadre de l'application en qualité de l'air, les prédicteurs sont les séries temporelles de concentrations prédites par des modèles, et le prédictand est la série observée à la station. Le prédicteur et le prédictand représentent donc la même variable physique. Les conditions (2) et (3) sont plus problématiques, et ne sont pas nécessairement vérifiées. Pour tenir compte de l'erreur modèle (condition (3)), l'utilisation d'un ensemble de modèles est une solution couramment utilisée.

Différentes méthodes statistiques ont été détaillées dans ce chapitre. Pour l'application à la qualité de l'air, les méthodes de régression linéaire ont été retenues, en utilisant ou non un pré-traitement pour rendre les régresseurs (ou prédicteurs) orthogonaux. Les méthodes utilisées dans le chapitre suivant ont été détaillées en partie 6.4.

6.5.2 Questions

Le chapitre 7 présente l'application à la qualité de l'air, pour la prévision des pics d'ozone à l'échelle continentale. Les questions qui y seront plus particulièrement abordées sont :

1. Quel est l'apport de l'ensemble, en fonction du nombre de modèles ? La colinéarité fait-elle apparaître des problèmes dans la performance en prévision ?
2. Quels sont les « bons » indicateurs pour montrer qu'il y a colinéarité ?
3. Y a-t-il stationnarité des coefficients de régression ?
4. Les résultats obtenus par la réduction d'échelle dépendent-ils du type de station, et donc de leur échelle de représentativité ?

Chapitre 7

Réduction d'échelle pour la qualité de l'air à l'échelle européenne

Ce chapitre présente l'application des méthodes de réduction d'échelle pour la prévision des pics d'ozone à l'échelle européenne. Dans un premier temps, la réduction d'échelle est effectuée sur un seul modèle, appelé « modèle de référence ». Il s'agit de Polair3D avec une configuration (paramétrisations physiques, mécanisme chimique...) standard. Pour ce modèle, la réduction d'échelle est effectuée sur deux réseaux de mesures, Pioneer et Airbase. Les résultats sont comparés pour les stations urbaines, péri-urbaines et rurales, afin de quantifier l'apport de la réduction d'échelle en fonction de l'échelle de représentativité des stations. Enfin, la question de la stationnarité des coefficients est abordée, notamment en comparant les résultats obtenus en fonction de la période au bout de laquelle les coefficients sont recalculés.
Dans un deuxième temps, on utilise un ensemble de modèles, construit en faisant varier les paramétrisations du modèle. La réduction d'échelle est effectuée sur un nombre N_s de modèles de l'ensemble, utilisés comme régresseurs, et la performance est étudiée en fonction du nombre de modèles inclus. Un sous-ensemble de modèles, « calibré » uniquement sur les stations de fond, est également utilisé. Sa dispersion est alors supposée représentative de l'« erreur modèle », différente de l'erreur de représentativité prise en compte en réduction d'échelle.
Les sorties de modèles utilisées comme régresseurs sont fortement colinéaires. Pour rendre les prédicteurs indépendants, des méthodes de prétraitement sont utilisées : la décomposition en composantes principales classique, ainsi que la décomposition en composantes principales ajustées. Les différentes méthodes sont comparées du point de vue de la performance en prévision.

Sommaire

7.1 Application : qualité de l'air à l'échelle européenne 197
7.1.1 Les réseaux de mesures 197
7.1.2 Les simulations 199
7.1.3 Méthode statistique utilisée : régression linéaire 199
7.2 Résultats par type de station 199
7.2.1 Régressions sur un seul modèle de référence (prévision à 1 jour) 200
7.2.2 Résultats pour la prévision à N jours et stationnarité des coefficients . . 202

7.3 Prise en compte de l'ensemble 206
7.3.1 Performance en fonction du nombre de modèles 206
7.3.2 Colinéarité 210
7.4 Réduction de dimension 213
7.4.1 Utilisation d'ensembles calibrés 213
7.4.2 Réduction en composantes principales (PCA) 214
7.4.3 Réduction en composantes principales ajustées (PFC) 217
7.4.4 Comparaison des différentes méthodes 223
7.5 Conclusion 223
7.5.1 Résultats aux stations 223
7.5.2 Représentativité des stations 223

Les méthodes de réduction d'échelle présentées dans le chapitre 6, et plus particulièrement dans la partie 6.4, sont appliquées à la prévision des pics d'ozone à l'échelle européenne. Les méthodes sont appliquées et comparées du point de vue de la performance aux stations d'observation, en particulier en terme de RMSE. La partie 7.1 décrit les configurations des simulations numériques réalisées, ainsi que les réseaux de mesure utilisés pour la comparaison aux observations. La partie 7.2 présente les résultats obtenus en réduction d'échelle pour un seul modèle, le « modèle de référence ». La partie 7.3 décrit les résultats obtenus en utilisant un ensemble de modèles, et met en évidence la colinéarité entre les prédicteurs et les problèmes qui peuvent en résulter. Enfin, la partie 7.4 introduit différentes méthodes pour réduire le nombre de régresseurs, ainsi que pour les rendre orthogonaux via une décomposition en composantes principales.

7.1 Application : qualité de l'air à l'échelle européenne

7.1.1 Les réseaux de mesures

Deux réseaux de stations sont utilisés pour la régression et la comparaison aux observations : Pioneer et Airbase.

Pioneer

Il contient 344 stations dont la majorité en France (179) et en Allemagne (82). Seules les stations dont le type était spécifié ont été retenues, afin de pouvoir effectuer des comparaisons entre les types de stations. De plus, afin d'avoir suffisamment de points pour la régression, une sélection est effectuée afin de ne retenir que les stations ayant plus de 200 observations. Cela correspond, au final, à 154 stations (figure 7.1). Les stations sont classées en trois types suivant la terminologie de l'ADEME (Agence de l'environnement et de la maîtrise de l'énergie) :

- une station **urbaine** possède un rayon de représentativité de 100 m à 1 km,
- une station **péri-urbaine** a un rayon de 1 km à 5 km,
- une station **rurale**, un rayon de plus de 5 km.

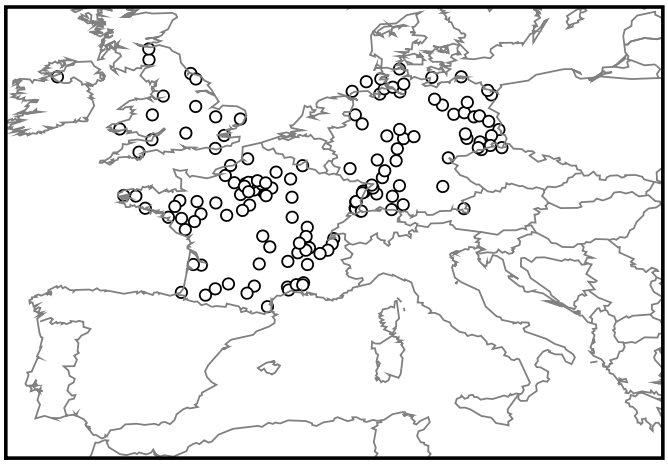


Fig. 7.1 – Emplacement des stations utilisées dans la régression, pour le réseau Pioneer (154 stations).

Airbase

Le réseau Airbase contient 6738 stations réparties sur toute l'Europe. Il s'agit du réseau de l'Agence Européenne de l'Environnement, qui regroupe les données de plus de 30 pays européens [1].

Les stations sont, en général, classées suivant l'occupation des sols (rural/urbain/péri-urbain) et le type de sources à proximité (industrielles/trafic). Cependant, comme l'ozone est un polluant secondaire, le classement en fonction des sources n'est plus très pertinent, et une autre classification est proposée.

Les stations sont classées en quatre types :

- une station **urbaine** est représentative de la pollution urbaine de fond (pas d'émissions locales), et située dans un site ouvert ou une zone résidentielle ou commerciale,
- une station **péri-urbaine** est située à une certaine distance de la région d'émissions maximales, plutôt sous les vents dominants,
- une station **rurale** est située en zone rurale, représentative pour l'ozone et éloignée de l'influence des émissions locales,
- une station **rurale de fond** est située en zone rurale, loin des émissions, et dans des endroits où la topographie et la météorologie permettent une large représentativité (pas d'inversion, pas de sommets montagneux ou de côtes).

Pour l'ozone, les stations urbaines donnent plutôt de faibles concentrations (titration par les NOx), et les stations péri-urbaines sont celles qui représentent les zones où la formation d'ozone est la plus importante. La plupart des stations sont représentatives des niveaux d'ozone sur quelques kilomètres carrés, voire quelques dizaines de kilomètres carrés, excepté les stations de fond qui représentent une zone de 1000 km^2 à 10 000 km^2.

Comme pour Pioneer, seules les stations dont le type était spécifié, et avec un suffisamment grand nombre d'observations, ont été utilisées, ce qui correspond à 1247 stations (figure 7.2).

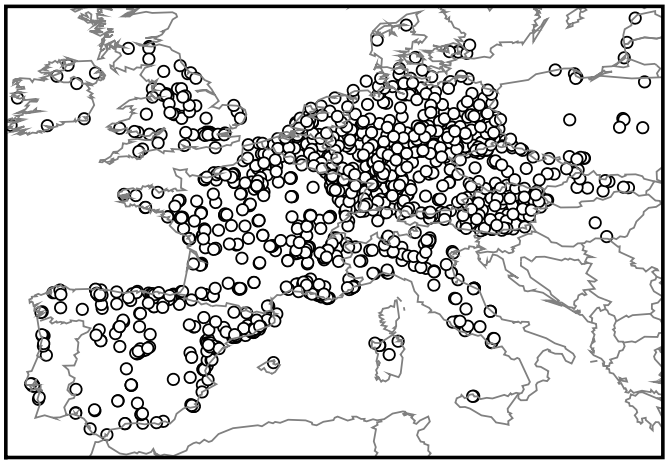


Fig. 7.2 – Emplacement des stations utilisées dans la régression, pour le réseau Airbase (1247 stations).

[1] http://air-climate.eionet.europa.eu/databases/airbase/airbasexml/index_html

7.1.2 Les simulations

Les simulations ont été effectuées sur l'Europe, avec une résolution de 0.5°, pour une durée d'un an (2001). Le polluant étudié est l'ozone, pour les pics journaliers. Au total nous disposons de 107 modèles qui ont été générés dans la plate-forme de modélisation Polyphemus. Ces modèles sont créés à partir de Polair3D et diffèrent par une ou plusieurs paramétrisations physiques, et/ou par les données d'entrée (émission, météorologie, données de sol) utilisées. La méthode utilisée pour construire l'ensemble est décrite dans Garaud et Mallet [2009b]. Dans un premier temps, la régression est effectuée avec le « modèle de référence ». Il s'agit du modèle que l'on aurait utilisé pour effectuer les prévisions, sans disposer d'un ensemble. Il ne s'agit pas forcément du meilleur modèle de l'ensemble pour le cas étudié, mais il est représentatif des performances d'un « bon » modèle. Ensuite, différents ensembles de modèles, choisis parmi les 107 modèles à notre disposition, sont utilisés.

Les valeurs simulées aux stations sont calculées avec une interpolation bilinéaire sur les mailles voisines de la station.

7.1.3 Méthode statistique utilisée : régression linéaire

La régression est effectuée station par station, sur les valeurs de pics d'ozone journaliers. L'erreur de représentativité que l'on cherche à corriger dépend en effet de chaque station. Les coefficients de régression sont propres à la station et correspondent à la correction systématique de cette erreur. La première moitié de l'année est utilisée uniquement comme période d'apprentissage, et la période de prévision s'étend sur les 150 derniers jours (5 mois). La comparaison aux observations s'effectue sur la RMSE aux stations, pour différents cas :

- Régression effectuée sur toute l'année, et RMSE calculée sur les 5 derniers mois. Il s'agit donc d'une RMSE a posteriori (les valeurs utilisées pour la comparaison ont été incorporées dans la régression).
- Prévision à N jours : la régression est effectuée sur une période donnée, puis les coefficients obtenus sont utilisés pour la prévision sur les N jours suivants. En général, dans les résultats présentés ici, $N = 1$.

La régression est effectuée de la façon décrite en partie 6.4.1.

7.2 Résultats par type de station

Le type de station est une information importante. En effet, l'enjeu de la réduction d'échelle est de compenser l'erreur de représentativité du modèle : en faisant des statistiques sur une période suffisamment longue, on compense les erreurs systématiques du modèle à une station. Si le modèle est, dans l'ensemble, débiaisé sur les stations de fond (c'est le cas de notre modèle pour les pics d'ozone), l'erreur systématique à une station est essentiellement due à l'erreur de représentativité. Celle-ci provient du fait que le modèle simule les phénomènes à une certaine échelle (de l'ordre de la taille d'une maille), et néglige les effets sous-maille. Ainsi, même un modèle « parfait » (prédisant parfaitement les valeurs moyennes dans une maille) aurait une erreur de représentativité (déjà évoquée en partie 1.3.5).

La variabilité locale est beaucoup plus forte autour d'une station urbaine, à cause des émissions (industrielles et trafic) et de l'influence des bâtiments. Le rayon de représentativité d'une station urbaine est beaucoup plus faible que pour une station rurale, et l'erreur de représentativité du modèle sera donc plus grande dans ce cas. L'un des enjeux de cette partie est de mettre en évidence l'apport de la régression en fonction de l'échelle de représentativité des stations, donc de leur type.

7.2.1 Régressions sur un seul modèle de référence (prévision à 1 jour)

Pioneer

Le tableau 7.1 donne les statistiques pour le réseau Pioneer, pour l'ensemble des stations ainsi que pour les trois types de stations décrits en partie 7.1.1. La RMSE est donnée pour les trois cas suivants :

1. *RMSE modèle* : le modèle de référence est utilisé seul, sans post-traitement statistique,
2. *RMSE a posteriori* : la régression est effectuée sur l'ensemble de l'année, y compris les 5 mois sur lesquels est calculée la RMSE,
3. *RMSE en prévision à 1 jour* : la régression est effectuée tous les jours, sur l'ensemble de la période passée, et les coefficients sont utilisés pour calculer les résultats au jour suivant.

Les résultats a posteriori sont meilleurs que ceux de la prévision, du fait que les observations sur lesquelles la RMSE est calculée sont incluses dans la régression. La prévision à un jour permet encore une amélioration significative des résultats par rapport à la performance du modèle sans régression (-6% sur la RMSE pour toutes les stations). L'erreur modèle est plus grande sur les stations urbaines que péri-urbaines et rurales :

$$RMSE_{\text{urbaine}} > RMSE_{\text{periurbaine}} > RMSE_{\text{rurale}}, \tag{7.1}$$

ce qui est cohérent avec la plus grande variabilité ainsi que l'échelle de représentativité plus faible des stations urbaines par rapport aux stations rurales. L'amélioration due à la méthode de réduction d'échelle suit la même tendance, y compris en considérant le pourcentage de diminution de la RMSE : -11% pour les stations urbaines, -4% pour les stations péri-urbaines et -2% pour les stations rurales.

Stations	RMSE modèle	RMSE a posteriori	RMSE prévision à 1 jour
Toutes (154)	22.4	20.3	21.1 (-6%)
Urbaines (35)	25.4	21.9	22.6 (-11%)
Péri-urbaines (47)	24.0	21.9	22.9 (-4%)
Rurales (72)	19.5	18.2	19.1 (-2%)

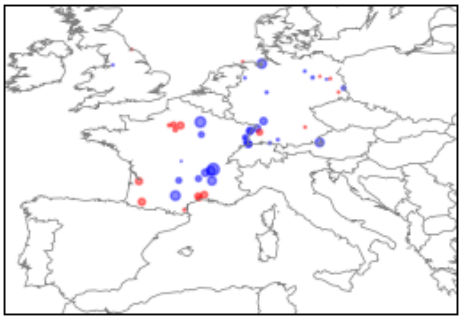
Tab. 7.1 – Statistiques sur le réseau Pioneer, en fonction du type de station : RMSE pour le modèle sans régression, RMSE a posteriori (régression sur toute la période) et RMSE pour la prévision à un jour. Les RMSE sont en $\mu g\,m^{-3}$. Le nombre de stations de chaque type est indiqué entre parenthèses. Dans la dernière colonne, la différence relative entre la RMSE du modèle et la RMSE en prévision à un jour est également mentionnée.

La figure 7.3 montre la différence entre la RMSE obtenue en prévision à un jour et la RMSE du modèle de référence, pour chacune des stations du réseau Pioneer. Les trois figures correspondent aux trois types de stations (urbaines, rurales et péri-urbaines). Les stations en bleu indiquent une différence de RMSE négative, donc une amélioration de la RMSE, et inversement le rouge est utilisé pour une différence positive.

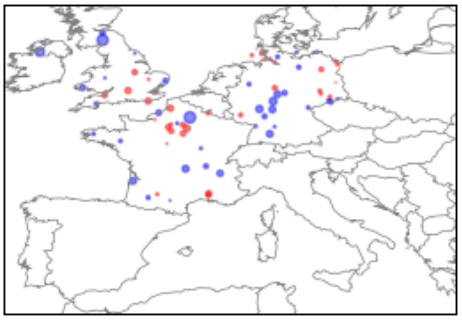
Pratiquement toutes les stations urbaines (figure 7.3(a)) montrent une amélioration due au traitement statistique, ce qui confirme les résultats précédents. Le nombre de stations en rouge (différence positive) est plus grand pour les stations péri-urbaines (figure 7.3(b)), et encore davantage pour les stations rurales (figure 7.3(c)). Cela signifie que la relation statistique entre le modèle et les observations, utilisée pour ces stations, ne modélise pas une erreur de représentativité, systématique, mais corrige des erreurs diverses (de modélisation, ou d'observation). Elle n'est donc pas fiable pour être utilisée en prévision d'un jour sur l'autre. Pour

les stations rurales notamment, il est souvent plus intéressant de prendre les valeurs données directement par le modèle, car il y a peu ou pas d'erreur de représentativité. Cependant, sur l'ensemble des stations, il est toujours intéressant de recourir à la réduction d'échelle. En effet, la différence de RMSE (en valeur absolue) est en général plus grande aux stations où elle est négative (amélioration) qu'aux endroits où elle est positive (dégradation). La RMSE globale est donc toujours améliorée par la réduction d'échelle.

(a) Pioneer, stations urbaines. Différences de RMSE (min : -9.6 ; max : 1.8)



(b) Pioneer, stations péri-urbaines. Différences de RMSE (min : -8.4 ; max : 2.8)



(c) Pioneer, stations rurales. Différences de RMSE (min : -7.8 ; max : 2.4)

FIG. 7.3 – Différence de RMSE aux stations, en $\mu g\,m^{-3}$, sur le réseau Pioneer : la surface des cercles est proportionnelle à la différence (en valeur absolue) entre la RMSE en prévision à un jour, et la RMSE du modèle de référence sans régression. Cercles bleus : différence négative (la régression améliore les résultats). Cercles rouges : différence positive (la régression détériore les résultats).

Airbase

Le tableau 7.2 montre les résultats pour le réseau Airbase, pour les mêmes cas que précédemment. Pour ce réseau beaucoup plus complet, la performance du modèle de référence sans traitement statistique est moins bonne que pour le réseau Pioneer, en particulier pour les stations urbaines et péri-urbaines. La RMSE est améliorée, là aussi, par le traitement statistique. La baisse de RMSE due au traitement statistique est de -20% en prévision à 1 jour par rapport à la perfor-

mance du modèle de référence, et jusqu'à −26% pour les stations urbaines. La relation 7.1 est encore vérifiée, et l'amélioration (en valeur absolue ou relative) due au traitement statistique suit la même tendance. En particulier, l'amélioration aux stations urbaines et péri-urbaines est beaucoup plus importante qu'aux stations rurales.

Stations	RMSE modèle	RMSE a posteriori	RMSE prévision à 1 jour
Toutes (1247)	27.4	21.0	21.9 (-20%)
Urbaines (646)	29.6	21.0	21.9 (-26%)
Péri-urbaines (251)	28.1	22.4	23.4 (-17%)
Rurales (350)	22.1	19.8	20.9 (-5%)
Rurales de fond (92)	21.4	19.2	20.4 (-5%)

Tab. 7.2 – Statistiques sur le réseau Airbase, en fonction du type de station : RMSE pour le modèle sans régression, RMSE a posteriori (régression sur toute la période) et RMSE pour la prévision à un jour. Les RMSE sont en $\mu g\,m^{-3}$. Le nombre de stations de chaque type est indiqué entre parenthèses. Les stations rurales incluent également les stations rurales de fond. Dans la dernière colonne, la différence relative entre la RMSE du modèle et la RMSE en prévision à un jour est également mentionnée.

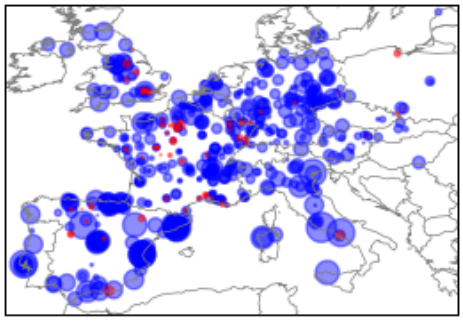
La figure 7.4 montre les différences de RMSE aux stations, pour les trois types différents de stations. Encore une fois, l'amélioration est notable pour une grande majorité de stations urbaines (figure 7.4(a)) alors que pour certaines stations péri-urbaines et rurales, la performance se dégrade légèrement avec le traitement statistique (stations en rouge). Dans tous les cas, il existe des stations auxquelles l'amélioration est très grande : la RMSE est améliorée au maximum de 57.2 $\mu g\,m^{-3}$ pour les stations urbaines, 51.6 $\mu g\,m^{-3}$ pour les stations péri-urbaines et 47.7 $\mu g\,m^{-3}$ pour les stations rurales. Ces différences très élevées proviennent de stations où la performance du modèle de référence est très mauvaise. Il est intéressant de remarquer que, pour les stations rurales (figure 7.4(c)), celles présentant l'amélioration la plus notable se situent près des côtes. Il s'agit donc de stations dont le rayon de représentativité est faible, et qui ne sont pas bien modélisées à cause de la topographie et des caractéristiques météorologiques mal représentées : l'erreur de représentativité est très bien corrigée par le traitement statistique à ces stations.

7.2.2 Résultats pour la prévision à N jours et stationnarité des coefficients

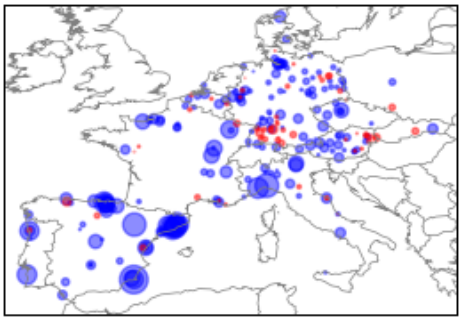
L'une des questions qui se posent est la stationnarité : est-ce que les poids, s'ils sont appris sur une période suffisamment grande, convergent ? Peuvent-ils dès lors être appliqués pour des prévisions sur de longues périodes ? Est-ce nécessaire, au contraire, de recalculer ces poids tous les jours, pour la prévision au jour suivant ?

Coefficients aux stations

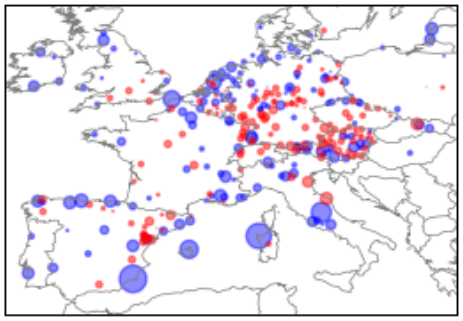
La régression sur un modèle fournit un coefficient de régression et une constante pour une station. Ces deux valeurs sont recalculées tous les jours en prévision, sur une période de plus en plus longue. Un premier aperçu de la stationnarité est donné par l'évolution des valeurs de ces coefficients au cours du temps, pour certaines stations. On en choisit quatre du réseau Pioneer. ST-NAZAIRE-BLUM est une station urbaine représentative de performances moyennes des modèles (RMSE modèle : 19.2, RMSE en prévision à 1 jour : 17.7), et LYON-GERLAND, également urbaine, a des performances plutôt mauvaises sans régression mais fortement améliorées par le traitement statistique (RMSE modèle : 34.8, RMSE en prévision à 1 jour : 28.4). Les deux



(a) Airbase, stations urbaines. Différences de RMSE (min : -57.2 ; max : 5.3)



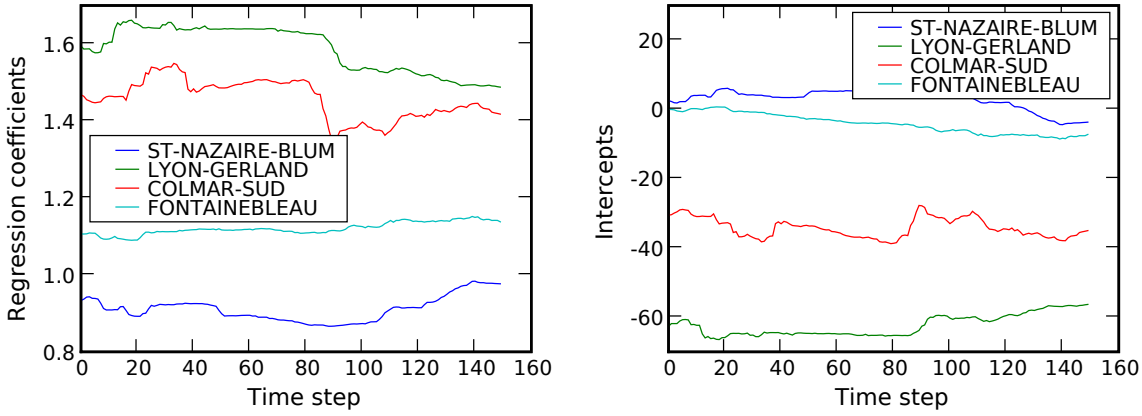
(b) Airbase, stations péri-urbaines. Différences de RMSE (min : -51.6 ; max : 4.4)



(c) Airbase, stations rurales. Différences de RMSE (min : -47.7 ; max : 8.9)

FIG. 7.4 – Différence de RMSE aux stations, en $\mu g\,m^{-3}$, sur le réseau Airbase : la surface des cercles est proportionnelle à la différence (en valeur absolue) entre la RMSE en prévision à un jour, et la RMSE du modèle de référence sans régression. Cercles bleus : différence négative (la régression améliore les résultats). Cercles rouges : différence positive (la régression détériore les résultats).

autres stations sont la station péri-urbaine COLMAR-SUD (RMSE modèle : 29.3, RMSE en prévision à 1 jour : 29.1), et la station rurale FONTAINEBLEAU (RMSE modèle : 18.0, RMSE en prévision à 1 jour : 19.6), où les performances de la régression sont moins bonnes. La figure 7.5 montre l'évolution des coefficients de régression au cours du temps, pour les quatre stations. Quel que soit le type de station et la performance de la régression, on constate une certaine stationnarité des coefficients et des valeurs de la composante constante. Il s'agit à présent de voir si les petites variations observées ont un impact sur la performance du modèle.



(a) Coefficient de régression (b) Constante

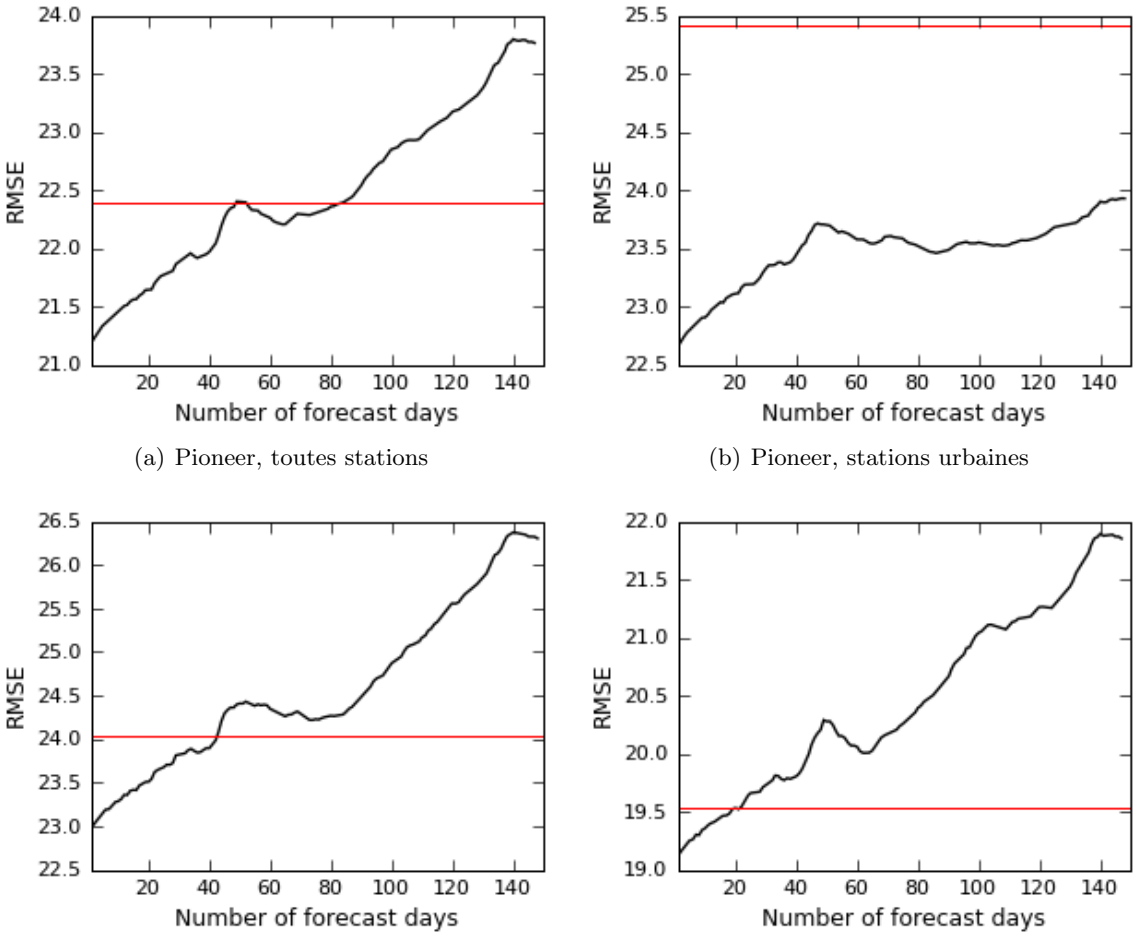
FIG. 7.5 – Coefficients de régression en fonction du temps, pour 4 stations du réseau Pioneer.

Une autre façon d'avoir un aperçu de la stationnarité (ou non) des coefficients est de calculer la RMSE en appliquant les coefficients de régression sur une période de prévision de plus en plus grande. Jusqu'à présent, les résultats présentés correspondent à une prévision à 1 jour : les coefficients sont calculés sur une période allant de $t = 0$ (début de l'année) à $t = i$ (jour i), puis appliqués pour calculer les concentrations au jour $i + 1$. Pour la prévision à N jours, les coefficients calculés au jour i sont utilisés en prévision pour les jours $i + 1$, $i + 2$, ..., $i + N$ sans être recalculés. La régression est donc effectuée tous les N jours. Dans la suite, les performances de la réduction d'échelle sont étudiées en faisant varier N entre 1 jour et 5 mois.

Pioneer

La figure 7.6 montre l'évolution de la RMSE en prévision en fonction du nombre de jours sur lesquels sont appliqués les mêmes coefficients. Sur chacune des figures, la RMSE du modèle de référence sans régression est également représentée. La figure 7.6(a) montre l'évolution de la RMSE globale, à toutes les stations. La RMSE a tendance à augmenter avec le nombre de jours de prévision, ce qui montre que les coefficients de régression ne sont pas complètement stationnaires. Cette tendance est visible également sur les courbes tracées séparément pour les différents types de station. La RMSE en prévision devient moins bonne que la RMSE de référence au bout d'un certain temps, qui varie suivant les types de stations : il est de 20 jours pour les stations rurales (figure 7.6(d)), 40 jours pour les stations péri-urbaines (figure 7.6(c)), et 45 jours pour la RMSE globale. La RMSE en prévision est toujours meilleure que la RMSE du modèle pour les stations urbaines (figure 7.6(b)). Cela provient surtout du fait que la différence de RMSE au départ (en prévision à 1 jour) est très grande dans le cas urbain, et qu'il faut donc plus de temps pour « revenir » à la RMSE de référence. L'augmentation en valeur absolue de la RMSE est la même pour les stations rurales et péri-urbaines. Pour les stations urbaines,

on observe une augmentation similaire entre 1 et 40 jours , puis une stabilisation de la RMSE qui devient quasiment constante, contrairement aux autres types de stations où l'augmentation continue.



(a) Pioneer, toutes stations

(b) Pioneer, stations urbaines

(c) Pioneer, stations péri-urbaines

(d) Pioneer, stations rurales

FIG. 7.6 – RMSE en fonction de la fréquence à laquelle sont recalculés les poids de la régression (entre 1 jour et 5 mois). RMSE calculée sur les 5 derniers mois de l'année. Réseau Pioneer. Résultats par type de station. La ligne rouge représente la RMSE du modèle de référence sans régression.

La réduction d'échelle est utilisée afin de compenser l'erreur de représentativité des stations. Celle-ci est en partie stationnaire, notamment en ce qui concerne l'erreur due à la localisation du capteur (capteur placé près d'une côte, ou près d'une route ou d'un bâtiment), mais dépend également des variations météorologiques. La prise en compte d'une période suffisamment longue dans la régression permet, théoriquement, de prendre en compte les variations météorologiques pour corriger l'erreur systématique sur l'ensemble des situations météorologiques possibles. Cependant, la régression traite également d'autres erreurs (erreurs modèle, erreurs d'observations) qui ne sont pas forcément stationnaires. De plus, l'erreur de représentativité peut varier suivant les saisons. L'absence de stationnarité n'est donc pas surprenante. Aux stations urbaines, où l'amélioration est la plus notable, l'erreur de représentativité représente une grande partie de l'erreur corrigée par la réduction d'échelle, ce qui peut expliquer une plus grande robustesse des

résultats. De plus, la réduction d'échelle continue à améliorer la RMSE, même en prévision à plusieurs mois.

Airbase

La figure 7.7 donne les mêmes résultats que précédemment, pour le réseau Airbase. Pour ce réseau, l'amélioration due au traitement statistique de réduction d'échelle est beaucoup plus importante que pour le réseau Pioneer. Ainsi, même en étendant la période de prévision sur laquelle sont utilisés les coefficients de régression, la RMSE est meilleure que la RMSE du modèle de référence dans pratiquement tous les cas. C'est notamment le cas pour les stations péri-urbaines (figure 7.7(b)), où l'amélioration est beaucoup plus notable que pour les stations péri-urbaines du réseau Pioneer, et la stationnarité des coefficients également. Seules les stations rurales (figure 7.7(c)) montrent une RMSE en prévision dépassant au bout d'un certain temps la RMSE de référence. Dans le réseau Airbase, le nombre de stations urbaines et péri-urbaines est beaucoup plus important que le nombre de stations rurales. C'est pourquoi la RMSE globale sur toutes les stations est très proche du comportement de la RMSE aux stations urbaines, à la fois en valeur absolue et en terme de stationnarité.

7.3 Prise en compte de l'ensemble

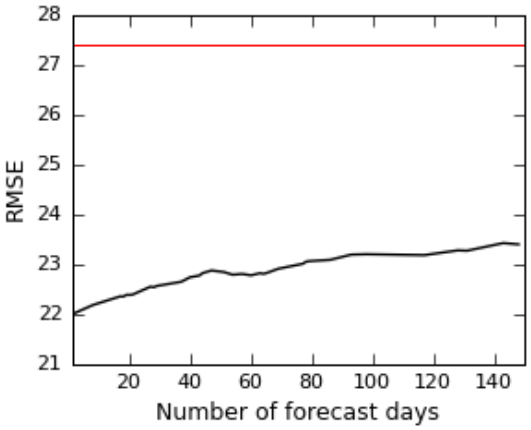
Nous disposons d'un ensemble de 107 modèles. Dans un premier temps, la régression linéaire multiple est effectuée aux stations sur une partie ou la totalité de l'ensemble. L'ensemble correspond à un très grand nombre de régresseurs, avec de fortes colinéarités (tous sont des sorties de modèles représentant les séries temporelles de concentrations aux stations). Il s'agit de savoir si l'utilisation d'un ensemble apporte une information supplémentaire par rapport à la régression sur un seul modèle. On cherche également à mettre en évidence la colinéarité, et à caractériser dans quelle mesure elle se répercute sur les résultats en prévision, avant d'utiliser des techniques de réduction du nombre de régresseurs. Dans cette partie, le réseau Pioneer est utilisé.

7.3.1 Performance en fonction du nombre de modèles

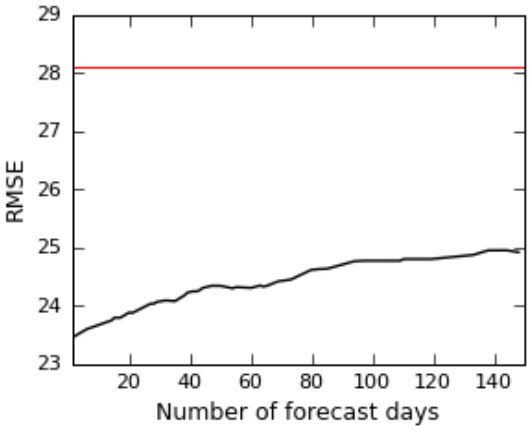
On effectue la régression à plusieurs stations, pour différents nombres de membres (sous-ensembles des 107 modèles à disposition). Tout d'abord, l'ensemble des modèles est trié de deux façons : (1) aléatoirement, (2) en incluant les modèles du meilleur (en terme de RMSE sur toute la période et toutes les stations) au moins bon. Ensuite, la régression linéaire multiple est effectuée sur un nombre croissant de membres, en prenant les N_s premiers modèles de l'ensemble trié ($N_s = 7$ à 107). Il est nécessaire, pour effectuer la régression multiple, de disposer d'au moins autant d'observations que de régresseurs (ici, jusqu'à 107). Comme on cherche à avoir la RMSE pour 150 jours, en prévision à 1 jour, on ne retient que les stations ayant au moins 258 observations (contre 200 dans la partie précédente), ce qui revient à 93 stations pour le réseau Pioneer. C'est pourquoi les valeurs de RMSE avec le modèle de référence ne sont pas exactement les mêmes que dans la partie précédente.

Performance pour différents ensembles

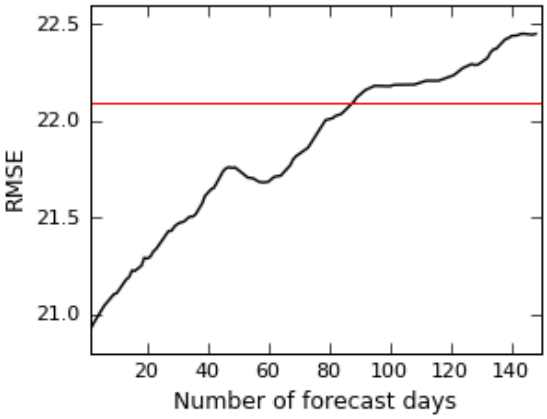
On calcule la RMSE en prévision à un jour sur les 5 derniers mois de l'année, sur toutes les stations (figure 7.8), pour différents ensembles. La performance des ensembles est comparée avec le modèle de référence (ligne rouge), ainsi qu'à la performance en utilisant ce modèle seul pour la régression (ligne bleue en pointillés). Cela permet de quantifier l'apport de l'ensemble, en fonction du nombre de membres. Deux façons de trier l'ensemble des modèles sont présentées : en



(a) Airbase, toutes stations



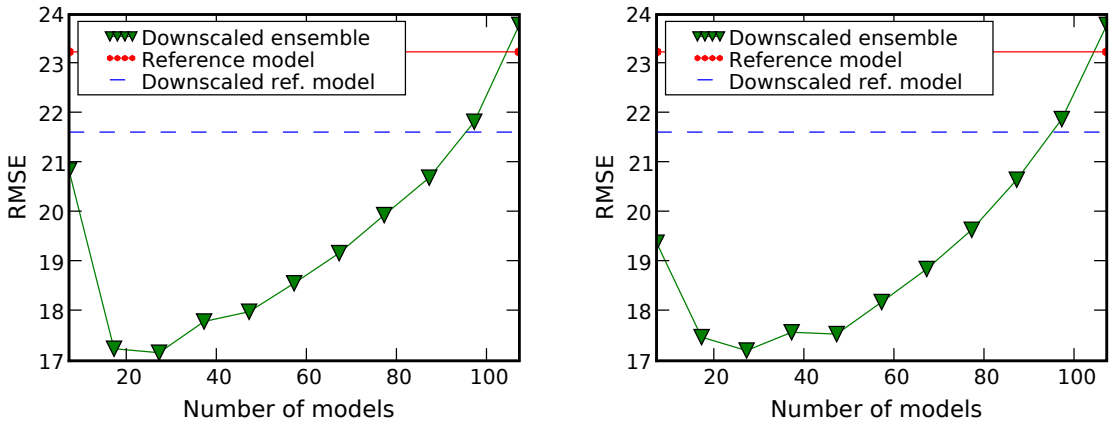
(b) Airbase, stations péri-urbaines



(c) Airbase, stations rurales

FIG. 7.7 – RMSE en fonction de la fréquence à laquelle sont recalculés les poids de la régression (entre 1 jour et 5 mois). RMSE calculée sur les 5 derniers mois de l'année. Réseau Airbase. Résultats par type de station. La ligne rouge représente la RMSE du modèle de référence sans régression.

triant par RMSE croissante (RMSE calculée sur toute l'année), et en prenant un ordre aléatoire. On note que la façon de trier les modèles n'influe que peu sur les résultats de la régression. Dans la suite, on prendra donc un seul ensemble (trié aléatoirement). Ces résultats montrent l'apport de l'ensemble par rapport à un seul modèle : la régression sur le modèle de référence permet de réduire la RMSE de 7% (de 23.2 à 21.6), tandis que l'utilisation d'une vingtaine de modèles permet de la réduire à nouveau de 20% environ. Les modèles apportent donc de l'information supplémentaire. Lorsque le nombre de modèles augmente encore, le résultat est moins bon (la colinéarité devient très importante, et il y a des risques de sur-ajustement). Il faut donc limiter le nombre de régresseurs utilisés.



(a) Ensemble 1 (ordre aléatoire)

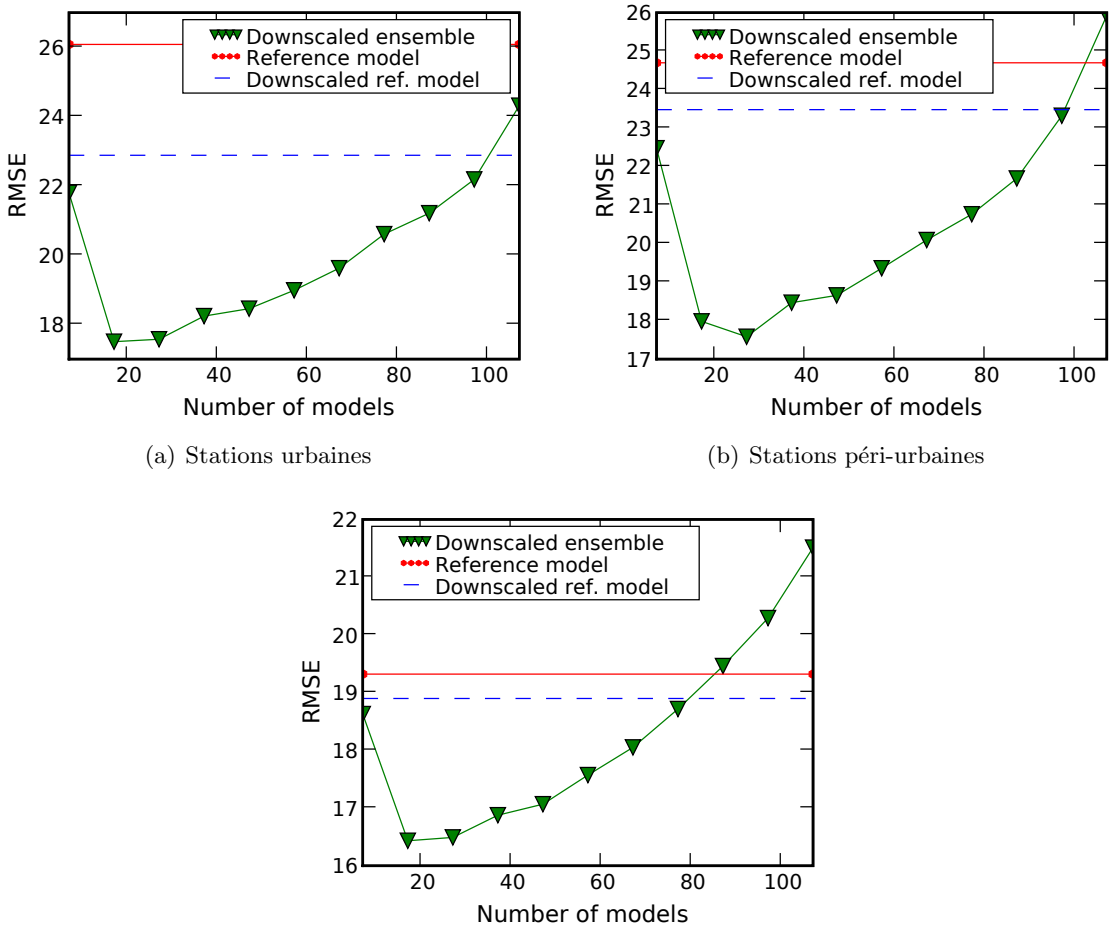
(b) Ensemble 2 (trié par RMSE croissante)

FIG. 7.8 – RMSE globale (prévision à 1 jour) sur 5 mois, en fonction du nombre de modèles, sur 93 stations (réseau Pioneer).

Performance aux stations

La figure 7.9 montre les résultats pour différents nombres de modèles et différents types de stations. Le comportement de la RMSE en fonction du nombre de modèles est le même pour tous les types de stations : diminution jusqu'à une vingtaine de modèles, puis augmentation. L'apport de l'ensemble est bien meilleur pour les stations urbaines et péri-urbaines, que pour les stations rurales. Ainsi, autour de 20 modèles, la RMSE est améliorée d'environ 20% pour les premières, contre 13% pour les dernières. Dans tous les cas, l'utilisation d'un ensemble de modèles permet d'améliorer les résultats du traitement statistique sur un seul modèle.

La figure 7.10 montre un exemple de performances en fonction du nombre de modèles aux quatre stations décrites en partie 7.2.2. La station ST-NAZAIRE-BLUM (figure 7.10(a)) a été choisie car elle est représentative d'une performance moyenne des modèles (sans réduction d'échelle). L'évolution de la performance de la réduction d'échelle en fonction du nombre de modèles inclus pour cette station est effectivement similaire à l'évolution de la RMSE globale. En revanche, le comportement à d'autres stations est différent : ainsi, LYON-GERLAND (figure 7.10(b)) est une station urbaine où les performances des modèles sont plutôt mauvaises. De ce fait, l'utilisation de la réduction d'échelle permet une amélioration très importante de la RMSE : la RMSE est diminuée de 28% par la régression sur le modèle de référence, et encore dans la même proportion en utilisant un ensemble de 20 modèles. De plus, si la RMSE augmente avec le nombre de modèles (au-delà d'une vingtaine de modèles), elle reste toujours bien inférieure à la RMSE du modèle de référence. Enfin, les figures 7.10(c) et 7.10(d) montrent des stations



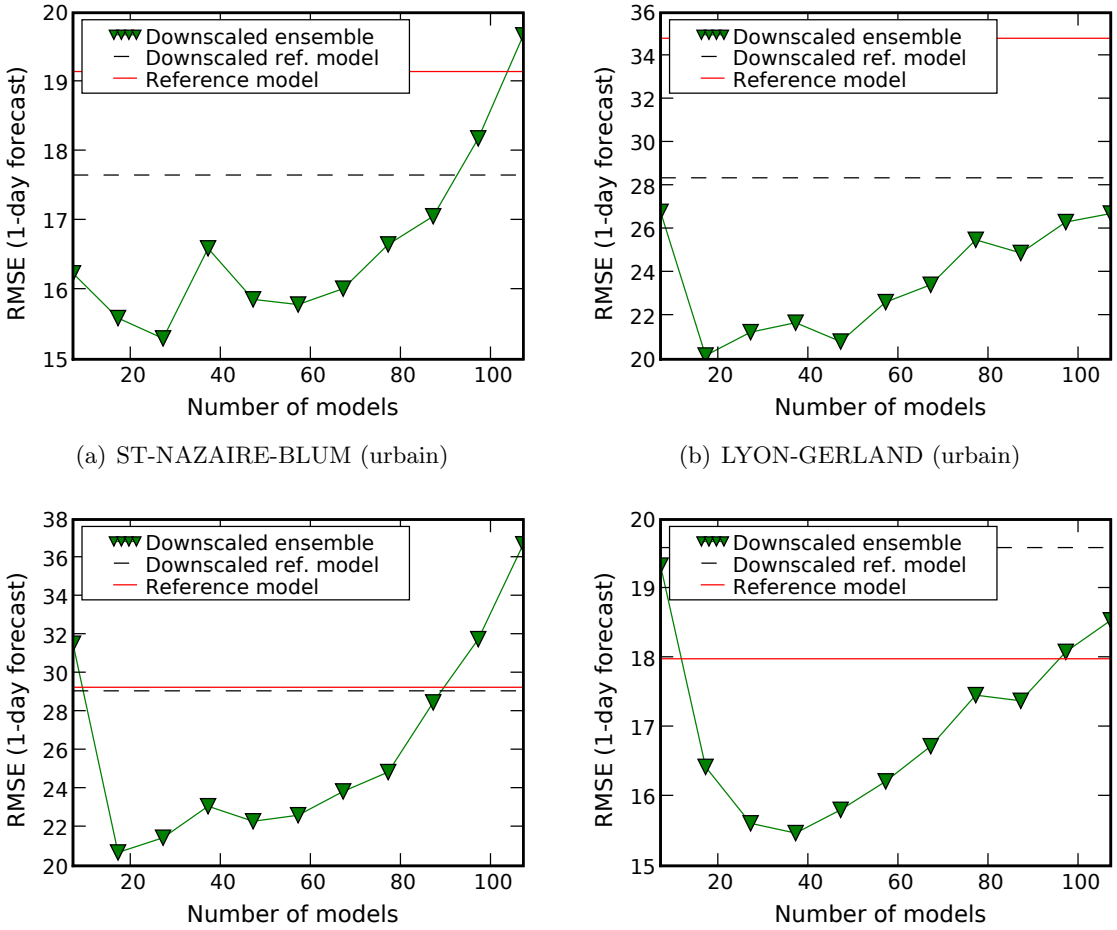
(a) Stations urbaines

(b) Stations péri-urbaines

(c) Stations rurales

FIG. 7.9 – RMSE globale (prévision à 1 jour) sur 5 mois, en fonction du nombre de modèles, pour différents types de stations. Réseau Pioneer, 93 stations.

péri-urbaine et rurale respectivement, où la régression pour un seul modèle n'améliore pas, voire dégrade les performances. Cela montre que l'utilisation d'un ensemble permet d'améliorer les performances dans de très nombreux cas, y compris aux stations où l'erreur de représentativité est moins importante.



(a) ST-NAZAIRE-BLUM (urbain)

(b) LYON-GERLAND (urbain)

(c) COLMAR-SUD (péri-urbain)

(d) FONTAINEBLEAU (rural)

FIG. 7.10 – RMSE (prévision à 1 jour) sur les 5 derniers mois en fonction du nombre de modèles inclus dans l'ensemble, pour 4 stations du réseau Pioneer.

7.3.2 Colinéarité

Les questions qui se posent sont :

- Peut-on constater la colinéarité ? Quels sont les bons indicateurs ?
- La colinéarité se traduit-elle par une inflation visible des coefficients de régression ?
- Cette inflation se traduit-elle dans les résultats en prévision ?

Pour évaluer la régression, on peut considérer le coefficient de détermination. Dans le cas d'une régression sur la variable y, avec l'estimation notée $\widehat{y}$ (équation 6.41), le coefficient de détermination est donné par

$$R^2 = \frac{\sum_i (\widehat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}. \tag{7.2}$$

Cela correspond à la part de variance de y expliquée par la régression. Ici, on utilise un coefficient ajusté, corrigé par le nombre de degrés de liberté de la régression, AR^2, qui permet de comparer les performances pour un nombre différent de régresseurs. En effet, R^2 croît mécaniquement en fonction du nombre de variables inclues dans la régression. On a

$$AR^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}, \tag{7.3}$$

avec n le nombre d'observations et p le nombre de variables (dans notre cas, le nombre de modèles dans le sous-ensemble utilisé). La valeur moyenne aux stations de AR^2 est de 0.73 pour 7 modèles, et monte jusqu'à 0.84 pour 107 modèles. Donc, augmenter le nombre de modèles permet d'expliquer environ 10% supplémentaires de la variance des observations.

Nombre de conditionnement

La partie 7.3.1 a montré que l'utilisation d'un ensemble de modèles est très prometteuse, et permet d'améliorer significativement les performances. Cependant, les variables utilisées pour la régression sont probablement très fortement colinéaires. On cherche à mettre en évidence cette colinéarité, et l'influence sur la régression. Les notations et définitions proviennent de Stewart [1987].

On définit la norme spectrale de la matrice X par

$$\| X \| = \max_{\|\boldsymbol{b}\|=1} \| X\boldsymbol{b} \| . \tag{7.4}$$

On peut montrer que cela correspond à la plus grande valeur singulière de X. De même, on définit

$$\inf(X) = \min_{\|\boldsymbol{b}\|=1} \| X\boldsymbol{b} \|, \tag{7.5}$$

qui correspond donc à la plus petite valeur singulière de X. D'autre part, la SVD de X montre que les valeurs singulières du pseudo-inverse de X, noté $X^\dagger$ sont les inverses des valeurs singulières de X. Cela conduit à

$$\inf(X) = \| X^\dagger \|^{-1} . \tag{7.6}$$

On définit enfin le nombre de conditionnement comme

$$\kappa = \| X \| \| X^\dagger \|, \tag{7.7}$$

ce qui donne la relation

$$\kappa^{-1} = \frac{\inf(X)}{\| X \|}. \tag{7.8}$$

Le nombre κ^{-1} est une mesure de la distance relative à la colinéarité, et représente le rapport entre la plus petite et la plus grande valeur singulière de X : si κ^{-1} est très petit, la distance à la colinéarité est faible. Certaines valeurs de la matrice $\left(C'_{XX}\right)^{-1}$, qui intervient dans le calcul de la variance des coefficients de régression (équation 6.47) seront alors très grandes, ce qui entraîne donc une inflation de la variance.

La figure 7.11 montre le nombre de conditionnement moyen aux stations, en fonction du nombre de modèles utilisés. La valeur de κ augmente rapidement avec le nombre de modèles.

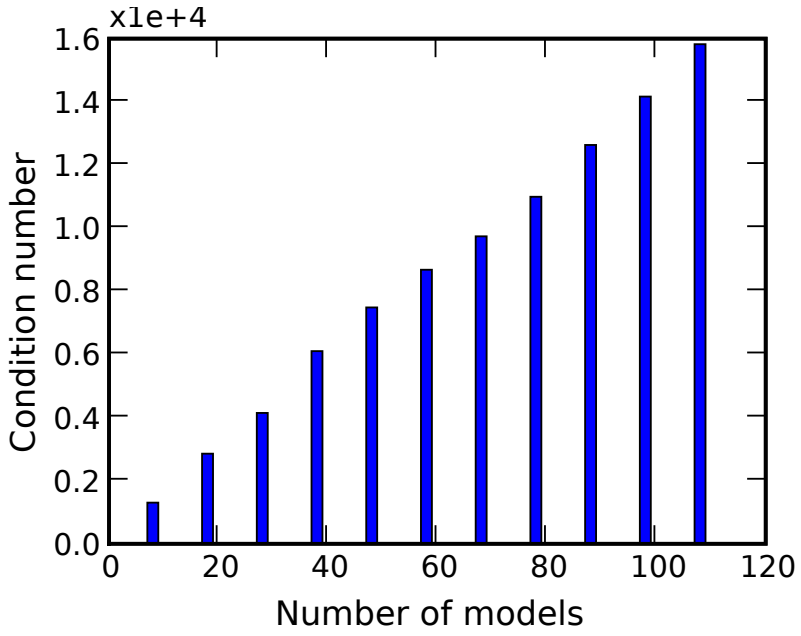


FIG. 7.11 – Nombre de conditionnement κ en fonction du nombre de modèles inclus dans l'ensemble. Moyenne sur toutes les stations.

Coefficient de corrélation multiple et VIF

Pour une mesure plus fine de la colinéarité, on peut calculer le coefficient de corrélation multiple R_j associé au j^{e} modèle. Son carré est le coefficient de détermination associé à la régression du modèle x_j sur tous les autres modèles (équation 7.2). Il s'agit de la part de variance des autres modèles expliquée par x_j, ou de la distance de x_j à une combinaison linéaire des autres modèles. On calcule le coefficient de corrélation multiple à toutes les stations, et pour chaque modèle. Les valeurs de R_j sont très élevées, y compris pour un ensemble de 7 modèles : la valeur moyenne (aux stations et sur les modèles) est de 0.96. La valeur minimale (aux stations et sur les modèles) descend à 0.59. Il y a donc une très forte corrélation entre les sorties des modèles.

On peut également calculer le facteur d'inflation de la variance (VIF) pour chaque modèle j. Il est défini par :

$$VIF_j = \frac{1}{1 - R_j^2} \tag{7.9}$$

Si R_j^2 est grand, VIF_j est grand et il y a risque de colinéarité. Ce facteur peut être relié à la variance des coefficients de régression (équation 6.47), qu'il amplifie. S'il y a une quasi colinéarité, la variance des coefficients de régression est très grande, et les coefficients ne seront pas significatifs. La figure 7.12 montre l'évolution de la valeur moyenne du VIF en fonction du nombre de modèles, sur toutes les stations. En moyenne, la variance des coefficients de régression est 14 à 800 fois plus élevée que si les régresseurs étaient indépendants.

Significativité des coefficients

On a montré que la colinéarité entre les modèles était importante, et que cela entraînait une inflation de la variance des coefficients de régression. Cette inflation est problématique. En effet, si la valeur des coefficients calculés b_j est très faible par rapport à l'écart type $\sigma_{b_j} = \sqrt{\text{var}(b_j)}$,

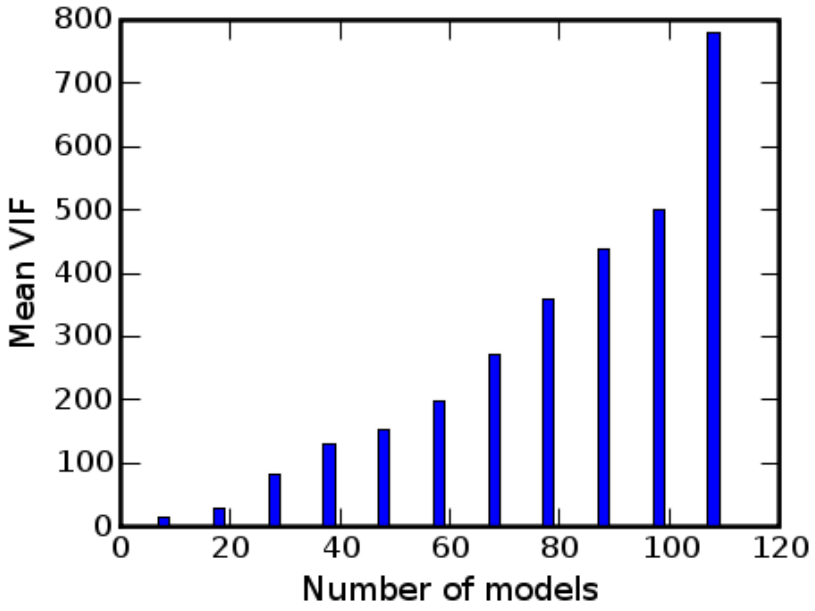


FIG. 7.12 – Facteur d'inflation de la variance en fonction du nombre de modèles inclus dans l'ensemble. Moyenne sur toutes les stations et tous les modèles.

cette valeur n'est pas fiable. Plus exactement, le coefficient b_j est considéré comme significatif au sens statistique si :

$$\left|\frac{b_j}{\sigma_{b_j}}\right| > 2. \tag{7.10}$$

On calcule donc la variance des coefficients de régression en fonction du nombre de modèles (équation 6.47), et l'on en déduit quels coefficients sont significatifs au sens de l'équation 7.10. Le nombre de coefficients significatifs est calculé en fonction du nombre de modèles inclus dans la régression, en moyenne sur les stations. Quel que soit le nombre de modèles, seuls 5 ou 6 sont réellement pris en compte dans la régression. La plupart des autres modèles ont des coefficients proches de 0. Notons que ce ne sont pas les mêmes modèles qui contribuent aux différentes stations. Il s'agit maintenant d'opérer une sélection afin de ne garder que quelques modèles, et/ou de rendre les régresseurs orthogonaux pour éliminer le problème de la colinéarité.

7.4 Réduction de dimension

7.4.1 Utilisation d'ensembles calibrés

Une façon de sélectionner les modèles utilisés dans l'ensemble pour la régression est de calibrer l'ensemble afin que la dispersion de l'ensemble soit représentative de l'erreur commise aux stations. La calibration est effectuée suivant certains critères, décrits dans Garaud et Mallet [2009a]. Si l'on se base sur toutes les stations pour la calibration, la dispersion de l'ensemble est alors représentative de l'erreur globale du modèle : non seulement l'erreur modèle (due aux paramétrisations et aux données d'entrée) mais aussi les erreurs de mesure et l'erreur de représentativité. Pour avoir un ensemble représentatif de l'erreur modèle, on effectue la calibration sur les stations rurales, où l'erreur de représentativité est beaucoup moins grande que pour les autres stations. Les ensembles ainsi calibrés sont donc représentatifs de l'erreur

aux stations rurales.

Pour le réseau Pioneer, la calibration sur les stations rurales permet d'obtenir un ensemble de 23 modèles. Le tableau 7.3 montre les performances de l'ensemble ainsi obtenu, en prévision à 1 jour, comparées aux performances en prévision à 1 jour du modèle de référence et à la RMSE du modèle de référence sans régression. Les résultats sont montrés pour les différents types de stations. L'amélioration due à l'ensemble est très nette pour tous les types de station, et suit encore une fois la tendance donnée par la relation 7.1. Cependant, la calibration ne permet pas une amélioration notable par rapport aux ensembles pris aléatoirement, si l'on compare aux performances avec une vingtaine de modèles montrées par exemple figure 7.9. Ainsi, si l'on prend 23 modèles au hasard, on obtient à peu près les mêmes résultats que dans le tableau 7.3 : 17.5 pour la RMSE globale, 17.4 pour les stations urbaines, 18.0 pour les stations péri-urbaines et 17.3 pour les stations rurales. La calibration permet donc, éventuellement, de séparer l'influence de l'erreur modèle et de l'erreur de représentativité dans l'analyse des résultats, mais non d'améliorer les performances.

Stations	RMSE référence	RMSE prévision (1 modèle)	RMSE prévision (ensemble calibré)
Toutes (154)	22.4	21.1 (-6%)	17.9 (-20%)
Urbaines (35)	25.4	22.6 (-11%)	17.5 (-31%)
Péri-urbaines (47)	24.0	22.9 (-4%)	18.8 (-22%)
Rurales (72)	19.5	19.1 (-2%)	17.4 (-11%)

Tab. 7.3 – Statistiques sur le réseau Pioneer, en fonction du type de station : RMSE pour le modèle sans régression, RMSE pour la prévision à un jour avec le modèle de référence, et RMSE en prévision à 1 jour pour l'ensemble calibré sur les stations rurales (23 modèles). Les RMSE sont en $\mu g\,m^{-3}$. Le nombre de stations de chaque type est indiqué entre parenthèses. Pour la prévision à 1 jour, la différence relative entre la RMSE du modèle et la RMSE en prévision est fournie entre parenthèses.

Au niveau de la colinéarité, les indicateurs (nombre de conditionnement, VIF), sont de l'ordre de ce que l'on obtenait pour une vingtaine de modèles dans la partie précédente. En effet, aucun traitement spécifique n'a été effectué sur les régresseurs pour les rendre orthogonaux. La calibration sélectionne un ensemble qui reproduit la dispersion des observations. Les modèles ainsi sélectionnés constituent un sous-ensemble des 107 modèles à disposition, mais sont toujours fortement colinéaires. Le nombre de conditionnement moyen aux stations est de plus de 4000. De même, le facteur d'inflation de la variance est en moyenne d'environ 50, ce qui est à peu près l'ordre de grandeur obtenu pour une vingtaine de modèles aléatoires. Il peut aller jusqu'à 179 pour certains modèles et certaines stations.

La figure 7.13 montre l'évolution des coefficients de régression pour les 23 membres de l'ensemble calibré, en fonction du temps. Pour chaque modèle, la moyenne est faite sur l'ensemble des stations. La moitié environ des modèles ont des moyennes très différentes de zéro, ce qui ne signifie pas que les autres modèles ne soient pas significatifs à certaines stations.

7.4.2 Réduction en composantes principales (PCA)

La calibration, voire l'utilisation d'un petit nombre de modèles choisis au hasard, permet d'obtenir de très bons résultats en régression et d'améliorer les performances par rapport à l'utilisation d'un seul modèle. Cependant, le problème de la colinéarité des régresseurs est toujours présent, et une façon efficace pour résoudre ce problème est de décomposer les régresseurs

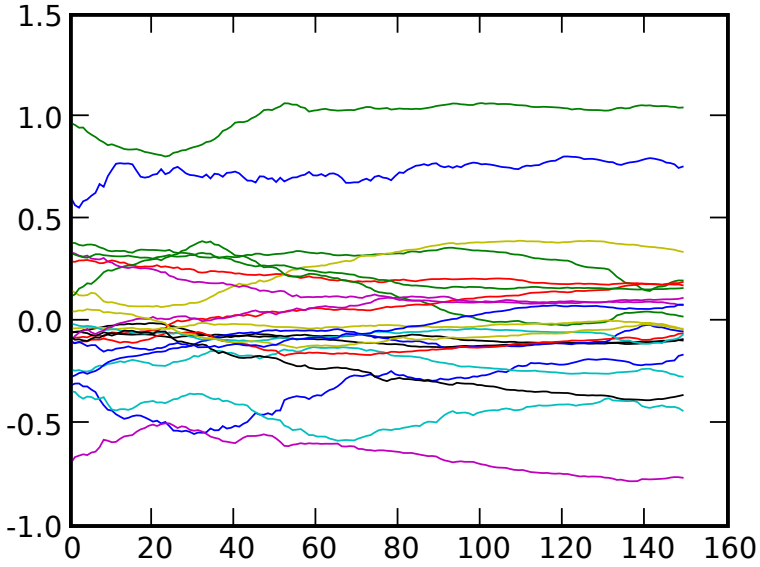


FIG. 7.13 – Coefficients de régression pour les 23 modèles de l'ensemble calibré. Moyenne sur l'ensemble des stations Pioneer utilisées (154 stations).

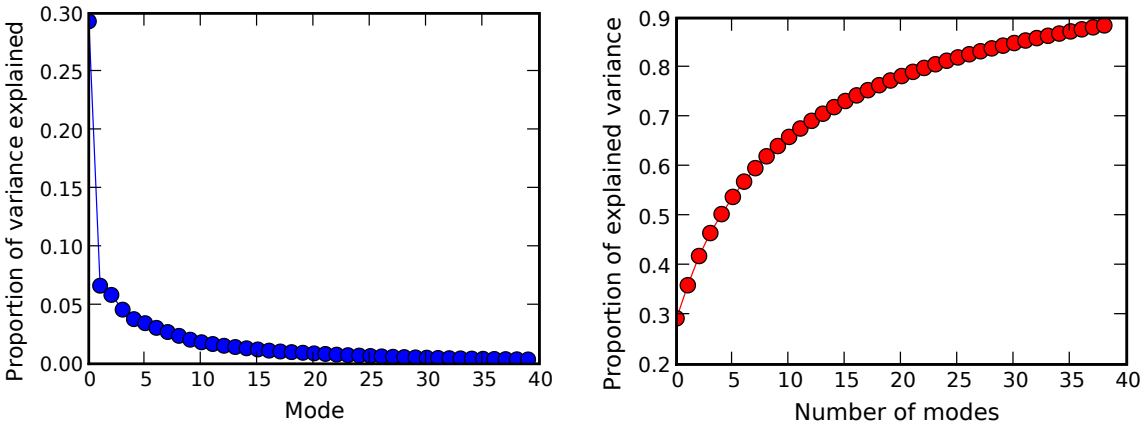
en composantes principales, de la façon décrite en partie 6.4.2. Notons que dans la suite, on suppose les prédicteurs centrés (équation 6.5), préalablement à la décomposition en composante principales.

Variance expliquée

Une réduction en composantes principale est donc effectuée sur les 107 modèles, préalablement à la régression, pour chacune des stations du réseau Pioneer. On utilise à nouveau 93 stations, comme dans la partie 7.3.1, pour avoir toujours davantage d'observations que de régresseurs pour effectuer la décomposition. Pour avoir une idée de la variance expliquée par chacune des composantes, donnée par l'équation 6.15, on fait la moyenne de la variance expliquée par chaque mode, pour toutes les stations. On peut également calculer la variance expliquée par les k premiers modes cumulés (somme de la variance expliquée par les k premiers modes). La figure 7.14(a) montre que le premier mode explique en moyenne 30% de la variance. Il faut donc inclure plusieurs directions afin de reconstituer l'ensemble de l'information. La direction suivante n'explique que 6% de la variance environ. La figure 7.14(b) montre qu'il faut inclure plus de 40 composantes pour expliquer 90% de la variance totale. Ce nombre est relativement élevé, notamment en comparant avec les performances des ensembles de modèles qui sont optimales autour de 20 modèles.

Régression sur les modes expliquant 90% de la variance

Dans un premier temps, nous effectuons la régression en retenant le nombre de modes nécessaire pour expliquer 90% de la variance (entre 40 et 50 suivant les stations). Comme dans les parties précédentes, les résultats sont présentés pour la RMSE a posteriori, et la RMSE en prévision à 1 jour. Le tableau 7.4 montre les résultats obtenus. Si la RMSE a posteriori est très bonne, la RMSE en prévision n'est pas très inférieure à celle obtenue avec un seul modèle. La



(a) Variance expliquée par chacun des modes

(b) Variance expliquée cumulée

FIG. 7.14 – Variance expliquée par les directions principales, pour les 40 premiers modes. Moyenne sur l'ensemble des stations Pioneer utilisées (93 stations).

performance est un peu moins bonne qu'en prenant directement le même nombre de modèles. Il y a probablement davantage de sur-ajustement dans la régression présentée ici, c'est-à-dire d'ajout de composantes qui devraient être négligées car elles représentent du « bruit ». La figure 7.14 montre en effet que de nombreuses composantes expliquent très peu de variance.

RMSE modèle	23.2
RMSE prévision (1 modèle)	21.6
RMSE a posteriori (PCA)	12.7
RMSE prévision (PCA)	20.6

TAB. 7.4 – Statistiques sur le réseau Pioneer (93 stations) : RMSE pour le modèle de référence avec et sans régression, RMSE a posteriori (régression sur toute la période) et RMSE pour la prévision à un jour, pour l'analyse en composantes principales (notée PCA). La régression est faite sur les k premiers modes expliquant 90% de la variance. Les RMSE sont en $\mu g\,m^{-3}$.

Performance en fonction du nombre de modes

On effectue à présent la régression sur un nombre fixé de modes entre 1 et 107, et l'on étudie l'influence du nombre de modes sur les performances, de façon similaire à l'étude réalisée en partie 7.3.1. La figure 7.15 montre la RMSE totale en fonction du nombre de modes inclus dans la régression. A partir de 10 modes inclus dans la régression, le minimum de la RMSE en prévision est à peu près atteint, et au-delà d'une trentaine de modes, il y a clairement un sur-ajustement qui conduit à une augmentation de la RMSE. Au final, l'utilisation de 15 à 20 modes permet d'obtenir une RMSE d'environ 19, ce qui est une amélioration par rapport à la réduction d'échelle sur le modèle de référence seul, mais moins bon que l'utilisation d'un ensemble avec le même nombre de membres.

Cependant, l'avantage de cette méthode est de s'assurer que les régresseurs ne sont pas colinéaires. Ainsi, le facteur d'inflation de la variance reste toujours à 1. De plus, le nombre de conditionnement reste relativement peu élevé : il reste égal à 200 (excepté pour 107 modes où il monte à 1200), contre 4000 environ avec l'ensemble calibré.

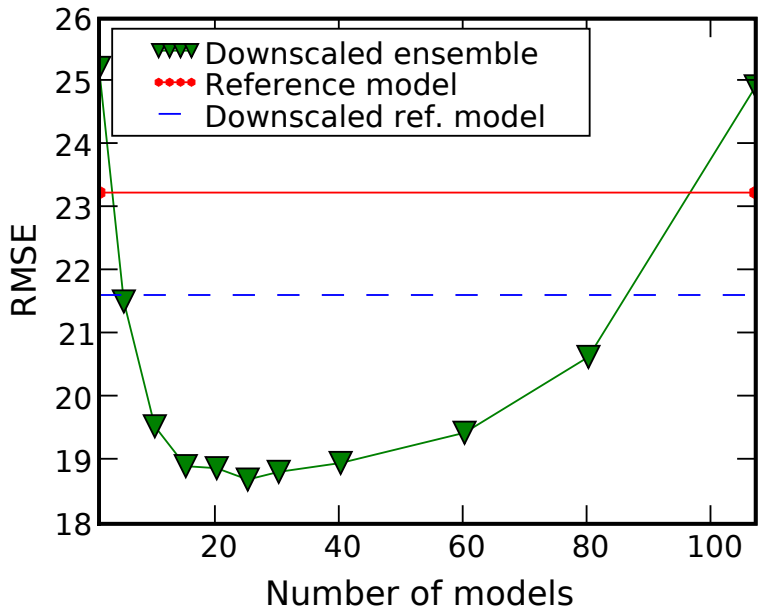


FIG. 7.15 – RMSE en fonction du nombre de modes pour 1, 5, 10, 15, 20, 25, 30, 40, 60, 80, et 107 modes. Moyenne sur l'ensemble des stations Pioneer utilisées (93 stations). Prévision à 1 jour. RMSE calculée sur 5 mois.

7.4.3 Réduction en composantes principales ajustées (PFC)

La décomposition en composantes principales permet de rendre les régresseurs orthogonaux, en choisissant les modes expliquant le plus de variance du prédicteur. Cependant, rien n'indique que ces modes sont les plus pertinents pour expliquer la relation entre les prédicteurs et le prédictand. La décomposition en composantes principales ajustées permet de déterminer des composantes basées à la fois sur le prédicteur et le prédictand, de la façon décrite en partie 6.4.3.

Détermination des fonctions F_y

L'idée est d'effectuer la décomposition en se basant non plus sur la matrice X de prédicteurs, mais sur une matrice $\widehat{X}$ ajustée. Celle-ci est une projection de X sur un espace engendré par les lignes d'une matrice F_y dépendant de y, selon l'équation 6.48. La matrice F_y est déterminée à partir de fonctions dépendant de $y(t)$ (variable mesurée aux stations, continue). Le vecteur $\boldsymbol{y}$ contient les valeurs discrètes de $y(t)$, mesurées à N_t pas de temps, $\boldsymbol{y} = (y_1, \cdots, y_{N_t})^T$. La première étape de la réduction en composantes principales ajustées est de déterminer une forme adéquate pour F_y. La matrice F_y est de taille $N_r \times N_t$, où N_r est le nombre de fonctions de y choisies, et détermine le nombre de composantes qui seront utilisées ($\leq N_r$).

Dans Cook [2007], il est recommandé d'effectuer une investigation préalable sur la forme de relation entre le prédicteur et le prédictand. Dans le cas où cela ne serait pas possible (nombre de régresseurs trop élevé), deux formes simples de F_y sont proposées, qui ne dépendent que de y :

- Prendre $F_y = \boldsymbol{y}^T$, préalablement centrée : $\sum_{k=1}^{n} y_k = 0$. On a alors $N_r = 1$.
- Utiliser N_r fonctions en escalier, construites en prenant N_r « tranches » parmi les valeurs

possibles de $y(t)$. Pour l'intervalle I_r, la fonction associée est

$$f_r(y_k) = \begin{cases} 1 - n_r/N_t & \text{si } y_k \in I_r, \\ -n_r/N_t & \text{sinon,} \end{cases} \tag{7.11}$$

avec n_r le nombre d'observations dans l'intervalle I_r. Chaque fonction f_r est donc une fonction constante par morceaux, centrée sur 0, dont les valeurs sont positives lorsque les observations sont dans l'intervalle I_r et négatives sinon. Il faut ensuite déterminer le nombre d'intervalles N_r, ainsi que la façon dont ils sont définis.

Dans la suite, les résultats sont présentés pour une matrice F_y calculée de trois façons différentes : (1) en prenant directement le vecteur d'observations centré, (2) en prenant des fonctions obtenues en étudiant la relation entre un régresseur (le modèle de référence) et les observations, et (3) en prenant les fonctions constantes par morceaux définies dans l'équation 7.11, pour $N_r = 10$.

Résultats pour $F_y = \boldsymbol{y}^T$

Il s'agit du cas où F_y est déterminée de la façon la plus simple. La matrice ajustée $\widehat{X}$ est donc déterminée d'après une régression de X sur $\boldsymbol{y}^T$. Dans le cas où le prédicteur et le prédictand ont une relation linéaire, cela peut donner de bons résultats. La partie non linéaire de la relation n'est donc pas prise en compte. De plus, on a $N_r = 1$: la régression n'est effectuée que sur une seule variable. Le tableau 7.5 donne les résultats obtenus pour les 93 stations du réseau Pioneer. Le résultat est comparable (légèrement meilleur) à celui obtenu en utilisant directement le modèle de référence. Il est meilleur que le résultat que l'on obtient pour une seule composante principale non ajustée (figure 7.15). Globalement, l'utilisation des PFC calculées de cette façon n'apporte pas une amélioration très notable aux résultats de la réduction d'échelle sur un seul modèle.

RMSE modèle	23.2
RMSE prévision (1 modèle)	21.6
RMSE a posteriori (PFC linéaires)	20.6
RMSE prévision (PFC linéaires)	21.5

Tab. 7.5 – Statistiques sur le réseau Pioneer (93 stations) : RMSE pour le modèle de référence avec et sans régression, RMSE a posteriori (régression sur toute la période) et RMSE pour la prévision à un jour, pour l'analyse en composantes principales ajustées, calculée avec $F_y = \boldsymbol{y}^T$ ($N_r = 1$), notée « PFC linéaires ». Les RMSE sont en $\mu\mathrm{g\,m^{-3}}$.

Résultats pour F_y déterminée avec le modèle de référence

On veut à présent déterminer les fonctions utilisées dans F_y d'après la relation entre le modèle de référence et les observations. La figure 7.16 montre diagramme de dispersion donnant les valeurs simulées en fonction des valeurs observées, sur l'ensemble des stations et des pas de temps. La relation montrée par cette figure est, en première approximation, affine. La composante associée, obtenue après régression, est de la forme

$$\widehat{\boldsymbol{x}} = a\ \boldsymbol{y} + b. \tag{7.12}$$

Pour déterminer une composante non linéaire, on effectue la même étude sur le résidu $\epsilon = \boldsymbol{x} - \widehat{\boldsymbol{x}}$. Celui-ci est tracé en fonction des observations, et approché par une forme polynomiale.

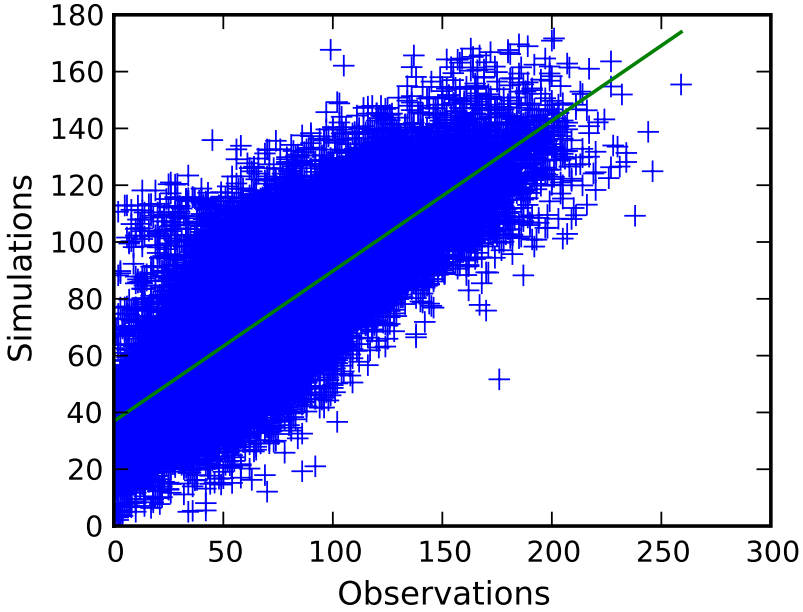


FIG. 7.16 – Diagramme de dispersion : valeurs simulées par le modèle de référence en fonction des observations, sur l'ensemble des stations du réseau Pioneer. Concentrations d'ozone en $\mu g\, m^{-3}$. Droite : résultat d'une régression linéaire.

La figure 7.17 montre ce résidu sous la forme d'un histogramme (valeurs moyennes par tranches) pour plus de clarté, ainsi que le polynôme d'ordre 2 qui permet de l'approcher

$$\widehat{\boldsymbol{\epsilon}} = a\ \boldsymbol{y}^2 + b\ \boldsymbol{y} + c \tag{7.13}$$

Ces résultats mettent en évidence l'intérêt d'inclure une composante quadratique dans la relation entre les modèles et les observations. La matrice F_y est donc construite avec

$$F_y = \begin{pmatrix} (a\ \boldsymbol{y} + b)^T \\ (\alpha\ \boldsymbol{y}^2 + \beta\ \boldsymbol{y} + \gamma)^T \end{pmatrix},$$

les coefficients a, b, α, β et γ déterminés ci-dessus, pour la station donnée et sur la période d'apprentissage. On a bien sûr $N_r = 2$. Le tableau 7.6 montre les résultats obtenus pour les PFC calculées de cette façon. L'ajout de la composante quadratique permet d'améliorer les résultats par rapport à l'utilisation de la composante linéaire seule (tableau 7.5), et par rapport à la réduction d'échelle sur le modèle de référence. Le résultat est également meilleur que celui obtenu par les PCA en utilisant moins de 5 composantes (figure 7.15). Cependant, il reste moins bon que ce que l'on pouvait obtenir avec un ensemble d'une vingtaine de modèles. Le facteur d'inflation de la variance est toujours égal à 1, et le nombre de conditionnement est environ égal à 200.

Résultats pour les fonctions constantes par morceaux

Les fonctions utilisées pour constituer F_y, décrites de façon générique par l'équation 7.11, dépendent en pratique des intervalles I_r choisis. Deux façons de déterminer ces intervalles sont mises en œuvre :

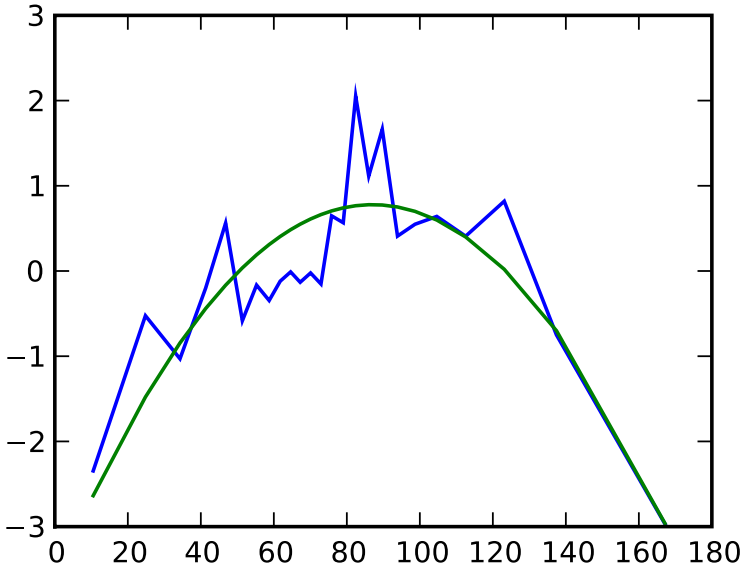


FIG. 7.17 – Résidu de la régression précédente en fonction des observations, sur l'ensemble des stations du réseau Pioneer. Concentrations d'ozone en $\mu g\,m^{-3}$. Courbe : résultat d'un ajustement avec un polynôme d'ordre 2.

RMSE modèle	23.2
RMSE prévision (1 modèle)	21.6
RMSE a posteriori (PFC polynomiales)	19.4
RMSE prévision (PFC polynomiales)	21.1

TAB. 7.6 – Statistiques sur le réseau Pioneer (93 stations) : RMSE pour le modèle de référence avec et sans régression, RMSE a posteriori (régression sur toute la période) et RMSE pour la prévision à un jour, pour l'analyse en composantes principales ajustées, calculée avec F_y déterminée à partir de la relation entre le modèle de référence et les observations. On a $N_r = 2$, avec une composante linéaire et une composante quadratique (on note ce cas « PFC polynomiales »). Les RMSE sont en $\mu g\,m^{-3}$.

1. En prenant, pour le r^{e} intervalle ($r \in [\![0, N_r[\![$),

$$I_r = \left[y_{\min} + r\,\frac{y_{\max} - y_{\min}}{N_r}, y_{\min} + (r+1)\,\frac{y_{\max} - y_{\min}}{N_r} \right], \tag{7.14}$$

   avec

$$y_{\min} = \min_k y_k \qquad \text{et} \qquad y_{\max} = \max_k y_k,$$

   de façon à ce que toutes les observations soient incluses dans les intervalles.

2. En déterminant les bornes des intervalles de façon à ce que tous les intervalles contiennent le même nombre d'observations (calculé sur l'ensemble des valeurs pouvant être prises par les observations sur la période de temps considérée)

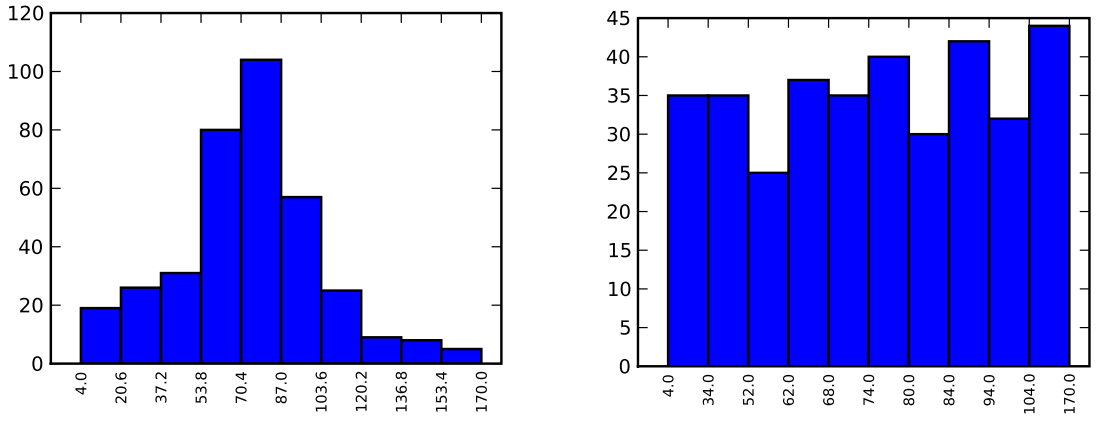
$$n_r = \frac{N_t}{N_r}. \tag{7.15}$$

   Les bornes des intervalles I_r sont définies de cette façon. Ensuite, le nombre d'observations contenu en pratique par ces intervalles peut varier légèrement, notamment du fait des « doublons » (valeurs observées identiques à plusieurs pas de temps).

La première méthode est appelée dans la suite PFC par morceaux à « intervalles constants », et la seconde PFC par morceaux à « nombre d'observations constant ». La figure 7.18 illustre le nombre d'observations contenu dans les intervalles I_r pour les deux cas, pour la station ST-NAZAIRE-BLUM. Dans le cas des intervalles constants (figure 7.18(a)), le nombre d'observations n_r incluses dans les intervalles I_r est très inégal : il y a très peu d'observations proches du minimum et du maximum, tandis que les intervalles proches de la moyenne (sur cette illustration, y n'est pas centré) contiennent beaucoup d'observations. La fonction f_r définie par l'équation 7.11 donnant des valeurs proches de zéro pour les intervalles contenant peu d'observations, cela peut poser des problèmes (matrice F_y mal conditionnée). La figure 7.18(b) montre que la seconde méthode (nombre d'observations constant) évite d'avoir des intervalles contenant trop peu d'observations.

Le tableau 7.7 montre les résultats pour ces deux variantes des PFC, pour $N_r = 10$, ainsi qu'un rappel de la RMSE du modèle de référence et de la réduction d'échelle sur un seul modèle. Avec les PFC par morceaux, la RMSE a posteriori est toujours meilleure que pour les autres résultats en PFC (linéaires et polynomiales). Dans le cas où l'on détermine les fonctions par palier avec les intervalles constants, la RMSE en prévision se dégrade, en restant toutefois légèrement meilleure que celle donnée par les PFC polynomiales (tableau 7.6). Cela provient très certainement du fait que des intervalles contenant très peu d'observations sont inclus. La forme choisie pour les fonctions f_y n'est donc pas bien représentative de la forme « réelle » de $y(t)$. Des composantes n'expliquant que peu de variance de $\widehat{X}$ sont alors incluses dans la régression, ce qui dégrade les résultats en prévision. En revanche, lorsque les fonctions par morceaux sont déterminées en prenant un nombre d'observations constant dans les intervalles, les résultats sont très bons : avec $N_r = 10$ (ce qui correspond en général à 8 ou 9 composantes principales effectivement utilisées dans la régression), la RMSE est aussi bonne que celle obtenue pour un ensemble d'une vingtaine de modèles. Le tableau 7.3 donne par exemple les résultats pour un ensemble calibré contenant 23 modèles, et 154 stations. Si l'on effectue la réduction d'échelle avec cet ensemble sur les 93 stations utilisées ici, la RMSE en prévision à un jour est alors de 17.5, comme pour les PFC par morceaux. Comme précédemment, en utilisant les PFC, le facteur d'inflation de la variance reste proche de 1, et le nombre de conditionnement autour de 200.

Les résultats sont montrés ici pour $N_r = 10$. Au-delà, le nombre d'intervalles est trop grand par rapport au nombre total d'observations à certaines stations, et certains intervalles



(a) Intervalle constant (b) Nombre d'observations constant

FIG. 7.18 – Nombre d'observations contenues dans les intervalles définis de deux façons : (a) en prenant des intervalles constants entre le minimum et le maximum des observations (équation 7.14) et (b) en définissant les intervalles de façon à avoir le même nombre de valeurs prises par les observations (le nombre n_r montré ici tient compte des doublons). Station ST-NAZAIRE-BLUM. En abscisse : valeurs de y définissant les intervalles ($\mu g\, m^{-3}$).

RMSE modèle	23.2
RMSE prévision (1 modèle)	21.6
RMSE a posteriori (PFC par morceaux, intervalle constant)	16.0
RMSE prévision (PFC par morceaux, intervalle constant)	20.5
RMSE a posteriori (PFC par morceaux, nombre d'observations constant)	15.7
RMSE prévision (PFC par morceaux, nombre d'observations constant)	17.5

TAB. 7.7 – Statistiques sur le réseau Pioneer (93 stations) : RMSE pour le modèle de référence avec et sans régression, RMSE a posteriori (régression sur toute la période) et RMSE pour la prévision à un jour, pour l'analyse en composantes principales ajustées, calculée avec F_y contenant des fonctions en escalier et $N_r = 10$ (notée « PFC par morceaux »). Les résultats pour les PFC par morceaux sont donnés pour les deux façons de déterminer les intervalles I_r (1) à intervalles constants, et (2) à nombre d'observations constant. Les RMSE sont en $\mu g\, m^{-3}$.

contiennent alors un trop faible nombre d'observations. Cela peut entraîner des erreurs lors de la projection (équation 6.48), car les valeurs de $f_r(y)$ sont proches de zéro pour les intervalles où n_r est faible, et la matrice F_y est alors mal conditionnée.

7.4.4 Comparaison des différentes méthodes

La figure 7.19 récapitule la plupart des résultats de cette partie : il s'agit de comparer les performances en prévision à un jour, sur le réseau Pioneer, en prenant les 93 stations ayant plus de 258 observations. La comparaison est effectuée sur la RMSE totale aux stations. Les résultats sont comparés pour

- L'ensemble calibré sur les stations rurales du réseau Pioneer, contenant 23 membres,
- La décomposition en composantes principales en prenant les modes expliquant 90% de la variance (40 environ), notée PCA1 (tableau 7.4),
- La décomposition en composantes principales pour les 20 premiers modes (PCA2),
- La décomposition en composantes principales ajustées avec les fonctions par morceaux à intervalles constants, pour $N_r = 10$ (PFC1),
- La décomposition en composantes principales ajustées avec les fonctions par morceaux à nombre d'observations constant, pour $N_r = 10$ (PFC2).

Les résultats sont également montrés pour le modèle de référence seul, ainsi que pour la prévision à un jour en réduction d'échelle avec le modèle de référence. Les performances obtenues pour l'ensemble calibré ainsi que pour PCA2 et PFC2 sont les meilleures : la RMSE décroît de -24.5% pour l'ensemble calibré et PFC2, et de -18% pour PCA2 (par rapport à la RMSE du modèle de référence sans réduction d'échelle). A titre de comparaison, l'amélioration obtenue en réduction d'échelle pour le modèle de référence est de -7%, et se situe autour de -11% pour PCA1 et PFC1. Les résultats pour les PFC linéaires et polynomiales décrites dans la partie précédente ne sont pas montrées ici, mais les performances sont similaires à celles de PCA1 et PFC1. L'ensemble calibré compte 23 membres, et PCA2 correspond à 20 modes, tandis que PFC2 utilise moins de dix composantes pour la régression. Les composantes principales ajustées donnent donc clairement les meilleurs résultats pour un plus faible nombre de prédicteurs. De plus, les PFC présentent l'avantage d'éliminer la colinéarité entre les régresseurs, contrairement à l'ensemble utilisé directement.

7.5 Conclusion

7.5.1 Résultats aux stations

Dans ce chapitre, des techniques statistiques de réduction d'échelle ont été appliquées à la modélisation de la qualité de l'air à l'échelle européenne. L'utilisation de ces méthodes pour la prévision à un jour des pics d'ozone a permis de réduire la RMSE aux stations, de façon significative : -6% à -20% sur l'ensemble des stations (pour les réseaux Pioneer et Airbase respectivement), en utilisant un seul modèle. Dans le cas où un ensemble d'une vingtaine de modèles est utilisé, de façon directe ou en utilisant un pré-traitement pour rendre les prédicteurs orthogonaux, ce résultat peut encore être amélioré, avec une diminution de près de -25% (Pioneer). L'utilisation des méthodes statistiques est donc très intéressante, en ce qui concerne les performances des modèles en prévision aux stations.

7.5.2 Représentativité des stations

L'apport de la réduction d'échelle a été comparé pour différents types de stations : rurales, péri-urbaines et urbaines. Cette classification donne, de façon grossière, l'échelle de représentativité

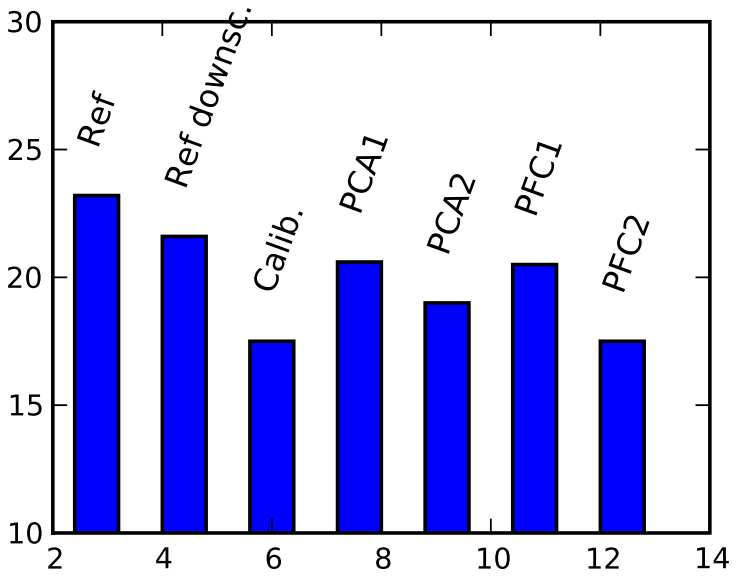


Fig. 7.19 – Comparaison des performances en prévision (à un jour) aux stations du réseau Pioneer (93 stations), pour : (1) le modèle de référence (« Ref »), (2) la réduction d'échelle appliquée au modèle de référence (« Ref. downsc. »), (3) l'ensemble de modèles calibré sur les stations rurales (23 membres) (« Calib. »), (4) la décomposition en composantes principales, avec les modes expliquant 90% de la variance (environ 40 modes) (« PCA1 »), (5) la décomposition en composantes principales avec les 20 premiers modes (« PCA2 »), (6) la décomposition en composantes principales ajustées avec les fonctions par morceaux à intervalles constants et $N_r = 10$ (« PFC1 ») et (7) la décomposition en composantes principales ajustées avec les fonctions par morceaux à nombre d'observations constant et $N_r = 10$ (« PFC2 »).

des stations concernées : pour le réseau Pioneer, le rayon de représentativité est estimé entre 100 m et 1 km pour les stations urbaines, 1 km et 5 km pour les stations péri-urbaines, et au-delà pour les stations rurales. Les simulations numériques ont été effectuées avec un maillage d'environ 50 km de résolution. L'erreur de représentativité est donc importante, puisque l'échelle de représentativité des stations est très inférieure à celle des mailles. Cela est particulièrement vrai pour les stations urbaines, pour lesquelles la réduction d'échelle apporte l'amélioration la plus notable, pour quasiment toutes les stations. Cette amélioration est moins grande pour les stations péri-urbaines, et moindre encore pour les stations rurales, excepté celles situées sur les côtes et dans les régions à orographie complexe. Ces dernières ont un rayon représentatif plus faible, et de nombreux processus locaux contribuent à leur erreur de représentativité (météorologie, topographie).

Par ailleurs, les performances décroissent lorsque l'on augmente l'intervalle de temps entre deux régressions, car l'erreur corrigée par la réduction d'échelle n'est pas entièrement stationnaire. L'amélioration due à la réduction d'échelle reste toutefois visible en utilisant les mêmes coefficients pendant plusieurs dizaines de jours, en particulier aux stations urbaines. Il est possible que la composante stationnaire de l'erreur de représentativité soit plus grande dans ce dernier cas, si les erreurs compensées par la réduction d'échelle ont des sources systématiques.

Apport de l'ensemble et réduction de dimension

L'utilisation d'un ensemble de modèles permet une amélioration des performances, par rapport à la réduction d'échelle sur un seul modèle. La prise en compte des concentrations fournies par plusieurs modèles permet donc d'expliquer une plus grande part de l'erreur de représentativité. Les meilleurs résultats sont obtenus pour 20 régresseurs environ. L'utilisation, de façon directe, des sorties de modèles comme régresseurs implique une forte colinéarité, et donc une inflation de la variance des coefficients de régression : le facteur d'inflation de la variance est de plus de 100 lorsque l'on utilise plus de 40 modèles, et d'environ 50 pour 20 modèles. Cela signifie que les coefficients de régression ont une variance 50 fois plus grande que si les régresseurs étaient orthogonaux. En pratique, cela pose le problème de la fiabilité de la régression et de la significativité des coefficients calculés, et peut entraîner une dégradation notable des performances en prévision.

L'utilisation de méthodes de pré-traitement pour rendre les régresseurs orthogonaux permet de remédier à ce problème. La décomposition en composantes principales (PCA) donne de bons résultats, mais ne permet pas d'atteindre les performances obtenues en utilisant directement une vingtaine de sorties de modèles comme régresseurs. La décomposition en composantes principales ajustées (PFC) permet d'améliorer cette méthode : les directions principales ne sont plus seulement choisies en fonction du prédicteur (sorties de modèles), mais également du prédictand (observations). De la sorte, les directions choisies sont plus à même d'expliquer la relation entre les deux variables. Différentes façon d'ajuster les composantes principales ont été étudiées. L'utilisation d'une fonction continue par morceaux, dépendant des observations, permet d'obtenir les meilleurs résultats pour un nombre de régresseurs relativement faible (≤ 10). Les PFC s'avèrent donc la meilleure méthode, en terme de performances aux stations et d'inflation de la variance des coefficients (aucune inflation), pour un nombre de régresseurs faible en comparaison des autres méthodes testées.

Conclusion

Cette thèse propose une revue des processus physiques mal représentés par les modèles couramment utilisés pour la simulation de la qualité de l'air. Dans le but de prendre en compte les changements d'échelle, elle explore plus spécifiquement deux approches :

1. Introduire une meilleure modélisation locale des émissions, en couplant un modèle gaussien pour l'échelle locale à un modèle eulérien. Cela permet de quantifier la variabilité due à la mauvaise représentation de l'échelle locale autour de la source,
2. Utiliser un traitement statistique de réduction d'échelle. Il s'agit de déterminer des relations entre les concentrations simulées et les observations aux stations dont l'échelle de représentativité est plus faible que l'échelle bien représentée par le modèle.

Ces deux approches sont complémentaires. La première permet d'estimer les sources d'incertitude dans la représentation des phénomènes sous-maille et d'améliorer la modélisation localement autour des sources. La seconde améliore les performances des prévisions aux stations d'observations, mais ne permet pas une estimation spatiale de la variabilité.

Modèles à l'échelle locale

Bilan Les modèles gaussiens présentés en partie I permettent de représenter un panache émis par une source ponctuelle ou linéique, à l'échelle locale (jusqu'à quelques dizaines de kilomètres autour de la source). Un modèle gaussien de panache stationnaire ainsi qu'un modèle à bouffées ont été développés au sein de la plate-forme Polyphemus, et évalués par comparaison à des observations. Les performances obtenues sont très satisfaisantes, notamment en comparaison avec d'autres modèles opérationnels couramment utilisés. Deux principales sources d'incertitude ont été identifiées dans ces modèles, à savoir la modélisation de la dispersion (écarts types gaussiens) et de la surhauteur. Différentes paramétrisations empiriques ont été utilisées pour représenter ces deux processus. Toutes choses étant par ailleurs égales dans les configurations utilisées, la variabilité due à ces paramétrisations en sortie de modèle est comparable à celle obtenue en utilisant un ensemble de modèles de la communauté. Un prolongement de ce travail serait donc l'utilisation de modélisation d'ensemble, effectuée grâce aux différentes paramétrisations disponibles dans les modèles, afin de mieux quantifier les incertitudes.

Météorologie et turbulence Parmi les perspectives d'évolution de ces modèles, différentes améliorations sont envisageables, et notamment une meilleure prise en compte de la variabilité locale des champs météorologiques. Cela peut inclure la reconstruction de profils verticaux à partir des observations météorologiques locale, et la scission des bouffées lorsque le vent ne peut plus être considéré comme homogène au sein d'une bouffée (cisaillement trop important). Par ailleurs, l'utilisation de deux panaches superposés pour représenter les courants ascendants et descendants dans les situations convectives peut permettre une meilleure représentation de la turbulence. Enfin, l'introduction de schémas de fermeture du second ordre pour les écarts types

est également possible, afin de mieux prendre en compte la turbulence et d'estimer la variance des concentrations qui en résulte.

Modélisation du trafic L'utilisation des modèles développés ici pour l'application aux routes est également un enjeu important, dans la mesure où le trafic contribue pour une grande part aux émissions de polluants ciblés en photochimie. L'extension des modèles à l'échelle locale, puis du couplage développé en partie II, à la modélisation des routes, permettrait donc de prendre en compte une grande part de la variabilité sous-maille due aux émissions. Dans ce cadre, des développements supplémentaires peuvent être envisagés, tels que la prise en compte de la turbulence due au trafic ainsi que celle des obstacles tels que les murs anti-bruit.

Panache sous-maille réactif et passif à différentes échelles

La partie II a présenté une modélisation sous-maille des émissions ponctuelles basée sur le couplage d'un modèle eulérien (Polair3D) et d'un modèle local (modèle à bouffées décrit dans la partie I). Ce modèle de panache sous-maille a été développé au cours de la thèse et évalué à différentes échelles.

Modélisation locale et incertitude L'utilisation d'un modèle local autour des sources permet une meilleure représentation de plusieurs processus « sous-maille » : la diffusion due à la turbulence est mieux représentée, notamment sur la verticale, et les émissions elles-mêmes sont modélisées avec plus de précision (prise en compte de la hauteur exacte et de la surhauteur). De plus, la dépendance à la résolution du maillage est réduite, qu'il s'agisse d'une dilution irréaliste du panache (grandes mailles), ou d'une surestimation des concentrations émises dans un petit volume (cas d'ETEX par exemple). Enfin, les réactions chimiques dans un panache réactif sont également modifiées en champ proche. Une bonne prise en compte de ces différents phénomènes dépend toutefois de la qualité des données d'entrée, et des incertitudes associées. Ainsi, une interpolation de champs météorologiques donnés à résolution de 1.5 $^\circ$ pour simuler la trajectoire des bouffées conduit à une grande incertitude sur la position du panache (ETEX). De plus, la méconnaissance des paramètres de la source et notamment de la surhauteur peuvent également limiter la performance du modèle local, bien qu'il reste meilleur que le modèle eulérien (Tchernobyl, région parisienne). Le couplage d'un modèle local pour la dispersion et de données d'entrées à haute résolution, ou d'un modèle météorologique de type diagnostique ou pronostique, est une solution envisageable pour améliorer encore le traitement sous-maille des émissions.

Échelle locale L'une des problématiques abordées consistait à définir l'échelle « locale », c'est-à-dire l'échelle autour de la source où l'erreur commise par le modèle eulérien seul est importante. Cette échelle peut être représentée par le « temps d'injection », qui détermine le temps pendant lequel le modèle à bouffées est utilisé pour décrire un panache avant de le transférer au modèle eulérien. En théorie, ce temps correspond au moment où l'erreur commise par le modèle eulérien devient faible, et où l'erreur du modèle local n'est pas trop importante. En pratique, cela dépend donc de la résolution du maillage eulérien, mais aussi de l'exactitude des champs météorologiques utilisés pour le modèle local. Cela dépend enfin des espèces transportées par le panache, car l'impact du panache sous-maille est transporté beaucoup plus loin dans le cas des espèces passives. Pour les espèces réactives, lorsque la composition du panache devient proche de celle du fond, l'impact devient négligeable et augmenter le temps d'injection apporte peu d'amélioration. Une solution est donc d'utiliser un critère d'injection basé sur la taille des bouffées pour un panache passif, et sur la composition chimique pour un panache réactif, voire de combiner les deux.

Échelle de l'impact et représentativité Une autre question concerne l'échelle de l'impact du traitement sous-maille. Celui-ci peut être transporté beaucoup plus loin que l'échelle où le modèle local est utilisé, notamment pour un panache non réactif. Dans le cas de l'expérience ETEX, l'impact du panache sous-maille est notable jusqu'à des capteurs situés à une centaine de kilomètres de la source. Dans le cas d'un panache réactif, l'échelle de l'impact est beaucoup plus petite. L'utilisation d'un modèle de panache sous-maille permet une meilleure représentation des sources de façon locale. Son impact sur les statistiques globales des modèles est variable, car il dépend de l'emplacement des points d'observations utilisés. Ainsi, dans le cas de Chernobyl, les stations de mesure les plus proches étant atteintes au bout de plusieurs jours, l'impact évalué aux stations est faible (notamment en regard de l'incertitude sur la source). Dans le cas de la photochimie à l'échelle régionale, l'impact en terme de performances aux stations (qui est l'un des objectifs recherchés, mais non le seul) est visible, bien qu'assez faible, lorsque l'on considère l'ensemble des stations sur une longue durée (six mois). L'utilisation du modèle de panache sous-maille est plus pertinente pour l'étude de certains épisodes particuliers. La pertinence d'utiliser un modèle local dépend de l'objectif recherché, qu'il s'agisse d'étuder l'impact spatial des concentrations ou les résultats aux stations. Dans ce dernier cas, il est préférable d'effectuer l'évaluation sur des observations en champ proche plutôt qu'en utilisant des mesures « de fond » (représentatives d'une large zone).

Réduction d'échelle

Réduction d'échelle pour la qualité de l'air Des techniques statistiques de réduction d'échelle ont été présentées et appliquées à la modélisation de la qualité de l'air à l'échelle européenne en partie III. L'utilisation de ces méthodes pour la prévision à un jour des pics d'ozone a permis de réduire la RMSE aux stations de façon significative (de l'ordre de -20% lorsqu'un seul modèle est utilisé, davantage pour un ensemble). L'amélioration est particulièrement notable dans les performances aux stations péri-urbaines et urbaines, où l'échelle de représentativité des stations est plus faible. Les coefficients de régression ne sont pas complètement stationnaires, mais conduisent tout de même à de bonnes performances pour la prévision à un, voire plusieurs jours à une station donnée. Ces performances décroissent lorsque l'on augmente l'intervalle de temps entre deux régressions (utilisation des mêmes coefficients sur une plus longue période de prévision). Cette décroissance est plus notable aux stations rurales.

Apport d'un ensemble de modèles L'utilisation d'un ensemble de modèles permet d'améliorer de façon notable les performances obtenues en réduction d'échelle. Le nombre optimal de régresseurs, pour minimiser la RMSE aux stations, se situe aux alentours de 20 membres. Lorsque les membres de l'ensemble sont utilisés directement pour la régression, il existe un problème de colinéarité qui entraîne une inflation de la variance des coefficients de régression. Cela peut entraîner une dégradation des performances, car il y a alors davantage d'incertitudes sur les coefficients obtenus. L'utilisation de méthodes de pré-traitement pour rendre les régresseurs orthogonaux permet de remédier à ce problème. La décomposition en composantes principales (PCA) en est une, qui donne des résultats satisfaisants mais ne permet pas de retrouver d'aussi bonnes performances qu'en utilisant directement un ensemble de 20 membres. Avec la décomposition en composantes principales ajustées (PFC), les directions principales ne dépendent plus seulement des prédicteurs (sorties de modèles), mais aussi du prédictand (observations aux stations). Cette méthode permet de retrouver des performances comparables aux meilleurs résultats obtenus avec l'ensemble, en utilisant un plus petit nombre de régresseurs (10 composantes) et en éliminant l'inflation de la variance des coefficients. Le principal problème

de ce type de méthode est de bien déterminer les fonctions qui permettent d'« ajuster » les directions principales en fonction des observations.

Prise en compte de l'erreur modèle Dans les méthodes de régression utilisées ici, l'erreur modèle n'est pas prise en compte de façon explicite (les régresseurs sont supposés parfaits). Un prolongement de ce travail serait donc d'introduire la part d'erreur modèle (estimée via l'utilisation d'un ensemble calibré) dans les équations de la régression, afin d'estimer son impact sur les résultats de la régression.

Échelle de représentativité Dans cette étude, l'échelle de représentativité des stations a été déterminée de façon discrète par une classification en types de stations (rurale, péri-urbaine, urbaine). Une estimation plus fine pourrait être obtenue par le biais de covariogrammes entre les stations, par exemple. Cela pourrait alors permettre d'estimer l'apport de la réduction d'échelle en fonction de l'échelle de représentativité. Une amélioration possible serait alors d'utiliser comme régresseurs les concentrations dans différentes mailles voisines des stations (le nombre de mailles pertinent étant déterminé par l'échelle de représentativité).

Une méthode alternative, connaissant l'échelle de représentativité des stations, serait d'utiliser un modèle de boîte local autour d'une station à cette échelle. Il s'agirait alors de résoudre l'équation d'advection-diffusion à l'intérieur de cette boîte, couplée avec le modèle eulérien.

Enfin, l'extension des méthodes de réduction d'échelle à l'estimation spatiale des concentrations peut être envisagée via des méthodes de krigeage.

Perspectives sur les changements d'échelle

Prolongements Les prolongements des travaux de cette thèse comprennent les perspectives d'évolution des modèles déjà évoquées ci-dessus, ainsi que l'extension des applications. On peut envisager en particulier l'utilisation du modèle de panache sous-maille et des techniques de réduction d'échelle pour d'autres polluants, et notamment pour les particules. Les concentrations locales autour des sources évoluent en effet très rapidement à cause de la sédimentation gravitationnelle (pour les grandes tailles de particules) ou de la coagulation (particules ultrafines), mais aussi de la condensation/évaporation et des réactions chimiques. Mieux représenter les phénomènes sous-maille est donc particulièrement intéressant pour ce type de polluants. On retiendra également, parmi les pistes les plus intéressantes :

- le couplage d'un modèle météorologique local avec le modèle gaussien utilisé en panache sous-maille, et
- l'utilisation d'un « modèle de boîte » local autour des stations de mesure pour tenir compte de l'échelle de représentativité.

Autres méthodes de changements d'échelle Le chapitre 1 a introduit différentes méthodes pour prendre en compte les changements d'échelle en modélisation de la qualité de l'air, et toutes n'ont pas été abordées au cours de la thèse. Parmi les autres méthodes qu'il serait envisageable de développer, et de comparer avec les solutions présentées ici, on peut citer l'utilisation d'un maillage adaptatif, structuré ou non. Cette solution est une alternative à la modélisation sous-maille des sources par les méthodes de couplage décrites ici. Enfin, l'utilisation de méthodes stochastiques, et notamment de techniques de type micro/macro, peut également s'avérer très intéressante pour une meilleure estimation de la contribution des différents processus (lessivage, chimie, diffusion) à la variabilité sous-maille.

Annexe A

Indicateurs statistiques

L'évaluation des performances des modèles présentés dans cette thèse, en comparaison à des observations, a fait intervenir différents indicateurs statistiques qui sont définis dans cette partie. Les formules pour calculer ces indicateurs sont données ci-dessous, en notant x_i et y_i les valeurs simulées et observées respectivement, pour une mesure i correspondant à un temps et un point d'observation donnés. Le nombre total de données est noté n. $\overline{\alpha}$ est la valeur moyenne de la variable α sur l'ensemble des données $\overline{\alpha} = \sum_{i=1}^{n} \alpha_i$.

A.1 Indicateurs pour les modèles à l'échelle locale

Les indicateurs utilisés dans le chapitre 3 sont :

1. Le biais fractionnel (« fractional bias ») :

$$\mathrm{FB} = \frac{2}{n} \sum_{i=1}^{n} \frac{y_i - x_i}{x_i + y_i}. \tag{A.1}$$

2. L'erreur quadratique normalisée (« normalized mean square error ») :

$$\mathrm{NMSE} = \frac{\sum_{i=1}^{n} (x_i - y_i)^2}{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}. \tag{A.2}$$

3. Le coefficient de corrélation :

$$\mathrm{Corr} = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \times \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2}}. \tag{A.3}$$

4. La proportion de valeurs simulées à moins d'un facteur 2 des valeurs observées :

$$\mathrm{FAC2} = \text{proportion de valeurs telles que } 0.5 \leq \frac{x_i}{y_i} \leq 2.0. \tag{A.4}$$

Notons que $\overline{y} - \overline{x}$ est utilisé pour le biais, et non $\overline{x} - \overline{y}$: les valeurs de FB sont donc négatives en cas de sur-estimation, et positives en cas de sous-estimation des valeurs.

A.2 Indicateurs à l'échelle régionale et continentale

En plus des indicateurs précédents, les indicateurs utilisés dans les chapitres 4, 5 et 7 sont :

1. Le biais moyen (« mean bias error ») :

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n} x_i - y_i. \tag{A.5}$$

2. Le biais fractionnel MFBE (« mean fractional bias error ») (opposée de FB) :

$$\mathrm{MFBE} = \frac{2}{n}\sum_{i=1}^{n} \frac{x_i - y_i}{x_i + y_i}. \tag{A.6}$$

3. La racine de l'erreur quadratique moyenne (« root mean square error ») :

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}. \tag{A.7}$$

4. La proportion de valeurs simulées à moins d'un facteur 5 des valeurs observées :

$$\mathrm{FAC5} = \text{proportion de valeurs telles que } 0.2 \leq \frac{x_i}{y_i} \leq 5.0. \tag{A.8}$$

5. la FMT (« figure of merit in time ») qui représente le pourcentage de valeurs simulées et mesurées qui sont superposées. On parle de FMT s'il s'agit d'une mesure en temps à un point donné, FMS (« figure of merit in space ») s'il s'agit d'une mesure en espace à un temps donné, et FM pour l'ensemble des temps et des stations. Dans cette thèse, l'indicateur est toujours noté FMT :

$$\mathrm{FMT} = \frac{\sum_{i=1}^{n} \min(x_i, y_i)}{\sum_{i=1}^{n} \max(x_i, y_i)} \tag{A.9}$$

A.3 Indicateurs utilisés pour l'ozone

On définit de plus des indicateurs spécifiques à l'ozone d'après EPA [2005] (chapitre 5) :

1. L'erreur moyenne absolue normalisée (« normalized mean gross error »)

$$\mathrm{MNGE} = \frac{1}{n}\sum_{i=1}^{n} \frac{|x_i - y_i|}{y_i}. \tag{A.10}$$

2. Le biais moyen normalisé (« mean normalized bias error ») :

$$\mathrm{MNBE} = \frac{1}{n}\sum_{i=1}^{n} \frac{x_i - y_i}{y_i}. \tag{A.11}$$

3. L'UPA (« unpaired peak accuracy »), indicateur de prédiction des pics. Cet indicateur n'est calculé que pour des valeurs supérieures à un certain seuil (ici, 30 $\mu\mathrm{g\,m^{-3}}$).

$$\mathrm{UPA} = \frac{x_{max} - y_{max}}{y_{max}}. \tag{A.12}$$

Annexe B

Performance de réseaux de détection à différentes échelles

L'étude présentée ici propose une méthode pour l'évaluation des capacités de détection de réseaux de capteurs utilisés pour la protection de zones sensibles. Les réseaux étudiés sont conçus pour détecter des émissions nucléaires, biologiques ou chimiques, quelle que soit la position de la source émettrice. L'objectif est de pouvoir détecter une émission potentiellement dangereuse, quel que soit son emplacement, avant que la zone protégée ne soit contaminée. On s'intéresse à deux types de configurations : des capteurs fixes, placés autours d'une zone contenant des bâtiments ou des installations à protéger (« échelle régionale »), ou des capteurs mobiles portés par des personnes dispersées dans un faible rayon (« échelle locale »). Il s'agit, dans les deux cas, de disposer les capteurs pour pouvoir détecter des émissions dans un large rayon, avec un temps de réponse assez faible. Pour cela, deux indicateurs de performance de réseaux sont proposés : la probabilité de détection, calculée sur un grand nombre de positions possibles pour les émissions, et le temps de saturation, qui est le temps où la détection de probabilité maximale est atteinte.

Les simulations ont été réalisées avec Polyphemus, en utilisant le modèle gaussien de panache pour l'échelle locale, et le modèle eulérien Polair3D pour l'échelle régionale. Les simulations comprennent différents scénarios en faisant varier l'emplacement de la source (961 cas), les espèces émises, et la situation météorologique. Cela permet de calculer les indicateurs de performance pour un capteur donné, ainsi que pour différents types de réseaux. Deux méthodes pour construire des réseaux de capteurs sont étudiées : (1) choisir un nombre donné de capteurs ayant la meilleure performance suivant un indicateur donné, et (2) placer les capteurs en cercles concentriques autour de la zone à protéger. Les réseaux constitués des capteurs ayant la meilleure détection de probabilité donnent de bonnes performances avec un faible nombre de capteurs. Le temps de saturation n'est pas un indicateur suffisamment robuste pour construire les réseaux sur cette base. Enfin, les réseaux formés de cercles concentriques donnent de bons résultats dans les cas étudiés, si le nombre de capteurs est suffisant.

Sommaire

B.1 Introduction **235**
B.1.1 Context and objectives 235
B.1.2 Methodology and outline 235
B.2 Test scenarios **236**
B.2.1 Domains and computing strategy 236
B.2.2 Sources 238
B.2.3 Meteorological scenarios 239
B.2.4 Simulations 239
B.3 Sensor and network statistics **241**
B.3.1 Statistical indicators for sensors 241
B.3.2 Constraints on the sensors 242
B.3.3 Statistical indicators for networks 243
B.4 Sensors performance **243**
B.4.1 Influence of scenarios and meteorology 243
B.4.2 Emitted quantities and loss processes 246
B.5 Networks performance **246**
B.5.1 Networks defined with respect to detection probability 246
B.5.2 Networks defined as concentric circles 249
B.5.3 Networks sensitivity to wind variations 253
B.6 Conclusions **253**

Ce chapitre est constitué de Korsakissok *et al.* [2010]. L'étude a été menée pour le compte de la DGA (Délégation Générale de l'Armement— `http://www.defense.gouv.fr/dga/`).

B.1 Introduction

B.1.1 Context and objectives

The design, evaluation and optimization of monitoring networks have already been investigated, both in the field of atmospheric chemistry for ozone and its precursors (Nychka et Saltzman [1998], Fuentes *et al.* [2007]), and to monitor accidental releases of radionuclides. Studies on the latter subject rely on the assumption of a few well-known possible source locations, corresponding to power plant locations. They have been carried out at several spatial scales, from local to global: near-field studies involve measurement sensors placed within 30 km of nuclear power plants, whether at fixed locations [Krysta *et al.*, 2006] or with mobile device [Jiang *et al.*, 2007], while the optimization of monitoring networks at mesoscale (up to 2000 km) is investigated in Abida *et al.* [2008]. The monitoring of radionuclides for military applications at global scale is thoroughly assessed by the CTBT organization ("Comprehensive nuclear Test-Ban Treaty", Ramaker [2003]). All these applications share two main purposes: the detection (of an accident or a pollution peak), and the reconstruction of the concentration fields on the basis of the measurements. Most studies focus on the networks ability to reconstruct the map of concentrations, which involves indicators such as the difference between the reconstructed concentrations and the *true* field (given by simulations, or measurements when available).

This paper addresses a similar issue, but in a military or homeland security context: a delimited area—a facility, a high-profile event, a combat unit, or dignitaries—has to be protected from any hazardous air pollutant. The pollutant may be released with harmful intent from a single ground location in the vicinity of the domain, and the source location and characteristics are unknown. Two sensor siting applications are addressed. In the first case, fixed sensors may be placed around a zone containing buildings to be protected. In the second case, people carrying detection devices are scattered within an area containing small units. In both cases, the *protection ability* of a network is determined by its ability to detect a release, wherever its location. In addition, a performance criterion related to the *response time* can be defined, in cases an advanced warning is needed (for instance, if the released species quickly reaches harmful concentrations with immediate health effects). Therefore, appropriate performance indicators should be designed, to quantify the network abilities in both respects. The aim of this study is to explore the new framework with (1) the design of performance indicators for such networks, (2) relevant tests in different meteorological and emission scenarios, and (3) the design of several networks based on the performance indicators. The chosen test scenarios are highly idealized (perfect sensors, simple domain shapes...) and should be adapted to real-life situations. The underlying objective is to issue general recommendations regarding the performance indicators and the methods to be retained in the design of a detection network for real cases.

B.1.2 Methodology and outline

In this study, two sets of simulations are performed: the first one corresponds to the *buildings protection scenario*, with a 50×50 km^2 domain, and the second one is the *small-units protection scenario*, with a 10×10 km^2 domain. The simulations are carried out with the air quality modeling system Polyphemus [Mallet *et al.*, 2007], using an Eulerian model in the former case and a Gaussian model in the latter. For both scenarios, many simulations are performed (1) to account for many sources locations and (2) to include several typical meteorological situations. This generates a large database from which detection capabilities are assessed. The simulation scenarios are described in Section B.2. Two statistical indicators are used to quantify the detection abilities: the detection probability and the saturation time—see Section B.3. The simulation database allows one to compute these indicators for any possible sensor location (Section B.4)

or network configuration (Section B.5). For both sensors and networks, Sections B.4 and B.5 address the sensitivity of the indicators to parameters like the meteorological conditions or the source characteristics. In Section B.5, two families of networks are considered. In the first one, the networks are composed of the best sensors according to the detection probability. In the second family, the networks are designed with concentric circles. This particular geometry is chosen because of its simple design, and it could provide good results without further optimization.

B.2 Test scenarios

B.2.1 Domains and computing strategy

In our context, a delimited area (the "protected domain" D_p) has to be protected from an hazardous air pollutant. A dangerous source is defined as a source that can contaminate the domain D_p, that is, triggers above-threshold concentrations in at least one point of D_p. This is one of the hypothesis made in this study, where idealized cases are presented. The main assumptions are:

1. Immediate contamination of the protected domain, as soon as one sensor within D_p detects the pollutant,
2. Square-shaped domains, while a real-case study might account for more complex shapes (oblong or L-shaped buildings) and land-use coverage (e.g., water might forbid siting sensors at some locations),
3. Perfect sensors, with immediate response, and a low detection threshold (see Section B.3.1).

In addition to the protected area, several domains are defined (see Figure B.1):

- D_c is the area where sensors may be placed (one per cell). It includes D_p. Some sensors can be placed within D_p.
- D_s contains all dangerous sources (one per cell). It includes D_c.
- D_M is the modeled domain. It includes all other domains.

In this section, the size of these domains is detailed, along with the simulation set-up, for the two base test scenarios: the buildings protection scenario ("regional scale"), and the small-units protection scenario ("local scale").

Buildings protection scenario

Here, the domain D_p is a facility (military base, for example). The sensors are placed at fixed locations, either within the base or in its surroundings. The main purpose in this case is advanced warning, so the alert must be raised *before* the pollutant reaches D_p. This early-warning criterion could be relaxed, for example in the case of a biological attack, where medicine could be given some time after the event. The simulations are carried out with the Eulerian model Polair3D [Boutahar *et al.*, 2004]. The domains D_p, D_c, D_s and D_M are defined in Table B.1. The simulation grid has a resolution of 0.01° in the horizontal, which is about 1 km. It includes ten vertical layers, up to 3060 m above ground. The height of the first layer is 30 m. The numerical time step is 60 s, and a simulation covers a time period of 3 hours. There are 50×50 cells in D_M, and meteorological conditions as well as land use can differ from one cell to another (although the terrain is flat and rural).

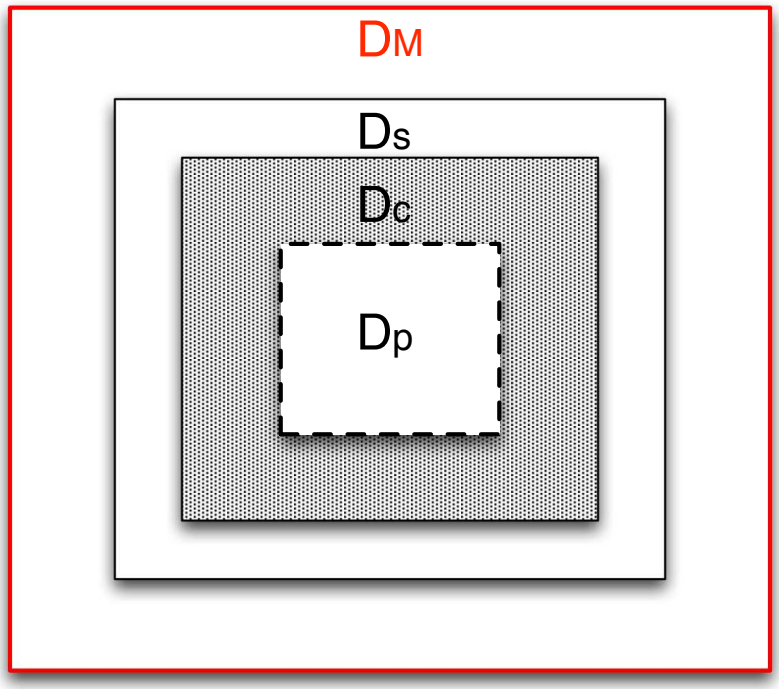


Figure B.1: Domains: D_p is the zone to be protected, D_c is the domain of sensors, D_s is the domain of sources, and D_M is the modeled domain.

Domain	Dimension
D_p	4×4 km^2
D_c	20×20 km^2
D_s	31×31 km^2
D_M	50×50 km^2

Table B.1: Size of the different domains at regional scale.

Small-units protection scenario

In this scenario D_p is a zone containing people (military units, dignitaries...), and the sensors are supposed to be scattered among them (typically, people carrying detection devices) so that $D_c = D_p$. Therefore, there is no advanced warning here, and it is assumed that the sensors threshold is lower than the minimum harmful concentrations, and that a rapid response is not a critical issue. The domains are defined in Table B.2. The simulations are carried out with the Gaussian puff model of the Polyphemus platform [Korsakissok et Mallet, 2009]. Contrary to the Eulerian model, the Gaussian model gives an analytical formula to compute the concentrations, and does not need a grid. The grid is used in order to consider a finite number of source and sensor locations. The concentrations are computed in each cell of the gridded domain D_c, at ground level only. The simulation duration is 3 hours, the time step is 100 s, and the space step in the horizontal is 200 m. Here, the meteorological data and the land type are supposed to be homogeneous.

Domain	Dimension
D_p	3×3 km^2
D_c	3×3 km^2
D_s	6.2×6.2 km^2
D_M	10×10 km^2

Table B.2: Size of the different domains at local scale.

B.2.2 Sources

The simulations are carried out for three species, the first with radioactive decay, the second with biological decay and the third with no decay at all (simply passive). The timescale of the decay is of the order of 8 days for the *radioactive species*, *gaseous iodine*, which makes it almost passive. The *biological species* decays with a characteristic time of ten minutes (during daytime) or twenty minutes (during nighttime). It is modeled as particulate matter (PM) with diameters of 0.5 µm, 5 µm and 10 µm. The third species is a purely passive gaseous species (GM for gaseous matter). Table B.3 gives the emitted quantities. Wet scavenging and dry deposition are taken into account for the three species. The dry-deposition velocity of the two gaseous species is supposed to be constant and equal to 5.10^{-3} m s^{-1}. The Belot parameterization [Belot *et al.*, 1988] is used to estimate their scavenging coefficients. For particulate matter, the dry-deposition velocities are computed for the several diameters (with gravitational settling), and the scavenging coefficients are computed according to the Slinn parameterization [Underwood, 2001].

Species	Diameter (µm)	Quantity (kg)
Iodine	gas	1.
GM	gas	5.
Biological	0.5	10.
Biological	5.	10.
Biological	10.	10.

Table B.3: Emitted quantities for the different species. The quantities are emitted instantaneously.

B.2.3 Meteorological scenarios

Because of the large number of simulations (one per possible source), running simulations for a full year is too high a computational burden. Therefore, we focused on a few typical meteorological situations. Low-dispersion cases were privileged, since high concentrations were supposed to be the most critical from a health response standpoint. Thus, the selected situations do not account for all possible wind directions. However, changing the source location is more or less equivalent to changing the wind direction, at least for a circular geometry (in the case of square domains, some "corner effects" can appear, see Section B.5.1 for instance). In addition, one can assume that the wind direction is known in the two kinds of sensors siting applications addressed here:

- Protecting a building: the most frequent wind direction(s) are well known and can be taken into account in a prospective study to place sensors,
- Protecting small units: there is no need to cover a wide range of meteorological situations since the sensors are placed in real-time.

Data for the Eulerian model were provided by a preliminary study of ECMWF meteorological data for year 2004 in the French region of Bourges (chosen because of its flat terrain). Four situations were retained. The first two cases, hereafter called "case" or "situation" 0 and 1 respectively, are winter cases of low dispersion. The third case, labelled "situation 2", is a summer case with very low dispersion during night, and substantial rain. The last case is a summer situation of medium dispersion. Table B.4 gives the averaged values of the wind speed, the boundary layer height, the rain, the cloudiness and the Pasquill stability class, for the four situations. Figure B.2 shows the maps of wind directions for the four meteorological situations at the beginning of the simulation (emission time). The maps give a good idea of the general wind direction, except for case 2 where the wind turns during the simulation, and partly comes from the south at the end of the three hours.

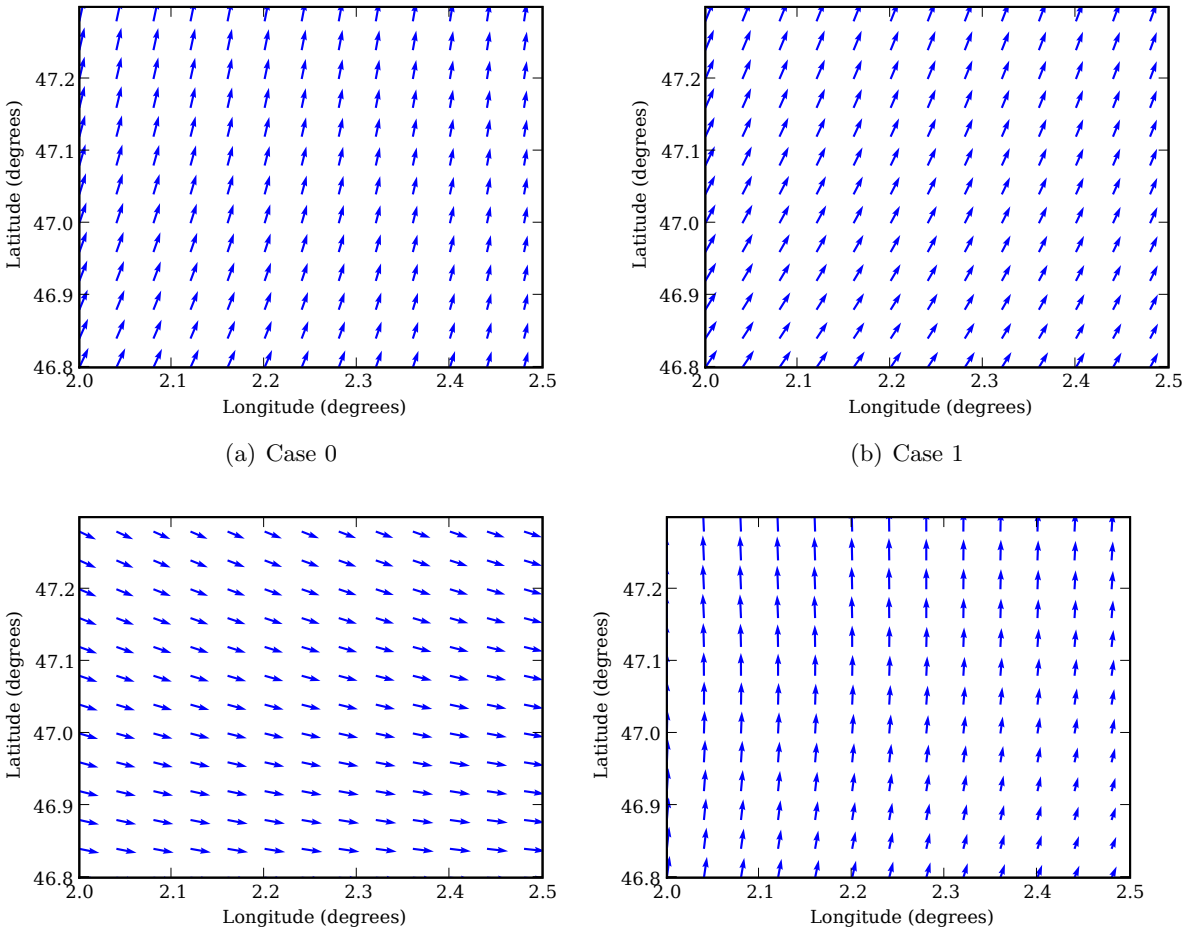
Date	Mean wind	Mean BLH	Mean rain	Cloudiness	Stability
2004-01-04 09h	1.05	175	0.	0.47	B
2004-01-04 15h	1.59	250	0.01	0.46	B
2004-08-10 03h	1.66	308	3.89	0.99	F
2004-08-23 09h	2.97	565	0.11	0.59	C

Table B.4: Wind speed in $\mathrm{m\,s^{-1}}$, boundary layer height (BLH) in m, rain intensity in $\mathrm{mm\,hr^{-1}}$, cloudiness (0 is clear sky, 1 is a total cloudiness) and Pasquill stability class for the four meteorological situations at regional scale (averaged values over the domain and simulation duration). The hours given here are local time.

In the Gaussian model, the meteorological data is assumed to be homogeneous and constant. Therefore, these are synthetic cases, designed to share the main dispersion characteristics (low or medium dispersion, rain) with the regional scale. The wind speed, the boundary layer height, the rain, the Pasquill stability class are taken in one cell of the regional domain, and summarized Table B.5. Only the wind directions have been shifted and therefore differ from the regional values (Figure B.3), in order to account for various wind directions.

B.2.4 Simulations

The simulations are carried out for all sources in the domain D_s, which in both test scenarios amounts to 961 sources: the regional domain D_s covers 31×31 km^2 and the discretization is



(a) Case 0

(b) Case 1

(c) Case 2

(d) Case 3

Figure B.2: Map of winds at the emission time for each meteorological situation at regional scale. Note that only wind direction is represented, the arrow lengths are not proportional to the speed.

Case	Wind	Boundary height	Rain	Stability
0	1.01	152	0.	B
1	1.66	244	0.003	B
2	1.42	300	4.19	F
3	3.46	623	0.18	C

Table B.5: Wind speed in $\mathrm{m\,s^{-1}}$, boundary layer height in m, rain intensity in $\mathrm{mm\,hr^{-1}}$ and Pasquill stability class for the four meteorological situations at local scale.

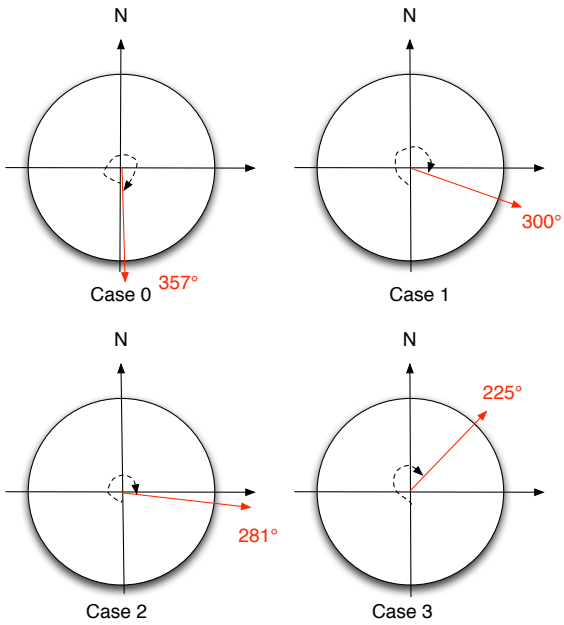


Figure B.3: Wind direction for each meteorological situation at local scale. The direction of the wind u is shown, and the wind angle from north is indicated in degrees.

1 km, and the local domain is $6.2 \times 6.2\,\mathrm{km}^2$ wide, with cells of size 200 m. There are four meteorological situations, two gaseous species and three particulate matter diameters, which corresponds to 20×961 simulations in total. Once the simulations are gathered into a "concentration database", one may compute the performance indicators for any network, any species and any meteorological situation. In Section B.5.3, additional simulations are carried out with variations of wind speed and wind angle, on the basis of meteorological situation 1 and for one species (iodine). There are 8 variations of wind speed and 12 variations of wind angle, so the number of extra runs required for the sensitivity study amounts to 20×961.

B.3 Sensor and network statistics

B.3.1 Statistical indicators for sensors

A sensor is defined by a position, the species it monitors, a detection threshold C_{min} (minimum concentration for the species to be detected), and a minimum duration t_{min}. A species is detected if and only if its concentration remains above the detection threshold C_{min} for at least t_{min}. In this study, $C_{min} = 0.01\ \mu\mathrm{g\,m^{-3}}$ and t_{min} is only limited by the simulation time step Δt. A sensor may be placed at the center of any cell in the domain D_c. Let k be the index of that cell for a given sensor. For this sensor k, we define the following indicators:

1. The arrival time of a given source s, $t^k_{arr}(s)$:

$$t^k_{arr}(s) = \min\{t | C(t) > C_{min} \text{ and } d(t) \geq t_{min}\}\ , \tag{B.1}$$

with $C(t)$ the concentration at the sensor location and at time t for the species emitted by the source s, and $d(t)$ the time during which $C(t)$ has been above the sensor threshold C_{min}. Thus, $t^k_{arr}(s)$ is the first date at which the source is detected by the sensor. It is

the sum of the propagation time from the source to the sensor, and of t_{min}. It is set to zero if the source s is never detected by the sensor.

2. The sensor detection probability at time t, $P^k(t)$. It is the proportion of sources that have been detected by sensor k at time t.

$$P^k(t) = \frac{1}{N_s} \sum_{s=0}^{N_s-1} P_s^k(t) \ , \tag{B.2}$$

where $P_s^k(t)$ is equal to zero if the source s has not been detected by the sensor at time t and 1 otherwise, that is:

$$P_s^k(t) = \begin{cases} 1 & \text{if } t_{arr}^k \neq 0 \text{ and } t \geq t_{arr}^k, \\ 0 & \text{otherwise.} \end{cases} \tag{B.3}$$

In Equation B.2, N_s is the number of sources, among all possible sources within D_s, that actually reach D_p (so-called *effective sources*). Thus, all sources downwind of D_p are ignored, which prevents a needless increase in N_s and an artificial decrease of the detection probabilities.

In order to quickly assess a sensor performance, it is useful to define more global indicators, that do not depend on time or source. In this study, we focus on the *maximum detection probability* P_{max}^k of a sensor k:

$$P_{max}^k = \max_t P^k(t) = P^k(+\infty) \ . \tag{B.4}$$

In addition, we also define the *saturation time* t_{sat}^k, which is the time when the maximum detection probability of sensor k has been reached:

$$t_{sat}^k = \inf\{t | P^k(t) = P_{max}^k\} \ . \tag{B.5}$$

Note that all indicators involving the arrival time are dependent on the simulation time step: the finer the time step, the better the discrimination between two sensors based on time constraints.

B.3.2 Constraints on the sensors

In the buildings protection scenario, there is an additional condition: the source must be detected before it contaminates D_p (this is irrelevant at local scale where $D_p = D_c$). Thus, we define t_{dp} as the time when the source first reaches D_p:

$$t_{dp} = \inf\{t_{arr}^k | k \in D_p\} \tag{B.6}$$

In that case, Equation B.3 is modified as follows:

$$P_s^k(t) = \begin{cases} 1 & \text{if } 0 < t_{arr}^k \leq t_{dp} \text{ and } t \geq t_{arr}^k. \\ 0 & \text{otherwise.} \end{cases} \tag{B.7}$$

This criterion could be modified, depending on the desired time interval between the warning and the contamination of D_p. In the following, a detection at time t_{dp} is considered successful, so sensors located within D_p can have a non-zero detection probability.

B.3.3 Statistical indicators for networks

A network is defined as a set of sensors. Its detection probability and arrival time can be defined similarly to the case of a single sensor:

$$P^{\text{network}}(t) = \frac{1}{N_s} \sum_{s=0}^{N_s-1} {P_s}^{\text{network}}(t) \ , \tag{B.8}$$

where ${P_s}^{\text{network}}(t)$ is equal to 0 if the source s has not been detected by the network and 1 otherwise. As soon as $P_s^k(t) = 1$ for a given sensor k of the network, ${P_s}^{\text{network}}(t) = 1$. It follows that

$${P_s}^{\text{network}}(t) = \max_k P_s^k(t) \ ,$$

where k is in $[\![0, N_{sensor}[\![$ if N_{sensor} is the number of sensors in the network.

The arrival time of the network for the source s is:

$$t_{arr}^{\text{network}}(s) = \min_k t_{arr}^k(s) \ . \tag{B.9}$$

Finally, the maximum detection probability and saturation time of a network are

$$P_{max}^{\text{network}} = \max_t P^{\text{network}}(t) = P^{\text{network}}(+\infty) \tag{B.10}$$

and

$$t_{sat}^{\text{network}} = \inf\{t | P^{\text{network}}(t) = P_{max}^{\text{network}}\} \ . \tag{B.11}$$

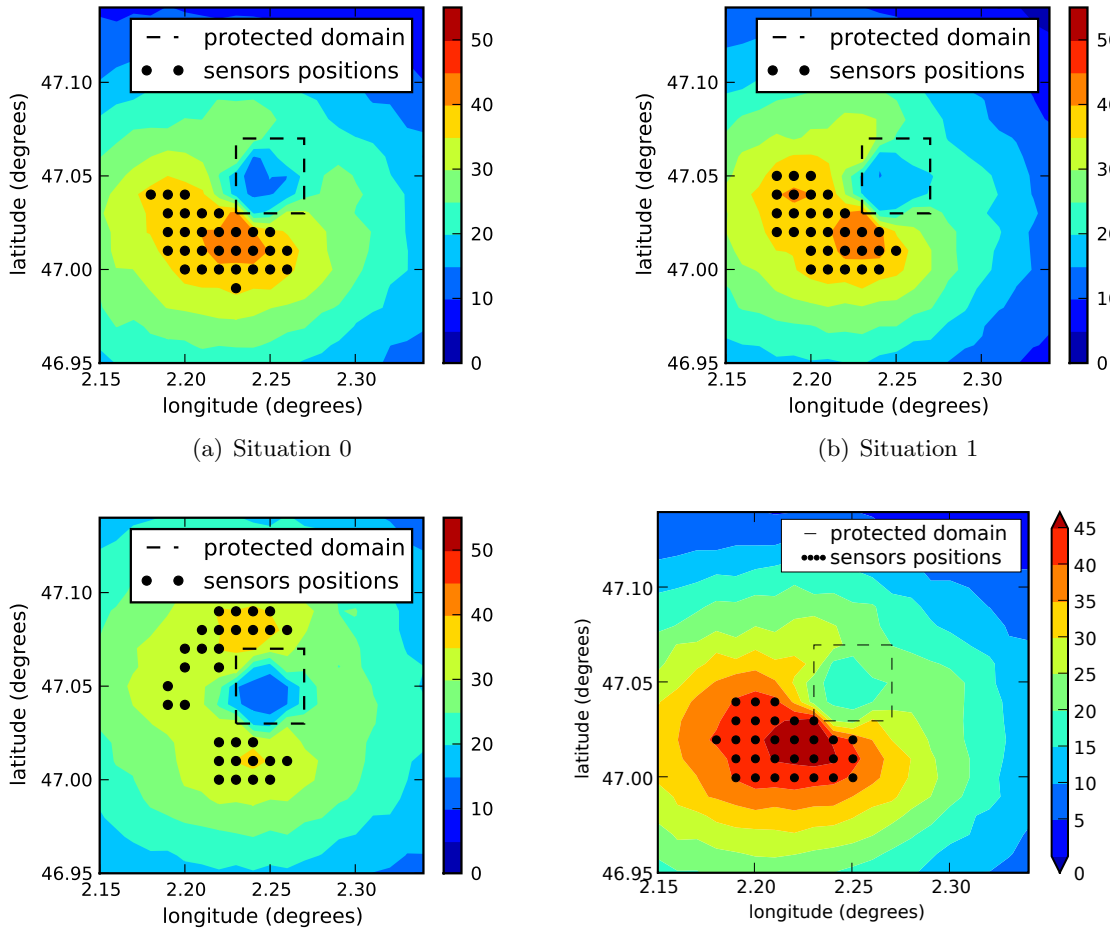
B.4 Sensors performance

Before defining any network, it is useful to have a glimpse of the performance of all individual sensors within the domain D_c: any network is a subset of this ensemble of sensors. In particular, the networks considered in Section B.5.1 are composed of the best sensors (according to a given indicator), whose positions can be inferred from the maps of detection probability P_{max} and saturation time t_{sat}. The maps also give insights on the sensors sensitivity to several parameters such as the emitted quantity, the dispersion, and the rain. They allow to address the following questions: to what extent do the conclusions of the study depend on the simulation scenarios? How to choose the best scenarios in order to be representative of the networks general performance?

B.4.1 Influence of scenarios and meteorology

Only D_c is shown on the maps, and the borders of D_p are indicated when necessary. Figure B.4 shows the maps of detection probabilities for iodine and all meteorological cases for the buildings protection scenario, and Figure B.5 for the small-units protection scenario. In the former case, the Eulerian simulations are carried out with non-homegenous conditions (meteorology and land-use), while the latter case corresponds to the Gaussian simulations, with constant and homogeneous conditions. The sensors performance depends, quite obviously, on the wind direction. While this is easily verified in the Gaussian results (Figure B.5), the "main wind direction" is not so clear with the Eulerian model, due to the inhomogeneous and unstationnary conditions (e.g., the wind turns from the west–north west to south direction in the case 2). Meanwhile, the other characteristics of the dispersion (wind speed and turbulence strength)

may have more complex effects on the results. The effects of a very low-dispersion case like situation 2 are twofold: (1) higher concentrations, and (2) lower horizontal spread of the emitted plume. The former has consequences similar to a small increase of the emitted quantity: there are more effective sources, and P_{max} is higher at the downwind part of the domain, which results in a more homogeneous map of detection probabilities. Conversely, the medium-dispersion case (situation 3) results in higher discrepancies between the sensors. On the other hand, a lower plume spread leads to a decrease in the number of effective sources as well as in detection probabilities (a narrower plume reaches fewer sensors).



(a) Situation 0

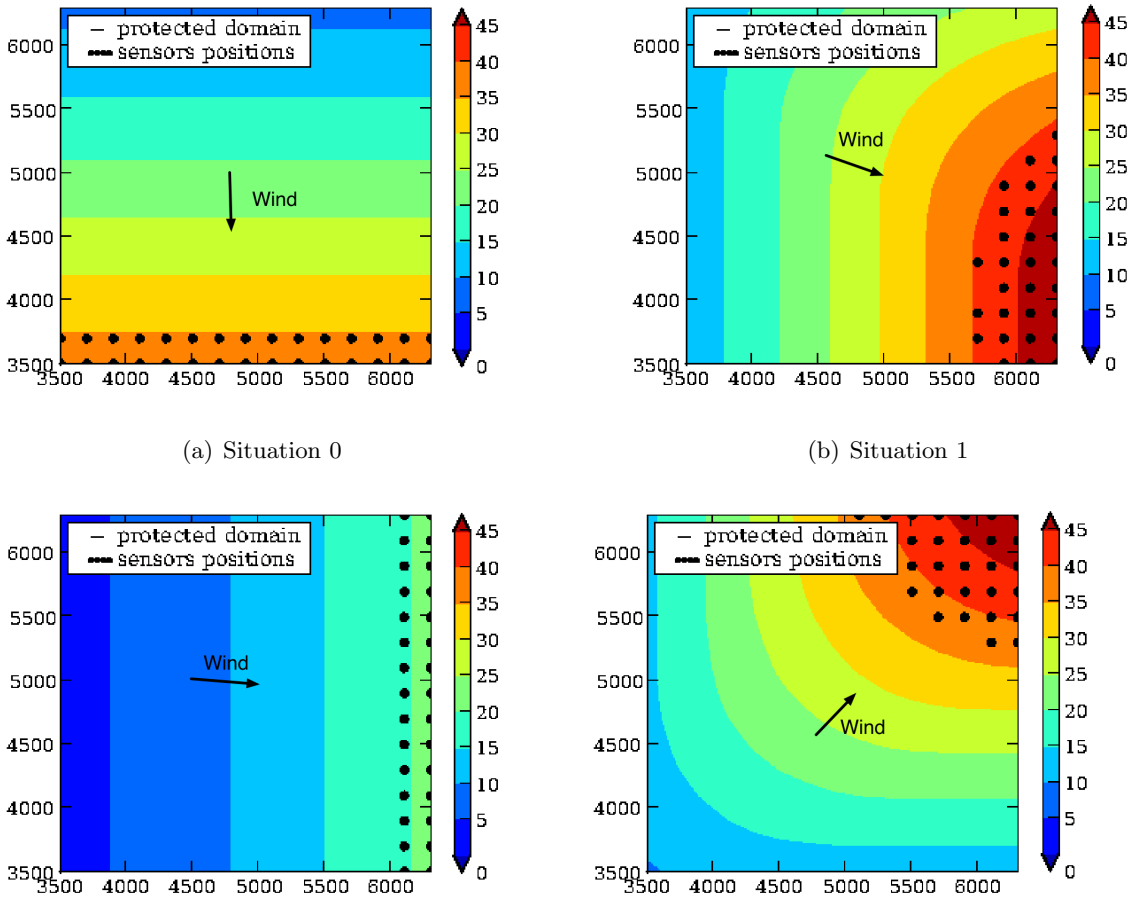
(b) Situation 1

(c) Situation 2

(d) Situation 3

Figure B.4: Regional scale: Map of maximum detection probability (percent) for all sensors in the domain D_c and the four meteorological situations. The black dots are the positions of the 30 sensors having the highest detection probability.

The main differences between the two scenarios are due to the additional constraint in the buildings protection case (Section B.3.2): a sensor successfully detects a source only if it does so before the pollutant reaches D_p. Thus, sensors within D_p show a very low detection probability, and the best sensors according to P_{max} are located near the boundary of D_p (more or less upwind). On the contrary, in the units protection scenario, $D_c = D_p$, and the sensors with the highest detection probability are placed at the farthest end of the domain, downwind. The maps of t_{sat}, not shown here, are exactly the "reverse" of the P_{max} maps in this case: the farthest sensors also have the highest saturation times. This suggests that coupling the two indicators



(a) Situation 0

(b) Situation 1

(c) Situation 2

(d) Situation 3

Figure B.5: Local scale: Map of maximum detection probability (percent) for all sensors in the domain D_c and the four meteorological situations. The black dots are the positions of the 30 sensors having the highest detection probability. The dashed line is the boundary of D_p.

could lead to a good compromise, if a time criterion had to be applied in this scenario.

B.4.2 Emitted quantities and loss processes

The maps of P_{max} and t_{sat} at regional scale are shown in Figure B.6 for the meteorological situation 1 and all species. It shows significant differences between the emitted species. There are two possible reasons for these differences: (1) the emitted quantities and (2) the loss processes. In meteorological situation 1, there is no wet scavenging: the only loss processes are dry deposition and linear decay. Since the two gaseous species have the same deposition velocities, and the timescale of radioactive decay for iodine is several days, the differences between the maps for GM and iodine are due to the emitted quantity (1 kg for iodine and 5 kg for GM, see Table B.3). The biological species has the highest emitted quantity (10 kg), but it undergoes a significant biological decay. Since all the processes involved in the dispersion are linear, the emitted quantity would not be an important parameter but for the sensors threshold. Increasing the quantity emitted by one source induces higher concentrations: the plume stays longer above the threshold, and can be detected by sensors at farther distances (downwind or crosswind). Thus, the maps of P_{max} become more homogeneous, since a source is detected by almost all sensors, even the farthest downwind (Figures B.6(c) and B.6(e)). This also leads to higher saturation times, since farther sources are detected. This raises the issue of the number of effective sources (as defined in Section B.3), which obviously increases with the emitted quantity. The size of D_s should therefore be modified accordingly, to account for all possible sources. Otherwise, the maps are *not* representative of the real sensors performance, and sensors placed downwind of D_p show an unrealistic detection probability. This condition is verified for the situation 1 and iodine, but not for all quantities tested here. In the rest of the paper, the reference species used is therefore iodine.

Figure B.7 allows to assess the influence of loss processes (wet scavenging and dry deposition) for the same emitted quantity: the maps of detection probability are shown for the biological species and three diameters, for the meteorological situation 2 (with rain). The detection probabilities of the coarsest particles are slightly lower, due to the strong scavenging, but there are no significant differences otherwise.

B.5 Networks performance

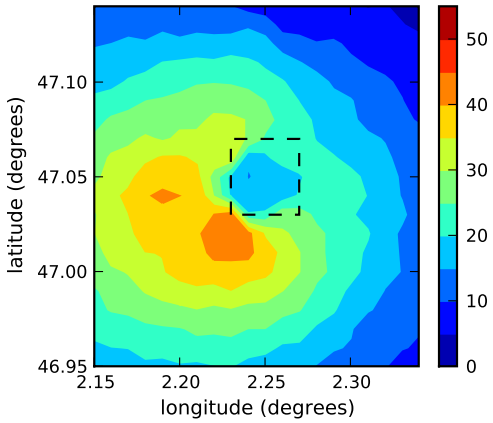
In this section the performance of two types of networks is evaluated:

1. Networks composed of the sensors with the highest detection probabilities,
2. Networks defined with a geometry of concentric circles.

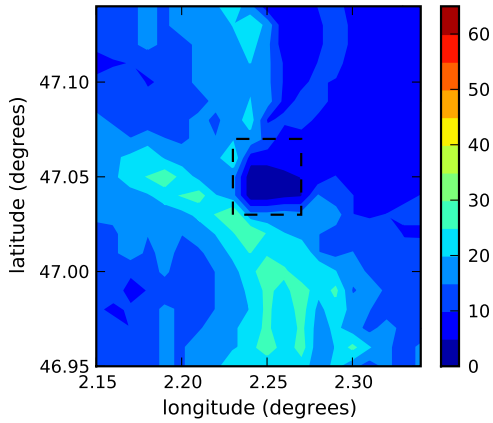
These networks are studied at local and regional scale, for all meteorological situations and for iodine. The results for networks composed of the sensors having the best saturation times are not shown, since they give an overall poor performance, except for an excessively large number of sensors (over 200). This indicator is still, however, used to evaluate the networks behavior with respect to time.

B.5.1 Networks defined with respect to detection probability

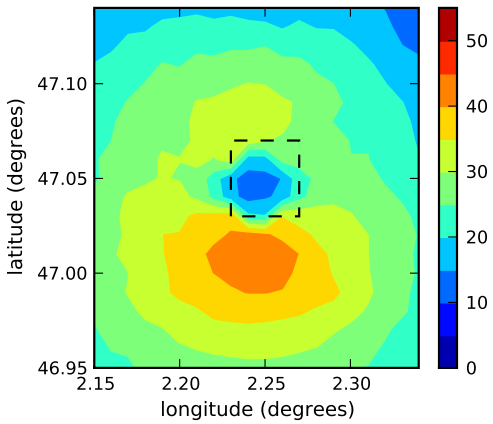
The networks evaluated in this section comprise the N_{sensor} sensors having the highest detection probabilities (see Figure B.4 and Figure B.5 for the sensors positions with $N_{sensor} = 30$). For 13 networks with $5 \leq N_{sensor} \leq 100$, the statistics described in Section B.3.3 are computed for all meteorological situations.



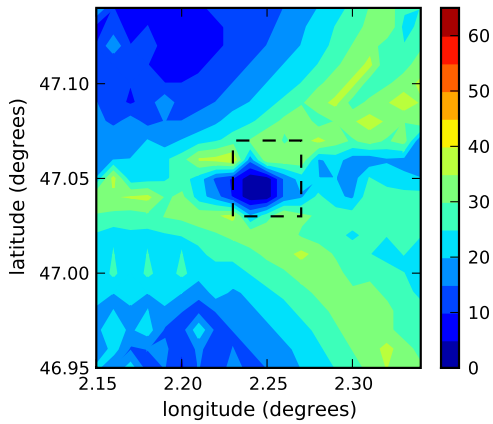
(a) P_{max} for iodine



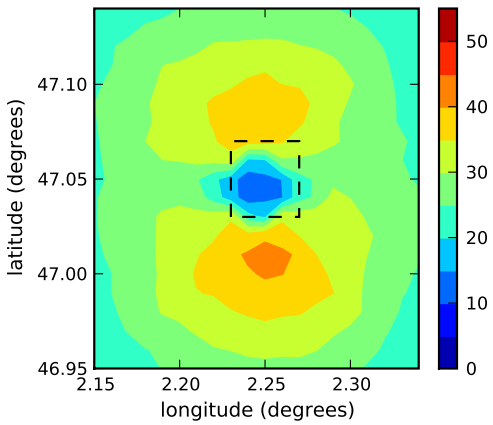
(b) t_{sat} for iodine



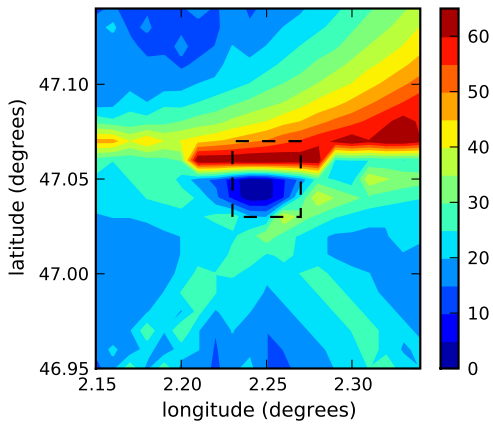
(c) P_{max} for GM



(d) t_{sat} for GM

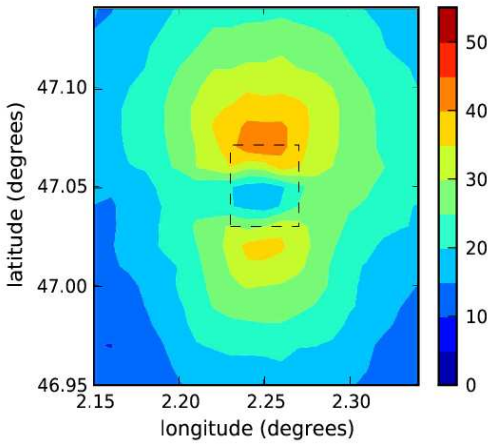


(e) P_{max} for biological PM - diameter 0.5 µm

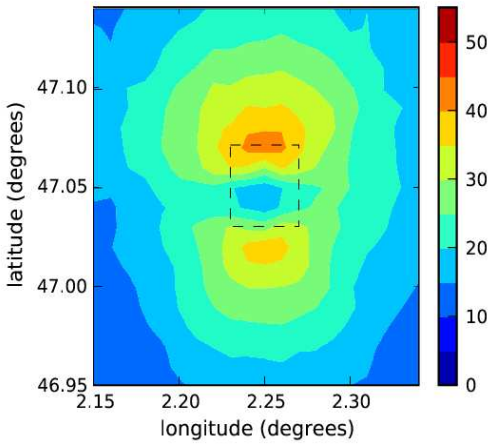


(f) t_{sat} for biological PM - diameter 0.5 µm

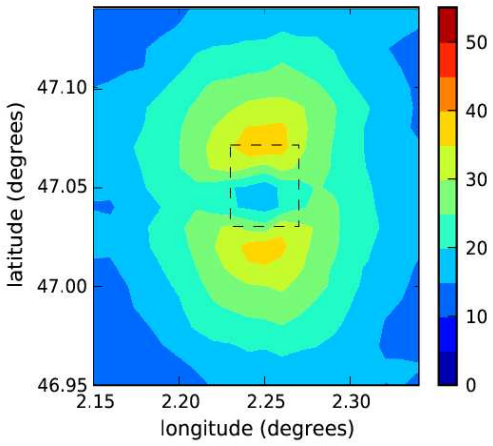
Figure B.6: Maps of maximum detection probability (in %) and saturation time (in minutes) for all sensors in the domain D_c and three different species – Meteorological situation 1 (low dispersion, winter, daytime). The emitted quantities are: 1 kg of iodine, 5 kg of GM and 10 kg of biological tracer for each diameter. The dashed line is the boundary of D_p.



(a) Biological PM – diameter 0.5 µm



(b) Biological PM – diameter 5 µm



(c) Biological PM – diameter 10 µm

Figure B.7: Map of maximum detection probability (percent) for all sensors in the domain D_c and biological particle of three different diameters – Meteorological situation 2 (low dispersion, summer, nighttime, rain). The dashed line is the boundary of D_p.

Figure B.8 shows the networks performance against the number of sensors, for the four meteorological cases. The high detection probability of networks comprising a few sensors suggests a good complementarity between the sensors (they do not detect the same sources), which is a nontrivial fact considering the small sensors spacing. In the units protection scenario, however, this complementarity is mitigated in cases when all the best sensors are located in one corner of the domain D_c (cases 1 and 3—see Figure B.5). This "corner effect" is avoided in the buildings protection scenario, since all sensors are placed around D_p—due to the time constraints—rather than at the border of D_c. The detection probabilities are not very sensitive to the meteorological situation. P_{max} for situation 2 is very high in both scenarios, despite the low detection probabilities of individual sensors highlighted in Section B.4, especially at local scale. This may be related to the narrow horizontal plume spread in this meteorological situation: each sensor detects few sources, but it does not detect the same sources as its neighbors. In this situation, P_{max} therefore increases very rapidly with the number of sensors in the network.

In terms of saturation times, Figure B.8 shows the good performance of networks at regional scale, compared to the values of t_{sat} for single sensors. This is consistent with the different patterns of the maps of detection probabilities and saturation times (Figure B.6): the sensors with good detection probabilities are not the same as the sensors having the highest saturation times. When increasing the number of sensors in the network, sensors with lower saturation times are rapidly included, contrary to the local-scale scenario which shows scarcely any improvement in t_{sat} as the number of sensors in the network increases. One may also notice significant differences between the networks saturation times in both scenarios—compared to regional scale, the values of t_{sat} at local scale highly depend on the situation and are much higher. This probably comes from a slower horizontal diffusion in the Gaussian model than in the Eulerian model.

The overall performance of such networks in terms of detection probability is quite good for all meteorological situations. The saturation times are relatively good (in comparison to single sensors) at regional scale, but this indicator depends too much on the situation to be reliable by itself. These results suggest that a practical way to design a detection network would be to choose a required minimum detection probability, and to adjust the number of sensors in the network in order to decrease the saturation time if necessary. The resulting network should then be tested for several dispersion conditions, and compared to the "best" networks for each situation, as done in Section B.5.3 where the robustness of such networks with respect to meteorological variations is assessed.

B.5.2 Networks defined as concentric circles

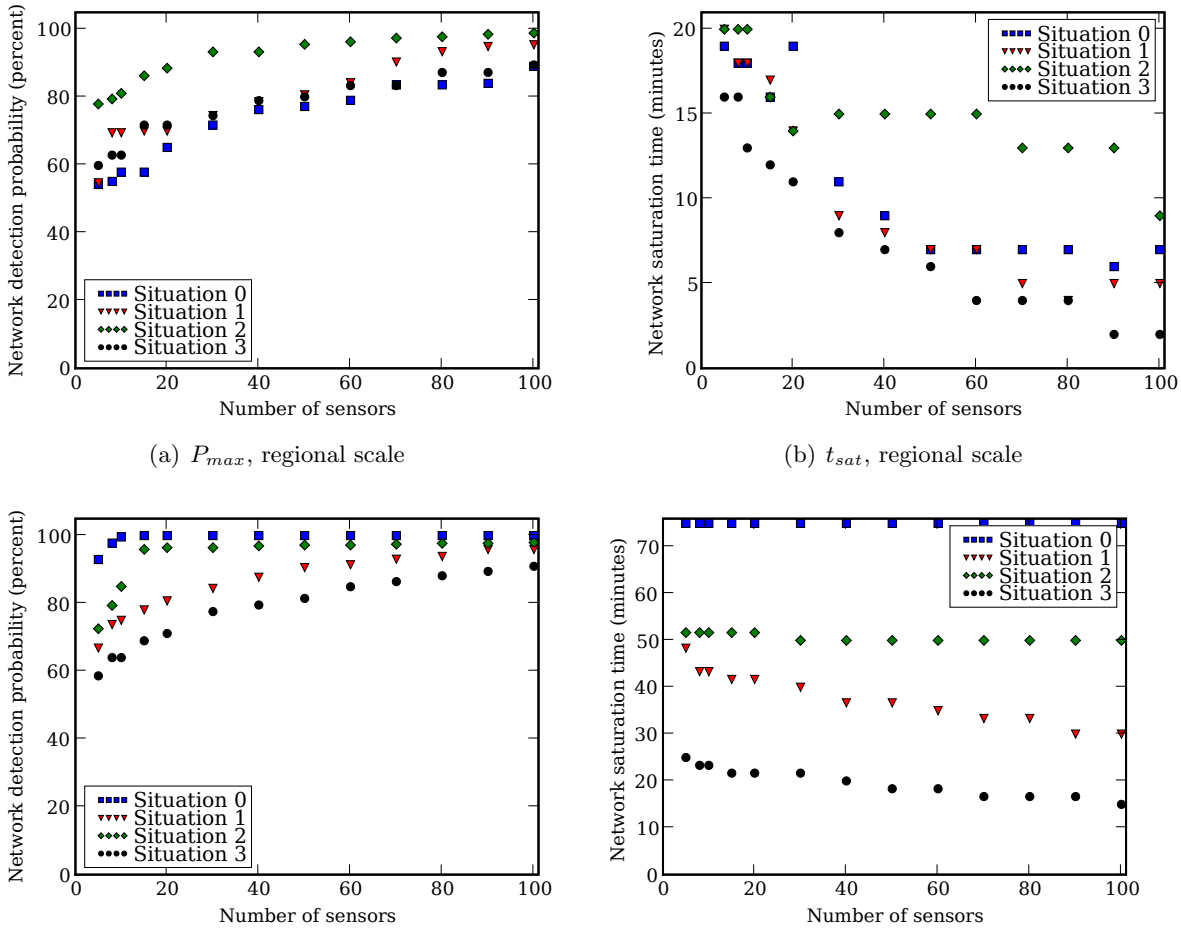
In the buildings protection scenario, one or more concentric circles are placed around D_p, with a constant distance between two circles. The first circle (internal circle) has a radius R_{min}, and the last one (external circle) has a radius R_{max}. The total number of circles is denoted N_{circle}. Either the number of sensors per circle N^0_{sensor} is constant, or, alternatively, the distance between two sensors l_{sensor} is fixed (then, the number of sensors increases with the circle radius). In the former case, the total number of sensors in the network is

$$N_{sensor} = N^0_{sensor} \times N_{circle} \ . \tag{B.12}$$

In the latter, the number of sensors is

$$N_{sensor} = \sum_{i=1}^{N_{circle}} 2 \ \pi \ R_i / l_{sensor} \ , \tag{B.13}$$

where R_i is the radius of the circle i: $R_i = R_{min}$ if $i = 1$ and $R_i = R_{max}$ if $i = N_{circle}$. For example, figure B.9 shows two possible configurations (for 3 circles), one with a given N^0_{sensor}



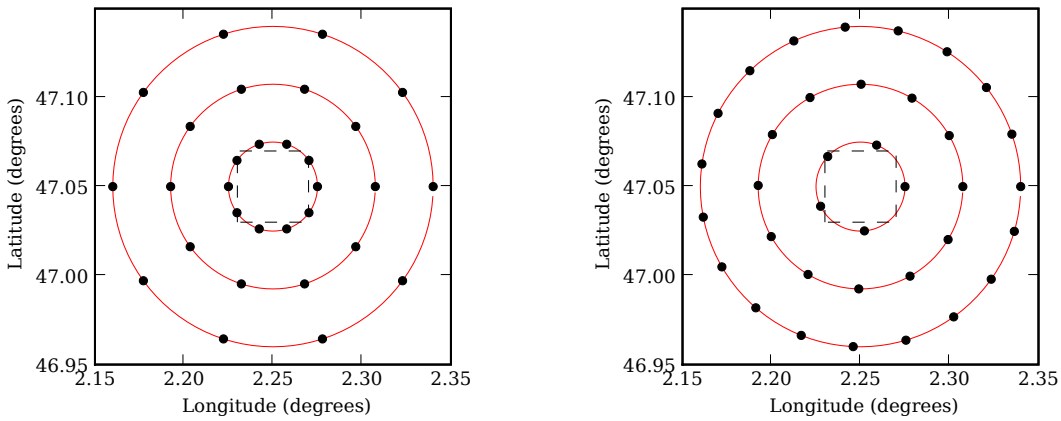
(a) P_{max}, regional scale

(b) t_{sat}, regional scale

(c) P_{max}, local scale

(d) t_{sat}, local scale

Figure B.8: Performance of the networks defined with the sensors having the highest detection probabilities against the number of sensors, for all meteorological situations, at local and regional scale. The species is iodine.



(a) Sensors position for 10 sensors per circle ($N^0_{sensor} = 10$.)

(b) Sensors position for a constant distance between 2 sensors equal to 3 km.

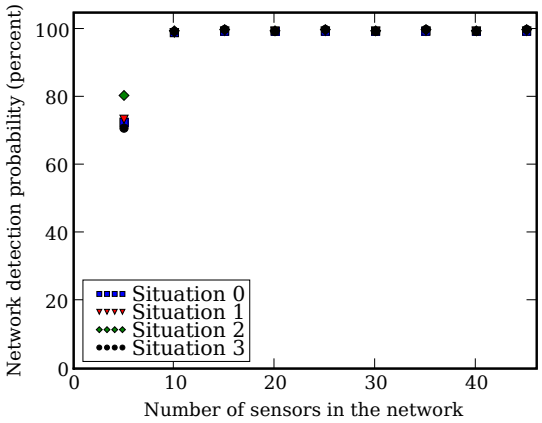
Figure B.9: Sensors positions for two network configurations with 3 concentric circles of radii 2.5 km, 5.75 km and 9 km. The dashed lines represent the boundaries of the protected domain D_p.

and one with a given l_{sensor}. To assess the performance and the robustness of such networks, we study the sensitivity to variations of three parameters: the number of circles, the number of sensors per circle, and the distance between two sensors. The circles are of radii comprised between $R_{min} = 2.5$ km and $R_{max} = 9$ km. The circles center is the center of D_p. Figure B.10 shows the results for the following variations:

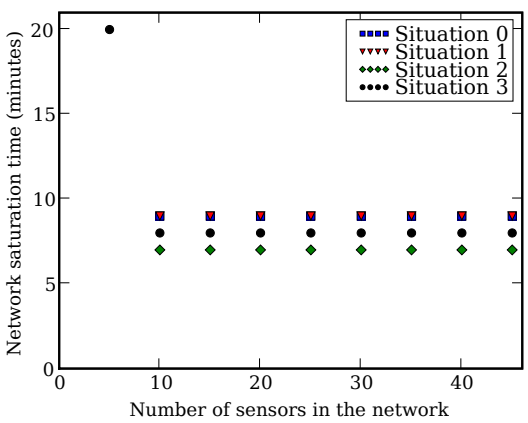
- Figures B.10(a) and B.10(b): the number of circles N_{circle} varies between 1 and 9 circles, with a number of sensors per circle equal to $N^0_{sensor} = 5$,
- Figures B.10(c) and B.10(d): the number of sensors per circle N^0_{sensor} varies between 1 and 14 sensors, with a number of circles equal to $N_{circle} = 3$,
- Figures B.10(e) and B.10(f): the distance between two sensors l_{sensor} varies betweent 2 km and 14 km, with a number of circles equal to $N_{circle} = 3$.

The evolution of P_{max} and t_{sat} is represented against the total number of sensors in the network, given by Equation B.12 in the first two cases, and by Equation B.13 otherwise. This allows to compare the performance in several configurations for a given number of sensors. It can hereby be observed that the detection probabilities reach 100% in all the configurations for 10 sensors or more. The detection probabilities also are very insensitive to the meteorological situation. Although the saturation times are more sensitive to the meteorology, the variation is very small with 12 sensors or more, which demonstrates a better robustness than the networks studied in Section B.5.1. The results are not sensitive to N_{circle}, as long as there are at least two circles (Figures B.10(a) and B.10(b)), and a large enough number of sensors per circle (here, 5 sensors). Indeed, the networks performance is much lower with less than 4 sensors per circle (Figures B.10(c) and B.10(d)). With more than 4 sensors, one may have at least one sensor placed in each direction, preventing the dependence to the wind direction, and allowing a sufficient scatter of the sensors. Placing the sensors at constant intervals induces slightly better performance (Figures B.10(e) and B.10(f)).

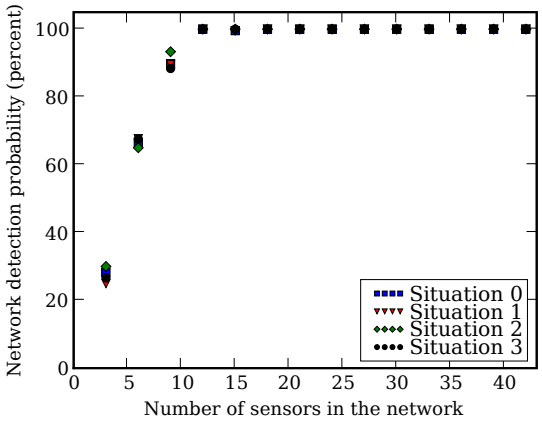
Whatever constraint may be put on the circles, the sensors are scattered throughout the domain and they are therefore very complementary, which enhances the network efficiency.



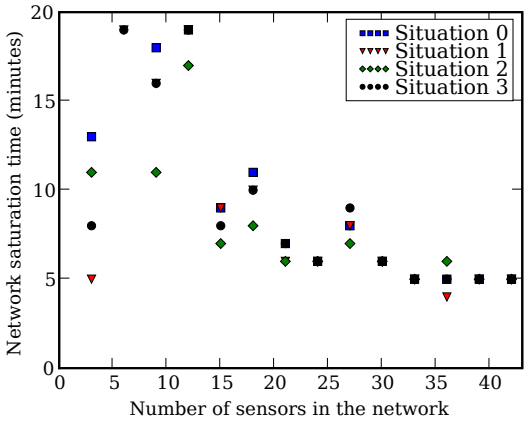
(a) P_{max} against $N_{sensor} = 5 \times N_{circle}$. N_{circle} varies between 1 and 9.



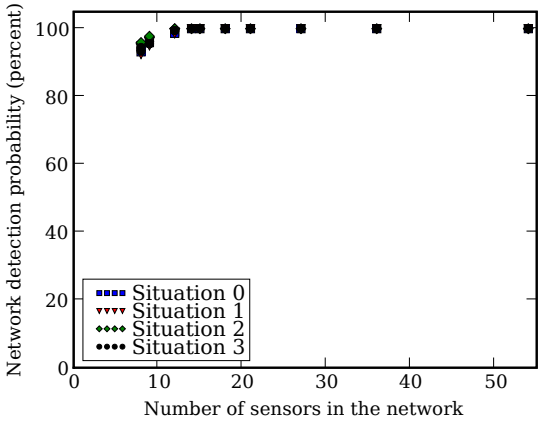
(b) t_{sat} against $N_{sensor} = 5 \times N_{circle}$. N_{circle} varies between 1 and 9.



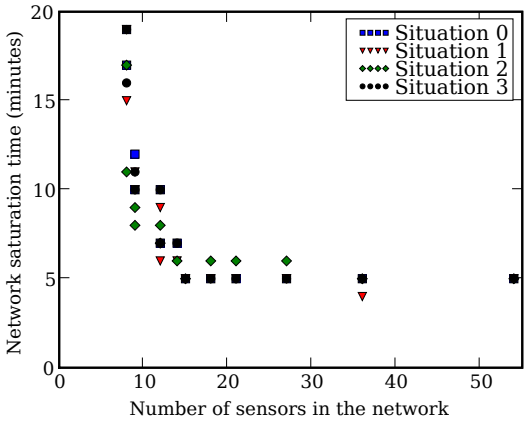
(c) P_{max} against $N_{sensor} = 3 \times N^0_{sensor}$. N^0_{sensor} varies between 1 and 14.



(d) t_{sat} against $N_{sensor} = 3 \times N^0_{sensor}$. N^0_{sensor} varies between 1 and 14.



(e) P_{max} against N_{sensor} given by Eq. B.13 ($N_{circle} = 3$). l_{sensor} varies between 2 and 15 km.



(f) t_{sat} against N_{sensor} given by Eq. B.13 ($N_{circle} = 3$). l_{sensor} varies between 2 and 15 km.

Figure B.10: Networks detection probability and saturation time with respect to N_{sensor} for several configurations in circles. The circles radii are comprised between 2.5 km and 9 km. In B.10(a) and B.10(b), the number of circles varies, for 5 sensors per circle. In B.10(c) and B.10(d), the number of sensors per circle varies, for 3 circles. In B.10(e) and B.10(f), the distance between two sensors l_{sensor} varies, for 3 circles.

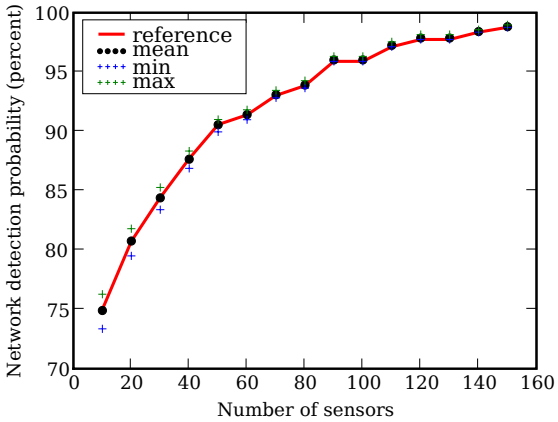
However, it is noteworthy that the performance is less dependent on the actual constraints on the circles than on the minimum number of sensors. One conclusion is that this method is efficient if enough sensors are available—otherwise, it is advocated to place the sensors according to their individual detection probabilities, like in Section B.5.1. Another conclusion is that other construction schemes, based on geometric constraints, may be efficient. For instance, simply enforcing a minimum distance between the sensors, without any shape constraint, could be an efficient scheme, combined with a selection based on the detection probability. While considering the inclusion of a new sensor in the network, the detection probability may also be redefined so as to account only for the sources not detected by the network (i.e., the sources already detected would be discarded in Equation B.7).

B.5.3 Networks sensitivity to wind variations

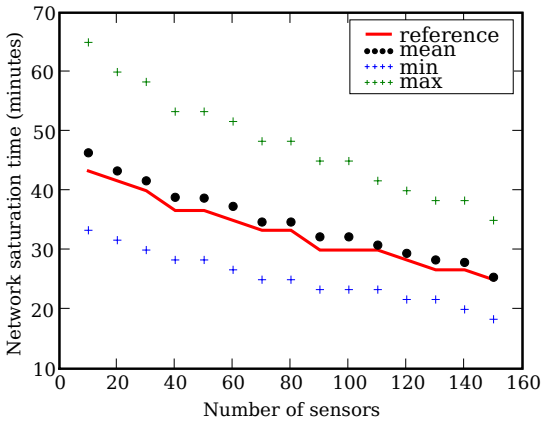
The effect of small variations in wind speed and direction is studied, to determine the networks sensitivity to meteorological measurement errors or to small variations in time. For that purpose, meteorological case 1 is taken as a reference, and we study variations of wind speed v between -35% and $+35\%$, with a 5% step, which should account for measurement errors. In the same way, the wind angle varies from $-25°$ to $+25°$ with a $5°$ step. The networks studied in Section B.5.1 for situation 1 are evaluated for each new wind speed or angle variation. The networks defined in Section B.5.2 would of course be insensitive to angle variations, although their performance could be affected by changes in the wind speed. Figure B.11 shows the influence of wind speed variations at local and regional scale. The mean (of P_{max} or t_{sat}) on all meteorological situations is shown, along with the range of variations (minimum and maximum values on all perturbed simulations). At both scales, the detection probability P_{max} is scarcely influenced by the wind speed variations. At local scale, t_{sat} increases by 40% when the wind speed decreases by 35%, for $N_{\text{sensor}} = 10$. The saturation time is, once again, the more sensitive parameter, especially at local scale. Figure B.12 shows that variations in the wind angle has similar effects. However, P_{max} is also affected by the variations at local scale: it may decrease by up to 10%. This sensitivity study is typical of the evaluation to be carried out in practical cases for a given network configuration and a given scenario. It shows the robustness of the networks to wind variations, especially in terms of detection probability, even when they are not placed in concentric circles.

B.6 Conclusions

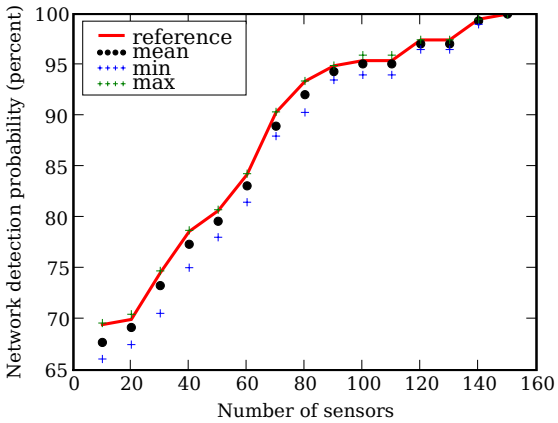
We investigated the use of monitoring networks for protection purposes, in a military or homeland security context, without constraint on the source location. Two scenarios were addressed, a buildings protection scenario and a small-units protection scenario. Indicators aimed at representing the protection ability of such networks, accounting for all possible harmful sources, were designed—the detection probability and the saturation time. They were evaluated for several meteorological situations and emission characteristics. The detection probability proves to be a reliable indicator, whereas the saturation time is too dependent on the dispersion conditions and the emitted quantity. Although it provides some measure of the behavior with respect to time, other indicators should also be tested, such as the averaged value of the arrival times on all sources (instead of the maximum value). In cases when the time of response is critical, one may also impose a minimum time between the source detection and the actual contamination of D_p. This study also highlights the need for appropriate sizes for the simulation domain, so as to ensure that all effective sources are taken into account for a given emitted quantity and meteorological situation. However, the networks should accommodate meteorological variations



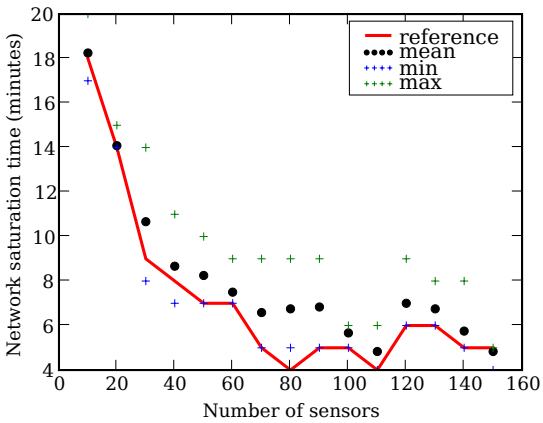
(a) Network detection probability – local scale



(b) Network saturation time – local scale

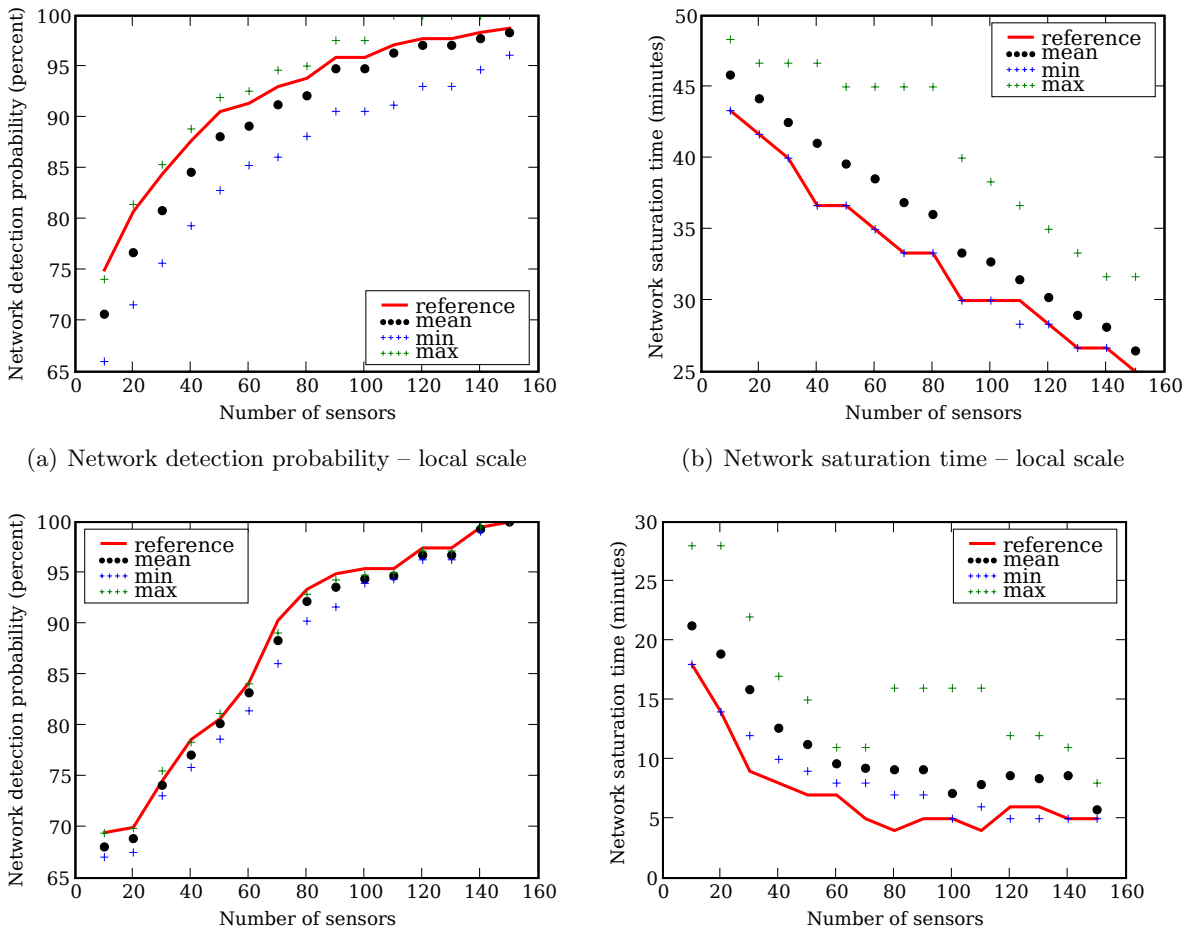


(c) Network detection probability – regional scale



(d) Network saturation time – regional scale

Figure B.11: Influence of wind speed variations: evolution of the mean detection probability and saturation time with a wind speed variation between −35% and +35%, at local and regional scale. Reference is the situation 1. Networks are defined with the sensors having the highest detection probability.



(a) Network detection probability – local scale

(b) Network saturation time – local scale

(c) Network detection probability – regional scale

(d) Network saturation time – regional scale

Figure B.12: Influence of wind angle variations: variations of detection probability and saturation time with a wind speed variation between $-25°$ and $+25°$ at local and regional scale. Reference is the situation 1. Networks are defined with the sensors having the highest detection probability.

(wind speed and direction, rain) without a total loss of performance, which ensures that a few well-chosen meteorological situations should be representative of a network protection ability. Finally, two methods to design an efficient network were evaluated: (1) taking the best sensors according to the detection probability, and (2) placing the sensors into circles around the protected domain. Placing the best sensors according to the saturation time gave poor networks performance, and was therefore not further investigated. While the first method depends on the wind direction, which makes it more suitable for the units protection scenario (real-time sensors siting), the second gives good results for the protection of larger areas containing buildings. Imposing a sufficient scattering of the sensors is a critical condition to ensure complementarity between the sensors and thus good performance for a reasonable number of sensors. This can be achieved by means of indicators coupling (e.g., choosing the sensors with constraints on both the detection probability and the saturation time), by imposing a minimum distance between two sensors, or by discarding any source already detected by a network when computing the statistics of a sensor to be included. A good scattering is also ensured by designing the network with concentric circles, if enough sensors are available.

The next step of such a study could be a network optimization that would maximize a statistical indicator on all sources and meteorological situations. Another perspective is the optimization of networks for the characterization of a source (place, date and quantity) through inversion methods. Finally, it would also be useful to extend this study to the prediction of the zones contaminated by a given source.

Bibliographie

Abida, R., Bocquet, M., Vercauteren, N. et Isnard, O. (2008). Design of a monitoring network over france in case of a radiological accidental release. *Atmos. Env.*, 42:5205–5219.

Arunachalam, S., Holland, A., Do, B. et Abraczinskas, M. (2006). A quantitative assessment of the influence of grid resolution on predictions of future-year air quality in North Carolina, USA. *Atmos. Env.*, 40:5010–5026.

Arya, S. (1999). *Air pollution meteorology and dispersion.* Oxford University press.

Auger, L. et Legras, B. (2007). Chemical segregation by heterogeneous emissions. *Atmos. Env.*, 41:2303–2318.

Barzyk, T. M., George, B. J., Vette, A. F., Williams, R. W., Croghan, C. W. et Stevens, C. D. (2009). Development of a distance-to-roadway proximity metric to compare near-road pollutant levels to a central site monitor. *Atmos. Env.*, 43(4):787–797.

Belot, Y., Caput, C. et Guenot, J. (1988). Etude bibliographique du lavage par la pluie des radionucléides particulaires et gazeux émis en situation accidentelle. Rapport technique, IRSN et EDF.

Benestad, R. (2001). A comparison between two empirical downscaling strategies. *Int. J. Climatol.*, 21(13):1645–1668.

Benestad, R. (2002). Empirically downscaled temperature scenarios for northern Europe based on a multi-model ensemble. *Clim. Res.*, 21(2):105–125.

Benestad, R., Hanssen-Bauer, I. et Chen, D. (2008). *Empirical-statistical downscaling.* World Scientific Publishing.

Bocquet, M. (2005). Reconstruction of an atmospheric tracer source using the principle of maximum entropy II : applications. *Quart. J. Roy. Meteor. Soc.*, 131 part B(610).

Boutahar, J., Lacour, S., Mallet, V., Quélo, D., Roustan, Y. et Sportisse, B. (2004). Development and validation of a fully modular platform for numerical modelling of air pollution : POLAIR. *Int. J. Env. and Pollution*, 22(1/2):17–28.

Brandt, J. (1998). *Modelling Transport, Dispersion and Deposition of Passive Tracers from Accidental Releases.* Thèse de doctorat, National Environmental Research Institute.

Brandt, J., Bastrup-Birk, A., Christensen, J., Mikkelsen, T., Thykier-Nielsen, S. et Zlatev, Z. (1998). Testing the importance of accurate meteorological input fields and parameterizations in atmospheric transport modelling using DREAM – validation against ETEX-I. *Atmos. Env.*, 32(24):4167–4186.

Brandt, J., Christensen, J. et Frohn, L. (2002). Modelling transport and deposition of caesium and iodine from the Chernobyl accident using the DREAM model. *Atmos. Chem. Phys.*, 2:397–417.

Brandt, J., Mikkelsen, T., Thykier-Nielsen, S. et Zlatev, Z. (1996). Using a combination of two models in tracer simulations. *Mathl. Comput. Modelling*, 23(10):99–115.

Bretherton, C., Smith, C. et Wallace, J. (1992). An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, 5(6):541–560.

Briant, R. (2009). Modélisation de l'impact des émissions du trafic routier sur la qualité de l'air. Rapport technique, CEREA.

Briggs, G. (1971). Some recent analyses of plume rise observations. *In* Englund, H.M. and Berry, W.T., éditeur : *Proceedings of the Second International Clean Air Congress*, pages 1029–1032, New York. Academic Press.

Briggs, G. A. (1969). Plume rise. Rapport technique T/D 25075, US Atomic Energy Commission Critical Review Series.

Brucher, W., Kessler, C., Kerschgens, M. et Ebel, A. (2000). Simulation of traffic-induced air pollution on regional to local scales. *Atmos. Env.*, 34(27):4675–4681. 8th International Symposium on Transport and Air Pollution, GRAZ, AUSTRIA, MAY 31-JUN 02, 1999.

Brummage, K. G. (1968). The calculation of atmospheric dispersion from a stack. *Atmos. Env.*, 2:197–224. Stichting Concawe study.

Byun, D. et Schere, K. L. (2006). Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.*, 59(1-6):51–77.

Canepa, E., Dallorto, L. et Ratto, C. (2000). About the plume rise description in the dispersion code SAFE_AIR. *Int. J. Env. and Pollution*, 14(1–6):235–245.

CERC (2007). ADMS4 Flat terrain validation : Kincaid, Indianapolis and Prairie Grass. Rapport technique, Cambridge Environmental Research Consultants. 15 pp.

Chang, J. et Hanna, S. (2004). Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, 87:167–196.

Chatfield, R. et Brost, R. (1987). A two-stream model of the vertical transport of trace species in the convective boundary layer. *J. Geophys. Res.*, 92:13263–13276.

Ching, J., Herwehe, J. et Swall, J. (2006). On joint deterministic grid modeling and sub-grid variability conceptual framework for model evaluation. *Atmos. Env.*, 40:4935–4945.

Chock, D., Winkler, S. et Sun, P. (2002). Effect of grid resolution and subgrid assumptions on the model prediction of a reactive bouyant plume under convective conditions. *Atmos. Env.*, 36(29):4649–4662.

Cimorelli, A., Perry, S., Venkatram, A., Weil, J., Paine, R., Wilson, R., Lee, R., Peters, W., Brode, R. et Paumier, J. (2004). AERMOD : Description of model formulation. Rapport technique EPA-454/R-03-004, EPA.

Cohan, D., Hu, Y. et Russell, A. (2006). Dependence of ozone sensitivity analysis on grid resolution. *Atmos. Env.*, 40(1):126–135.

Constantinescu, E. M., Sandu, A. et Carmichael, G. R. (2008). Modeling atmospheric chemistry and transport with dynamic adaptive resolution. *Comput. Geosci.*, 12(2):133–151.

Cook, R. D. (2007). Fisher Lecture : Dimension Reduction in Regression. *Stat. Sci.*, 22(1):1–26.

Corrsin, S. (1974). Second 25 years of statistical-theory of turbulent-diffusion. *Trans. Am. Geophys. Union*, 55(3):135.

Crane, R. et Hewitson, B. (1998). Doubled CO2 precipitation changes for the susquehanna basin : Down-scaling from the genesis general circulation model. *Int. J. Climatol.*, 18(1):65–76.

Cuijpers, J. et Holtslag, A. (1998). Impact of skewness and nonlocal effects on scalar and buoyancy fluxes in convective boundary layers. *J. Atmos. Sci.*, 55:151–162.

Davoine, X. et Bocquet, M. (2007). Inverse modelling-based reconstruction of the Chernobyl source term available for long-range transport. *Atmos. Chem. Phys.*, 7(6):1549–1564.

Deardorff, J. (1970). Convective velocity and temperature scales for the unstable planetary boundary layer and rayleigh convection. *J. Atmos. Sci.*, 27:1211–1213.

Deguillaume, L., Beekmann, M. et Derognat, C. (2008). Uncertainty evaluation of ozone production and its sensitivity to emission changes over the Ile-de-France region during summer periods. *J. Geophys. Res.*, 113(D2).

Demaël, E. et Carissimo, B. (2008). Comparative evaluation of an eulerian CFD and gaussian plume models based on prairie grass dispersion experiment. *J. Applied Meteor.*, 47(3):888–900.

Doury, A. (1976). Une méthode de calcul pratique et générale pour la prévision numérique des pollutions véhiculées par l'atmosphère. Rapport technique 4280, CEA, France.

EPA (1991). Guideline for regulatory application of the urban airshed model. Rapport technique EPA-450/4-91-013, US EPA.

EPA (2005). Guidance on the use of models and other analyses in attainment demonstrations for the 8-hr ozone NAAQS. Rapport technique EPA-450/R-99-004, US EPA.

EPA (2009). Primary national ambient air quality standard for nitrogen dioxide. Rapport technique EPA-HQ-OAR-2006-0922, US EPA.

Esler, J. (2003). An integrated approach to mixing sensitivities in tropospheric chemistry : a basis for the parameterization of subgrid-scale emissions for chemistry transport models. *J. Geophys. Res.*, 108(D20).

Fraigneau, Y., Gonzalez, M. et Coppalle, A. (1996). The influence of turbulence upon the chemical reaction of nitric oxide released from a ground source into ambient ozone. *Atmos. Env.*, 30(9):1467–1480.

Fuentes, M., Chaudhuri, A. et Holland, D. M. (2007). Bayesian entropy for spatial sampling design of environmental data. *Environ. Ecol. Stat.*, 14:323–340.

Galmarini, S., Vinuesa, J. et Martilli, A. (2008). Modeling the impact of sub-grid scale emission variability on upper-air concentration. *Atmos. Chem. Phys.*, 8:141–158.

Garaud, D. et Mallet, V. (2009a). Automatic calibration of an ensemble for uncertainty estimation. *J. Geophys. Res.* In preparation.

Garaud, D. et Mallet, V. (2009b). Automatic generation of large ensembles for air quality forecasting using the polyphemus system. *Geosci. Model Dev. Discuss.*, 2(2):889–933.

Garratt, J. (1992). *The atmospheric boundary layer.* Cambridge University Press – Atmospheric and Space Sciences Series.

Gauthier, A. (1999). *Analyse d'une modélisation lagrangienne des panaches dans un modèle régional de pollution atmosphérique.* Thèse de doctorat, Mines de Paris.

Ghorai, S., Tomlin, A. et Berzins, M. (2000). Resolution of pollutant concentrations in the boundary layer using a fully 3D adaptive gridding technique. *Atmos. Env.*, 34(18):2851–2863.

Gilbert, N., Woodhouse, S., Stieb, D. et Brook, J. (2003). Ambient nitrogen dioxide and distance from a major highway. *Sci. Total Environ.*, 312(1–3):43–46.

Gillani, N. (1986). Ozone formation in pollutant plumes : Development and application of a reactive plume model with arbitrary crosswind resolution. Rapport technique EPA-600/S3-86-051, US Environmental Protection Agency, Research Triangle Park, N.C.

Gillani, N. et Godowitch, J. (1999). Plume-in-grid treatment of major point source emissions. *In* Byun, D. et Ching, J., éditeurs : *Science algorithms of the EPA models-3 Community Multi-scale Air Quality (CMAQ) modeling system*, chapitre 9, pages 9.1–9.37. EPA.

Golder, D. (1972). Relations among stability parameters in the surface layer. *Boundary-Layer Meteor.*, 3:47–58.

Gonzalez, M. (1997). Analysis of the effect of microscale turbulence on atmospheric chemical reactions by means of the pdf approach. *Atmos. Env.*, 31(4):575–586.

Grotch, S. et MacCracken, M. (1991). The use of general-circulation models to predict regional climatic-change. *J. Climate*, 4(3):286–303.

Guillas, S., Bao, J., Choi, Y. et Wang, Y. (2008). Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta. *Atmos. Env.*, 42(6):1338–1348.

Guillas, S., Rhomari, N. et Zhang, J. (2000). Bilan de l'existant en matière de prévision statistique des pics de pollution. Rapport technique, Université Paris VI.

Guillas, S., Tiao, G., Wuebbles, D. et Zubrow, A. (2006). Statistical diagnostic and correction of a chemistry-transport model for the prediction of total column ozone. *Atmos. Chem. Phys.*, 6:525–537.

Han, J. et Byun, D. (2005). An improvement of the two-stream model for vertical mixing of passive tracer in the convective boundary layer. *Atmos. Env.*, 39:1775–1788.

Hanna, S. (1984). Applications in air pollution modeling. *In* Niewstadt, F. et van Dop, H., éditeurs : *Atmospheric turbulence and Air Pollution modeling*, pages 275–309. D. Reidel Publishing Co.

Hanna, S., Briggs, G. et Hosker Jr., R. (1982). *Handbook on atmospheric diffusion.* Technical Information center, US Department of energy.

Hanna, S., Hansen, O. et Dharmavaram, S. (2004). FLACS CFD air quality model performance evaluation with Kit Fox, MUST, Prairie Grass, and EMU observations. *Atmos. Env.*, 38:4675–4687.

Hanna, S. et Paine, R. (1989). Hybrid plume dispersion model (HPDM) development and evaluation. *J. Applied Meteor.*, 28:206–224.

Hanna, S. R., Lu, Z., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M. et Hansen, D. A. (2001). Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain. *Atmos. Env.*, 35(5):891–903.

Hessami, M., Gachon, P., Ouarda, T. B. M. J. et St-Hilaire, A. (2008). Automated regression-based statistical downscaling tool. *Env. Mod. Soft.*, 23(6):813–834.

Hewitt, C. (1991). Spatial variations in nitrogen dioxide concentrations in an urban area. *Atmos. Env.*, 25(B):429–434.

Holland, J. Z. (1953). A meteorological survey of the oak ridge area. Rapport technique ORO-99, US AEC, Oak Ridge, Tennessee.

Honore, C., Vautard, R. et Beekmann, M. (2000). Low and high NOx chemical regimes in an urban environment. *Env. Mod. Soft.*, 15(6-7, Sp. Iss. SI):559–564. Meeting on Air Pollution Modeling and Simulation, PARIS, FRANCE, OCT, 1998.

Horowitz, L. W., Walters, S., Mauzerall, D. L., Emmons, L. K., Rasch, P. J., Granier, C., Tie, X., Lamarque, J.-F., Schultz, M. G., Tyndall, G. S., Orlando, J. J. et Brasseur, G. P. (2003). A global simulation of tropospheric ozone and related tracers : description and evaluation of MOZART, version 2. *J. Geophys. Res.*, 108(D24).

Huth, R. (1999). Statistical downscaling in central Europe : evaluation of methods and potential predictors. *Clim. Res.*, 13(2):91–101.

Irwin, J. (1979). Scheme for estimating dispersion parameters as a function of release height. Rapport technique 600 :4-79-062, US EPA. 26 pp.

Isakov, V., Touma, J. S., Burke, J., Lobdell, D. T., Palma, T., Rosenbaum, A. et Ozkaynak, H. (2009). Combining Regional- and Local-Scale Air Quality Models with Exposure Models for Use in Environmental Health Studies. *J Air Waste Manage Assoc.*, 59(4):461–472.

Isakov, V. et Venkatram, A. (2006). Resolving neighborhood scale in air toxics modeling : A case study in Wilmington, CA. *J Air Waste Manage Assoc.*, 56(5):559–568.

Jacob, D. J. (1999). *Introduction to atmospheric chemistry.* Princeton University Press.

Jacobson, M. et Seinfeld, J. (2004). Evolution of nanoparticle size and mixing state near the point of emission. *Atmos. Env.*, 38(13):1839–1850.

Jakobs, H., Feldmann, H., Hass, H. et Memmesheimer, M. (1995). The use of nested models for air-pollution studies - An application of the EURAD model to a SANA episode. *J. Applied Meteor.*, 34(6):1301–1319.

Jiang, Z., de Bruin, S., Heuvelink, G. et Twenhöfel, C. (2007). Optimization of mobile radioactivity monitoring networks. *In Fifth International Symposium on Spatial Data Quality.*

Karamchandani, P., Koo, A. et Seigneur, C. (1998). Reduced gas-phase kinetics mechanism for atmospheric plume chemistry. *Environ. Sci. Technol.*, 32:1709–1720.

Karamchandani, P., Lohman, K. et Seigneur, C. (2009). Using a sub-grid scale modeling approach to simulate the transport and fate of toxic air pollutants. *Environ. Fluid Mech.*, 9(1, Sp. Iss. SI):59–71.

Karamchandani, P., Santos, L., Sykes, I., Zhang, Y., Tonne, C. et Seigneur, C. (2000). Development and evaluation of a state-of-the-science reactive plume model. *Environ. Sci. Technol.*, 34:870–880.

Karamchandani, P., Seigneur, C., Vijayaraghavan, K. et Wu, S. (2002). Development and application of a state-of-the-science plume-in-grid model. *J. Geophys. Res.*, 107(D19).

Karamchandani, P., Vijayaraghavan, K., Yen, S., Seigneur, C. et Edgerton, E. (2006). Plume-in-grid modeling for particulate matter. *Atmos. Env.*, 40(38):7280–7297.

Kidson, J. et Thompson, C. (1998). A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *J. Climate*, 11(4):735–753.

Kim, Y., Sartelet, K. et Seigneur, C. (2009). Comparison of two gas-phase chemical kinetic mechanisms of ozone formation over Europe. *J. Atmos. Chem.* Submitted.

Kolmogorov, A. (1991). The local-structure of turbulence in incompressible viscous-fluid for very large Reynolds numbers. *Proc. R. Soc. A*, 434:9–13.

Korsakissok, I. et Mallet, V. (2009). Comparative study of Gaussian dispersion formulas within the Polyphemus platform : evaluation with Prairie Grass and Kincaid experiments. *J. Applied Meteor.*, 48(12):2459–2473.

Korsakissok, I. et Mallet, V. (2010a). Development and application of a reactive plume-in-grid model : evaluation over Greater Paris. *Atmos. Chem. Phys. Discuss.*, 10(2):5091–5134.

Korsakissok, I. et Mallet, V. (2010b). Subgrid-scale treatment for major point sources in an Eulerian model : A sensitivity study on the European Tracer Experiment (ETEX) and Chernobyl cases. *J. Geophys. Res.*, 115(D03303).

Korsakissok, I., Mallet, V., Quélo, D. et Sportisse, B. (2006). Modeling of dispersion and scavenging in the polyphemus platform. Applications to passive tracers. Rapport technique 30, CEREA.

Korsakissok, I., Mallet, V. et Sportisse, B. (2010). Assessment of the detection abilities of monitoring networks for passive tracers at local and regional scales. *Atmos. Env.*, 44(3):408–420.

Krol, M., Molemaker, M. et Vilà-Guerau de Arellano, J. (2000). Effects of turbulence and heterogeneous emissions on photochemically active species in the convective boundary layer. *J. Geophys. Res.*, 105:6871–6884.

Krysta, M., Bocquet, M., Sportisse, B. et Isnard, O. (2006). Data assimilation for short-range dispersion of radionuclides : An application to wind tunnel data. *Atmos. Env.*, 40:7267–7279.

Kumar, N. et Russell, A. (1996). Development of a computationally efficient, reactive subgrid-scale plume model and the impact in the northern United States using increasing levels of chemical details. *J. Geophys. Res.*, 101:16737–16744.

Lagzi, I., Karman, D., Turanyi, T., Tomlin, A. et Haszpra, L. (2004). Simulation of the dispersion of nuclear contamination using an adaptive Eulerian grid model. *J. Environ. Radioactivity*, 75(1):59–82.

Lagzi, I., Turanyi, T., Tomlin, A. S. et Haszpra, L. (2009). Modelling photochemical air pollutant formation in Hungary using an adaptive grid technique. *Int. J. Env. and Pollution*, 36(1-3):44–58.

Liang, J. et Jacobson, M. (2000). Effects of subgrid segregation on ozone production efficiency in a chemical model. *Atmos. Env.*, 34(18):2975–2982.

Lipphardt, M. (1997). *Prise en compte d'une source ponctuelle dans un modèle régional de pollution atmosphérique.* Thèse de doctorat, Grenoble I.

Louis, J.-F. (1979). A parametric model of vertical eddy fluxes in the atmosphere. *Boundary-Layer Meteor.*, 17:187–202.

Luhar, A. et Sawford, B. (2005). Micromixing modelling of mean and fluctuating scalar fields in the convective boundary layer. *Atmos. Env.*, 39(35):6673–6685.

Lumley, J. et Panofsky, H. (1964). *The Structure of Atmospheric Turbulence.* Wiley-Interscience.

Mallet, V. (2005). *Estimation de l'incertitude et prévision d'ensemble avec un modèle de chimie-transport. Application à la simulation numérique de la qualité de l'air.* Thèse de doctorat, ENPC.

Mallet, V., Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M. et Foudhil, H. (2007). Technical Note : The air quality modeling system Polyphemus. *Atmos. Chem. Phys.*, 7(20):5479–5487.

Mallet, V. et Sportisse, B. (2005). A comprehensive study of ozone sensitivity with respect to emissions over Europe with a chemistry-transport model. *J. Geophys. Res.*, 110(D22).

Maryon, R. et Buckland, A. (1995). Tropospheric dispersion : The first ten days after a puff release. *Quart. J. Roy. Meteor. Soc.*, 121(528, Part B):1799–1833.

Mastorakos, E. (2003). The conditional moment closure approach for atmospheric pollution problems. *In Air Pollution XI*, pages 125–134. WIT Press.

Moeng, C. et Wyngaard, J. (1984). Statistics of conservative scalars in the convective boundary layer. *J. Atmos. Sci.*, 41:3161–3169.

Molemaker, M. et Vilà-Guerau de Arellano, J. (1998). Turbulent control of chemical reactions in the convective boundary layer. *J. Atmos. Sci.*, 55:568–579.

Morris, R., Yocke, M., Myers, T. et Kessler, R. (1991). Development and testing of UAM-V : A nested-grid version of the urban airshed model. *In Proceedings of the AWMA conference : Tropospheric ozone and the environment II*, Pittsburgh, PA.

Mosca, S., Graziani, G., Klug, W., Bellasio, R. et Bianconi, R. (1998). A statistical methodology for the evaluation of long-range dispersion models : An application to the ETEX exercise. *Atmos. Env.*, 32(24):4307–4324.

Murphy, J. (1999). An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Climate*, 12(8):2256–2284.

Najac, J., Boe, J. et Terray, L. (2009). A multi-model ensemble approach for assessment of climate change impact on surface winds in France. *Clim. Res.*, 32(5):615–634.

Nasstrom, J. et Pace, J. (1998). Evaluation of the effect of meteorological data resolution on Lagrangian particle dispersion simulations using the ETEX experiment. *Atmos. Env.*, 32(24):4187–4194.

Nodop, K., Connolly, R. et Girardi, F. (1998). The field campaigns of the European Tracer Experiment (ETEX) : Overview and results. *Atmos. Env.*, 32(24):4095–4108.

Nychka, D. et Saltzman, N. (1998). Design of air quality monitoring networks. *In* Nychka, D., Cox, L. et Piegorsch, W., éditeurs : *Case Studies in Environmental Statistics*, Lecture Notes in Statistics 120, pages 51–76. Springer-Verlag.

Odman, M. et Khan, M. (2004). Evaluation of the adaptive grid air pollution model. *In* Borrego, C and Incecik, S, éditeur : *Air pollution modeling and its application XVI*, volume 16 de *NATO - Challenges of modern society*, pages 635–636. 26th NATO/CCMS International Technical Meeting on Air Pollution Modelling and Its Application, Istanbul, TURKEY, MAY, 2003.

Odman, M. et Russel, A. (1991). Multiscale modeling of pollutant transport and chemistry. *J. Geophys. Res.*, 96:7363–7370.

Olesen, H. (1995). The model validation exercise at mol : overview of results. *Int. J. Env. and Pollution*, 5(4–6):761–784.

Olesen, H. et Chang, J. (2005). Consolidating tools for model evaluation. *In Proceedings of the 10th international conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, pages 263–267, Sissi, Crete.

Panofsky, H. et Dutton, J. (1984). *Atmospheric Turbulence.* Wiley.

Pasquill, F. (1961). The estimation of the dispersion of windborne material. *Meteorol. Mag.*, 90(1063):33–49.

Perkins, R., Soulhac, L., Mejean, P. et Rios, I. (2002). Modélisation de la dispersion des émissions atmosphériques d'un site industriel – vers un guide de l'utilisateur. 1ère partie : état de l'art. Rapport technique, Laboratoire de Mécanique des Fluides et d'Acoustique. UMR CNRS 5509 – UCBL1 – ECL.

Pleijel, H., Karlsson, G. et Gerdin, E. (2004). On the logarithmic relationship between NO2 concentration and the distance from a highroad. *Sci. Total Environ.*, 332(1–3):261–264.

Pope, S. (2000). *Turbulent flows.* Cambridge University Press.

Promeyrat-Qotbi, S. (2001). *Contribution a l'étude de la pollution atmosphérique autoroutière. Caractérisation des flux de déposition et de la contamination metallique de l'environnement de proximité.* Thèse de doctorat, Université de Metz.

Pryor, S., Schoof, J. et Barthelmie, R. (2005). Climate change impacts on wind speeds and wind energy density in northern Europe : empirical downscaling of multiple AOGCMs. *Clim. Res.*, 29(3):183–198.

Pryor, S. C. (1998). A case study of emission changes and ozone responses. *Atmos. Env.*, 32(2):123–131.

Quélo, D., Krysta, M., Bocquet, M., Isnard, O., Minier, Y. et Sportisse, B. (2007). Validation of the Polyphemus platform on the ETEX, Chernobyl and Algeciras cases. *Atmos. Env.*, 41(26):5,300–5,315.

Ramaker, J. (2003). The final test a history of the comprehensive nucelar-test-ban treaty negotiations. Rapport technique, Provisional Technical Secretariat of the Preparatory Commission for the CTBTO.

Reponen, T., Grinshpun, S., Trakumas, S., Martuzevicius, D., Wang, Z., LeMasters, G., Lockey, J. et Biswas, P. (2003). Concentration gradient patterns of aerosol particles near interstate highways in the Greater Cincinnati airshed. *J. Environ. Monit.*, 5:557–562.

Rodes, C. et Holland, D. (1981). Variations of NO, NO2 and O3 concentrations downwind of a Los Angeles freeway. *Atmos. Env.*, 15:243–250.

Ryall, D. et Maryon, R. (1998). Validation of the UK Met. Office's NAME model against the ETEX dataset. *Atmos. Env.*, 32(24):4265–4276.

Sagaut, P. (2002). *Large Eddy Simulation for Incompressible Fows. An Introduction.* Springer.

Schneider, C., Kessler, C. et Moussiopoulos, N. (1997). Influence of emission input data on ozone level predictions for the Upper Rhine valley. *Atmos. Env.*, 31(19):3,187–3,205.

Seigneur, C., Tesche, T., Roth, P. et Liu, M. (1983). On the treatment of point source emissions in urban air quality modeling. *Atmos. Env.*, 17:1655–1676.

Seinfeld, J. H. et Pandis, S. N. (1998). *Atmospheric chemistry and physics : from air pollution to climate change.* Wiley-Interscience.

Sillman, S., Logan, J. et Wofsy, S. (1990). A regional scale model for ozone in the United States with subgrid representation of urban and power plant plumes. *J. Geophys. Res.*, 95:5731–5748.

Sillman, S., Vautard, R., Menut, L. et Kley, D. (2003). O-3-NOx-VOC sensitivity and NOx-VOC indicators in Paris : Results from models and atmospheric pollution over the Paris area (ESQUIF) measurements. *J. Geophys. Res.*, 108(D12).

Sorensen, J. (1998). Sensitivity of the DERMA long-range Gaussian dispersion model to meteorological input and diffusion parameters. *Atmos. Env.*, 32(24):4195–4206.

Soulhac, L., Puel, C., Duclaux, O. et Perkins, R. (2003). Simulations of atmospheric pollution in Greater Lyon an example of the use of nested models. *Atmos. Env.*, 37(37):5147–5156.

Spak, S., Holloway, T., Lynn, B. et Goldberg, R. (2007). A comparison of statistical and dynamical downscaling for surface temperature in North America. *J. Geophys. Res.*, 112(D8).

Sportisse, B. (2007a). *Pollution atmosphérique. Des processus à la modélisation.* Springer-Verlag.

Sportisse, B. (2007b). A review of parameterizations for modelling dry deposition and scavenging of radionuclides. *Atmos. Env.*, 41:2683–2698.

Sportisse, B. et Du Bois, L. (2002). Numerical and theoretical investigation of a simplified model for the parameterization of below-cloud scavenging by falling raindrops. *Atmos. Env.*, 36:5719–5727.

Srivastava, R., McRae, D. et Odman, M. (2001). Simulation of dispersion of a power plant plume using an adaptive grid algorithm. *Atmos. Env.*, 35(28):4801–4818.

Stein, A. F., Isakov, V., Godowitch, J. et Draxler, R. R. (2007). A hybrid modeling approach to resolve pollutant concentrations in an urban area. *Atmos. Env.*, 41(40):9410–9426.

Stewart, G. (1987). Collinearity and least squares regression. *Stat. Sci.*, 2:68–100.

Stockwell, W. R., Kirchner, F., Kuhn, M. et Seefeld, S. (1997). A new mechanism for regional atmospheric chemistry modeling. *J. Geophys. Res.*, 102(D22):25847–25879.

Stohl, A. (1998). Computation, accuracy and applications of trajectories - A review and bibliography. *Atmos. Env.*, 32(6):947–966.

Stohl, A. et Koffi, N. (1998). Evaluation of trajectories calculated from ECMWF data against constant volume balloon flights during ETEX. *Atmos. Env.*, 32(24):4151–4156.

Stohl, A., Wotawa, G., Seibert, P. et Krompkolb, H. (1995). Interpolation errors in wind fields as a function of spatial and temporal resolution and their impact on different types of kinematic trajectories. *J. Applied Meteor.*, 34(10):2149–2165.

Stümke, H. (1963). Vorschlag einer empirischen Formel für die Schornsteinüberhöhung. *Staub*, 23:549.

Sykes, I. (2000). Lagrangian puff dispersion modeling and uncertainty estimation using second-order closure. *In* Boybeyi, Z., éditeur : *Mesoscale atmospheric dispersion*, pages 257–285. WITpress.

Sykes, R. I., Parker, S., Henn, D., Cerasoli, C. et Santos, L. (2000). PC-SCIPUFF Version 1.3 Technical Documentation. Rapport technique ARAP Report No. 725, Titan Corporation, ARAP Group.

Taghavi, S. M., Carissimo, B., Sartelet, K., Muller, K., Korsakissok, I. et Milliez, M. (2009). Étude par modélisation de la dispersion et du dépôt de la pollution particulaire a proximité d'un axe routier, à l'aide du code de CFD atmosphérique Mercure_Saturne. Rapport technique 2009-11, CEREA.

Taylor, G. (1921). Diffusion by continuous movements. *Proc. Lond. Math. Soc.*, 20:196–211.

Thykier-Nielsen, S., Deme, S. et Mikkelsen, T. (1998). Description of the Atmospheric Dispersion Module RIMPUFF. Rapport technique RODOS(WG2)-TN(98)-02, Riso National Laboratory.

Tombette, M. et Sportisse, B. (2007). Aerosol modeling at a regional scale : Model-to-data comparison and sensitivity analysis over Greater Paris. *Atmos. Env.*, 41:6941–6950.

Troen, I. et Mahrt, L. (1986). A simple model of the atmospheric boundary layer ; sensitivity to surface evaporation. *Boundary-Layer Meteor.*, 37:129–148.

Turner, D. B. (1969). Workbook of atmospheric diffusion estimates. Rapport technique 999-AP-26, US EPA, Washington, DC.

Underwood, B. (2001). Review of deposition velocity and washout coefficient. Rapport technique, AEA Technology, Harwell.

UNSCEAR (2000). Annex J : exposure and effects of the Chernobyl accident. Rapport technique, UNSCEAR.

Valari, M. et Menut, L. (2008). Does an Increase in Air Quality Models' Resolution Bring Surface Ozone Concentrations Closer to Reality ? *J. Atmos. Oceanic Technol.*, 25(11):1955–1968.

Vautard, R., Maidi, M., Menut, L., Beekmann, M. et Colette, A. (2007). Boundary layer photochemistry simulated with a two-stream convection scheme. *Atmos. Env.*, 41:8275–8287.

Venkatram, A. et Horst, T. (2006). Approximating dispersion from a finite line source. *Atmos. Env.*, 40:2401–2408.

Venkatram, A., Isakov, V., Seila, R. et Baldauf, R. (2009). Modeling the impacts of traffic emissions on air toxics concentrations near roadways. *Atmos. Env.*, 43(20):3191–3199.

Venkatram, A., Strimaitis, D. et Cicristofaro, D. (1984). A semiempirical model to estimate dispersion of elevated releases in the stable boundary layer. *Atmos. Env.*, 18:923–928.

Vijayaraghavan, K., Karamchandani, P. et Seigneur, C. (2006). Plume-in-grid modeling of summer air pollution in Central California. *Atmos. Env.*, 40(26):5097–5109.

Vilà-Guerau de Arellano, J. et Cuijpers, J. (2000). The chemistry of a dry cloud : the effects of radiation and turbulence. *J. Atmos. Sci.*, 57:1573–1584.

Vilà-Guerau de Arellano, J., Dosio, A., Vinuesa, F., Holtslag, A. et Galmarini, S. (2004). The dispersion of chemically reactive species in the atmospheric boundary layer. *Meteor. Atmos. Phys.*, 87:23–38.

Vinuesa, J.-F. et Vilà-Guerau de Arellano, J. (2003). Fluxes and (co-)variances of reacting scalars in the convective boundary layer. *Tellus*, 55(B):935–949.

Vinuesa, J.-F. et Vilà-Guerau de Arellano, J. (2005). Introducing effective reaction rates to account for the inefficient mixing of the convective boundary layer. *Atmos. Env.*, 39:445–461.

von Storch, H. (1999). On the use of "inflation" in statistical downscaling. *J. Climate*, 12(12): 3505–3506.

von Storch, H., Zorita, E. et Cubasch, U. (1993). Downscaling of global climate-change estimates to regional scales - an application to Iberian rainfall in wintertime. *J. Climate*, 6(6):1161–1171.

Wang, Y., McElroy, M., Jacob, D. et Yantosca, R. (2004). A nested grid formulation for chemical transport over Asia : Applications to CO. *J. Geophys. Res.*, 109(D22).

Webster, H. et Thomson, D. (2002). Validation of a lagrangian model plume rise scheme using the kincaid data set. *Atmos. Env.*, 36(1–6):5031–5042.

Weil, J. (1988). Dispersion in the convective boundary layer. *In* Venkatram, A. et Wyngaard, J., éditeurs : *Lectures on Air Pollution modeling*. American Meteorological Society.

Weinroth, E., Stockwell, W., Koracin, D., Kahyaoglu-Koracin, J., Luria, M., McCord, T., Podnar, D. et Gertler, A. (2008). A hybrid model for ozone forecasting. *Atmos. Env.*, 42:7002–7012.

Wilby, R., Charles, S., Zorita, E., Timbal, B., Whetton, P. et Mearns, L. (2004). Guidelines for use of climate scenarios developed from statistical downscaling methods. Rapport technique, Data distribution center of the intergovernmental panel on climate change.

Wilby, R., Tomlinson, O. et Dawson, C. (2003). Multi-site simulation of precipitation by conditional resampling. *Clim. Res.*, 23(3):183–194.

Wilby, R., Wigley, T., Conway, D., Jones, P., Hewitson, B., Main, J. et Wilks, D. (1998). Statistical downscaling of general circulation model output : A comparison of methods. *Water Resour. Res*, 34(11):2995–3008.

Yamartino, R. (2008). *AIR QUALITY MODELING - Theories, Methodologies, Computational Techniques, and Available Databases and Software. Vol III-Special Issues.* EnviroComp Institute and the Air & Waste Management Association.

Zhu, Y., Hinds, W., Kim, S. et C., S. (2002). Concentration and size distribution of ultrafine particles near a major highway. *J. Environ. Monit.*, 52:1032–1042.