



HAL
open science

Quantification des sources de méthane en Sibérie par inversion atmosphérique à la méso-échelle

Antoine Berchet

► **To cite this version:**

Antoine Berchet. Quantification des sources de méthane en Sibérie par inversion atmosphérique à la méso-échelle. Sciences de la Terre. Université de Versailles-Saint Quentin en Yvelines, 2014. Français. NNT : 2014VERS0058 . tel-01222485

HAL Id: tel-01222485

<https://theses.hal.science/tel-01222485v1>

Submitted on 30 Oct 2015

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

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

ECOLE DOCTORALE SCIENCES DE L'ENVIRONNEMENT D'ÎLE-DE-FRANCE
MÉTÉOROLOGIE, OCÉANOGRAPHIE ET PHYSIQUE DE L'ENVIRONNEMENT

THÈSE

EN VUE DE L'OBTENTION DU GRADE DE

DOCTEUR

DE L'UNIVERSITÉ DE VERSAILLES

SAINT-QUENTIN-EN-YVELINES

(arrêté ministériel du 7 août 2006)

PRÉSENTÉE PUBLIQUEMENT

À GIF-SUR-YVETTE

LE 19 DÉCEMBRE 2014 PAR

ANTOINE BERCHET

QUANTIFICATION DES SOURCES DE MÉTHANE EN SIBÉRIE PAR INVERSION ATMOSPHÉRIQUE À LA MÉSO-ÉCHELLE

SOUTENUE DEVANT LE JURY COMPOSÉ DE :

Directeur de thèse :	Frédéric Chevallier	Directeur de recherche (LSCE/CEA)
Encadrante de thèse :	Isabelle Pison	Enseignante-chercheuse (LSCE/UVSQ)
Président du jury :	Guy Cernogora	Professeur émérite (UVSQ/IPSL)
Rapporteurs :	Marc Bocquet	Chercheur/Professeur (CEREA/ENPC)
	Philippe Ricaud	Directeur de recherche (CNRS/CNRM)
Examinatrice :	Fatima Laggoun-Défarage	Chargée de recherche (ISTO/INSU/CNRS)

PRÉPARÉE AU SEIN DU
LABORATOIRE DES SCIENCES DU CLIMAT ET DE L'ENVIRONNEMENT

Quantification des sources de méthane en Sibérie par inversion atmosphérique à la méso-échelle

Les émissions anthropiques et naturelles de méthane en Sibérie contribuent de manière notable, mais mal quantifiée au budget mondial de méthane (3–11% des émissions mondiales). Au Sud de la région, les émissions anthropiques sont liées aux grands centres urbains. Au Nord, l'extraction de gaz et de pétrole en Sibérie occidentale induit d'importantes sources anthropiques ponctuelles. Ces régions sont aussi couvertes de vastes zones humides naturelles émettant du méthane durant l'été (typiquement de mai à septembre). Nous utilisons des inversions atmosphériques régionales à la méso-échelle pour mieux comprendre les contributions de chaque processus dans le budget sibérien. Les inversions souffrent des incertitudes dans les observations, dans la simulation du transport et dans l'amplitude et la distribution des émissions. Pour prendre en compte ces incertitudes, je développe une nouvelle méthode d'inversion basée sur une marginalisation des statistiques d'erreurs. Je teste cette méthode et documente sa robustesse sur un cas test. Je l'applique ensuite à la Sibérie. À l'aide de mesures de concentrations atmosphériques de méthane collectées par des sites d'observation de surface en Sibérie, j'estime le budget régional de méthane sibérien à 5–28 TgCH₄.a⁻¹ (1–5% des émissions mondiales), soit une réduction de 50% des incertitudes par rapport aux précédentes études dans la région. Grâce à cette méthode, je suis de plus en mesure de détecter des structures d'émissions par zones de quelques milliers de km² et leur variabilité à une résolution de 2–4 semaines.

Quantification of methane sources in Siberia by meso-scale atmospheric inversions

Anthropogenic and natural methane emissions in Siberia significantly contribute to the global methane budget, but the magnitude of these emissions is uncertain (3–11% of global emissions). To the South, anthropogenic emissions are related to big urban centres. To the North, oil and gas extraction in West Siberia is responsible for conspicuous point sources. These regions are also covered by large natural wetlands emitting methane during the snow-free season, roughly from May to September. Regional atmospheric inversions at a meso-scale provide a mean for improving our knowledge on all emission process. But inversions suffer from the uncertainties in the assimilated observations, in the atmospheric transport model and in the emission magnitude and distribution. I developp a new inversion method based on error statistic marginalization in order to account for these uncertainties. I test this method on case study and explore its robustness. I then apply it to Siberia. Using measurements of methane atmospheric concentrations gathered at Siberian surface observation sites, I found a regional methane budget in Siberia of 5–28 TgCH₄.a⁻¹ (1–5% of global emissions). This implies a reduction of 50% in the uncertainties on the regional budget. With the new method, I also can detect emission patterns at a resolution of a few thousands km² and emission variability at a resolution of 2–4 weeks.



Remerciements

Je tiens à remercier chaleureusement Isabelle Pison pour le temps qu'elle m'a consacré, pour m'avoir encadré, parfois recadré pendant ces trois années de thèse, pour m'avoir accueilli dans son bureau pour des discussions scientifiques et souvent plus informelles, pour m'avoir transmis son enthousiasme pendant les creux de moral. Je remercie également Frédéric Chevallier pour son encadrement scientifique rigoureux mais cordial, pour les orientations subtiles de mon travail malgré mon entêtement, pour sa patience aussi.

Merci à Jean-Daniel Paris pour m'avoir accueilli la première fois au LSCE en 2010, pour m'avoir fait découvrir à plusieurs reprises la Sibérie et pour m'avoir incité à rester en thèse.

Merci aux collègues russes, japonais et allemands – Mishail Arshinov, Motoki Sasakawa, Jost Lavric, Jan Winderlich – qui m'ont aidé dans mon étude sur la Sibérie.

Un grand merci à Julien Bruna, à François Marabelle et à l'équipe informatique pour leur soutien et pour avoir réglé les bugs parfois non-conventionnels que j'ai pu générés en trois années. Merci à Jean-Yves Peterschmidt pour l'aide sur Python et pour les biscuits.

Merci à Fatima Laggoun-Défarage et à Laurent Menut pour le temps qu'ils m'ont consacré au cours des trois années de thèse à l'occasion des comités de thèse. Merci pour leurs remarques et les discussions scientifiques enrichissantes. Merci à Marc Bocquet pour la relecture de mon manuscrit et pour m'avoir introduit aux subtilités de l'assimilation de données à l'École d'été des Houches en 2012. Merci à Guy Cernogora et à Philippe Ricaud pour leur implication dans mon jury de thèse. Merci à l'équipe informatique de la station Concordia pour la connection internet pour la soutenance conjointe entre la France et l'Antarctique.

Merci à mes relecteurs et à tous ceux qui m'ont aidé dans la préparation de ma soutenance et de la partie gustative de l'évaluation de thèse.

Merci à Jean-Louis, Lamia, Pauline, Lucie, Annemiek, aux doctorants, à mes co-bureaux, CDDs, ... pour les pauses café, repas et autres séances de complaints pendant ces trois années.

Merci Mai-Lan.



Table des matières

Table des matières	i
Table des figures	vii
Liste des tableaux	ix
Introduction	1
Partie I Présentation générale	5
I Le cycle biogéochimique du méthane	7
1 Le méthane atmosphérique	8
1.1 Processus de production et de destruction du méthane	8
1.1.1 La méthanogenèse	8
1.1.2 La métagenèse	8
1.1.3 La pyrolyse végétale et l'irradiation foliaire	9
1.1.4 L'oxydation du méthane	9
1.2 Présence naturelle et perturbée dans l'atmosphère	9
1.3 Rôle dans le système atmosphérique	10
1.3.1 Dégradation de la qualité de l'air	10
1.3.2 Modification de la composition stratosphérique	11
1.3.3 Accentuation de l'effet de serre	11
2 Le budget atmosphérique de méthane	13
2.1 Budget total actuel du méthane	13
2.2 Les sources avérées de méthane atmosphérique	13
2.2.1 Les zones inondées naturelles	13
2.2.2 Les rizières	16
2.2.3 Les feux de forêts	16

2.2.4	Les combustibles fossiles	16
2.2.5	Les autres émissions notables	18
2.3	Les hydrates de méthane et le pergélisol	19
II	La Sibérie dans le cycle du méthane	21
1	Histoire et géographie du méthane en Sibérie	22
1.1	Histoire géologique et accumulation de matière organique	22
1.1.1	Le plateau de Sibérie centrale	23
1.1.2	La plaine de Sibérie occidentale	23
1.1.3	Le plateau continental arctique	23
1.2	Climat et environnements en Sibérie	24
1.2.1	Le climat continental sibérien	24
1.2.2	Les écosystèmes sibériens	24
1.3	Colonisation et peuplement de la Sibérie	24
2	Le budget régional sibérien de méthane	25
2.1	Approches « montantes » de quantification par type d'émission	26
2.1.1	Les zones inondées naturelles	26
2.1.2	Les feux de forêt	29
2.1.3	Les fuites de gaz et autres émissions anthropiques	29
2.1.4	Les hydrates et la végétation	30
2.2	Importantes incertitudes sur le budget sibérien	31
2.3	Approches atmosphériques	31
2.3.1	La composition isotopique	31
2.3.2	La co-émission de traceurs atmosphériques	32
2.3.3	La modélisation inverse	32
	Partie II Matériels et méthodes	35
III	La problématique inverse	37
1	Transport atmosphérique de traceurs passifs	39
1.1	Dynamique et mélange atmosphériques	39
1.1.1	Phénomènes synoptiques	39
1.1.2	Méso-échelle et échelle locale	40
1.2	Export et mélange de traceurs	42
1.2.1	Les traceurs passifs	42
1.2.2	Transport à grande distance	42
1.2.3	Mélange local vertical de traceurs	42
2	Formulation de l'inversion atmosphérique	44
2.1	Modélisation et simulation du transport atmosphérique	44
2.1.1	Formulation mathématique du monde réel	44
2.1.2	Résolution numérique de la formulation mathématique	45
2.2	Lien réciproque entre sources et concentrations	45
2.2.1	Modélisation directe et inverse	45
2.2.2	Modélisation inverse sous-contrainte	46
2.2.3	Nombreuses incertitudes dans l'inversion	47
3	Résolution du problème inverse atmosphérique	48

3.1	Implémentation des incertitudes et régularisation du problème	48
3.1.1	Le formalisme bayésien	48
3.1.2	Ajout d'informations a priori	50
3.2	Estimation de la densité de probabilité a posteriori	50
3.2.1	L'hypothèse gaussienne	51
3.2.2	Les densités de probabilités nécessaires à l'inversion	51
3.2.3	Les méthodes classiques de résolution d'inversion	53
3.3	L'incertitude statistique manquante	54
3.3.1	Une lacune associée au formalisme bayésien	54
3.3.2	Estimations des densités de probabilités a priori	54
4	Évaluation des résultats d'une inversion	55
4.1	Évaluations <i>ad hoc</i>	55
4.1.1	Confrontation à des données indépendantes	55
4.1.2	Tests de sensibilité	56
4.2	Critères algébriques et statistiques	56
4.2.1	Le critère du χ^2	56
4.2.2	Cohérence de matrices de covariances	57
4.2.3	Le maximum de vraisemblance	58
IV	Les ingrédients d'une inversion atmosphérique	61
1	Les observations	62
1.1	Les sites de surface	62
1.1.1	Les mesures flacons	63
1.1.2	Les mesures in situ	63
1.2	Les campagnes intensives et mesures ponctuelles	63
1.2.1	Les campagnes aéroportées YAK-AEROSIB	63
1.2.2	Le laboratoire mobile ferroviaire TROICA	64
1.3	Les satellites	64
2	Les flux a priori	65
2.1	Les inventaires anthropiques	65
2.2	Les modèles de processus biogéochimiques	67
2.3	Les bases de données de feux de biomasse	67
3	Les modèles de transport	67
3.1	Les modèles lagrangiens	67
3.2	Les modèles eulériens	68
3.2.1	Les eulériens directs et adjoints	68
3.2.2	Les eulériens mondiaux	68
3.2.3	Eulériens régionaux	69
Partie III	Applications et développements scientifiques	71
V	Facteurs influençant l'atmosphère sibérienne	73
1	Étude du transport par une campagne aéroportée	74
1.1	Le transport : prérequis de l'approche atmosphérique	74
1.2	La campagne YAK-AEROSIB d'avril 2010	74
1.3	L'ozone comme traceur du transport atmosphérique	74

1.4	Une estimation lagrangienne des structures de transport	75
2	<i>Tropospheric ozone over Siberia in spring 2010</i>	75
2.1	<i>Introduction</i>	77
2.2	<i>Methods</i>	78
2.2.1	<i>Campaign description and sampling area</i>	78
2.2.2	<i>Instruments used and data processing</i>	79
2.2.3	<i>Air mass origin and back trajectories</i>	80
2.3	<i>Results and discussion</i>	81
2.3.1	<i>Average, clean and polluted air masses composition</i>	81
2.3.2	<i>Influence of combustion processes on ozone concentrations</i>	84
2.3.3	<i>Upper troposphere ozone excess and stratospheric intrusion</i>	87
2.3.4	<i>Widespread upper tropospheric low ozone concentrations</i>	90
2.4	<i>Conclusions</i>	91
3	Structures météorologiques et contributions distantes en Sibérie	92
3.1	Structures de transport caractérisées	92
3.2	Identification de sources ponctuelles	94
3.3	Estimation de l'influence des hydrates de méthane en mer des Laptev	94
VI	Adaptation de méthodes de quantification d'erreurs	97
1	Choix des observations et erreurs associées	99
1.1	Types d'observations pertinentes pour contraindre les flux en Sibérie	99
1.2	Traduction dynamique des informations des observations de surface	100
2	Critères mathématiques objectifs d'ajustement des erreurs	101
2.1	Mauvaise connaissance identifiée des matrices de covariance	101
2.2	Ajustement objectif des matrices	102
2.3	Application à un cas d'étude	102
3	<i>Error statistics for atmospheric inversion</i>	102
3.1	<i>Introduction</i>	104
3.2	<i>Methods</i>	106
3.2.1	<i>Inversion system</i>	106
3.2.2	<i>Error configuration: description of the algorithms</i>	110
3.3	<i>Results</i>	114
3.3.1	<i>Patterns in the error variances for the 3 methods</i>	115
3.3.2	<i>Temporal variability and diurnal cycle of the observation errors</i>	116
3.3.3	<i>Correlations in the observation errors</i>	119
3.3.4	<i>Correlations in the background errors</i>	120
3.4	<i>Flux inversion: the Elgin leak case</i>	123
3.4.1	<i>Physical relevance of the inverted fluxes</i>	123
3.4.2	<i>Using the optimised tuples (\mathbf{R}_{ND}, \mathbf{B}_{ND}) for flux inversion</i>	123
3.5	<i>Discussion</i>	125
3.5.1	<i>Limitations and hypothesis probation</i>	125
3.5.2	<i>Implications for data selection</i>	126
3.6	<i>Conclusions</i>	129
4	Des erreurs objectivées mais encore incertaines	130
4.1	Choix de la méthode d'ajustement de matrices	130
4.2	Incertitudes de la méthode	130
4.3	Acquis sur le système d'inversion	131

VII Développement d'une inversion bayésienne marginalisée	133
1 Construction heuristique de la méthode générale	135
1.1 Approche pour l'estimation des incertitudes sur les matrices d'erreurs .	135
1.2 Définition d'un problème adapté à la théorie	135
1.3 Cas d'étude sibérien	136
2 <i>Objectified uncertainty quantification in Bayesian inversions</i>	137
2.1 <i>Introduction</i>	139
2.2 <i>Marginalized Bayesian inversion</i>	141
2.2.1 <i>Context and motivation for the marginalization</i>	141
2.2.2 <i>Marginalization of the inversion</i>	144
2.3 <i>Informed definition of the problem</i>	146
2.3.1 <i>Principle for problem reduction</i>	148
2.3.2 <i>Representation choice</i>	149
2.3.3 <i>Numerical artefacts</i>	151
2.4 <i>Validation experiments</i>	152
2.4.1 <i>Required tests</i>	152
2.4.2 <i>OSSE evaluation</i>	154
2.5 <i>Set up of the OSSEs</i>	155
2.5.1 <i>Virtual true state x^t</i>	156
2.5.2 <i>Simulation of the observation operator \mathbf{H}</i>	158
2.5.3 <i>Synthetic observations y^o</i>	159
2.6 <i>Results and discussion</i>	160
2.6.1 <i>Robustness of the method</i>	160
2.6.2 <i>Spatial evaluation</i>	163
2.6.3 <i>Limitations and benefits</i>	165
2.7 <i>Conclusions</i>	166
3 Une méthode générale et cohérente	167
3.1 Vers un système opérationnel	167
3.2 Améliorations à prévoir	167
VIII Application de l'inversion marginalisée en Sibérie en 2010	169
1 Mise en place de l'inversion marginalisée en Sibérie	171
1.1 Choix d'une fenêtre d'inversion	171
1.2 Une étude unique	172
2 <i>Natural and anthropogenic methane fluxes in Eurasia</i>	172
2.1 <i>Introduction</i>	174
2.2 <i>Marginalized inversion framework</i>	176
2.2.1 <i>Motivations towards marginalizing</i>	176
2.2.2 <i>Method outline</i>	178
2.2.3 <i>Output analysis</i>	179
2.2.4 <i>Size reduction and filters</i>	180
2.3 <i>Set up for an Eurasian domain</i>	182
2.3.1 <i>The observation network: y^o</i>	182
2.3.2 <i>Estimates of the network footprints</i>	183
2.3.3 <i>Prior fluxes and state vector: x</i>	184
2.3.4 <i>The observation operator: \mathbf{H}</i>	186
2.3.5 <i>Independent observations for evaluation</i>	187

2.4	<i>Diagnostics of the marginalized inversion</i>	188
2.4.1	<i>Observation weight in the inversion</i>	188
2.4.2	<i>Constrained regions</i>	191
2.4.3	<i>Solved time and space resolution</i>	192
2.5	<i>Results of the marginalized inversion</i>	194
2.5.1	<i>Inverted fluxes</i>	194
2.5.2	<i>Siberian Lowland CH₄ budget</i>	196
2.6	<i>Evaluation of the inversion</i>	198
2.6.1	<i>Performance on filtered out data</i>	198
2.6.2	<i>GOSAT evaluation</i>	198
2.6.3	<i>Toward using satellite measurements in regional frameworks</i>	200
2.7	<i>Conclusions</i>	202
3	<i>Des contraintes objectivées sur le budget sibérien</i>	203
3.1	<i>Des observations insuffisantes</i>	203
3.2	<i>Incertitude réduite sur le budget annuel de méthane en Sibérie</i>	204
	Partie IV Conclusions et perspectives	205
	Bibliographie	217
	Annexes	239
A	Acronymes et notations courantes	241
B	Liste des communications	245
C	Précisions techniques	251
D	Estimation des émissions de méthane en mer de Laptev	255
1	<i>Introduction</i>	256
2	<i>Materials and Methods</i>	257
2.1	<i>Observation sites</i>	257
2.2	<i>Polar CHIMERE transport model</i>	257
2.3	<i>Transport inputs and emission scenarios</i>	258
3	<i>Results</i>	259
3.1	<i>Model–observation comparisons at four Arctic sites</i>	259
3.2	<i>Summer isotopic observations in the Arctic</i>	261
3.3	<i>Estimation of methane emissions from the ESAS</i>	262
4	<i>Conclusions</i>	262



Table des figures

I.1	Séries temporelles des concentrations atmosphériques de CH ₄ au cours des 1000 dernières années	10
I.2	Forçages radiatifs anthropiques pour la période 1750–2011	12
I.3	<i>Evolution of the atmospheric global mole fraction, growth rate and budget of methane for the past three decades</i>	14
I.4	Répartition des zones inondées dans le monde	15
I.5	Processus influençant les émissions de méthane par les zones inondées	17
I.6	Émissions de méthane géologique par voie naturelle ou anthropique	18
I.7	Distribution spatiale du pergélisol dans l'hémisphère nord	19
II.1	Topographie de la Sibérie et principales formations géologiques	22
II.2	Principaux écosystèmes en Sibérie	25
II.3	Densité de population en Sibérie, Transsibérien et oléoducs	26
II.4	Distributions de zones inondées dans la plaine de Sibérie occidentale	28
III.1	Échelles spatio-temporelles du transport atmosphérique	40
III.2	Transport dans un cyclone de moyennes latitudes	41
III.3	Hauteur de couche limite moyenne en Sibérie en hiver	43
III.4	Exemple d'erreurs de représentation dans l'inversion	47
III.5	Incertitudes dans l'inversion	49
III.6	Distributions a posteriori du vecteur d'état et « réalité »	50
III.7	Représentation géométrique de l'inversion	58
IV.1	Distribution des mesures de méthane en Eurasie	62
IV.2	Plan de vol en 3 dimensions de la campagne YAK-AEROSIB d'avril 2010	64
IV.3	Répartition spatiale des émissions anthropiques en 2010	66
IV.4	Représentation en trois dimensions d'un domaine à aire limitée de simulations eulériennes	69
V.1	<i>Vertical profiles of CO₂, CO, O₃ and CH₄ for each flight</i>	81

V.2	<i>Species-species scatter plots</i>	84
V.3	<i>Back trajectories for four specific plumes</i>	86
V.4	<i>Stratospheric intrusion</i>	88
V.5	<i>Measured O₃ and air fraction originating from the North Atlantic lower troposphere</i>	90
V.6	<i>Influence à longue distance du volcan Eyjafjöll.</i>	93
V.7	<i>Transport des émissions des hydrates de méthane en mer de Laptev</i>	95
VI.1	<i>Contributions régionales sibériennes aux colonnes GOSAT</i>	99
VI.2	<i>Comparaison de l'orographie dans ETOPO et dans LMDz</i>	101
VI.3	<i>Spatial extension of the simulations with the CTM CHIMERE</i>	107
VI.4	<i>Comparison of the observation errors</i>	115
VI.5	<i>Statistics of the errors projected along PBL heights</i>	118
VI.6	<i>Mean temporal auto-correlation</i>	119
VI.7	<i>Correlations between the errors on the LBC elements of the background vector</i>	121
VI.8	<i>Correlations between the errors of the background aggregated regions</i>	122
VI.9	<i>Increments on the emissions</i>	124
VI.10	<i>Sensitivity of the inversion to each site</i>	127
VI.11	<i>Total sensitivity per site</i>	128
VII.1	<i>Incertitude statistique dans l'inversion bayésienne</i>	136
VII.2	<i>Choix des régions d'agrégation en fonction des observations</i>	137
VII.3	<i>Distribution of one component of the Monte Carlo posterior ensemble</i>	143
VII.4	<i>Block diagram of the method</i>	147
VII.5	<i>Topographic map of the domain of interest</i>	156
VII.6	<i>Distribution of the scaling factors applied to the emission databases</i>	158
VII.7	<i>Score comparison on fossil fuel and wetland regions for all OSSEs</i>	161
VII.8	<i>Map of the average scores</i>	164
VIII.1	<i>Disponibilité des données en Eurasie et fenêtre d'inversion</i>	171
VIII.2	<i>Topographic map of the domain of interest</i>	185
VIII.3	<i>Observation availability and contribution to the inversion for each site</i>	188
VIII.4	<i>Maps of the constraints on the fluxes</i>	190
VIII.5	<i>Temporal and spatial resolutions solved by the inversion</i>	193
VIII.6	<i>Inversion results for the twelve months of 2010</i>	195
VIII.7	<i>Time series of the posterior emissions on the Siberian Lowlands</i>	197
VIII.8	<i>Distribution of the differences between the observed and simulated prior and posterior GOSAT total columns</i>	200
VIII.9	<i>Observed and simulated contributions of the local Siberian sources to GOSAT total columns</i>	201
D.1	<i>Map of the domain with the emission zone from ESAS</i>	258
D.2	<i>Time series of observed and simulated methane mole fractions at four Arctic sites in 2012</i>	260
D.3	<i>Taylor diagrams using varying emission scenarios for ESAS</i>	263



Liste des tableaux

V.1	<i>YAK-AEROSIB campaign description</i>	78
V.2	<i>Instrument characteristics</i>	79
V.3	<i>Selected single plumes and specific events</i>	83
VI.1	<i>European observation site characteristics</i>	109
VI.2	<i>Correlations between calculated errors and the simulated PBL heights</i>	117
VII.1	<i>OSSEs summary</i>	153
VII.2	<i>Eurasian site characteristics</i>	160
VIII.1	<i>Sites characteristics</i>	184
VIII.2	<i>Correlations of observed and simulated mixing ratios</i>	199
A.1	Acronymes et abréviations utilisés dans le rapport	241
A.2	Notations scientifiques et mathématiques	243
D.1	<i>Observation site characteristics</i>	257



Introduction

Le méthane, un gaz trace naturellement présent dans l'atmosphère terrestre, représente environ 0,0002% de la masse atmosphérique, une proportion qui a plus que doublé à un rythme variable depuis le milieu du XVIII^e siècle. En dépit de ses teneurs négligeables, le méthane est le troisième contributeur à l'effet de serre terrestre après la vapeur d'eau et le dioxyde de carbone. Les concentrations croissantes de méthane dans l'atmosphère sont de plus en partie responsables de la dégradation mondiale de la qualité de l'air du fait d'une production indirecte d'ozone et d'une réduction de la capacité oxydante de l'atmosphère.

Or, nos sociétés sont sensibles aux modifications de l'effet de serre, et donc du climat, et à la dégradation de la qualité de l'air. En effet, un climat en réchauffement avec un effet de serre plus marqué pourrait induire une nette élévation du niveau de la mer, menaçant les populations et l'activité économique côtières, ainsi que des événements météorologiques extrêmes plus nombreux comme des sécheresses, des ouragans ou des inondations, perturbant le fonctionnement des économies locales. En outre, la santé publique et les rendements agricoles peuvent aussi pâtir d'une qualité de l'air dégradée.

Par conséquent, au-delà de la curiosité scientifique, il est important de comprendre les facteurs pilotant l'augmentation des concentrations de méthane dans l'atmosphère afin de prévoir et d'anticiper leur évolution future. De la sorte, des projections réalistes et fiables sur le climat et la qualité de l'air futurs pourront être effectuées.

L'augmentation des concentrations de méthane dans l'atmosphère est induite par un déséquilibre entre les émissions naturelles et anthropiques de méthane d'un côté et les puits de méthane par oxydation à la surface de la Terre et dans l'atmosphère de l'autre. De ce fait, la détermination de la contribution relative et absolue de chacun des processus de production et de destruction impliqués dans le budget mondial de méthane est nécessaire pour expliquer la variabilité passée et future des concentrations de méthane dans l'atmosphère. En outre, les contributions anthropiques aux émissions de méthane doivent être rigoureusement quantifiées et leur origine déterminée dans une optique de vérification du respect des

réglementations sur les émissions par les futurs protocoles internationaux post-Kyoto.

Les émissions par les zones humides naturelles et par l'exploitation anthropique de combustibles fossiles représentent la majeure partie des émissions mondiales de méthane. Toutefois, ces deux types d'émissions sont documentés de manière incertaine. En effet, les zones humides dans les Tropiques et aux hautes latitudes ne sont pas associées à des données de cadastres précises. De plus, les émissions dans ces régions sont spatialement et temporellement très sensibles à des paramètres difficilement quantifiables à haute résolution tels que la température dans le sol ou la hauteur de la nappe phréatique. Les émissions anthropiques par les combustibles fossiles sont de leur côté principalement liées à des fuites très variables en intensité et dont la localisation est connue de manière imprécise.

En conséquence, les émissions à l'échelle mondiale sont connues avec des incertitudes importantes. Ces incertitudes doivent être réduites afin de comprendre quantitativement l'augmentation des concentrations de méthane dans l'atmosphère.

Il existe deux approches à même d'apporter des informations sur le budget mondial de méthane : l'approche dite « montante » et l'approche « descendante ».

L'approche montante vise à comprendre les émissions à l'échelle des processus et à extrapoler à grande échelle chaque émission par processus. Par exemple, des modèles de processus biogéochimiques simulent les émissions de méthane des zones inondées pour chaque écosystème et en fonction de paramètres physiques, météorologiques et hydrologiques. Les émissions sont ensuite extrapolées à l'échelle mondiale suivant des cartes de zones inondées. Suivant une approche aussi montante, des inventaires d'émissions comptabilisent les sources anthropiques de méthane à partir de données sur la démographie ou sur l'activité économique et de facteurs d'émissions par secteur. Ces approches montantes sont à même de proposer des cartes d'émissions à diverses échelles (depuis l'échelle de pays ou de régions à des résolutions très fines dans certaines métropoles).

L'approche descendante s'appuie sur l'analyse du signal atmosphérique des concentrations de méthane pour contraindre les puits et les sources atmosphériques. En effet, les concentrations variant en fonction des sources et des puits, il est possible a priori d'inverser la chaîne de causalité et de déduire par exemple les émissions de méthane en surface. C'est pourquoi l'approche descendante est aussi appelée inversion atmosphérique. Des inversions utilisent la variabilité moyenne des concentrations atmosphériques afin de déduire l'équilibre général entre les sources et les puits de méthane. D'autres proposent des bilans d'émissions à l'échelle continentale ou régionale en corrigeant des cartes d'émissions issues d'approches montantes.

Nous disposons donc de cartes dérivées d'approches montantes et de bilans d'émissions déduits des approches descendantes par grandes régions. Cependant, les modèles de processus et les facteurs d'émissions des approches montantes nécessitent d'être étalonnés. De leur côté, les approches descendantes ne savent concilier fine résolution et bilan précis d'émission. De plus, ni les approches montantes à l'échelle mondiale, ni les approches descendantes à des échelles au mieux continentales ne sont à l'heure actuelle en mesure d'identifier la variabilité

temporelle des émissions à une échelle de quelques jours et leur distribution à une résolution inférieure à quelques dizaines de kilomètres. Or, de telles résolutions sont nécessaires afin de comprendre et quantifier chaque processus d'émission. Pour cette raison, les travaux de quantification les plus récents s'attachent à caractériser les échelles régionales, voire plus fines.

L'objectif de la régionalisation est d'évaluer pour une région donnée les contributions de chaque processus d'émission en présence, ainsi que la distribution et la variabilité des flux. Parmi les régions clés pour la quantification du budget du méthane, la Sibérie occupe une place d'importance. Cette masse continentale est couverte d'immenses étendues de zones humides naturelles parsemées de nombreux gisements de pétrole et de gaz exploités de manière intensive. De plus, le pergélisol et les hydrates de méthane présents aux hautes latitudes dans la région sont susceptibles d'induire une importante rétro-action positive au réchauffement du climat. En effet, la déstabilisation et le dégel de ces formations géologiques dans un climat réchauffé pourrait accentuer les émissions mondiales de méthane et favoriserait en retour un réchauffement climatique plus fort.

Dans le but d'estimer les émissions de méthane en Sibérie, il est envisageable d'adopter une approche montante. Néanmoins, les informations de cadastre, sur l'hydrologie, ou sur les taux de fuites de gaz sont particulièrement lacunaires en Sibérie. De ce fait, les approches montantes souffrent d'importantes incertitudes dans la quantification de ce budget régional.

En conséquence, nous choisissons d'utiliser les observations de concentrations de méthane disponibles en Sibérie dans une inversion atmosphérique à la méso-échelle, à savoir à des résolutions typiques de 50–100 km. Notre étude consiste à séparer les sources potentielles de méthane en Sibérie et à estimer leur variabilité temporelle et leur répartition spatiale à partir d'une analyse statistique du signal atmosphérique mesuré dans la région.

L'analyse des émissions à partir des concentrations dépend de notre faculté à reproduire le transport atmosphérique qui relie émissions et concentrations. Comme nous utilisons un modèle de transport numérique pour évaluer ce lien, la résolution de l'inversion atmosphérique est nécessairement entachée d'incertitudes. Évaluer objectivement et précisément ces incertitudes est essentiel pour obtenir des informations crédibles sur les émissions à partir des concentrations.

Par conséquent, notre travail s'est articulé en quatre étapes distinctes.

Dans un premier temps, nous avons identifié les sources de méthane et les structures de transport qui influencent la composition atmosphérique en Sibérie. Pour cela, nous avons examiné les observations collectées à l'occasion d'une campagne intensive de mesures aéroportées du projet YAK-AEROSIB¹ en avril 2010. Cette analyse doublée de l'utilisation d'un modèle de transport atmosphérique nous a permis de localiser les sources locales et distantes modifiant la composition atmosphérique sibérienne. Nous avons en outre été en mesure de distinguer certaines des erreurs faites au cours de la simulation du transport et de documenter

¹Tous les acronymes et abréviations utilisés dans ce rapport sont définis dans l'annexe A page 241

notre difficulté à quantifier celles-ci.

Dans un deuxième temps, nous avons adapté des méthodes développées par la communauté de l'assimilation de données pour les géosciences dans le but de quantifier de manière objective les erreurs dont souffre l'inversion. Pour cela, nous avons choisi des critères mathématiques et statistiques (χ^2 , maximum de vraisemblance, etc.) objectifs et calculables de manière systématique. Nous les avons testés sur un cas réel en Europe, plus favorable que la Sibérie du fait d'observations plus nombreuses et de flux mieux documentés. Ce travail nous a permis de sélectionner un critère privilégié adapté au contexte des inversions atmosphériques à la méso-échelle.

Dans un troisième temps, nous avons développé un système complet d'inversion basé sur des erreurs quantifiées objectivement, moyennant une sélection d'hypothèses statistiques. Le critère sélectionné précédemment est déterminé avec une inévitable incertitude due aux erreurs dans les calculs numériques et au manque d'information dans le système. Le nouveau système d'inversion que nous avons mis en place quantifie à l'aide d'une marginalisation les incertitudes dans la caractérisation des statistiques d'erreur. Pour cela, les équations fondamentales de l'inversion atmosphérique gouvernant le calcul de la densité de probabilité des flux optimisés à l'aide des observations atmosphériques ont été remaniées. La densité de probabilité ainsi obtenue de manière théorique a ensuite été estimée à l'aide d'une méthode de Monte-Carlo. Ce système d'inversion nouvellement développé a été évalué dans un cas réaliste à l'aide d'une « réalité » virtuelle connue et à travers des tests de sensibilité.

Dans un quatrième temps, notre système complet d'inversion par marginalisation a été utilisé pour caractériser la variabilité temporelle et la répartition spatiale des flux de méthane en Sibérie. Pour cela, des données de concentrations atmosphériques de méthane in situ ont été sélectionnées et regroupées dans un unique système. La capacité du système à réduire de manière notable les incertitudes sur le budget régional sibérien a alors été évaluée.

À l'issue de cette démarche scientifique, nous avons été en mesure d'améliorer notre connaissance du budget de méthane en Sibérie et donc du budget mondial. À cette occasion, nous avons de surcroît développé une nouvelle méthode d'inversion qui pourra être appliquée dans d'autres contextes que la Sibérie à la méso-échelle.

Première partie

Présentation générale

Le cycle biogéochimique du méthane

Sommaire

1	Le méthane atmosphérique	8
1.1	Processus de production et de destruction du méthane	8
1.1.1	La méthanogenèse	8
1.1.2	La métagenèse	8
1.1.3	La pyrolyse végétale et l'irradiation foliaire	9
1.1.4	L'oxydation du méthane	9
1.2	Présence naturelle et perturbée dans l'atmosphère	9
1.3	Rôle dans le système atmosphérique	10
1.3.1	Dégradation de la qualité de l'air	10
1.3.2	Modification de la composition stratosphérique	11
1.3.3	Accentuation de l'effet de serre	11
2	Le budget atmosphérique de méthane	13
2.1	Budget total actuel du méthane	13
2.2	Les sources avérées de méthane atmosphérique	13
2.2.1	Les zones inondées naturelles	13
2.2.2	Les rizières	16
2.2.3	Les feux de forêts	16
2.2.4	Les combustibles fossiles	16
2.2.5	Les autres émissions notables	18
2.3	Les hydrates de méthane et le pergélisol	19

1 Le méthane atmosphérique

Le méthane a été identifié au XVIII^e siècle par l'italien Alessandro Volta comme le gaz inflammable composant les bulles s'échappant des marécages. La molécule de méthane est constituée d'un atome de carbone (C) et de quatre atomes d'hydrogène (H) formant un tétraèdre régulier : CH₄. Sous les conditions de température et de pression qui règnent dans l'atmosphère terrestre, il se présente sous forme gazeuse. Lorsque ses concentrations sont comprises entre 5 et 15%, le méthane peut s'enflammer dans l'air.

1.1 Processus de production et de destruction du méthane

Le méthane est l'alcane le plus simple. Il est majoritairement produit à partir de la dégradation de molécules organiques plus complexes (par exemple les hydrocarbures). Cette dégradation peut se faire par voie biologique active (la méthanogenèse), géologique (la métagenèse) ou par une dégradation directe de végétaux (pyrolyse et irradiation des feuilles par des rayonnements ultra-violet). Les molécules de méthane ainsi créées peuvent ensuite être détruites par oxydation.

1.1.1 La méthanogenèse

Les méthanogènes sont les seuls êtres vivants en mesure de produire activement du méthane. Ces micro-organismes appartiennent au domaine des archées, au côté des eucaryotes, des bactéries et des champignons.

Les méthanogènes vivent exclusivement dans des environnements anaérobies dans lesquels de la matière organique dégradée est présente. Ces archées sont au bout de la chaîne alimentaire anaérobie (Reay et al., 2010). Dans les environnements anaérobies, les molécules organiques les plus complexes sont d'abord fermentées, puis dégradées en acétate et en formate. Les méthanogènes utilisent enfin l'acétate et le formate comme source d'énergie et rejettent du méthane. La dégradation de l'acétate et du formate par les méthanogènes se fait à l'aide d'enzymes et de catalyseurs agissant à des échelles de temps inférieures au jour.

Les méthanogènes interagissent avec des bactéries fermentatives et acétatogènes. Les interactions entre ces différents micro-organismes sont complexes et concurrentielles. Par exemple, des bactéries peuvent être plus efficaces dans certaines conditions pour consommer les aliments de base des méthanogènes. Selon la matière organique disponible dans un milieu anaérobie, les espèces de bactéries en présence, la température, etc., l'équilibre de la chaîne alimentaire anaérobie change et induit des taux de production de méthane très différents.

1.1.2 La métagenèse

Les mêmes chaînes de dégradation que dans la méthanogenèse peuvent se reproduire dans des conditions de très hautes pression et température et à des rythmes beaucoup plus lents de plusieurs millions d'années (Tissot and Welte, 1984). Ces conditions sont remplies lorsque des couches sédimentaires sont ensevelies dans la croûte terrestre. Au sein de ces couches sédimentaires, on trouve souvent de la matière organique piégée, alors appelée kérogène. Le kérogène perd dans un premier temps son eau et gaz carbonique lors de la diagenèse (à faible température). Ses molécules sont ensuite craquées thermiquement lorsqu'elles atteignent des profondeurs où la température dépasse 60°C. Cette étape, nommée catagenèse,

donne en général lieu à la formation de pétrole. Enfin, lors de la métagenèse, dans des conditions de forte pression et de température élevée, les molécules de pétrole sont à leur tour craquée pour aboutir à la production de gaz naturel (essentiellement composé de méthane). Selon les matériaux organiques initiaux, les conditions de sédimentation et le temps d'ensevelissement, la dégradation du kérogène aboutit à la formation de charbon, de pétrole ou de gaz naturel. En général, dans une zone de métagenèse donnée, ces trois types d'hydrocarbures peuvent être trouvés.

1.1.3 La pyrolyse végétale et l'irradiation foliaire

La dernière voie de production de méthane vient de la dégradation directe de matière végétale vivante.

Lors de feux de biomasse, la matière organique subit une pyrolyse à haute température. La pyrolyse fait perdre son humidité à la matière organique, puis la transforme en charbon, bitumes et différents gaz, dont le méthane.

Le rayonnement ultra-violet (UV) est aussi susceptible de dégrader les végétaux et est à l'origine d'émissions de méthane. Le processus exact menant à des émissions de méthane en présence d'UV n'est pas encore totalement identifié (Vigano et al., 2008; Keppler et al., 2008). Les taux de production de méthane par ce processus, localement très faibles, concernent tous les végétaux.

1.1.4 L'oxydation du méthane

Le méthane est une espèce chimiquement très stable. Le principal moyen de détruire une molécule de méthane est de lui arracher un atome d'hydrogène, ce qui produit un radical méthyle (CH_3). Les molécules de méthane sont donc détruites par oxydation.

Cette oxydation peut se faire par combustion du méthane en présence d'oxygène, ce qui libère une grande quantité d'énergie. Les radicaux hydroxyles (OH) naturellement présents dans l'atmosphère ont aussi la faculté d'oxyder le méthane. Une troisième voie d'oxydation en milieu naturel est accomplie par des bactéries.

En effet, certaines bactéries, les méthanotrophes, possèdent des enzymes particulières leurs permettant d'oxyder activement les molécules de méthane. Ces bactéries sont présentes dans tous les types d'environnement et utilisent le méthane comme seule source d'énergie et de carbone (Hanson and Hanson, 1996). Les enzymes que ces bactéries sont capables de produire catalysent l'oxydation du méthane. Le résidu principal de la chaîne d'oxydation du méthane par les méthanotrophes est le dioxyde de carbone.

1.2 Présence naturelle et perturbée dans l'atmosphère

Du fait de ces voies de production et de destruction, le méthane est naturellement présent en faibles quantités dans l'atmosphère. Ces niveaux de concentrations sont en nette augmentation depuis le début de l'ère industrielle en 1750 (voir figure I.1). L'analyse de micro-bulles d'air capturées pendant la formation de la glace au niveau des calottes polaires en Antarctique et au Groenland (Petit et al., 1999) a permis de déduire que, au cours des 400 000 dernières années, les concentrations de méthane ont varié entre 300 et 700 parties par milliards (ppb). Depuis 1750, les concentrations de méthane sont passées à 1850 ppb, soit une augmentation

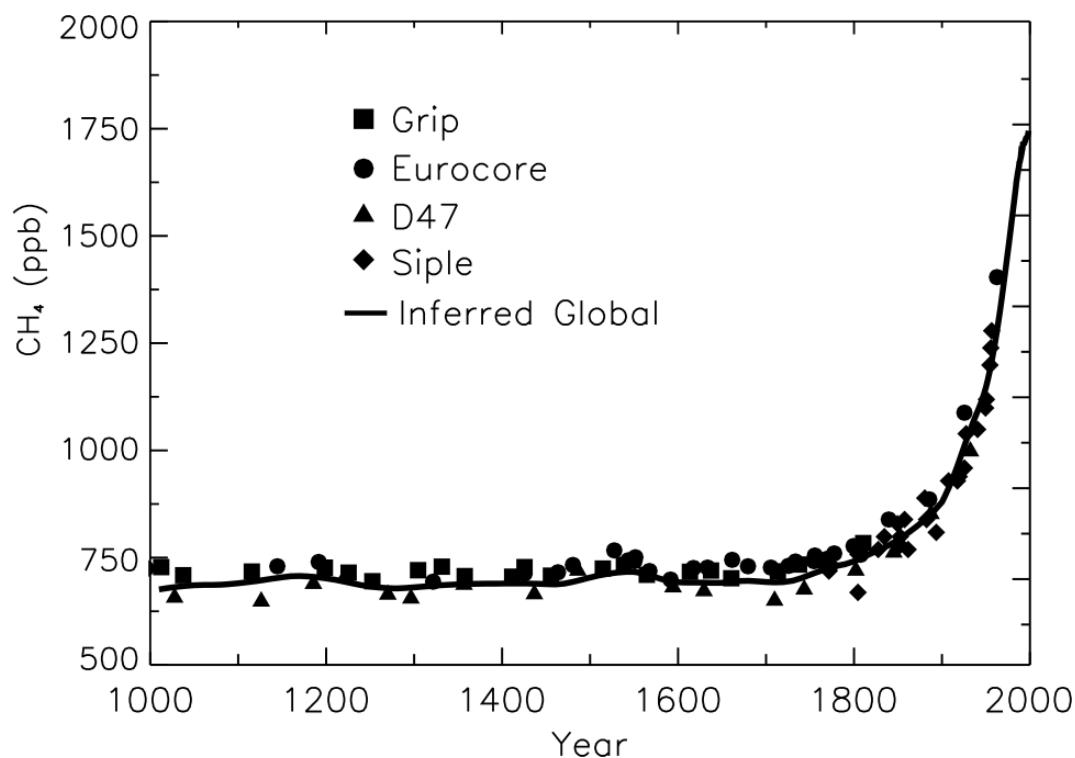


FIGURE I.1: Séries temporelles des concentrations atmosphériques de CH₄ au cours des 1000 dernières années. Les concentrations sont déduites de l'analyse de bulles d'air capturées dans des carottes de glaces prélevées en Antarctique et au Groenland (Etheridge et al., 1998).

de 150%. Cette augmentation est corrélée avec l'augmentation des concentrations du dioxyde de carbone (CO₂) de 280 à 400 ppm sur la même période et est attribuée principalement à l'activité humaine, notamment à l'usage de combustibles fossiles.

1.3 Rôle dans le système atmosphérique

Du fait des très faibles teneurs en méthane dans l'atmosphère, celui-ci a longtemps été simplement considéré comme une source d'énergie lorsqu'on le brûle. Un intérêt nouveau a été porté au méthane pendant la seconde moitié du XX^e siècle avec la découverte de son rôle dans le système atmosphérique. Ce rôle est multiple de part une dégradation indirecte de la qualité de l'air, une modification de la composition naturelle de la stratosphère et une contribution notable à l'effet de serre.

1.3.1 Dégradation de la qualité de l'air

Le méthane dans l'atmosphère est principalement oxydé par les radicaux hydroxyles, la réaction d'oxydation produisant des radicaux méthyles et de la vapeur d'eau. La consommation d'OH et la production de radicaux méthyle a deux effets : la production indirecte d'ozone (O₃) et la réduction du pouvoir oxydant de la troposphère.

Premièrement, en présence d'oxydes d'azote, le radical méthyle, très réactif dans l'atmosphère, initie une chaîne de réactions aboutissant à la production d'O₃ (Cicerone and Orem-

land, 1988). La famille des oxydes d'azote, notée NO_x , est composée des monoxyde et dioxyde d'azote. Les NO_x sont présents dans les régions « polluées ». Ils sont typiquement émis en grande quantité par le trafic automobile et l'industrie. Leurs concentrations sont donc élevées à proximité des grands centres urbains et industriels, en particulier en Europe, en Amérique du Nord et en Chine (p. ex. Boersma et al., 2008). L'ozone dans la basse troposphère est nocif pour tout les tissus vivants car il est un oxydant très fort. En particulier, les voies respiratoires humaines et les stomates des plantes sont sensibles à la pollution à l'ozone. L'ozone est donc une espèce problématique en termes de santé publique et de sécurité alimentaire, avec des surcoûts induits très importants (Brandt et al., 2013) et des baisses de rendements agricoles (Feng and Kobayashi, 2009; Van Dingenen et al., 2009). Le méthane contribue en grande partie aux niveaux de fond d'ozone et à leur augmentation, induisant des expositions chroniques potentiellement dangereuses pour les tissus vivants.

Deuxièmement, l'atmosphère, via les radicaux hydroxyles, est en mesure d'oxyder une grande partie des espèces chimiques présentes dans l'atmosphère. En particulier, sont oxydés le monoxyde de carbone (CO), le méthane, les Composés Organiques Volatils (COV), les NO_x , etc. Les COV et les NO_x sont considérés comme des polluants, en eux-mêmes et en tant que précurseurs de l'ozone. L'oxydation du méthane par les radicaux OH dans l'atmosphère réduisant les concentrations atmosphériques de ces derniers, les polluants tels que les COV et les NO_x peuvent s'accumuler plus facilement dans l'atmosphère (Wang and Jacob, 1998; Holmes et al., 2013). Par conséquent, la détérioration de la qualité de l'air induite est plus importante qu'en l'absence de méthane.

1.3.2 Modification de la composition stratosphérique

La stratosphère, située au dessus de la tropopause (vers 18 km aux tropiques et 8 km dans les hautes latitudes), est très sèche et naturellement riche en ozone. L'ozone stratosphérique protège la surface de la Terre contre les UV du soleil, nocifs pour les cellules vivantes. Le méthane dans la stratosphère est dégradé de la même manière que dans la troposphère : par oxydation, en grande majorité par des radicaux hydroxyles. La vapeur d'eau résultant de la dégradation du méthane est alors à son tour oxydée et induit, dans la stratosphère, une chaîne de réaction détruisant l'ozone. Les niveaux actuels de méthane pourraient ainsi induire la disparition de jusqu'à quelques pourcents de l'ozone stratosphérique (p. ex. Portmann et al., 2012).

1.3.3 Accentuation de l'effet de serre

L'effet de serre dans l'atmosphère est principalement causé par la vapeur d'eau naturellement présente dans l'atmosphère (de l'ordre de 1% de la masse atmosphérique). Il est bénéfique à la vie car il permet de maintenir les températures moyennes de surface à $\sim 16^\circ\text{C}$ au lieu de -18°C pour une atmosphère terrestre privée d'effet de serre.

Néanmoins, les températures moyennes de surface pourraient croître de quelques degrés en quelques décennies sous l'influence d'un effet de serre en augmentation. Un changement brutal de la température pourrait menacer les écosystèmes les plus fragiles et aussi potentiellement la civilisation telle que nous la connaissons (p. ex. Field et al., 2014). En effet, les variations naturelles de température inter-glaciaires sont de $2-3^\circ\text{C}$ en plusieurs milliers d'années et induisent pourtant des climats terrestre radicalement différents.

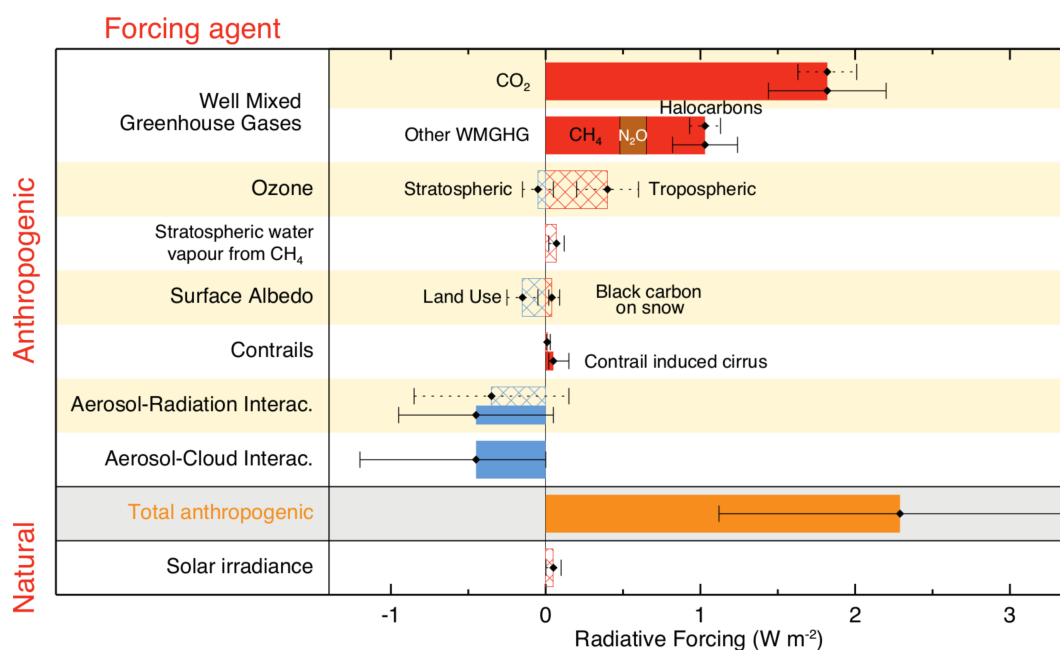


FIGURE I.2: Forçage radiatif anthropique et forçage radiatif effectif pour la période 1750–2011 (tiré de Myhre et al., 2013). Les incertitudes (intervalle de confiance de 5 à 95%) sont spécifiées en lignes pointillées pour le forçage radiatif et en lignes pleines pour le forçage radiatif effectif.

De par ses liaisons carbone-hydrogène agencées selon une forme tétraédrique, le méthane est un fort gaz à effet de serre. Des modèles de transfert radiatif montrent qu'il faut environ 100 fois moins de CH₄ que de CO₂ pour obtenir le même effet radiatif dans l'atmosphère (Myhre et al., 1998). Du fait de l'oxydation par les radicaux OH, le méthane réside dans l'atmosphère pendant un temps plus court que le CO₂ (typiquement 8–10 ans contre plus de 100 pour le CO₂; Dentener et al., 2003). Le pouvoir de réchauffement d'une molécule de méthane à un horizon de 20 ans est donc 86 fois plus élevé que celui du CO₂ et 23 fois plus élevé à un horizon de 100 ans.

La figure I.2, tirée du V^e Rapport d'Évaluation du Groupe d'experts Intergouvernemental sur l'Évolution du Climat (GIEC) publié en 2014, résume les impacts des activités humaines sur le bilan radiatif atmosphérique. Ainsi, l'augmentation des concentrations de méthane attribuables à l'homme est responsable du second forçage radiatif anthropique. Ce forçage est environ quatre fois moindre que celui du CO₂ alors que les augmentations de concentration de méthane sont plus de 100 fois inférieures à celles du dioxyde de carbone. De plus, la vapeur d'eau importée dans la stratosphère par oxydation du méthane et la production troposphérique d'ozone ajoutent encore une contribution à l'effet de serre mondial (Shindell et al., 2009).

Pour ces raisons, des études ont montré l'importance de réduire prioritairement les émissions anthropiques de méthane dans l'optique de diminuer à court terme le réchauffement climatique (p. ex. Reay et al., 2010; Harvey et al., 2012). Bien que non dominant au total, l'impact radiatif élevé du CH₄ rendrait les efforts de réduction des émissions de méthane beaucoup plus rentables que pour le CO₂ (chaque tonne de CH₄ non émise équivalant à 86

tonnes de CO₂ à un horizon de 20 ans).

2 Le budget atmosphérique de méthane

Du fait de l'impact du méthane sur le système atmosphérique, il est important de comprendre pourquoi les concentrations de méthane atmosphérique ont augmenté de 150% depuis 1750. Il est primordial de quantifier les contributions naturelles et anthropiques à cette augmentation afin d'anticiper les augmentations ou stagnations futures de concentrations de méthane et ainsi effectuer des projections climatiques fiables. Nous détaillons ici le budget total de méthane atmosphérique (Sect. 2.1), les sources avérées de méthane atmosphérique (Sect. 2.2) et les sources potentielles futures (Sect. 2.3).

2.1 Budget total actuel du méthane

L'augmentation des concentrations atmosphériques de méthane depuis 1750 ne s'est pas faite de manière constante. La variation moyenne des concentrations de méthane représente le bilan des sources et des puits dans l'atmosphère. Les augmentations ou diminutions des concentrations de méthane dans l'atmosphère sont donc attribuables à un déséquilibre entre sources et puits. La variabilité de la concentration atmosphérique peut facilement se convertir en tonnes de méthane ajoutées ou ôtées à l'atmosphère. Kirschke et al. (2013) ont compilé différentes études afin de déduire la variabilité inter-annuelle du budget total de méthane (voir figure I.3).

Dans cette compilation, la somme des puits de méthane atmosphérique s'élève à 500–600 Tg.a⁻¹ (1 Tg = 10¹² g). Le méthane est en très grande partie oxydé dans la troposphère par les radicaux hydroxyle (plus de 90% des puits). Il est aussi diffusé dans les couches superficielles des sols secs où il est oxydé par des méthanotrophes (Curry, 2007) responsables de 5–10% des puits.

Les études compilées par Kirschke et al. (2013) trouvent des émissions totales de 500 à 800 Tg.a⁻¹ selon les années. Ces émissions se répartissent typiquement en 35% venant des zones inondées, 35% de l'agriculture et 30% d'autres sources. Les émissions moyennes totales par décennies sont connues avec un niveau de confiance de l'ordre de 10%.

Les concentrations d'OH varient d'une année à l'autre, ce qui modifie le puits atmosphérique de méthane (Dentener et al., 2003; Montzka et al., 2011). Néanmoins, la variabilité des concentrations d'OH dans l'atmosphère ne peut seule expliquer les différences inter-annuelles dans le budget du méthane. Il est donc nécessaire de comprendre, de situer et de quantifier les différents types d'émissions afin d'expliquer et potentiellement prévoir l'évolution du budget du méthane.

2.2 Les sources avérées de méthane atmosphérique

2.2.1 Les zones inondées naturelles

Des milliers de petagrammes (1 Pg = 10¹⁵ g) de matière organique sont disponibles dans les sols. Les zones inondées (marécages, tourbières, etc.) recouvrent une partie de cette matière organique disponible. Les surfaces de zones inondées sont évaluées entre 5 et 20 millions de km² selon les définitions et les méthodes utilisées (Matthews and Fung, 1987; Melton

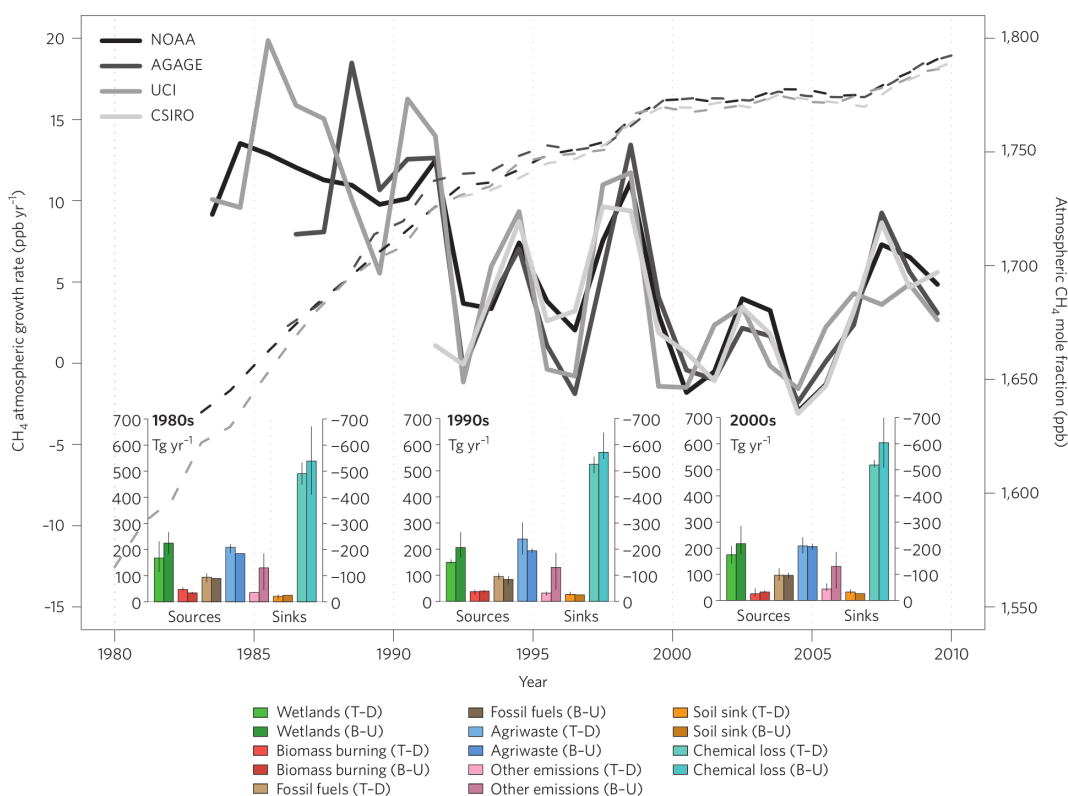


FIGURE I.3: Evolution of the atmospheric global mole fraction, growth rate and budget of methane for the past three decades (from Kirschke et al., 2013). The mole fraction (dashed lines) and growth rate (solid lines) from NOAA, AGAGE, UCI and CSIRO networks are shown in varying shades of black/grey. Bar charts show global decadal surface emissions and sinks calculated from top-down (T-D, light-coloured bars) and bottom-up (B-U, dark-coloured bars) approaches. Categories are split into : natural wetlands, biomass burning, fossil fuels, agriculture and waste, other sources (see Table 1 in Kirschke et al., 2013), soil uptake and chemical loss by OH oxidation. Error bars spread between minimum and maximum values.

et al., 2013), soit 3–14% des terres émergées. Les méthanogènes disposent ainsi d'immenses réservoirs de matière organique potentiellement dégradable en méthane. Dans la figure I.4, la répartition spatiale et les types de zones inondées potentiellement émettrices de méthane sont représentées. Ces zones sont particulièrement denses dans les tropiques, en Afrique centrale et en Amazonie, mais aussi aux hautes latitudes, au Canada, en Scandinavie et en Sibérie.

Des émissions de méthane sont susceptibles d'être détectables dans toutes ces zones. Mais d'une région à l'autre, les émissions vers l'atmosphère sont très différentes. En effet, comme le résume la figure I.5, le méthane produit sous terre par les méthanogènes est amené à la surface par trois vecteurs : l'ébullition, qui a permis à Alessandro Volta de découvrir le méthane, la diffusion à travers l'eau en surface et le pompage à travers les racines des plantes en présence. Lors de l'ébullition et de la diffusion, le méthane peut être oxydé en conditions aérobies à proximité de la surface par des bactéries méthanotrophes, limitant ainsi les émissions nettes de méthane de la zone marécageuse vers l'atmosphère.

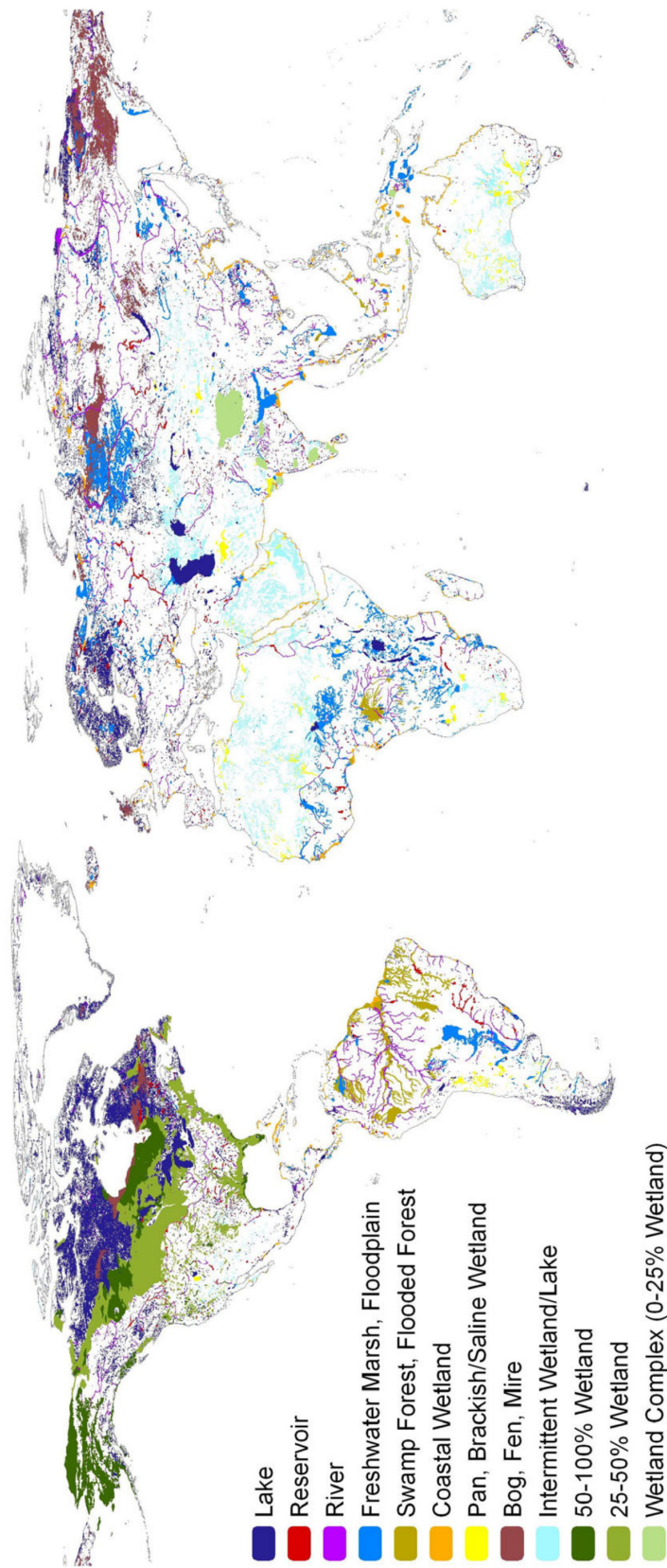


FIGURE I.4: Répartition des zones inondées dans le monde (tiré de Lehner and Döll, 2004). Les zones inondées sont regroupées en catégories en fonction de la densité régionale inondée, la végétation et la structure de la zone inondée.

Les émissions de méthane par les marécages dépendent donc du niveau de la nappe phréatique, permettant ou non l'activation des méthanogènes, mais aussi de l'épaisseur d'eau contenant de l'oxygène à proximité de la surface et de l'efficacité des plantes à pomper le méthane. Ces différents processus dépendent de la température, des précipitations, de la saison, du type de végétation, du drainage de la zone, donc de la topographie et du type de sol. Dans les zones froides, les zones en partie ou entièrement inondées peuvent partiellement ou totalement geler pendant une partie de l'année, inhibant l'activité des archées. Les émissions de méthane par les zones inondées sont donc très hétérogènes spatialement et temporellement, allant jusqu'à quelques dizaines $\mu\text{gCH}_4\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.

Les émissions totales de méthane par les 5–20 millions de km^2 de zones inondées naturelles sont estimées à 150–400 $\text{TgCH}_4\cdot\text{a}^{-1}$ (Denman et al., 2007; Kirschke et al., 2013).

2.2.2 Les rizières

Les rizières sont des zones inondées artificiellement. Les processus d'émission de méthane sont sensiblement semblables à ceux agissant dans les zones inondées naturelles. Les deux différences principales avec les zones inondées naturelles sont une inondation maîtrisée et continue pendant typiquement 4 mois dans l'année et l'usage de fertilisants et autres produits phytosanitaires. L'apport de produits phytosanitaires modifie l'équilibre de l'environnement anaérobie. Selon les produits et les autres micro-organismes présents, les méthanogènes peuvent être favorisés ou non, augmentant ou réduisant les émissions de méthane (Matthews et al., 1991). L'irrigation peut par ailleurs être maîtrisée afin de trouver un niveau d'eau limitant les émissions de méthane (Reay et al., 2010).

Environ 1,6 millions de km^2 de rizières sont cultivés dans le monde, principalement en Inde et en Chine (> 50% de la superficie). Ainsi, les émissions de méthane par la riziculture sont estimées à 33–40 $\text{TgCH}_4\cdot\text{a}^{-1}$ (Ciais et al., 2013).

2.2.3 Les feux de forêts

Le méthane produit par pyrolyse dans les feux de biomasse est souvent directement brûlé, alimentant ainsi le feu. Mais il peut aussi être relâché directement dans l'atmosphère. Une combustion lente aura tendance à émettre plus de méthane qu'une combustion rapide et complète. En moyenne, 0,5% de la matière végétale est transformée en méthane atmosphérique (Lobert et al., 1991).

Les feux de biomasse naturels se déclenchent ponctuellement et peuvent durer de quelques heures à quelques jours. Les émissions de méthane par les feux de biomasse sont intenses au niveau des feux individuels. À l'échelle du globe, ces émissions de méthane très hétérogènes s'élèvent à une trentaine de $\text{TgCH}_4\cdot\text{a}^{-1}$ (van der Werf et al., 2010).

2.2.4 Les combustibles fossiles

Les émissions de méthane venant de sources géologiques sont résumées dans la figure I.6.

Le méthane formé par métagenèse dans les couches sédimentaires peut se diffuser par porosité de la roche ou être transporté dans des micro-fissures, pour atteindre la surface terrestre ou le fond de l'océan. Ainsi, des émissions diffuses ont été détectées au niveau des bassins de formation d'hydrocarbures et de sédiments marins (Etiope et al., 2008). Ces suintements de

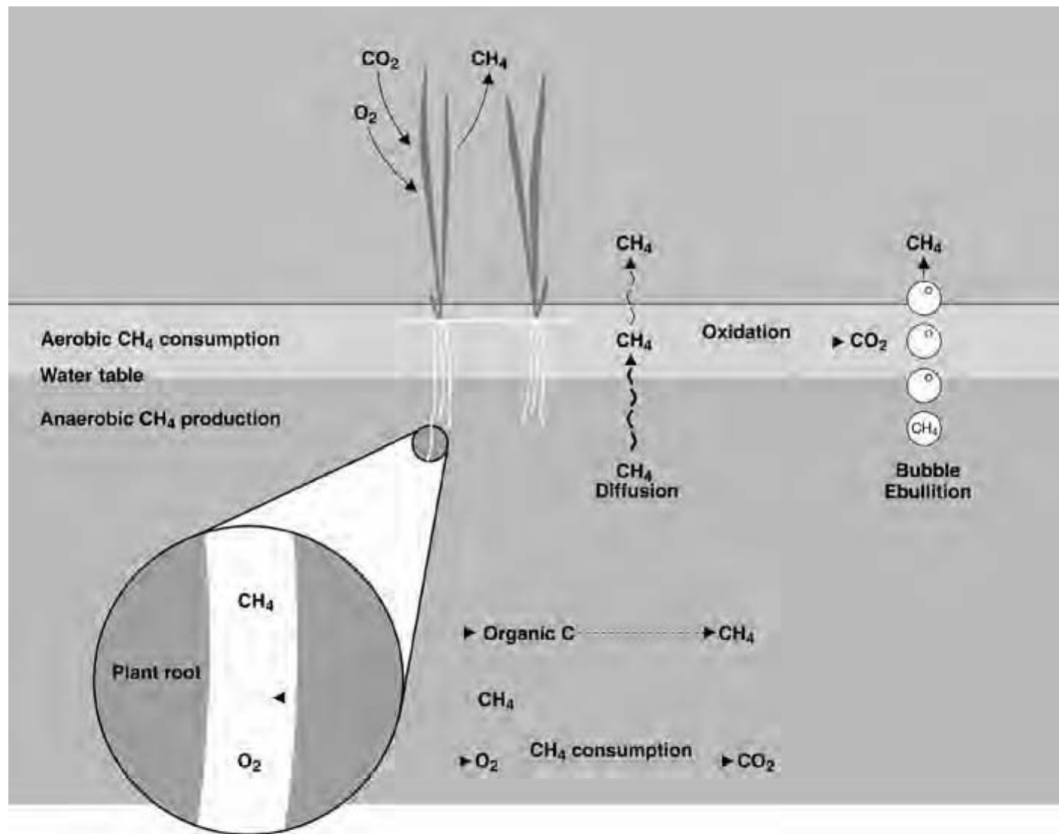


FIGURE I.5: Processus influençant les émissions de méthane par les zones inondées (tiré de Joabsson et al., 1999). Le méthane est produit dans le milieu anaérobie et est transporté jusqu'à l'atmosphère à travers la portion d'eau aérobie où il peut être oxydé. Le transport se fait par diffusion, ébullition et pompage racinaire.

méthane sont parfois réduits par l'intervention de méthanotrophes dans les couches proches de la surface. Les incertitudes sont grandes sur ces émissions naturelles de méthane géologique du fait de flux faibles et diffus sur de grandes zones, et donc difficilement détectables individuellement. Les estimations mondiales de ces sources géologiques naturelles vont de 10 à 80 $\text{TgCH}_4 \cdot \text{a}^{-1}$.

Les émissions liées à l'industrie d'extraction d'hydrocarbures sont plus certaines. Elles sont de deux types. Dans le cas des gisements de charbon et de pétrole, des poches de gaz sont présentes au milieu du gisement. De ce fait, elles sont la plupart du temps considérées comme un déchet, voire un danger (par exemple le grisou dans les mines de charbon). Dans ces cas, le méthane est relâché directement dans l'atmosphère, ou éventuellement brûlé à la source. La production de pétrole mondiale est de 90 millions de barils par jour, soit 4000 Tg par an. Pour chaque centaine de barils de pétrole produit, l'équivalent en masse d'environ un demi baril s'échappe dans l'atmosphère sous forme de méthane. À l'échelle mondiale, cela représente donc 20 $\text{TgCH}_4 \cdot \text{a}^{-1}$. Le taux de méthane dans les gisements de charbon est semblable à celui des nappes de pétrole. C'est donc environ 22 $\text{TgCH}_4 \cdot \text{a}^{-1}$ qui sont émis dans l'atmosphère sur les 7500 Tg de charbon produits chaque année.

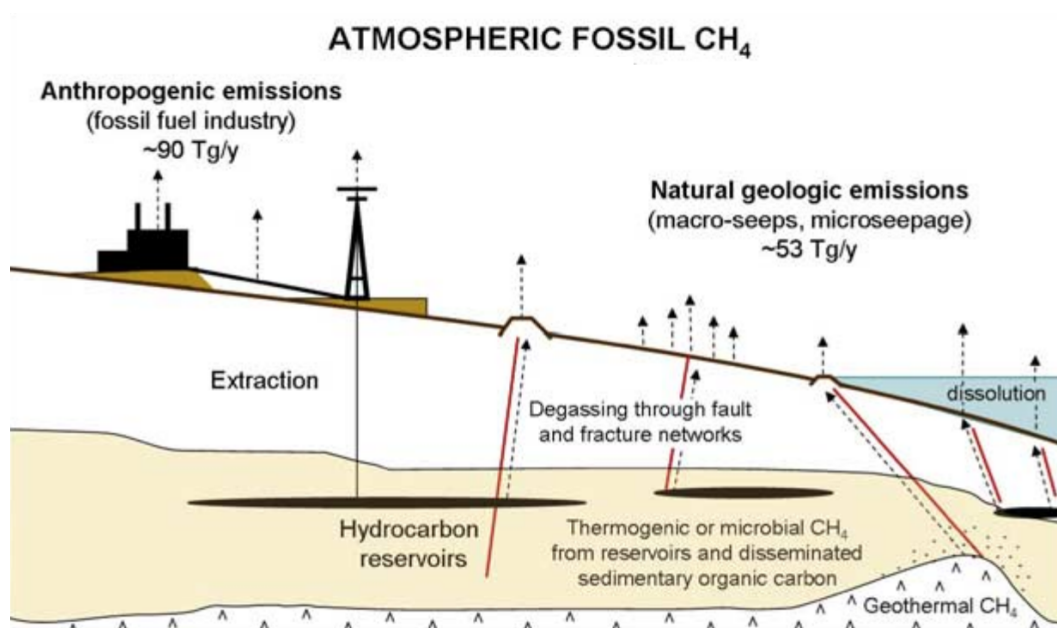


FIGURE I.6: Émissions de méthane géologique par voie naturelle ou anthropique (tiré de Etiope et al., 2008). En rouge, les vecteurs naturels du transport de méthane depuis les réservoirs d'hydrocarbures jusqu'à la surface : les micro-failles du sol et la porosité du sol.

Dans le cas des exploitations de gaz naturel, le but est d'extraire le plus de gaz possible. Mais le caractère volatil de cette ressource le rend difficile à puiser entièrement. Des fuites indésirables sont présentes tout au long de la chaîne d'extraction et de distribution du gaz naturel. Ainsi, selon les gisements, les techniques d'extraction et de distribution, l'entretien des infrastructures, etc., de quelques dixièmes de pourcents à quelques pourcents des ressources extraites sont relâchées dans l'atmosphère, principalement sous forme de méthane (E.P.A., 2011). La production mondiale de gaz naturel est de 3000 Tg chaque année. Entre 1 et 10% s'échappent dans l'atmosphère, représentant des émissions d'au moins $30 \text{ TgCH}_4 \cdot \text{a}^{-1}$. Les fuites de gaz naturel sont souvent sous-estimées et certaines études proposent par exemple des émissions de méthane dans l'atmosphère venant du réseau de distribution de gaz naturel de l'ordre de 10–15% des émissions totales anthropiques (Wunch et al., 2009).

2.2.5 Les autres émissions notables

On trouve des archées méthanogènes dans d'autres environnements anaérobies que les zones inondées. Les émissions significatives de méthane des environnements anaérobies autres que les zones inondées proviennent : des estomacs des ruminants, des décharges à ciel ouvert, des stations d'épurations et des systèmes digestifs des termites.

Les émissions venant du système digestif des termites sont individuellement infinitésimales. Mais selon les estimations, les termites représentent jusqu'à 25% de la masse animale terrestre, soit 450 millions de tonnes (Sanderson, 1996). Ainsi, les émissions venant des termites, principalement confinées au niveau des régions tropicales, sont estimées à jusqu'à $20 \text{ TgCH}_4 \cdot \text{a}^{-1}$ (Ciais et al., 2013).



FIGURE I.7: Distribution spatiale du pergélisol dans l'hémisphère nord (tiré de Schuur et al., 2008). Les nuances de violet dénotent la continuité et la densité de pergélisol.

Les émissions dans les systèmes digestifs des ruminants, depuis les décharges à ciel ouvert et les stations d'épurations sont liées aux centres de population humains. Au total, ces émissions atteignent $100\text{--}200 \text{ TgCH}_4.\text{a}^{-1}$.

Les émissions liées à la dégradation de matière foliaire, localement très faibles, concernent tous les végétaux. Elles pourraient ainsi atteindre $230 \text{ TgCH}_4.\text{a}^{-1}$ selon les estimations (Houweling et al., 2006; Reay et al., 2010).

2.3 Les hydrates de méthane et le pergélisol

En plus des émissions actuelles prouvées de méthane, d'autres émissions sont incertaines et pourraient devenir significatives dans un climat plus chaud que le climat actuel.

Dans les régions les plus froides de la planète (les régions de montagne ou aux hautes latitudes), le sol peut être gelé en permanence sur une profondeur allant de quelques mètres à quelques centaines de mètres. Toute portion de sol gelée plus de deux hivers d'affilée est considérée comme du pergélisol. Dans la figure I.7 page 19, la distribution du pergélisol dans l'hémisphère nord est représentée. La grande majorité de ces zones est gelée depuis des milliers d'années sans interruption.

Dans ces conditions, l'activité biologique dans le sol est très ralentie, voire éliminée. Le dégel du pergélisol sous des conditions climatiques plus chaudes rendrait disponible la matière organique gelée dans de nouvelles zones inondées. Ainsi, un réservoir estimé à environ 1000 Pg de carbone serait rendu disponible pour les méthanogènes et les autres micro-organismes (Zimov et al., 2006; Tarnocai et al., 2009). Des rétro-actions positives au réchauffement climatique entreraient en action, accentuant le réchauffement climatique en augmentant la quantité de méthane contenue dans l'atmosphère.

Outre le pergélisol, les accumulations d'hydrates de méthane dans les couches sédimentaires sous-marines pourraient induire d'importantes émissions de méthane dans un climat plus chaud qu'aujourd'hui. L'hydrate de méthane est un mélange stabilisé d'eau et de méthane sous forme de glace. Sous des conditions de pression et de température satisfaisantes (au niveau des fonds marins à basse température), le méthane émis au niveau des sédiments se stabilise dans ces hydrates et s'accumule. Ces conditions sont souvent présentes aux hautes latitudes, mais des hydrates ont aussi été trouvés dans des zones tempérées.

Les hydrates sont sensibles à des changements de températures dans l'océan. Un réchauffement climatique pourrait déstabiliser de tels hydrates (Giustiniani et al., 2013).

De plus, des projets d'exploitation des hydrates de méthane comme combustibles fossiles existent avec de possibles déstabilisations artificielles et des fuites identiques à celles trouvées dans l'industrie gazière classique (Kvenvolden, 1993).

Un cycle biogéochimiques du méthane incertain

Les sources de méthane atmosphérique sont diverses et quantifiées de manière encore incertaine (Kirschke et al., 2013). Le budget total de méthane est de plus sensible au changement climatique en cours. Afin d'améliorer la connaissance des sources de méthane atmosphérique, nous identifions dans la suite une région clé du cycle biogéochimique de méthane, la Sibérie, et nous efforçons d'en quantifier les émissions.

La Sibérie dans le cycle du méthane

Sommaire

1	Histoire et géographie du méthane en Sibérie	22
1.1	Histoire géologique et accumulation de matière organique	22
1.1.1	Le plateau de Sibérie centrale	23
1.1.2	La plaine de Sibérie occidentale	23
1.1.3	Le plateau continental arctique	23
1.2	Climat et environnements en Sibérie	24
1.2.1	Le climat continental sibérien	24
1.2.2	Les écosystèmes sibériens	24
1.3	Colonisation et peuplement de la Sibérie	24
2	Le budget régional sibérien de méthane	25
2.1	Approches « montantes » de quantification par type d'émission . . .	26
2.1.1	Les zones inondées naturelles	26
2.1.2	Les feux de forêt	29
2.1.3	Les fuites de gaz et autres émissions anthropiques	29
2.1.4	Les hydrates et la végétation	30
2.2	Importantes incertitudes sur le budget sibérien	31
2.3	Approches atmosphériques	31
2.3.1	La composition isotopique	31
2.3.2	La co-émission de traceurs atmosphériques	32
2.3.3	La modélisation inverse	32

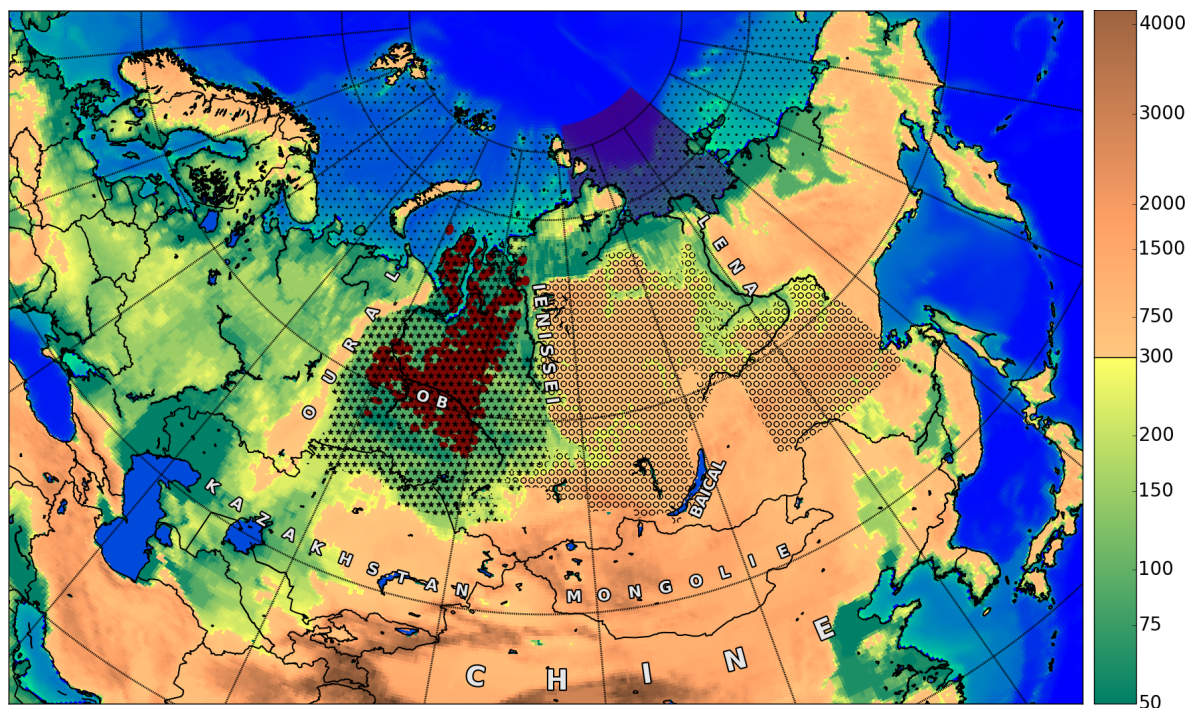


FIGURE II.1: Topographie de la Sibérie et principales formations géologiques. La topographie est issue de la base de donnée ETOPO1 (Amante and Eakins, 2009). Les étoiles représentent la plaine de Sibérie occidentale, les cercles le plateau de Sibérie centrale et les points le plateau continental arctique. La mer des Laptev est située au niveau de la zone violette. Les zones rouges montrent les gisements avérés de pétrole et de gaz dans la plaine de Sibérie occidentale (à partir de la base de données PETRODATA; Lujala et al., 2007).

La Sibérie est la portion du continent eurasien s'étendant à l'Est des monts Oural et au Nord du Kazakhstan, de la Mongolie et de la Chine. Elle s'étale sur 13 millions de km², soit environ 10% des terres émergées.

Nous montrons dans ce chapitre que la plupart des processus d'émissions sont actifs dans cette immense étendue. Nous mettons aussi en évidence les grandes incertitudes sur le budget sibérien de méthane qui fait de celui-ci l'un des principaux contributeurs aux incertitudes sur le budget mondial. Nous détaillons d'abord, dans la sect. 1, l'histoire géologique et démographique de la Sibérie, ainsi que son climat et les types d'environnements rencontrés. Ensuite, en sect. 2, nous explicitons la place tenue par la Sibérie dans le cycle biogéochimique du méthane et les approches utilisées pour quantifier les contributions de chaque processus d'émission au budget régional.

1 Histoire et géographie du méthane en Sibérie

1.1 Histoire géologique et accumulation de matière organique

La Sibérie est constituée de deux complexes géologiques principaux : le plateau de Sibérie centrale et la plaine de Sibérie occidentale (voir figure II.1). À ces deux ensembles, il faut ajouter le plateau continental arctique le long des côtes Nord du continent eurasien.

1.1.1 Le plateau de Sibérie centrale

La plus ancienne composante géologique de la Sibérie est délimitée par les fleuves Iénissey à l'Ouest et Léna à l'Est et par le lac Baïkal au Sud. Il s'agit du plateau de Sibérie centrale, formé par le craton sibérien (Metelkin et al., 2011), caractérisé par des plateaux d'altitudes supérieures à 300 m. Un craton est un ensemble de la croûte terrestre ayant résisté à la plupart des mouvements des plaques tectoniques depuis sa formation. Le craton sibérien s'est formé à partir de l'époque néo-archéenne (-2,5 milliards d'années).

Les principales roches le constituant sont ainsi vieilles de 1–2 milliards d'années. Les roches de cette époque sont caractérisées par de nombreux gisements de minerais. Cette zone est donc très riche en nickel, cuivre, uranium, platine, fer, terres rares, etc., et diamants. Par conséquent, ces ressources sont extraites de manière intensive en Sibérie, ce qui fait de la Russie l'un des principaux producteurs mondiaux de ces ressources.

L'activité minière et industrielle est ponctuellement très intense autour de ces gisements, favorisant les émissions anthropiques de méthane.

1.1.2 La plaine de Sibérie occidentale

Le deuxième complexe géologique est formé par la plaine de Sibérie occidentale. Cette plaine est une étendue extrêmement plate de 2,6 millions de km² coincée entre les monts de l'Oural et le plateau de Sibérie centrale. Plus de 50% de cette étendue est à une altitude inférieure à 100 m.

La plaine de Sibérie occidentale est alimentée par les fleuves Ob et Iénissey chariant de grandes quantités de sédiments. Les mouvements tectoniques lents de la région ont permis l'accumulation de sédiments depuis 65 millions d'années au gré de la formation de mers ou lacs géants dans la région. Ainsi, de 4 à 10 km de couches sédimentaires recouvrent la plaine de Sibérie occidentale (Cherepanova et al., 2013).

L'accumulation de telles quantités de sédiments a favorisé les processus de formation d'hydrocarbures sous la plaine de Sibérie. De nombreux gisements de gaz et de pétrole ont été découverts depuis l'époque soviétique (Grace, 2005) et sont exploités en Sibérie occidentale (voir figure II.1). Les ressources avérées dans la région sont estimées à au moins $37\,000 \times 10^9 \text{ m}^3$ de gaz naturel et 80×10^9 barils de pétrole, soit respectivement l'équivalent de 8,5 et 7,5 années de production mondiale. Actuellement, l'extraction de gaz et de pétrole se fait respectivement au rythme de $623 \times 10^9 \text{ m}^3$ et 3×10^9 barils par an.

Ces ressources sont acheminées via des gazoducs et oléoducs dans toute la Russie et en Europe occidentale (voir figure II.3). Cette distribution est accompagnée d'inévitables fuites de méthane. De plus, le méthane emprisonné dans les gisements de pétrole n'est pas toujours complètement brûlé en Sibérie. Alors que chaque année, environ $40 \times 10^9 \text{ m}^3$ de gaz (= $29 \text{ TgCH}_4 \cdot \text{a}^{-1}$), soit la consommation moyenne française, sont brûlés au niveau des gisements de pétrole, des quantités indéterminées de méthane s'échappent dans l'atmosphère.

1.1.3 Le plateau continental arctique

Le plateau continental arctique s'étend au Nord des côtes arctiques du continent eurasiatique. Certaines régions de ce plateau continental à de faibles profondeurs (0–300 m) sont riches

en sédiments. En particulier, au niveau de la mer des Laptevs (voir figure II.1), le fleuve Lena dépose de grandes quantités de sédiments qui s'accumulent au large de son delta. Du méthane est naturellement produit de manière biogénique ou géologique dans ces sédiments. Le méthane produit peut alors s'accumuler au sein d'hydrates ou s'échapper du fond marin.

1.2 Climat et environnements en Sibérie

1.2.1 Le climat continental sibérien

La Sibérie est au cœur de la masse continentale eurasiennne. Hormis l'Océan glacial arctique pris dans les glaces la majeure partie de l'année, l'océan le plus proche est situé à 3000 km de la Sibérie. Le climat sibérien est ainsi un climat continental très marqué. Les températures moyennes sont de -15°C en hiver et 20°C en été, avec des températures descendant ponctuellement jusqu'à -70°C et s'élevant au-dessus de 30°C . La Sibérie est ainsi essentiellement couverte de neige en hiver, alors que les hautes températures estivales peuvent favoriser les feux de forêts.

Une large formation anticyclonique recouvre la Sibérie en hiver (voir encadré 43). Cet anticyclone sibérien a la capacité de réduire la force des perturbations arrivant au-dessus de la Sibérie et favorise un climat sec en hiver (précipitations inférieures à 20 mm.mois^{-1}).

1.2.2 Les écosystèmes sibériens

Le climat sibérien marqué par de grands écarts de température favorise le développement de steppes herbueses et de taïgas peuplées de conifères (voir figure II.2). La Sibérie est ainsi couverte de 4 millions de km^2 de forêts boréales (soit 10% de la surface forestière dans le monde). Des dizaines de milliers de km^2 brûlent chaque année en Sibérie, engendrant ponctuellement d'importantes émissions de méthane.

L'accumulation récente de matière organique, le climat et la topographie ont permis la formation de tourbières et de marécages sur une grande partie de la plaine de Sibérie occidentale (Kremenetski et al., 2003). Ce sont entre 50 et 75% de la surface de la plaine de Sibérie occidentale qui sont considérés comme des zones inondées (voir figure II.2 et Peregon et al., 2009). Ces tourbières et marécages sont gelés en hiver (typiquement d'octobre à mai) et inondés pendant le court été sibérien (de juin à septembre). Ces vastes étendues marécageuses émettent donc du méthane au moins une partie de l'année.

1.3 Colonisation et peuplement de la Sibérie

La Sibérie est peuplée par des groupes nomades depuis au moins 40 000 ans. Aujourd'hui, ces peuples sont largement minoritaires. Depuis le XVI^e siècle et jusqu'au XIX^e siècle, l'empire russe s'est développé grâce à la conquête et la colonisation de la Sibérie au détriment des peuples autochtones. La colonisation moderne de la Sibérie s'est faite au moyen du chemin de fer transsibérien reliant Moscou à Vladivostok (voir tracé en figure II.3). L'exploitation des ressources minières, fossiles et forestières se fait aujourd'hui à partir de noyaux de population établis le long du Transsibérien ou sur des sous-branches de celui-ci.

La Sibérie est actuellement peuplée de 30 millions de citoyens de la Fédération de Russie principalement groupés dans les centres urbains du Sud de la région le long du Transsibérien

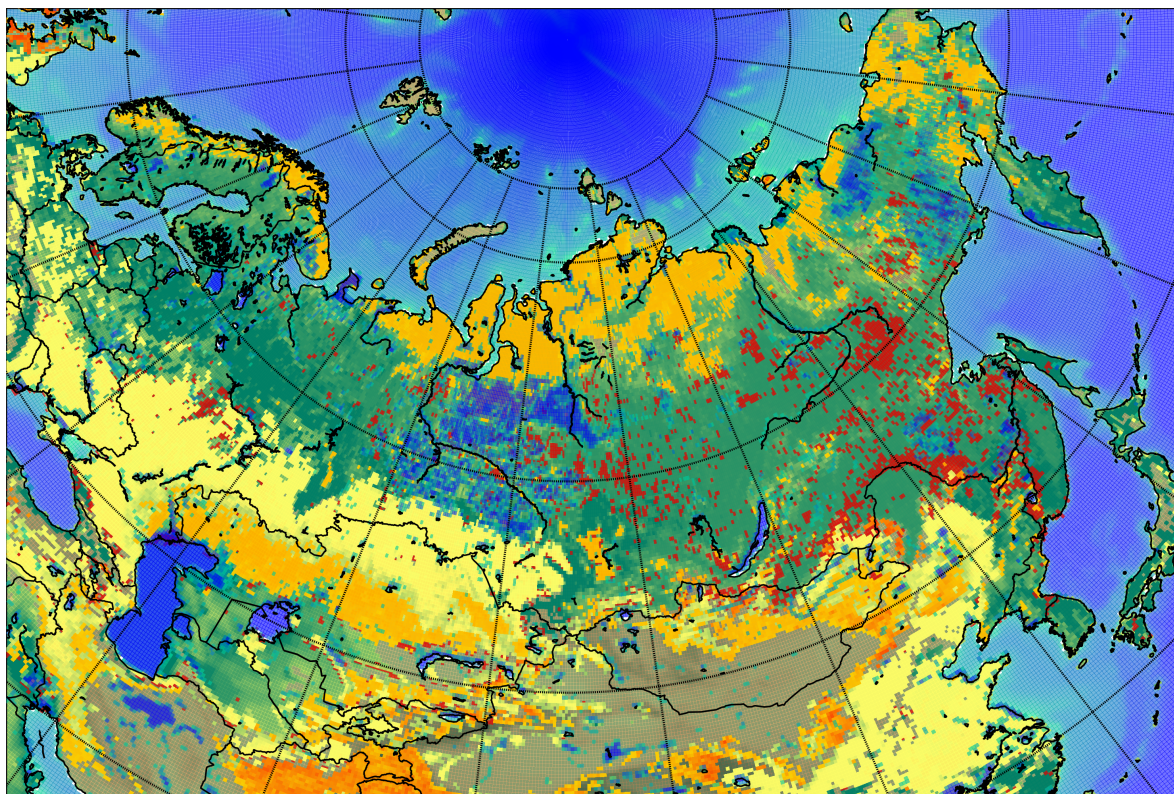


FIGURE II.2: Principaux écosystèmes en Sibérie. En vert, les forêts boréales, en orange les taïgas et steppes, en jaune les plaines cultivables et pâturages. Les points rouges localisent les feux détectés par l'instrument MODIS entre 2009 et 2013 sur des zones forestières (Giglio et al., 2013). Les zones inondées estimées par la base de données GLWD (voir détails en figure II.4) sont repérées en bleuté.

(voir figure II.3). Les principales villes de Sibérie sont Novossibirsk (1,5 millions d'habitants), Omsk (1,1 million hbts), Tomsk (500 000 hbts), Yakutsk (270 000 hbts), avec des colonies ponctuelles d'exploitation intensives de ressources telles que Norilsk (175 000 hbts ; exploitations de nickel) ou encore Myrnii (35 000 hbts ; mine de diamant). Ces centres urbains génèrent des émissions de méthane dues aux décharges, aux stations d'épurations, à la distribution du gaz naturel et à l'élevage.

2 Le budget régional sibérien de méthane

De par l'étendue de la Sibérie, son relief, son histoire géologique et démographique, tous les processus émettant du méthane dans l'atmosphère à l'exception des émissions venant des termites et des rizières se rencontrent dans la région. Les émissions de méthane en Sibérie sont tant naturelles, avec de vastes étendues de marécages et de tourbières et des feux de forêts, qu'anthropiques, avec l'extraction intensive de ressources gazières et pétrolières et de grands centres urbains. Plusieurs questions se posent quant à ces émissions. Quels sont les processus dominants ? Avec quelle distribution spatiale et variabilité temporelle ? Quels facteurs (économiques, physiques, climatiques, etc.) influencent ces émissions ? Quelles sont

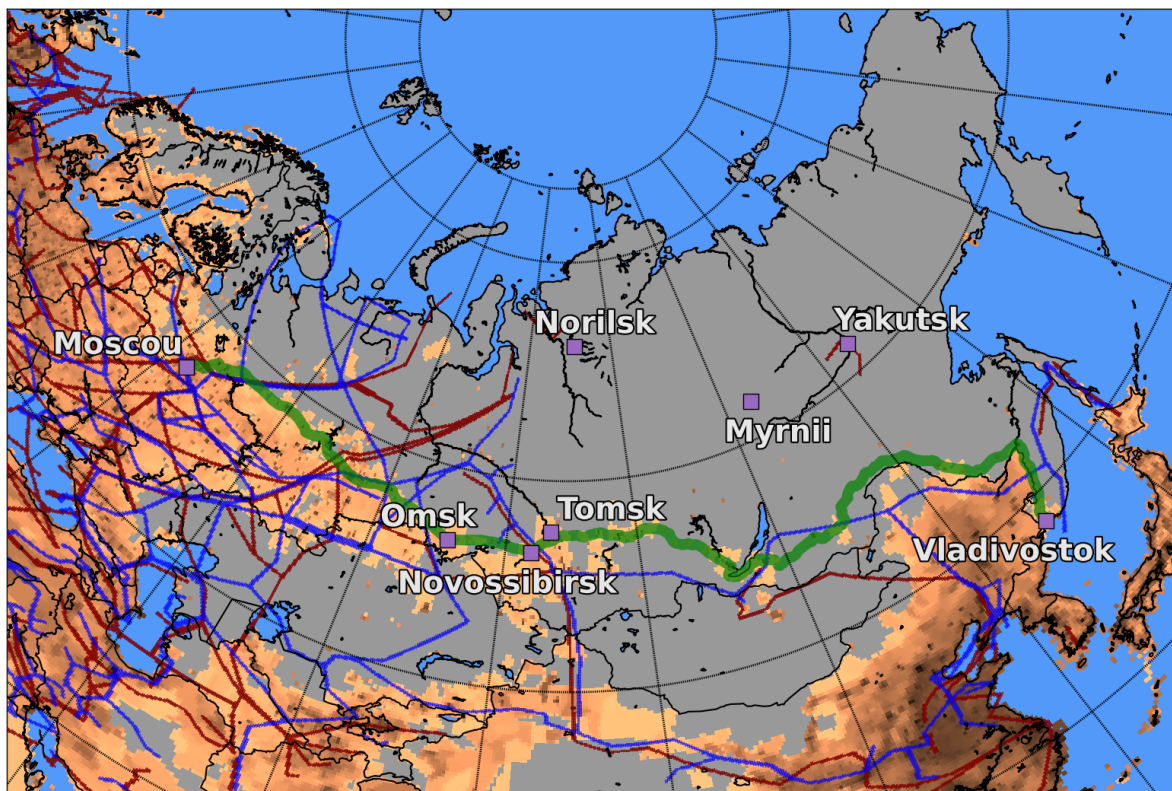


FIGURE II.3: Densité de population en Sibérie, tracé du chemin de fer transsibérien (vert) et des gazoducs (rouge) et oléoducs (bleu) en Eurasie.

les émissions totales de la région dans le budget mondial du méthane ? Nous résumons dans la suite les études disponibles dans la littérature qui sont consacrées à la quantification des émissions en Sibérie.

2.1 Approches « montantes » de quantification par type d'émission

En premier lieu, nous détaillons les travaux se basant sur une approche dite « montante ». L'approche montante consiste à comprendre chaque processus d'émission localement. Une fois les flux quantifiés localement, des extrapolations sont effectuées afin de déduire les émissions à plus grande échelle.

2.1.1 Les zones inondées naturelles

Chaque été, une grande partie de la plaine de Sibérie occidentale dégèle et est alors inondée, entraînant l'activation des méthanogènes dans les sols. L'activité des méthanogènes est très sensible à la température du sol, à l'humidité et aux nutriments disponibles.

En théorie, il suffit de connaître la fonction de réponse décrivant le rendement en méthane des méthanogènes en fonction des différents paramètres physiques auxquels ils sont sensibles. En connaissant la distribution des sols inondés, leur composition et les conditions météorologiques et hydrologiques locales, il est possible de modéliser les flux de méthane d'une région donnée. Cependant, la fonction de réponse recherchée est complexe et ne peut

être explicitée. Les méthanogènes peuvent être observés en laboratoire afin de déduire des points de la fonction de réponse. Néanmoins, les expériences en laboratoire ne reproduisent pas nécessairement la complexité d'un sol naturel. Pour cette raison, les projets d'observation in situ sont favorisés pour caractériser la réponse des écosystèmes à différents paramètres physiques.

Les observations in situ se font généralement à l'aide de chambres à flux. Ces chambres recouvrent hermétiquement une parcelle de sol (de l'ordre du m^2). Il est ensuite possible de mesurer l'accumulation de méthane dans la chambre et donc le flux local. De telles observations ont été effectuées dans des environnements de hautes latitudes afin de mesurer l'impact de la température, du pompage de méthane par les racines des plantes, du niveau de la nappe, etc. (Macdonald et al., 1998; Nakano et al., 2000; Godwin et al., 2013; Noyce et al., 2013).

Ces expériences sont extrêmement locales car elles ne représentent que quelques m^2 de sol inondé alors que les émissions sont très hétérogènes. Pour s'affranchir de la sensibilité à l'emplacement de la chambre à flux, des tours à flux sont utilisées pour mesurer les flux de méthane sur un rayon de quelques centaines de mètres (Hargreaves and Fowler, 1998; Park et al., 2014). Les mesures par tours à flux utilisent le fait qu'à très haute fréquence, le périmètre contraint peut être considéré comme une chambre à flux. Les variations rapides des concentrations de méthane sont alors dues à la turbulence locale et aux flux de méthane que l'on cherche. L'utilisation locale de profils verticaux de concentrations de méthane amène par le même principe que les tours à flux à une estimation des flux de méthane (Winderlich et al., 2014).

Les études in situ permettent donc de documenter la sensibilité de chaque type d'environnement à différents paramètres physiques. Glagolev et al. (2011) ont installé de nombreuses chambres à flux dans différents environnements sibériens sur de longues périodes. Ces données leur ont permis d'extrapoler directement les émissions de méthane de la plaine de Sibérie occidentale. Leur estimation des émissions régionales était de $2-5 \text{ TgCH}_4 \cdot \text{a}^{-1}$.

Néanmoins, en été, le sol sibérien est majoritairement spongieux et imbibé d'eau, ce qui restreint toute opération logistique. Par conséquent, la plupart des régions inondées de Sibérie sont très difficiles d'accès et il est très coûteux d'installer des sites de mesures en des endroits suffisamment nombreux et de manière pérenne. Les études in situ sont donc difficiles à reproduire et sont sujettes à d'importantes erreurs d'extrapolation.

C'est pourquoi des modèles numériques sont utilisés afin de déduire la distribution et la variabilité des flux de méthane. Les observations in situ fournissent un étalonnage pour ces modèles pour un ensemble de conditions hydrologiques et d'écosystèmes. Les modèles utilisent des bases de données sur la distribution de la végétation et des modèles hydrologiques et météorologiques (Melton et al., 2013; Wania et al., 2013) afin d'extrapoler les observations in situ. De la sorte, un modèle tel que le modèle LPJ (Lund-Postdam-Jena) fournit des simulations de flux de CH_4 à 1° de résolution spatiale et à une fréquence mensuelle (Spahni et al., 2011).

Les modèles numériques de processus simulent des émissions de $5-15 \text{ TgCH}_4 \cdot \text{a}^{-1}$ venant des zones inondées en Sibérie. Ces chiffres varient d'un modèle à l'autre suivant les données météorologiques et hydrologiques utilisées, les cartes de zones inondées, la paramétrisation

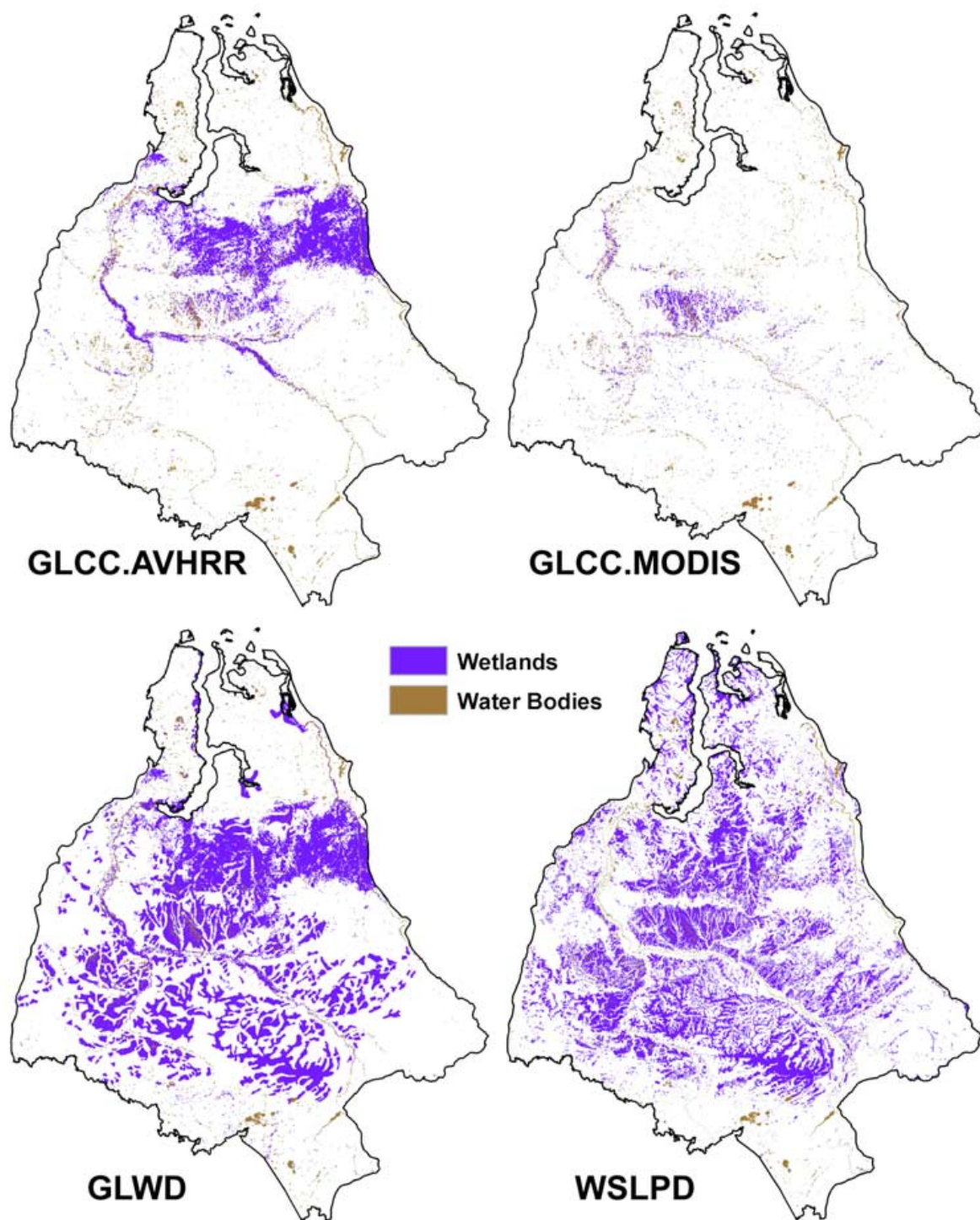


FIGURE II.4: Distribution des zones inondées dans la plaine de Sibérie occidentale (voir figure II.1 page 22) déduites de plusieurs méthodes de calcul (Frey and Smith, 2007). GLCC.AVHRR et GLCC.MODIS font partie de la base de donnée de l'USGS *Global Land Cover Characterization*. Elles sont établies à partir d'observations satellitales. GLWD (*Global Lake and Wetland Database*) s'appuie sur différentes bases de données de cadastres pour établir la répartition de lacs et de zones inondées. WSLPD (*West Siberian Lowland Peatland Database*) se base sur des observations et référencements de terrain par le gouvernement russe.

de la diffusion du méthane dans les sols, etc.

Deux des paramètres les plus critiques dans ces modèles pour la Sibérie sont l'étendue des zones inondées et le début de la période d'émission au printemps. La Sibérie étant immense et peu peuplée, les informations de cadastre n'existent pas et l'estimation de la superficie des zones inondées peut varier du simple au double (voir figure II.4 et Ringeval et al., 2010). Le début de la période d'émission dépend de la température dans le sol et du niveau de la nappe. A priori, tant que les sols sont enneigés, les émissions sont négligeables. Cependant, des travaux d'observations montrent que l'activité méthanogénique hivernale pourrait être notable (Panikov and Dedysh, 2000; Arshinov et al., 2009; Winderlich, 2012), ce qui n'est pas pris en compte dans les modèles.

2.1.2 Les feux de forêt

Malgré une forte humidité dans les sols en général, certaines zones de Sibérie sont victimes de sécheresse en été. Ces conditions favorisent le déclenchement et la propagation des feux de forêts. Ainsi, plusieurs dizaines de milliers de km² de forêt brûlent chaque année en Sibérie (voir figure II.2), relâchant du méthane dans l'atmosphère.

Selon le même principe que pour les zones inondées, il est possible de quantifier les émissions régionales de méthane en extrapolant les émissions locales. Des expériences de terrain sont effectuées afin de comprendre les mécanismes de propagation des feux et les conditions de combustion de la masse végétale (McRae et al., 2006). D'autres évaluations en laboratoire ont aussi été effectuées afin de compléter les mesures de terrain (Lobert et al., 1991).

Tout comme pour les zones inondées, ces mesures locales sont ensuite extrapolées à plus grande échelle en s'appuyant sur une estimation des surfaces brûlées et de l'intensité des feux. De cette manière, des modèles sont en mesure d'estimer les émissions des feux de forêts à l'aide de cartes de végétation, de zones brûlées et de quantité de matière organique disponible (van der Werf et al., 2010). Pour les zones reculées telles que la Sibérie, cette estimation se fait à l'aide d'images satellitales. Par exemple, la base de donnée GFED (Global Fire Emission Database) propose des estimations mondiales des émissions venant des feux (Giglio et al., 2003, 2010).

Pour la Sibérie, les estimations des émissions de méthane varient entre 1 et 6 TgCH₄.a⁻¹ (Kasischke and Bruhwiler, 2002; Soja et al., 2004; Kukavskaya et al., 2012; Vasileva and Moiseenko, 2013). Simuler les quantités de matière organique brûlées dans chaque feu est imprécis et les bases de données ne sont pas cohérentes quant aux superficies brûlées chaque année en Sibérie. Ceci explique les incertitudes sur les émissions dues aux feux de biomasse.

2.1.3 Les fuites de gaz et autres émissions anthropiques

Les émissions anthropiques sont en général liées aux centres de population. En Sibérie, ces émissions sont ainsi majoritairement confinées au Sud autour du tracé du transsibérien. Néanmoins, l'extraction intense et le transport de gaz et de pétrole au niveau de la plaine de Sibérie occidentale (voir figures II.2 et II.3) contribue aussi aux émissions régionales anthropiques de méthane loin des centres urbains.

La quantification des émissions anthropiques de méthane est essentiellement statistique. À partir d'émissions typiques par secteur, les émissions sont déduites et répertoriées dans des

inventaires (Olivier et al., 2005) en fonction de cartes d'activité économiques, de population, ou autres. Par exemple, des facteurs d'émission sont calculés en fonction de la population pour les émissions venant des décharges et des stations d'épuration. Plus un centre de population est peuplé, plus les émissions des décharges sont élevées. Les facteurs d'émissions peuvent changer d'une région à l'autre en fonction, entre autres, des normes imposées aux différents secteurs.

Les facteurs d'émissions sont généralement déduits de mesures sur le terrain et en laboratoire (E.P.A., 2011). Pour l'industrie gazière et pétrolière, par exemple, Dedikov et al. (1999) et Lechtenböhmer et al. (2005) ont effectué des campagnes intensives de mesures de concentrations de méthane à proximité d'installations gazières en Sibérie occidentale. Les fuites ont été directement détectées et estimées. D'autres études utilisent des mesures éloignées des sources (de quelques centaines de mètres à quelques km; Krings et al., 2013; Mobbs et al., 2013; Allen et al., 2013). Dans ces études, une représentation du panache créé par la source ponctuelle est utilisée et des observations d'une section du panache permettent de déduire l'amplitude de la source.

Les estimations des fuites de méthane ainsi répertoriées s'étalent sur plusieurs ordres de grandeurs. De plus, certaines fuites sont intermittentes, soit qu'elles sont liées à des opérations de maintenance, soit qu'elles sont détectées et réparées. Par exemple, en été, les compagnies pétrolières et gazières en Sibérie profitent d'une diminution de la demande pour effectuer des opérations de maintenance, ce qui peut créer des émissions supplémentaires. En conséquence, les émissions de méthane par l'industrie gazière et pétrolière ne sont pas nécessairement proportionnelles à la production. Les extrapolations à des régions d'émission et à des complexes gaziers et pétroliers sont donc incertaines.

Les émissions de méthane par l'industrie gazière et pétrolière et liées à la distribution de ces ressources en Sibérie sont estimées à 5–35 TgCH₄.a⁻¹ (Reshetnikov et al., 2000; Olivier et al., 2005). Les autres émissions anthropiques (décharges, ruminants principalement) s'élèvent, elles, à environ 2–5 TgCH₄.a⁻¹ (Olivier et al., 2005).

2.1.4 Les hydrates et la végétation

La quantification des émissions de méthane dans l'atmosphère par les hydrates ou par dégradation de la matière foliaire par le rayonnement UV est aussi difficile.

Les hydrates de méthane sur le plateau continental arctique appartiennent à des régions sous-marines peu connues du fait de la présence de la banquise et de l'absence d'installations portuaires à proximité. Par conséquent, la production de méthane dans le fond marin, ainsi que les mécanismes d'oxydation dans la colonne d'eau y sont mal quantifiés. Néanmoins, au Nord, sur le plateau continental arctique, de rares expéditions maritimes ont permis de repérer des hydrates de méthane, ainsi que des zones de suintement de méthane. Par exemple, d'après des mesures de saturation d'eau de mer en méthane, il a été estimé que la mer des Laptevs, alimentée en sédiments par la Léna, pourrait être responsable de 5–10 TgCH₄.a⁻¹ (Shakhova et al., 2010).

Concernant la végétation, les émissions sont faibles mais étalées sur d'immenses surfaces. De surcroît, les expériences en laboratoire proposent des flux ponctuels très différents et donc difficiles à extrapoler. Les émissions végétales pourraient ainsi s'élever à quelques TgCH₄.a⁻¹

en Sibérie. En effet, bien que 10% des zones forestières se trouvent en Sibérie, on s'attend à des émissions végétales régionales inférieures à 10% du total mondial car les espèces dominantes sont des conifères épineux et la zone subit un ensoleillement limité en hiver du fait des hautes latitudes.

2.2 Importantes incertitudes sur le budget sibérien

Les différentes approches « montantes » décrites en sect. 2.1 permettent d'estimer les émissions en Sibérie. Nous disposons ainsi de cartes d'émissions par processus à une résolution de typiquement $0,5 \times 0,5^\circ$. Si l'on somme les contributions de tous les processus d'émission en Sibérie, nous obtenons des émissions de méthane de 20–75 TgCH₄.a⁻¹.

L'intervalle de confiance des émissions sibériennes s'étend donc sur environ 50 TgCH₄.a⁻¹, soit 10% des émissions mondiales de méthane. Ces grandes incertitudes sont liées à des erreurs d'extrapolation, une paramétrisation incertaine des processus de fine échelle et des bases de données incomplètes en Sibérie.

De plus, la variabilité intra- et inter-annuelle des émissions sibériennes est très importante. La production de gaz et de pétrole n'est pas constante au cours des années et des saisons en fonction de la demande et de la capacité d'extraction. Les installations ne sont pas non plus maintenues de la même manière selon les régions. Les émissions anthropiques de méthane varient donc sans que les modèles statistiques d'extrapolation puissent en rendre compte. Il en va de même pour les émissions de zones inondées et des feux présentant une forte variabilité temporelle en fonction des paramètres météorologiques et hydrologiques.

La reproduction des variabilité intra- et inter-annuelle est critique dans la compréhension du cycle du méthane sibérien. Comprendre les mécanismes sous-jacents de ces variabilités est essentiel pour prévoir les émissions sibériennes futures et leur contribution au budget mondial des prochaines décennies. En effet, on peut s'attendre à une augmentation des émissions régionales dans un climat réchauffé avec des étés plus longs et un pergélisol en dégel. Le réchauffement du climat dans la région pourrait favoriser des feux de forêts plus intenses, ainsi que l'activité méthanogénique (Bohn et al., 2007; McGuire et al., 2009) et anthropique avec des projets d'exploitation des gisements de gaz et de pétrole de plus en plus nombreux dans la région et visant jusqu'à l'océan arctique.

Les approches « montantes » décrites précédemment nécessitent donc une validation et un étalonnage extérieurs de leurs paramétrisations pour être fiables et projetables dans le futur.

2.3 Approches atmosphériques

Les émissions contribuent à la variabilité du signal atmosphérique. Analyser le signal atmosphérique dans les régions d'émission et autour peut permettre d'estimer les quantités de méthane émises et les contributions relatives des processus d'émission. On appelle aussi cette approche l'approche « descendante ».

2.3.1 La composition isotopique

Des mesures isotopiques atmosphériques ont été utilisées pour caractériser les émissions de méthane sibériennes. Les variations de teneurs isotopiques en ¹³C et ¹⁴C informent directement sur le processus d'émission (p. ex. Quay et al., 1999).

Le ^{14}C est très majoritairement produit dans l'atmosphère et est absorbé par la matière organique lors de la photosynthèse. C'est un isotope instable du carbone. De faibles teneurs en ^{14}C indiquent donc une origine organique ancienne.

La matière organique dégradée dans les feux de forêt et les zones humides est une matière récente (quelques dizaines d'années), alors que le méthane issu des sources géologiques a des millions d'années. Les émissions de méthane venant de l'utilisation d'énergie fossile diminue donc les teneurs atmosphériques en ^{14}C alors que les zones inondées et les feux les augmentent.

Bergamaschi et al. (1998), Tarasova et al. (2006) et Tarasova et al. (2009) ont mesuré les concentrations de méthane isotopique lors de campagne au sud de la Sibérie sur le tracé du transsibérien. Leurs calculs confirment une importante source biogénique venant des zones inondées en été et des émissions anthropiques sous-estimées dans certaines bases de données.

Sugawara et al. (1996), Yamada et al. (2005) et Umezawa et al. (2012) ont utilisé des mesures aéroportées ponctuelles de teneurs isotopiques au-dessus de zones d'extraction de gaz et de pétrole et de zones inondées. Leurs travaux confirment la source dominante anthropique en hiver (décembre–avril) et biogénique en été (juin–octobre). Le début de la saison d'émission de méthane par les zones inondées reste difficile à estimer du fait des fortes émissions anthropiques régionales.

2.3.2 La co-émission de traceurs atmosphériques

Une approche similaire à l'analyse du signal isotopique est l'analyse d'espèces co-émises par différents types d'émission. Par exemple, les zones inondées sont des zones d'émissions de méthane, mais aussi de croissance végétale. La photosynthèse y est donc active avec un puits net de dioxyde de carbone l'été en journée lié à la photosynthèse et une source la nuit et en hiver du fait de la respiration.

Sasakawa et al. (2012) ont mesuré les variations des concentrations de méthane et de dioxyde de carbone au milieu de zones inondées. Les différences de rapport entre les concentrations de méthane et de dioxyde de carbone le jour et la nuit informent sur les émissions relatives de méthane des zones inondées. Leurs calculs ont permis de conclure à un flux maximum de méthane en juillet dans les zones étudiées.

2.3.3 La modélisation inverse

Les émissions peuvent être mélangées dans l'atmosphère et transportées à de grandes distances. La variabilité des concentrations de méthane en un point donné peut par conséquent être reliée aux émissions et au changement d'origine des masses d'air. Une analyse statistique de la variabilité du signal atmosphérique doublée de l'utilisation d'un modèle de transport à même de déterminer l'origine des masses d'air observées informe sur les émissions des régions survolées par ces masses d'air. Cette approche est appelée modélisation inverse ou inversion atmosphérique.

Winderlich (2012) et Kim et al. (2011) ont utilisé une telle approche avec des sites de mesure continue en surface. Winderlich (2012) a trouvé que les émissions de méthane dues à l'industrie des hydrocarbures étaient sous-estimées en Sibérie par certaines bases de don-

nées (EDGAR v4.2). Kim et al. (2011) ont établi un budget d'émission des zones inondées et d'extraction de gaz et de pétrole, mais avec de grandes incertitudes.

Cette approche est quantitative mais ne permet pas toujours de séparer les différents processus d'émission.

Une nécessité de mieux contraindre le budget sibérien

La Sibérie est une région clé du budget mondial de méthane. Les émissions dans la région pourraient représenter jusqu'à plus de 10% des émissions mondiales de méthane. De plus, ces émissions risquent de s'accroître dans un climat plus favorable à l'activité biologique à l'exploitation des ressources par l'Homme en Sibérie. Il est donc nécessaire de documenter précisément les émissions de la région. Les approches « montantes » ne sont pas suffisantes dans la région. L'approche atmosphérique est prometteuse, en particulier la modélisation inverse quantitative. Nous allons nous efforcer de développer un système analysant le signal atmosphérique afin d'obtenir des estimations fiables des émissions de méthane en Sibérie.

Deuxième partie

Matériels et méthodes

La problématique inverse

Sommaire

1	Transport atmosphérique de traceurs passifs	39
1.1	Dynamique et mélange atmosphériques	39
1.1.1	Phénomènes synoptiques	39
1.1.2	Méso-échelle et échelle locale	40
1.2	Export et mélange de traceurs	42
1.2.1	Les traceurs passifs	42
1.2.2	Transport à grande distance	42
1.2.3	Mélange local vertical de traceurs	42
2	Formulation de l'inversion atmosphérique	44
2.1	Modélisation et simulation du transport atmosphérique	44
2.1.1	Formulation mathématique du monde réel	44
2.1.2	Résolution numérique de la formulation mathématique	45
2.2	Lien réciproque entre sources et concentrations	45
2.2.1	Modélisation directe et inverse	45
2.2.2	Modélisation inverse sous-contrainte	46
2.2.3	Nombreuses incertitudes dans l'inversion	47
3	Résolution du problème inverse atmosphérique	48
3.1	Implémentation des incertitudes et régularisation du problème	48
3.1.1	Le formalisme bayésien	48
3.1.2	Ajout d'informations a priori	50
3.2	Estimation de la densité de probabilité a posteriori	50
3.2.1	L'hypothèse gaussienne	51
3.2.2	Les densités de probabilités nécessaires à l'inversion	51
3.2.3	Les méthodes classiques de résolution d'inversion	53
3.3	L'incertitude statistique manquante	54
3.3.1	Une lacune associée au formalisme bayésien	54

	3.3.2	Estimations des densités de probabilités a priori	54
4		Évaluation des résultats d'une inversion	55
	4.1	Évaluations <i>ad hoc</i>	55
		4.1.1 Confrontation à des données indépendantes	55
		4.1.2 Tests de sensibilité	56
	4.2	Critères algébriques et statistiques	56
		4.2.1 Le critère du χ^2	56
		4.2.2 Cohérence de matrices de covariances	57
		4.2.3 Le maximum de vraisemblance	58

L'approche « montante » ne suffit pas à caractériser les émissions de méthane que l'on souhaite étudier. Des mesures de concentrations atmosphériques de méthane sont disponibles en différents points du globe. L'objectif de l'approche d'inversion atmosphérique est d'obtenir des informations sur les émissions à partir de l'analyse des concentrations atmosphériques, via le transport des émissions. L'inversion atmosphérique se base sur la capacité de l'atmosphère à mélanger et transporter les émissions de méthane sur des grandes distances et sur notre faculté à représenter précisément ce transport.

Nous présentons d'abord, dans la sect. 1, le transport atmosphérique de traceurs passifs et voyons en quoi l'inversion est une approche pertinente dans le cadre de l'objectif recherché. Dans la sect. 2, nous formulons de manière précise le problème de l'inversion. Ensuite, dans la sect. 3, nous montrons comment il est possible de résoudre le problème inverse ainsi formulé. Enfin, nous détaillons, dans la sect. 4, des approches pour évaluer les résultats d'une inversion atmosphérique.

1 Transport atmosphérique de traceurs passifs

1.1 Dynamique et mélange atmosphériques

Des mouvements permanents animent la troposphère et la stratosphère. Les mouvements atmosphériques sont générés par les gradients de température et de pression. Ils sont aussi influencés par la force de Coriolis.

Les courants atmosphériques uniformisent la composition atmosphérique verticalement et horizontalement. Cette uniformisation se fait à plusieurs échelles de temps et d'espace, depuis les circulations climatiques à l'échelle planétaire aux turbulences de fine échelle en passant par les formations cycloniques ou anticycloniques à l'échelle synoptique (voir figure III.1 page 40).

Nous détaillons les schémas de transports synoptiques et locaux cruciaux dans le mélange régional des émissions de méthane.

1.1.1 Phénomènes synoptiques

Les ondes de très grande échelle de la circulation générale génèrent des phénomènes synoptiques, c'est-à-dire à des échelles de typiquement 1000 km ou plus. Les phénomènes synoptiques les plus communs dans le continent eurasiatique sont les anticyclones et les dépressions, accompagnées de cyclones extra-tropicaux et de zones frontales.

Les anticyclones sont des zones de hautes pressions en surface caractérisées par des vents horizontaux faibles et de légers courants descendants. Ils peuvent s'étendre sur plusieurs milliers de km comme l'anticyclone sibérien en hiver (voir encadré 43).

Les cyclones extra-tropicaux accompagnent les zones de basse pression ou dépressions. Ils se caractérisent par des gradients de température et de pression plus prononcés que ceux des anticyclones. La circulation dans les cyclones extra-tropicaux est par conséquent plus rapide et complexe que dans les anticyclones. En particulier, lorsque le cyclone extra-tropical est assez développé, des zones frontales peuvent se développer.

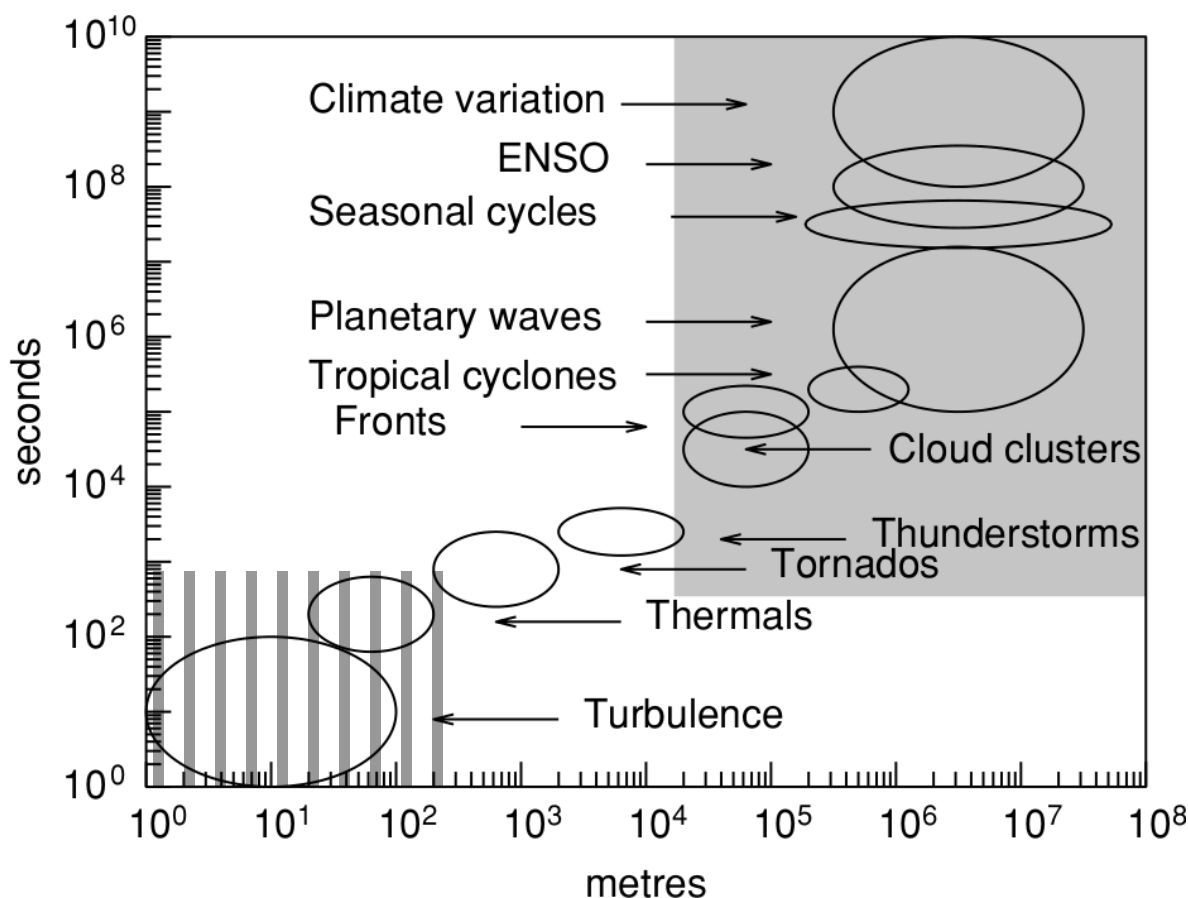


FIGURE III.1: Échelles spatio-temporelles des principaux processus de mélange et de transport dans l'atmosphère (modifié depuis Enting, 2002). En grisé, les processus assez bien représentés dans le cadre du transport à longue distance de traceurs passifs. En hachuré, les échelles sources d'erreurs notables dans le transport de traceurs ; les grilles des modèles de transport atmosphérique et la résolution temporelle de ceux-ci sont en effet généralement plus grand que ces échelles ; ces phénomènes sous-mailles doivent de ce fait être paramétrisés.

Dans les zones frontales, de l'air chaud est importé de la surface vers des altitudes supérieures à 8 km. Il s'agit du front chaud ou *Warm Conveyor Belt* (WCB). En parallèle, un front froid pompe de l'air froid et sec stratosphérique vers la troposphère : c'est l'intrusion froide.

Les fronts sont les principaux moteurs du mélange vertical troposphérique aux moyennes latitudes et des échanges entre la stratosphère et la troposphère (Holton et al., 1995). Ils peuvent aussi transporter horizontalement des masses d'air en quelques heures sur des centaines de km.

Le transport horizontal et vertical dans les zones frontales est illustré par la figure III.2.

1.1.2 Méso-échelle et échelle locale

En dessous de l'échelle synoptique, les déséquilibres énergétiques et la dissipation de l'énergie des phénomènes de grande échelle peuvent engendrer des phénomènes à des échelles

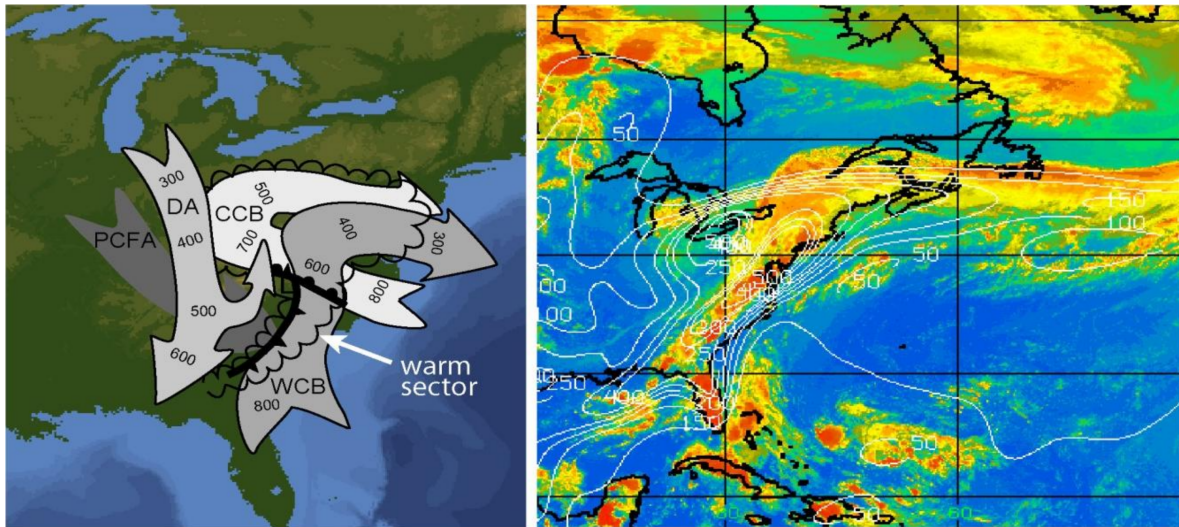


FIGURE III.2: *Mid-latitude cyclone transport. (Left) Conceptual model of the airstreams within a mid-latitude cyclone : WCB, cold conveyor belt (CCB), dry airstream (DA), and the post-cold-front airstream (PCFA). The relationships of the airstreams to the centre of the surface low (L) and the warm and cold fronts are shown, as is the edge of the cloud shield formed by the WCB and CCB (scalloped lines). The numbers on the WCB and CCB indicate the pressure (hPa) at the top of these airstreams, while numbers on the DA indicate the pressure at the bottom of the airstream. From Cooper et al. (2002). (Right) GOES-EAST infrared image on July 27, 2004, with yellows and reds indicating cold, high altitude cloud tops, and blues indicating the Earth's surface. The cloud band along the U.S. east coast is a weak summertime WCB advecting polluted air over the western North Atlantic Ocean, indicated by the white contours which depict an anthropogenic carbon monoxide tracer ($\text{mg}\cdot\text{m}^{-2}$). This polluted air mass was transported to Europe in the upper troposphere over the next four days. From Methven et al. (2006).*

de l'ordre de 50 km. C'est la méso-échelle.

On peut citer parmi ces processus à la méso-échelle la brise entre terre et mer et les formations orageuses. En Sibérie, du fait du climat continental très marqué, des formations orageuses violentes se forment en été. Les orages sont liés à l'ascension d'air instable. Ils induisent de forts courants ascendants de la surface vers la tropopause lors de convections profondes.

À l'échelle locale, la turbulence domine. À proximité de la surface, la vitesse de vent tend vers 0 alors que la circulation moyenne horizontale est non nulle dans l'atmosphère. Le cisaillement de vent engendre de la turbulence localement accentuée par les obstacles tels que les arbres, les bâtiments, etc. À cette turbulence de cisaillement, il faut ajouter une turbulence de convection thermique. La convection thermique de surface est créée par l'apport de chaleur latente ou sensible à l'air par le sol. L'air chauffé monte naturellement et accentue la turbulence de cisaillement.

La turbulence au sol a tendance à uniformiser localement les masses d'air sur la verticale. La couche mélangée à proximité de la surface est appelée couche limite atmosphérique ou couche de mélange. Elle peut être épaisse de quelques dizaines de mètres à 2–3 km. La turbu-

lence thermique est plus active le jour, lorsque la surface est chauffée par le Soleil. La hauteur de la couche limite suit donc un cycle diurne en fonction de l'ensoleillement.

L'encadré 43 détaille l'impact de l'anticyclone sibérien sur le mélange vertical local dans la couche limite.

1.2 Export et mélange de traceurs

La dynamique et le mélange atmosphériques permettent de transporter les traceurs atmosphériques.

1.2.1 Les traceurs passifs

Un traceur passif est un élément transporté dans un fluide sans en influencer l'écoulement.

Le méthane dans l'atmosphère peut être considéré comme un traceur passif. En moyenne, les concentrations de méthane influencent le climat comme nous l'avons vu dans le chapitre I et donc indirectement la circulation atmosphérique. Néanmoins, seul l'effet de serre intégré de toute la masse de méthane atmosphérique est significative. L'effet de serre des masses d'air individuelles a un impact négligeable sur l'écoulement atmosphérique.

Il ne faut pas confondre traceur passif et traceur réactif. L'oxydation du méthane par les radicaux hydroxyles en fait un traceur réactif, bien que passif. Dans la suite, nous négligeons la chimie du méthane car le temps de vie du méthane est long comparé aux échelles de temps du transport atmosphérique régional (8–10 ans contre quelques jours).

1.2.2 Transport à grande distance

Le transport de traceurs peut se faire sur de très longues distances suivant l'écoulement atmosphérique. Par exemple, des accidents nucléaires, tels que l'explosion de la centrale nucléaire ukrainienne de Tchernobyl en 1986 ou de celle de Fukushima en 2011, ont rejeté ponctuellement de grandes quantités d'éléments radioactifs absents de l'atmosphère naturelle. Ces éléments radioactifs ont ensuite été détectés dans tout l'hémisphère nord à des milliers de km du lieu d'émission (Brandt et al., 2002; Quélo et al., 2007; Bocquet, 2012; Winiarek et al., 2012). Il en est de même pour les poussières et gaz relâchés lors d'éruptions volcaniques telles que celle de l'Eyjafjöll en Islande en 2010 qui a perturbé la circulation aérienne civile jusqu'en Europe (p. ex. Stohl et al., 2011).

Connaissant le transport atmosphérique, on peut attribuer une augmentation de concentration à une source. L'analyse du signal atmosphérique est par conséquent un outil pertinent pour augmenter le nombre d'informations disponibles sur les émissions de surface.

1.2.3 Mélange local vertical de traceurs

À l'échelle locale, la variabilité du signal atmosphérique en surface est en partie régulée par le mélange dans la couche limite. En effet, la turbulence locale, ainsi que le mélange thermique (dépendant du gradient thermique vertical) uniformisent les masses d'air dans la couche limite. La hauteur de la couche limite est un paramètre intégrant la turbulence locale et le mélange thermique à proximité de la surface, en fonction de conditions météorologiques et tenant compte du mélange vertical à l'échelle locale. Plus la couche limite est élevée,

L'anticyclone Sibérien

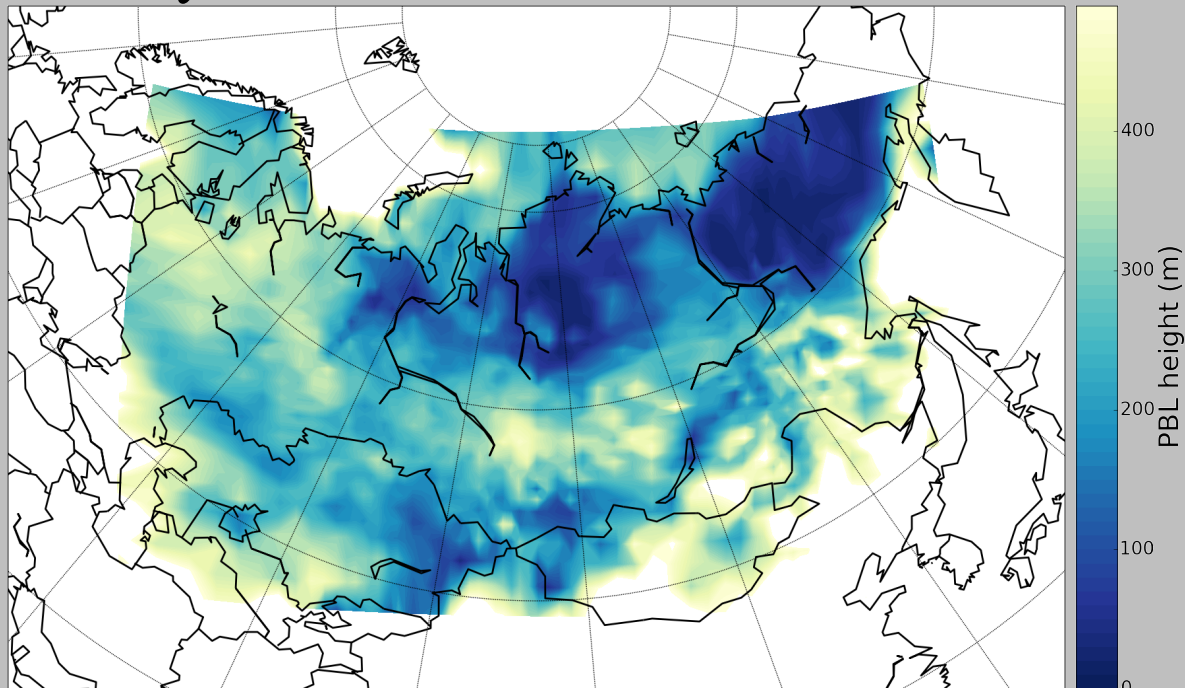


FIGURE III.3: Hauteur de couche limite moyenne en janvier-février entre 2010 et 2014 calculée à partir des données du CEPMMT ERA-INTERIM (Uppala et al., 2005).

En hiver, le faible ensoleillement aux hautes latitudes génère en Sibérie une accumulation de masses d'air froid à proximité de la surface. Cette accumulation favorise la formation d'un immense anticyclone qui accentue à son tour les faibles températures (Cohen et al., 2001). Cet anticyclone sibérien se forme au début de l'hiver (en novembre-décembre) et recouvre en grande partie la plaine et le plateau sibériens. Il ne disparaît qu'à la fin de l'hiver, typiquement en avril.

Cette formation anticyclonique d'échelle synoptique influence significativement la trajectoire des perturbations et des masses d'air dans la région. Elle a la faculté de détourner ou de réduire la force des perturbations arrivant en Sibérie et favorise un climat sec en hiver (précipitations inférieures à $20 \text{ mm} \cdot \text{mois}^{-1}$).

Cet anticyclone agit aussi sur le mélange vertical local dans la couche limite atmosphérique. Sous l'influence de l'anticyclone sibérien, l'air froid s'accumule et les couches d'air en contact avec le sol froid ont tendance à se stratifier plutôt que de s'élever du fait d'un gradient thermique vertical inversé au niveau de la surface. La couche de mélange disparaît alors quasiment ou est limitée à quelques dizaines de mètres. Les gaz émis au niveau du sol sont alors enfermés à proximité de la surface. Ces émissions s'accumulent sur une mince couche dans laquelle les concentrations augmentent fortement.

plus les émissions locales sont mélangées sur une grande hauteur et ont donc une moindre influence sur les concentrations.

Comme la couche limite suit un cycle diurne, à émissions constantes, un cycle diurne de concentrations pourra être relié aux émissions locales (p. ex. Schmidt et al., 2001b; Lopez et al., 2012).

2 Formulation de l'inversion atmosphérique

La circulation atmosphérique relie les émissions aux concentrations de méthane. Nous formalisons ici le problème de l'inversion atmosphérique de méthane, à savoir la déduction des flux de méthane à partir d'observations des concentrations atmosphériques de méthane. Les considérations suivantes se veulent générales et peuvent s'appliquer à d'autres espèces que le méthane.

2.1 Modélisation et simulation du transport atmosphérique

L'approche atmosphérique repose sur le transport atmosphérique. La quantification des émissions par ce moyen dépend donc intrinsèquement de notre capacité à modéliser et simuler le transport de traceurs dans l'atmosphère. Dans la suite, nous utiliserons le verbe « modéliser » pour parler de la transcription du monde réel dans un monde mathématique. Le verbe « simuler » dénotera l'utilisation de modèles numériques pour résoudre les équations du modèle mathématique. Il faudra donc distinguer le monde réel, le monde mathématique et le monde numérique. Ces distinctions sont importantes pour résoudre rigoureusement le problème que l'on se pose.

2.1.1 Formulation mathématique du monde réel

Dans le monde mathématique, on modélise le transport atmosphérique en utilisant la conservation de la matière en tout point. Pour un traceur donné, en tout point \mathbf{r} de l'espace et à tout instant t , le champ de masse de celui-ci est conservé suivant l'équation :

$$\frac{\partial m(\mathbf{r}, t)}{\partial t} = \mathfrak{T}[m(\mathbf{r}, t), t] + s(\mathbf{r}, t) \quad (\text{III.1})$$

où \mathfrak{T} représente le transport et s les « sources ». Ici, on appelle « sources » tant les processus de création de traceurs que les puits. Pour le méthane, s sera donc non nul dans toute l'atmosphère.

L'opérateur de transport \mathfrak{T} peut être modélisé de plusieurs manières. En général, pour le transport atmosphérique, on prend en compte l'advection, c'est-à-dire l'entraînement des molécules de traceur par l'écoulement du fluide, et la diffusion.

L'advection dépend des courants du fluide. Une molécule de traceur se déplace à la même vitesse et dans la même direction que le fluide la transportant.

La diffusion est un terme représentant les phénomènes qu'on sait mal représenter. Dans l'atmosphère, le terme de diffusion prend en compte l'agitation thermique moléculaire à l'origine du mouvement brownien, ainsi que toutes les turbulences de fine échelle que nous ne

savons pas représenter avec précision. En général, nous manquons d'information pour modéliser avec précision ce qu'il se passe dans les turbulences, les thermiques, la convection d'air dans les nuages (voir zone hachurée de la figure III.1). Le terme de diffusion permet de représenter ces phénomènes en moyenne.

2.1.2 Résolution numérique de la formulation mathématique

Une fois ce modèle mathématique admis, vient l'étape de résolution. La formulation du problème est complexe avec des équations aux dérivées partielles dont la résolution formelle n'est pas possible. Nous utilisons donc un modèle numérique pour obtenir une approximation du champ de masse $m(\mathbf{r}, t)$ solution de l'équation de conservation de la matière.

Les modèles numériques reposent sur la discrétisation du monde mathématique. Le monde mathématique continu est en général approximé par un ensemble fini de mailles de modèle. Certains modèles numériques utilisent aussi une représentation spectrale du monde. Nous nous concentrons sur les modèles basés sur un maillage de l'espace et du temps. Pour chaque modèle mathématique, il existe une infinité de maillages différents, suivant les résolutions spatiales et temporelles choisies, la forme des mailles, etc. Cet ensemble de choix est appelé la représentation du problème et est noté ω (suivant le formalisme introduit par Bocquet et al., 2011).

Le choix de la représentation ω est crucial pour bien représenter les différents types de circulation atmosphérique. Par exemple, la circulation frontale est souvent filamenteuse et rapide. Une résolution spatiale et temporelle fine est donc nécessaire pour bien représenter ces phénomènes.

Une fois la représentation ω choisie, il existe plusieurs schémas numériques à même de résoudre de manière discrète l'équation de conservation de la masse (p. ex. Hourdin and Armengaud, 1999). Ces schémas résolvent de manière itérative l'équation de conservation de la masse, c'est-à-dire que la résolution du problème se propage itérativement dans l'espace et le temps à partir de conditions aux bords spécifiées. Les schémas numériques disponibles sont de précision, de complexité et de stabilité numérique variables. Une meilleure précision implique presque toujours une plus grande complexité. Dans certaines conditions, les schémas numériques peuvent devenir instables, une erreur infinitésimale étant amplifiée indéfiniment et faussant ainsi la résolution du problème. Pour éviter ces instabilités, des conditions contraignantes sur les résolutions spatiales et temporelles doivent être imposées.

Les schémas numériques sont choisis en fonction des objectifs de précisions et surtout des moyens de calcul à disposition. Le calcul est enfin effectué sur un ordinateur avec les erreurs numériques inévitables du calcul numérique dues aux troncatures et arrondis effectués par la machine.

2.2 Lien réciproque entre sources et concentrations

2.2.1 Modélisation directe et inverse

L'atmosphère, que ce soit dans le monde réel, mathématique ou numérique, a un comportement déterministe. Pour un champ de sources donné, un champ de composition initiale et un état de l'atmosphère, il existe un unique champ de concentrations possible. Si on note \mathfrak{A} l'opérateur représentant l'atmosphère réelle, on peut ainsi relier les concentrations c aux

sources s :

$$\mathfrak{A}(s) = c \quad (\text{III.2})$$

On dispose de deux autres opérateurs \mathfrak{M} représentant l'atmosphère modélisée, et $\mathfrak{S}_{\mathfrak{M},\omega}$ l'atmosphère simulée en fonction du modèle mathématique et de la représentation ω choisis.

En pratique, on sait calculer $\mathfrak{S}_{\mathfrak{M},\omega}(s)$. Il s'agit d'un champ discret de concentrations modélisées. Ce champ discret nous donne un aperçu des concentrations que l'on peut attendre pour un champ d'émissions donné. Le calcul de ce champ est appelé simulation directe.

La modélisation inverse revient réciproquement à trouver tous les champs de sources possibles remplissant la condition de l'équation III.2 (dans le cas d'observations parfaites). Comme le comportement de l'atmosphère est déterministe, ce problème admet au moins une solution.

2.2.2 Modélisation inverse sous-contrainte

Le problème de la modélisation inverse théorique de l'équation III.2 admet un ensemble de solutions $s \in \mathcal{S}$. Dans l'idéal, on veut trouver une solution unique décrivant ce problème. En pratique, une multitude de solutions existent.

La première des raisons de cette multitude de solutions est la perte d'informations dans l'atmosphère. L'atmosphère agit naturellement comme un intégrateur et mélange de manière irréversible les masses d'air.

La deuxième raison est le manque d'information pour résoudre le problème. Le champ des concentrations est un champ continu dans le monde réel. La documentation des concentrations de méthane et des autres traceurs est coûteuse et fastidieuse. Les moyens d'observations actuels ne permettent même pas une caractérisation quasi-continue du champ de concentrations atmosphériques de méthane. Le champ de concentrations réelles est uniquement échantillonné en différents points de l'espace et du temps. Cet échantillonnage est de plus parcellaire et lacunaire avec seulement quelques dizaines de sites d'observations pérennes (p. ex. Dlugokencky et al., 2009). On note ici \hat{c} l'échantillonnage disponible des concentrations.

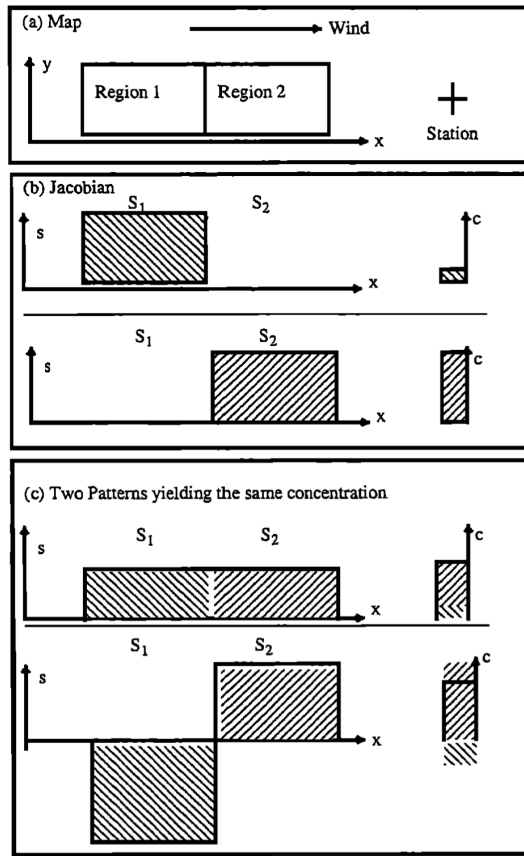
Il s'agit donc de résoudre le problème :

$$\hat{\mathfrak{A}}(s) = \hat{c} \quad (\text{III.3})$$

Le champ de sources s a une très grande dimension : il contient des informations sur la distribution spatiale et la variabilité temporelle des sources auxquelles on s'intéresse. Le problème inverse consistant à quantifier les émissions en fonction des concentrations est donc largement sous-contraint.

Le problème théorique de l'équation III.3 admet donc une infinité de solutions. Des puits et des sources en différents emplacements seront donc compatibles avec les concentrations échantillonnées (voir figure III.4). Cependant, les emplacements de ces sources et puits peuvent être physiquement incompatibles avec d'autres informations extérieures sur les sources : par exemple d'importantes émissions de méthane par des zones inondées dans un désert. Il s'agira dans la suite de distinguer parmi l'ensemble des solutions de l'équation III.3 lesquelles sont les plus pertinentes.

FIGURE III.4: Exemple d'erreurs de représentation dans l'inversion. Cette figure est tirée de Kaminski et al. (2001). (a) An idealized situation of two regions upwind of an observational site. Since region 1 is farther upwind than region 2, for equal emissions S_1 and S_2 , the source S_2 in region 2 has a higher impact on the concentration at the station. (b) Jacobian that quantifies this impact. (c) Two choices of emission patterns that yield the same concentration at the station. While for the first pattern the aggregation over both regions a net source, the second is a net sink. Hence from the atmospheric information alone it is not possible to distinguish a net source from a net sink. This is determined the additional information by provided through choosing a pattern.



2.2.3 Nombreuses incertitudes dans l'inversion

L'inversion atmosphérique telle que l'on se la pose est un problème inverse sous-contraint. Des contraintes insuffisantes dans un système créent des incertitudes. Néanmoins, ces incertitudes ne sont pas les seules à restreindre l'apport d'information sur les émissions via une inversion atmosphérique. Des incertitudes supplémentaires viennent des deux membres de l'équation III.3 et sont résumées dans la figure III.5.

Dans le terme de gauche, on cherche à résoudre $\mathcal{A}(s)$ alors qu'on ne dispose que de $\mathfrak{S}_{\mathfrak{M},\omega}(s)$. Le modèle mathématique choisi \mathfrak{M} nécessite des approximations telles que la diffusion moléculaire et turbulente que l'on n'arrive pas à représenter précisément. Ensuite, le modèle numérique discrétise le monde, ce qui engendre des erreurs supplémentaires sur le calcul.

Dans le terme de droite, on cherche des émissions compatibles avec un échantillonnage de concentrations \hat{c} sujet à des erreurs de mesures. Les erreurs de mesures faites par les instruments utilisés pour déterminer les concentrations de méthane proviennent des techniques de mesures, du bruit instrumental, des protocoles d'étalonnage, etc. (p. ex. Vermeulen et al., 2005; Winderlich et al., 2010; Ramonet et al., 2011).

De plus, des incertitudes viennent de la comparaison de $\mathfrak{S}_{\mathfrak{M},\omega}(s)$ dans le monde numérique discret à l'échantillonnage \hat{c} de concentrations réelles. Dans un modèle numérique,

les mailles sont considérées comme uniformes. Tous les processus internes à la maille sont moyennés. De la sorte, une turbulence locale influençant une composante de \hat{c} ne sera pas nécessairement prise en compte dans $\mathfrak{S}_{\mathfrak{M},\omega}(s)$.

Pour résumer, l'inversion atmosphérique s'appuie sur l'équation suivante :

$$\mathfrak{A}(s) = \mathfrak{S}_{\mathfrak{M},\omega}(s) + \epsilon_{phys} + \epsilon_{num} + \epsilon_{discr} = \hat{c} + \epsilon_{obs} \quad (\text{III.4})$$

où ϵ_{phys} , ϵ_{num} , ϵ_{discr} et ϵ_{obs} sont respectivement les erreurs (par définition inconnues) dans le modèle physique, le calcul numérique, dans la discrétisation et la comparaison aux observations, et dans les mesures.

3 Résolution du problème inverse atmosphérique

3.1 Implémentation des incertitudes et régularisation du problème

Les incertitudes sont nombreuses dans l'approche de l'inversion atmosphérique. Nous choisissons d'exprimer notre manque d'informations exactes sur le système en des termes statistiques. L'objectif de l'inversion dans ce contexte est d'interpréter statistiquement les informations contenues dans les observations et d'en déduire une estimation statistique des émissions. Pour cela, nous nous appuyons sur le formalisme bayésien.

3.1.1 Le formalisme bayésien

La règle de Bayes, énoncée par Pierre Simon Laplace au XVIII^e siècle et attribuée au père Thomas Bayes, fournit un formalisme pour l'utilisation de probabilités conditionnelles.

En inversion atmosphérique, on cherche à estimer la densité de probabilité de l'état du système en connaissant les observations dont on dispose : $p(\mathbf{x}|\mathbf{y}^o)$, appelée densité a posteriori. Ici, \mathbf{x} est un vecteur appelé vecteur d'état représentant l'état du système auquel on s'intéresse, typiquement le champ des sources. \mathbf{y}^o est le vecteur d'observation regroupant les observations dont nous disposons. Ils appartiennent respectivement aux espaces vectoriels d'état et d'observation.

L'état réel du système est représenté par \mathbf{x}^t . Il peut se situer n'importe où dans la densité de probabilité $p(\mathbf{x}|\mathbf{y}^o)$. Néanmoins, il est attendu que, en moyenne, les composantes de \mathbf{x}^t soient proches des modes (à savoir au maximum de probabilité) des densités de probabilités partielles de $p(\mathbf{x}|\mathbf{y}^o)$ (voir exemple en figure III.6).

Le formalisme bayésien nous permet alors d'écrire la densité de probabilité de l'état du système en fonction des observations ainsi :

$$p(\mathbf{x}|\mathbf{y}^o) = \frac{p(\mathbf{y}^o|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y}^o)} \quad (\text{III.5})$$

- \mathbf{y}^o est connu et fixé. $p(\mathbf{y}^o)$ est donc une constante qu'il n'est pas nécessaire de connaître pour déterminer la forme de $p(\mathbf{x}|\mathbf{y}^o)$.
- $p(\mathbf{y}^o|\mathbf{x})$ représente la densité de probabilité d'obtenir un vecteur d'observation en fonction d'un état spécifique du système. Cette densité de probabilité dépend du modèle atmosphérique $\mathfrak{S}_{\mathfrak{M},\omega}$ et des erreurs ϵ_{phys} , ϵ_{num} , ϵ_{discr} et ϵ_{obs} , décrits en sect. 2.2.3.

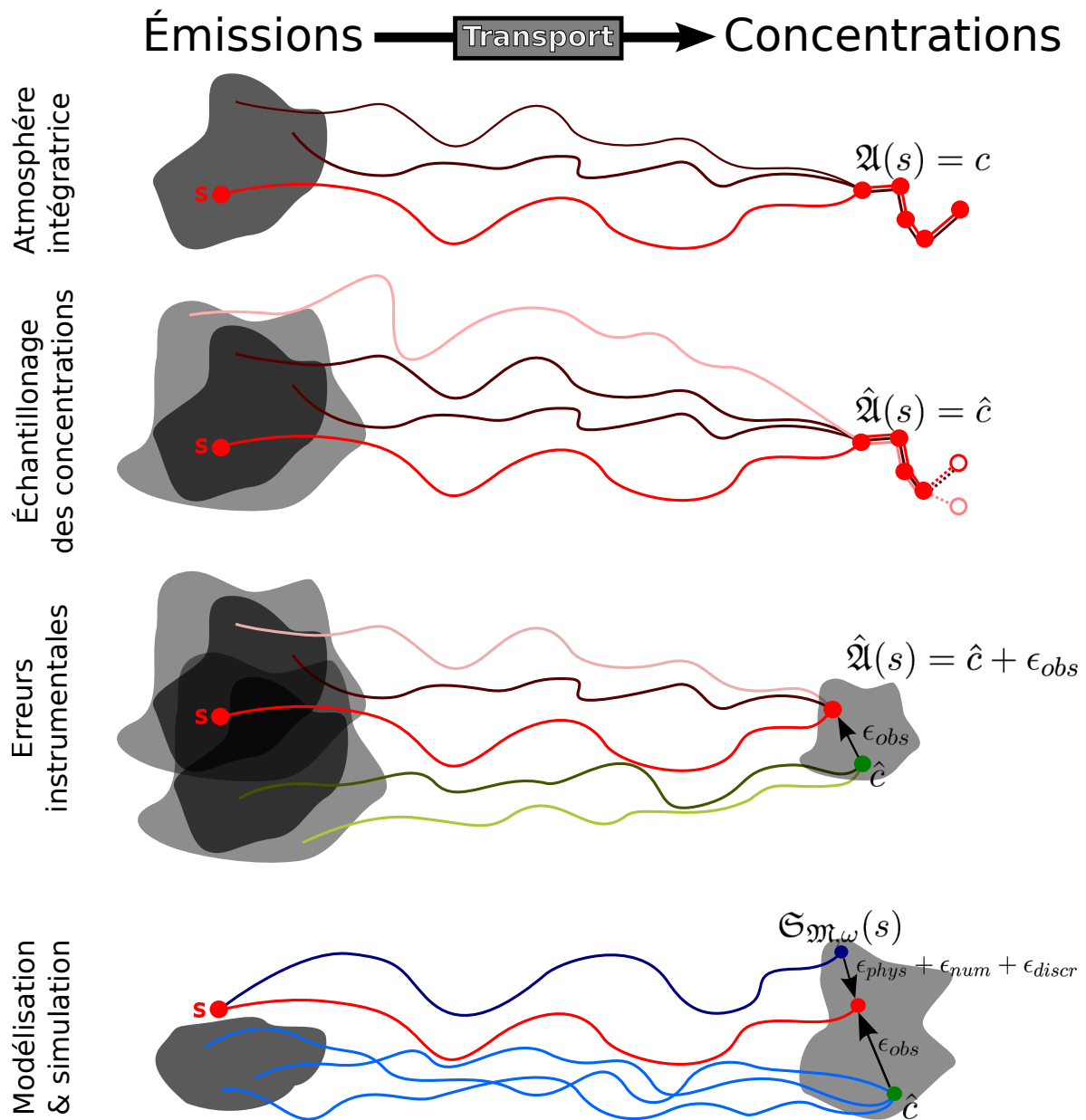


FIGURE III.5: Incertitudes dans l'inversion. Dans la colonne de gauche l'espace des émissions ; à droite, l'espace des concentrations ; les deux espaces étant reliés par le transport atmosphérique. En rouge, la « réalité ». Atmosphère intégratrice : du fait du caractère intégrateur, plusieurs champs d'émissions sont compatibles avec un champ de concentrations donné. Échantillonnage des concentrations : seule une partie du champ des concentrations est connue, ce qui augmente le nombre de champs d'émissions compatibles ; des champs d'émissions peuvent être compatibles avec les concentrations échantillonnées \hat{c} (disque rouges pleins), mais pas avec le champ complet de concentrations (cercles rouges). Erreurs instrumentales : les concentrations sont connues avec une incertitude instrumentale ; un ensemble de concentrations est possible avec les champs d'émissions associés. Modélisation et simulation : le transport est simulé, en particulier, en discrétisant le monde ; le champ de sources « réel » est transporté différemment dans le modèle (ligne bleu foncé) et dans la réalité ; les émissions compatibles avec les observations (lignes bleu clair) n'incluent donc pas nécessairement la « réalité ».

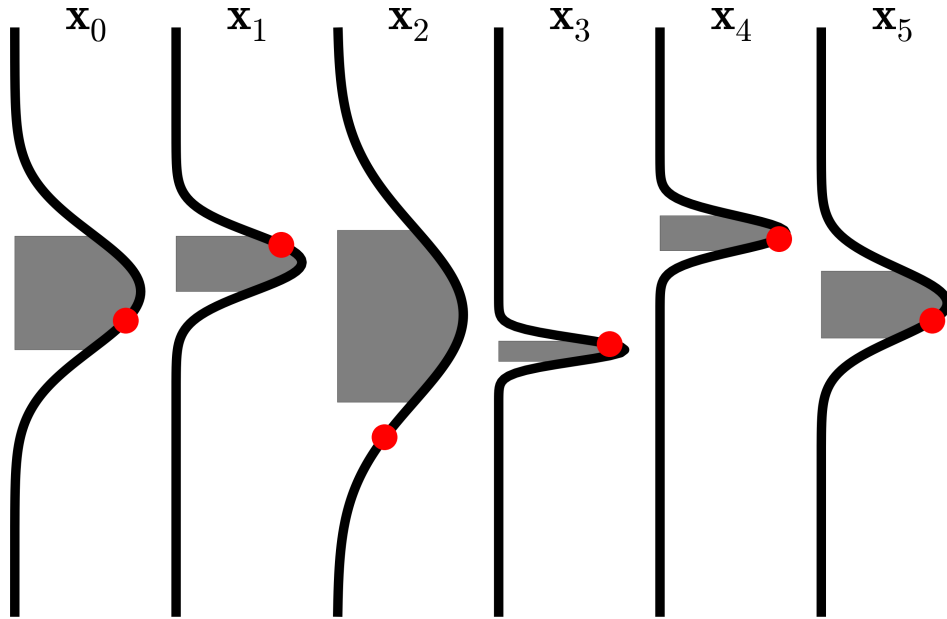


FIGURE III.6: Distributions a posteriori de six composantes x_i du vecteur d'état $\mathbf{x} \propto \mathcal{N}(\mathbf{x}^a, \mathbf{P}^a)$. Les zones grisées représentent l'intervalle de confiance à 66% de la composante x_i : $[x_i^a \pm (\mathbf{P}_i^a)^{1/2}]$. En rouge, les composantes x_i^t de la « réalité » \mathbf{x}^t . Dans la plupart des cas, x_i^t est proche du mode de la gaussienne, mais pas systématiquement (par exemple x_2^t).

- $p(\mathbf{x})$ représente la probabilité générale d'obtenir un état plutôt qu'un autre a priori.

3.1.2 Ajout d'informations a priori

À ce niveau de la définition du problème, $p(\mathbf{x})$ n'est pas connu, ce qui bloque la résolution de l'inversion.

Dans la sect. 2.1 du chapitre II, nous avons décrit des approches directes de quantifications des sources de méthane. Ainsi, grâce à ces approches, nous disposons d'un certain savoir sur les flux de méthane (distributions possibles, quantités émises, etc.). Nous pouvons transformer ce savoir en le modélisant par une densité de probabilité a priori pour les émissions $p(\mathbf{x})$.

Cette densité de probabilité a priori assure une certaine cohérence physique pour les flux déduits de l'inversion bayésienne. Des flux biogénique de méthane depuis une zone désertique seront par exemple très improbables dans $p(\mathbf{x})$.

3.2 Estimation de la densité de probabilité a posteriori

Le formalisme bayésien nous fournit un outil conceptuel efficace pour estimer la densité de probabilité des sources a posteriori. Le calcul de $p(\mathbf{x}|\mathbf{y}^o)$ se fait par simple multiplication des densités de probabilité $p(\mathbf{y}^o|\mathbf{x})$ et $p(\mathbf{x})$. Estimer une densité produit de plusieurs densités peut se faire à l'aide d'un échantillonnage de Monte-Carlo. Cependant, la grande dimension des vecteurs en présence empêche une telle approche.

Des hypothèses sont généralement faites quant aux formes des densités de probabilité afin de simplifier les calculs et d'explicitier la densité de probabilité a posteriori $p(\mathbf{x}|\mathbf{y}^o)$.

3.2.1 L'hypothèse gaussienne

De nombreux phénomènes physiques aléatoires peuvent être approximés par des variables aléatoires normales suivant des lois gaussiennes. Ce fait découle du théorème de la limite centrale (énoncé par Pierre Simon Laplace) qui spécifie qu'une somme d'un grand nombre de variables aléatoires identiques tend vers une loi normale. Or, les erreurs physiques de l'inversion (dans les observations et les flux a priori) sont générées par un grand nombre de petites causes. Dans toute la suite, nous faisons donc l'hypothèse que toutes les incertitudes présentes dans le problème inverse que l'on cherche à résoudre sont gaussiennes. Ainsi, ϵ_{phys} , ϵ_{num} , ϵ_{discr} et ϵ_{obs} peuvent être représentées par des variables aléatoires suivant une loi normale.

Cette hypothèse est parfois mise en doute dans les inversions bayésiennes (Chevallier et al., 2006; Michalak and Kitanidis, 2005). Les distributions d'erreurs ont une forme qui ne suit pas toujours une gaussienne. En particulier, des limites physiques existent sur les variables considérées (p. ex. des concentrations ne peuvent pas être négatives), alors que les lois normales sont non bornées. Des travaux ont été menés afin d'estimer des densités de probabilité plus représentatives de la réalité (par exemple des gaussiennes tronquées; Lauvernet et al., 2009). Néanmoins, l'approximation gaussienne permet de simplifier grandement les calculs et de rendre possible un certain nombre de manipulations algébriques sans pour autant détériorer radicalement la qualité de l'inversion (Tarantola, 1987).

Dans l'hypothèse gaussienne, un vecteur aléatoire peut être entièrement décrit par le mode de la distribution et par la matrice de covariances. La loi normale étant une loi symétrique, le mode de la distribution n'est autre que l'espérance de la variable aléatoire. La matrice de covariance décrit les variances de chaque composante du vecteur aléatoire et les covariances entre composantes. Il s'agit d'une matrice symétrique définie positive, c'est-à-dire une matrice inversible et dont toutes les valeurs propres sont positives.

Pour une variable aléatoire $\mathbf{c} \in \mathbb{R}^n$ suivant une loi gaussienne d'espérance $\bar{\mathbf{c}} = \mathbb{E}[\mathbf{c}] \in \mathbb{R}^n$ et de matrice de covariance $\mathbf{C} = \mathbb{E}[\mathbf{c}\mathbf{c}^T] \in \mathcal{S}_n^{++}(\mathbb{R})$, on peut écrire :

$$p(\mathbf{c}) = \mathcal{N}(\bar{\mathbf{c}}, \mathbf{C}) \propto \exp \left\{ -\frac{1}{2}(\mathbf{c} - \bar{\mathbf{c}})^T \mathbf{C}^{-1}(\mathbf{c} - \bar{\mathbf{c}}) \right\} \quad (\text{III.6})$$

Ici, l'opérateur $\mathbb{E}[\cdot]$ désigne l'espérance statistique d'une variable aléatoire et $(\cdot)^T$ l'opérateur transposé.

3.2.2 Les densités de probabilités nécessaires à l'inversion

Nous avons vu que la densité de probabilité a posteriori est proportionnelle au produit de deux autres densités de probabilité : $p(\mathbf{y}^o|\mathbf{x})$ et $p(\mathbf{x})$.

La densité de probabilité $p(\mathbf{x})$ dépeint le savoir a priori dont nous disposons sur l'état du système. Dans le contexte gaussien, le mode de sa distribution représente la meilleure estimation de l'état avant l'inversion.

On note :

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}^b, \mathbf{B}) \quad (\text{III.7})$$

où \mathbf{x}^b est le vecteur a priori (ou *background*).

Dans la suite, on posera :

$$\epsilon^b = \mathbf{x}^b - \mathbf{x}^t \propto \mathcal{N}(\mathbf{0}, \mathbf{B}) \quad \text{et} \quad \mathbf{B} = \mathbb{E}[\epsilon^b(\epsilon^b)^T] \quad (\text{III.8})$$

Le calcul de $p(\mathbf{y}^o|\mathbf{x})$ nécessite de relier observations et état du système. Pour cela, on introduit l'opérateur d'observation \mathcal{H} . Cet opérateur prend en compte le modèle mathématique utilisé \mathfrak{M} , la représentation ω , le modèle numérique $\mathfrak{S}_{\mathfrak{M},\omega}$ et l'échantillonnage $\hat{\cdot}$.

Ainsi, on voit que :

$$p(\mathbf{y}^o|\mathbf{x}) \propto p(\mathbf{y}^o - \mathcal{H}(\mathbf{x})) \propto \mathcal{N}(\eta, \mathbf{R}) \quad (\text{III.9})$$

η dénote les biais systématiques dans les observations et les modèles. Les efforts de la communauté d'observation et de modélisation se concentrent en partie sur la suppression de tels biais. Il existe aussi des moyens statistiques de détecter de tels biais (Dee, 2005). Bien que l'objectif de suppression des biais ne soit pas toujours parfaitement atteint en inversion atmosphérique, nous ne cherchons pas à diagnostiquer précisément les biais et considérons dans la suite qu'ils sont négligeables :

$$\eta = \mathbf{0} \quad (\text{III.10})$$

On notera donc :

$$\mathbf{y}^o - \mathcal{H}(\mathbf{x}) = \epsilon^o \propto \mathcal{N}(\mathbf{0}, \mathbf{R}) \quad \text{et} \quad \mathbf{R} = \mathbb{E}[\epsilon^o(\epsilon^o)^T] \quad (\text{III.11})$$

La variable aléatoire ϵ^o inclue les erreurs ϵ_{phys} , ϵ_{num} , ϵ_{discr} et ϵ_{obs} .

L'opérateur d'observation est souvent considéré comme linéaire : si on double les émissions, les concentrations sont doublées. Cette hypothèse, en général justifiée, n'est plus valide pour le méthane lorsque l'on prend en compte la chimie puisque les taux de réaction dépendent des concentrations. Aux échelles de transport régional considérées, on néglige la chimie du méthane, ce qui justifie l'hypothèse linéaire. Par ailleurs, bien que l'hypothèse linéaire soit réaliste physiquement pour la plupart des traceurs, certains schémas numériques de résolution du transport atmosphérique ne sont pas linéaire (par exemple le schéma de van Leer). Nous négligeons dans la suite les non linéarités numériques des schémas de transport.

Pour les traceurs passifs inertes, \mathcal{H} est approximé par sa matrice jacobienne \mathbf{H} telle que :

$$\mathbf{H}_{i,j} = \frac{\partial \mathbf{y}_i^o}{\partial \mathbf{x}_j} \quad (\text{III.12})$$

En plus du caractère non-biaisé des erreurs, nous admettons que les erreurs d'observations et dans les flux a priori sont décorréliées. Cela implique que :

$$\mathbb{E}[\epsilon^o(\epsilon^b)^T] = \mathbf{0} \quad \text{et} \quad \mathbb{E}[\epsilon^b(\epsilon^o)^T] = \mathbf{0} \quad (\text{III.13})$$

3.2.3 Les méthodes classiques de résolution d'inversion

Sous les conditions de gaussianité des densités de probabilité, de la linéarité de l'opérateur \mathbf{H} et du caractère non-biaisé des erreurs, la densité de l'état du système a posteriori s'écrit :

$$p(\mathbf{x}|\mathbf{y}^o) \propto \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - \frac{1}{2}(\mathbf{y}^o - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}) \right\} \quad (\text{III.14})$$

Cette densité de probabilité suit une loi normale en tant que produit d'un nombre fini de lois normales. Il existe donc un vecteur \mathbf{x}^a et une matrice de covariance \mathbf{P}^a qui la décrivent entièrement :

$$p(\mathbf{x}|\mathbf{y}^o) \propto \mathcal{N}(\mathbf{x}^a, \mathbf{P}^a) \quad (\text{III.15})$$

\mathbf{x}^a est appelé analyse du problème. Il représente l'estimation de l'état du système après l'inversion. La matrice de covariance \mathbf{P}^a décrit les incertitudes sur l'état du système a posteriori.

Il existe trois manières principales d'évaluer \mathbf{x}^a et \mathbf{P}^a : l'inversion analytique, l'inversion variationnelle et les méthodes d'ensembles.

L'inversion analytique L'inversion analytique, parfois appelée *Best Linear Unbiased Estimator* ou *BLUE*, s'appuie sur des manipulations algébriques simples. On peut en effet montrer que le vecteur d'analyse et la matrice de covariances a posteriori valent :

$$\begin{cases} \mathbf{x}^a &= \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \\ \mathbf{P}^a &= \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B} \end{cases} \quad (\text{III.16})$$

avec $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$, appelée matrice de gain de Kalman.

L'inversion analytique se base sur des produits matriciels et des inversions de matrices simples. Cependant, les problèmes d'inversion atmosphérique sont généralement de grande dimension. Le calcul de produits et d'inverses de matrices peut devenir très lourd et limitant en grande dimension.

L'inversion variationnelle La deuxième approche de résolution de l'inversion permet d'éluider ce problème. Dans cette approche, on remarque que le vecteur d'analyse est co-localisé avec le minimum de la fonction objective ou fonction coût suivante :

$$\begin{aligned} \mathcal{J}(\mathbf{x}) &= -\ln p(\mathbf{x}|\mathbf{y}^o) \\ &= \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}) + cte \\ &\equiv \mathcal{J}^b(\mathbf{x}) + \mathcal{J}^o(\mathbf{x}) \end{aligned} \quad (\text{III.17})$$

Il suffit alors d'implémenter un algorithme de minimisation plus léger que le calcul explicite des produits matriciels et des inverses de matrices.

En revanche, on ne peut obtenir la matrice de covariances a posteriori par ce moyen. Théoriquement, la matrice de covariance \mathbf{P}^a est égale à l'inversion de la hessienne de la fonction de coût \mathcal{J} en \mathbf{x}^a . Il est possible d'explicitier cette hessienne et le calcul nous mène à une expression compatible avec l'équation III.16.

Les méthodes d'ensembles Cependant, le problème des produits matriciels et inverses en grandes dimensions réapparaît immédiatement. Des approches par échantillonnage de Monte Carlo sont parfois utilisées pour ne pas avoir à calculer la hessienne (p. ex. Chevallier et al., 2007; Broquet et al., 2011).

Ces méthodes d'ensembles s'appuient principalement sur la formulation du filtre de Kalman d'ensemble (Houtekamer and Mitchell, 1998; Evensen, 2003). Il consiste principalement à calculer les matrices de covariances à partir des covariances des échantillons.

3.3 L'incertitude statistique manquante

3.3.1 Une lacune associée au formalisme bayésien

Le formalisme bayésien nous permet de résoudre au mieux le problème de l'inversion atmosphérique si l'on s'affranchit de contraintes techniques posées par la taille du problème. Ce formalisme agence de manière optimale les informations incomplètes dont on dispose afin de caractériser objectivement le savoir que les observations atmosphériques nous apportent sur les sources.

Néanmoins, dans cette formulation du problème, les incertitudes statistiques ne sont pas prises en compte. En effet, les méthodes de résolution de l'inversion reposent sur une connaissance parfaite des densités de probabilité implémentées dans le système. Or, ces densités décrivent des erreurs faites au fil du problème par rapport à une réalité que l'on ne connaît pas.

Les variables indispensables à la définition de notre problème sont donc mal documentées, ce qui peut détériorer la qualité des résultats de l'inversion.

Les études se basant sur des inversions atmosphériques ont très vite noté cette lacune (p. ex. Rayner et al., 1999). Cette faiblesse n'a rien d'anodin car les bases de données utilisées pour construire les flux a priori ne permettent pas de construire la matrice de covariance \mathbf{B} et notre connaissance de l'atmosphère et de sa reproduction par un modèle de transport est trop faible pour expliciter \mathbf{R} . Ainsi, les matrices choisies le sont en fonction du « jugement expert » que l'on possède (p. ex. Bousquet et al., 1999a; Bergamaschi et al., 2010).

L'incertitude statistique est d'autant plus importante que le problème que l'on cherche à résoudre est sous-contraint. Dans ces conditions, le résultat de l'inversion est très sensible aux statistiques que l'on prescrit a priori. Des études d'inversion atmosphérique ont testé différentes matrices de covariances crédibles physiquement, mais aboutissant à des flux a posteriori sensiblement différents (Kaminski et al., 1999; Bergamaschi et al., 2010; Peylin et al., 2013). Engelen et al. (2002) ont exploré de manière systématique les impacts de la mauvaise reproduction des matrices de covariance et ont estimé que l'incertitude induite sur les flux a posteriori pouvait atteindre 100%.

3.3.2 Estimations des densités de probabilités a priori

Du fait de la sensibilité des résultats d'une inversion aux statistiques d'erreurs prescrites dans le système, des efforts ont été faits pour améliorer notre connaissance du comportement des modèles utilisés pour le transport atmosphérique et de leur possible comparaison à des observations.

Des comparaisons de différents modèles de transport ont été effectuées afin de caractériser les incertitudes spécifiquement attribuables à une mauvaise représentation du transport (Peylin et al., 2002; Baker et al., 2006; Geels et al., 2007; Prather et al., 2008). Par exemple, suivant les modèles utilisés, une masse d'air particulière pourra voir sa trajectoire changer légèrement. Du point de vue d'un échantillonnage de concentration, une masse d'air placée au mauvais endroit provoquera une inadéquation entre le signal observé (qui voit par exemple la masse d'air) et le signal simulé (qui ne la voit pas). Les erreurs de transport viennent aussi d'une mauvaise représentation des cycles diurnes, des situations d'accumulation de traceur près de la surface et des zones à forts gradients de concentrations comme les zones frontales.

De telles études suggèrent des erreurs pouvant dépasser 1% du signal. Cela représente pour le méthane des erreurs de transport de 20 ppb. À titre de comparaison, l'Organisation Météorologique Mondiale (OMM) impose des spécificités de précisions instrumentales de l'ordre de 1 ppb (WMO/GAW, 2011). Les erreurs de transport sont donc dominantes par rapport aux erreurs instrumentales dans le problème de l'inversion des flux de méthane à partir de mesures de surface de concentrations de méthane.

En outre, l'inversion repose sur la comparaison de deux mondes : le monde réel des concentrations échantillonnées et le monde simplifié des modèles. Cette comparaison se base sur des approximations assez fortes comme la non prise en compte des processus internes aux mailles. Gerbig et al. (2003b) et Tolk et al. (2008) ont estimé de telles erreurs en comparant des modèles à différentes résolutions. Kaminski et al. (2001) et Bocquet et al. (2011) ont formalisé plus précisément les erreurs induites par la discrétisation du problème sur le résultat de l'inversion. Ces erreurs pourraient atteindre 100% des estimations des flux a posteriori.

On en déduit que, dans l'équation III.4, les erreurs de simulation ϵ_{num} et de discrétisation ϵ_{discr} dominant probablement sur les autres sources d'erreur. Néanmoins, des informations plus précises sont nécessaires pour réduire l'incertitude statistique de l'inversion atmosphérique bayésienne.

4 Évaluation des résultats d'une inversion

Le formalisme bayésien est robuste uniquement si l'on néglige l'incertitude statistique. Comme nous l'avons vu, cette incertitude peut avoir un impact significatif sur les inversions. Une inversion atmosphérique bayésienne nécessite donc une évaluation de ses résultats et de la robustesse de ceux-ci.

4.1 Évaluations *ad hoc*

Des évaluations sont effectuées de manière *ad hoc* après l'inversion en confrontant les résultats à des observations indépendantes ou en estimant la sensibilité des résultats de l'inversion à différents paramètres jugés critiques.

4.1.1 Confrontation à des données indépendantes

De nombreuses inversions atmosphériques recourent à cette méthode d'évaluation peu coûteuse à conditions de disposer d'observations indépendantes. Le principe est d'effectuer l'inversion atmosphérique puis de comparer les concentrations simulées a priori et a poste-

riori à des mesures non utilisées par l'inversion. L'inversion atmosphérique améliore nécessairement les concentrations simulées pour les sites assimilés, mais des artefacts de la méthode peuvent détériorer les concentrations simulées à d'autres endroits (voir l'exemple de la figure III.4).

4.1.2 Tests de sensibilité

Une autre méthode d'évaluation consiste à estimer la sensibilité du système d'inversion aux observations assimilées et aux flux a priori.

En principe, si l'on ôte des observations au système, les flux seront moins contraints, c'est-à-dire que les incertitudes a posteriori seront plus grande. Mais les incertitudes a posteriori dans le système complet et le système réduit doivent rester compatibles. De plus, les observations ôtées du système d'inversion peuvent être considérées comme des données indépendantes, ce qui permet l'application du test de la sect. 4.1.1. Dans le cas contraire, cela indique que des artefacts dominent le système qui doit alors être modifié. Cette méthode d'évaluation s'appelle la validation croisée ou *leave-one-out validation*.

Par exemple, Law et al. (2003) et Bergamaschi et al. (2005) ont testé un système d'inversion avec différents sous-jeux d'observations et ont évalué la cohérence entre les différentes expériences. Dans leur cas, le système semblait supporter de manière cohérente le retrait d'un ensemble de données.

Il est aussi possible d'implémenter des bases de données de flux différentes dans le système d'inversion (p. ex. Bergamaschi et al., 2010). Bien qu'en désaccord, différentes bases de données peuvent être tout à fait cohérentes physiquement. Aucune de ces différentes bases ne doit être favorisée dans l'inversion. Si le système d'inversion est cohérent, les résultats à partir des différentes bases de données doivent être compatibles statistiquement.

4.2 Critères algébriques et statistiques

Les méthodes *ad hoc* sont un bon moyen de détecter des problèmes dans un système d'inversion. Mais elles ne permettent pas de circonscrire ou de corriger le problème détecté. Des critères mathématiques objectifs de cohérence statistique existent pour évaluer un système d'inversion. Ces critères ont été développés par la communauté des mathématiciens appliqués des problèmes inverses. Ils ont été adapté par et pour la communauté de prévision météorologique (p. ex. Dee, 1995) pour mettre en évidence des failles d'un système d'inversion, mais aussi pour optimiser de tels systèmes. Ils sont encore assez marginalement utilisés par la communauté d'inversion de traceurs atmosphériques.

4.2.1 Le critère du χ^2

Un loi de probabilité de χ^2 est une loi formée par une somme de carrés de plusieurs lois normales. La fonction de coût \mathcal{J} (définie dans l'équation III.17) suit donc une loi de χ^2 .

En admettant que le système d'inversion est défini de manière optimale, c'est-à-dire que les matrices de covariances \mathbf{R} et \mathbf{B} représentent bien les covariances des erreurs dans le système, il est possible de démontrer par des manipulations algébriques (Talagrand, 1998; Ménard et al., 2000; Desroziers and Ivanov, 2001; Chapnik et al., 2004) que pour tout sous-ensemble j d'observations ou de composantes du vecteur d'état, l'espérance de la fonction de

coût sur le sous-ensemble $\mathbb{E}[\mathcal{J}_j]$ respecte l'équation suivante :

$$\mathbb{E}[\mathcal{J}_j] = \frac{1}{2} \left\{ m_j - \text{tr}(\Gamma_j^T \mathbf{S}_j^{-1} \Gamma_j \mathbf{P}^a) \right\} \quad (\text{III.18})$$

avec m_j , Γ_j et \mathbf{S}_j respectivement la taille de l'ensemble j , le sous-opérateur de \mathbf{H} associé et la matrice de covariance dans le sous-ensemble.

Lorsque l'on applique cette formule aux espaces des observations et d'état dans leur ensemble, on obtient :

$$\begin{cases} \mathbb{E}[\mathcal{J}^o] &= \frac{1}{2} \text{tr}(\mathbf{I}_p - \mathbf{H}\mathbf{K}) \\ \mathbb{E}[\mathcal{J}^b] &= \frac{1}{2} \text{tr}(\mathbf{K}\mathbf{H}) \end{cases} \quad (\text{III.19})$$

avec p le nombre d'observations disponibles et \mathbf{I}_p la matrice identité de dimension $p \times p$.

Dans ces équations, les matrices $\mathbf{H}\mathbf{K}$ et $\mathbf{K}\mathbf{H}$ jouent un rôle particulier. Il s'agit respectivement des matrices de sensibilité et d'influence (Cardinali et al., 2004). La matrice de sensibilité dénote la sensibilité de l'inversion à un changement élémentaire dans le vecteur des observations alors que la matrice d'influence est à relier à l'influence qu'a l'inversion sur les flux.

En sommant ces deux équations, on voit ainsi que la somme de $\mathbb{E}[\mathcal{J}^o]$ et de $\mathbb{E}[\mathcal{J}^b]$ a une valeur fixe proportionnelle au nombre d'observations :

$$\mathbb{E}[\mathcal{J}] = p/2 \quad (\text{III.20})$$

La loi de χ^2 associée à la fonction de coût \mathcal{J} est donc une loi de χ^2 avec p degrés de liberté. De surcroît, cela confirme que plus la sensibilité aux observations est élevée, plus l'inversion aura une grande influence sur les flux, et inversement.

En outre, lorsque p est grand (dès quelques dizaines), la loi χ^2 est très reserrée sur son mode. On peut donc en déduire que :

$$\mathcal{J}(\mathbf{x}^a) \sim \mathbb{E}[\mathcal{J}] = p/2 \quad (\text{III.21})$$

Estimer la fonction de coût en son minimum et comparer celui-ci au nombre d'observation est donc un diagnostic peu coûteux de la bonne définition du problème. Ce test, dit du χ^2 , est utilisé après l'inversion dans plusieurs études pour contrôler la cohérence du système (p. ex. Bergamaschi et al., 2005). Il est aussi parfois utilisé pour optimiser le système avant inversion (p. ex.)

4.2.2 Cohérence de matrices de covariances

Le test du χ^2 s'appuie sur la statistique de la fonction coût et du vecteur d'analyse. Ce critère fournit une contrainte unidimensionnelle sur les matrices de covariances \mathbf{R} et \mathbf{B} , donc relativement faible. Il est possible de déduire des propriétés plus fortes à l'aide de manipulations algébriques sur les vecteurs d'innovation du type $\mathbf{y}^o - \mathbf{H}\mathbf{x}$.

On peut construire trois tels vecteurs d'innovation :

$$\begin{cases} \mathbf{d}_b^o &= \mathbf{y}^o - \mathbf{H}\mathbf{x}^b \\ \mathbf{d}_a^o &= \mathbf{y}^o - \mathbf{H}\mathbf{x}^a \\ \mathbf{d}_a^b &= \mathbf{H}\mathbf{x}^b - \mathbf{H}\mathbf{x}^a \end{cases} \quad (\text{III.22})$$

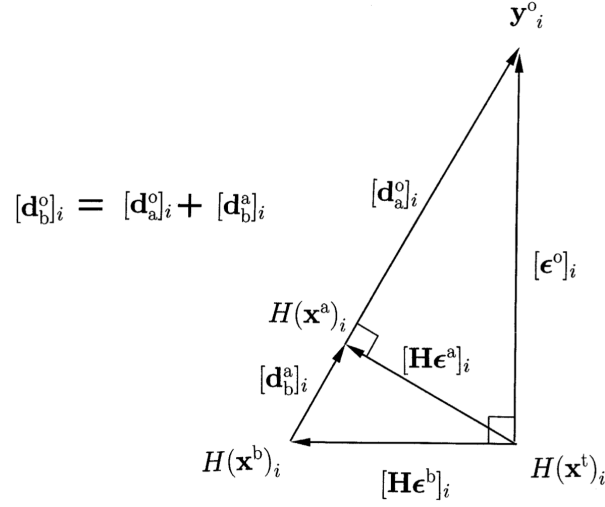


FIGURE III.7: Représentation géométrique de l'analyse projetée sur les vecteurs propres de la matrice \mathbf{HK} (tiré de Desroziers et al., 2005).

Il est alors possible de montrer (Desroziers et al., 2005) que :

$$\begin{cases} \mathbb{E} \left[\mathbf{d}_b^o (\mathbf{d}_b^o)^T \right] = \mathbf{R} + \mathbf{HBH}^T \\ \mathbb{E} \left[\mathbf{d}_b^a (\mathbf{d}_b^a)^T \right] = \mathbf{HBH}^T \\ \mathbb{E} \left[\mathbf{d}_a^o (\mathbf{d}_a^o)^T \right] = \mathbf{R} \\ \mathbb{E} \left[\mathbf{d}_b^a (\mathbf{d}_a^o)^T \right] = \mathbf{HP}^a \mathbf{H}^T \end{cases} \quad (\text{III.23})$$

Desroziers et al. (2005) proposent une interprétation géométrique marquante de ces relations. On définit, pour 2 variables aléatoires ϵ_1 et ϵ_2 , le produit scalaire statistique suivant :

$$\langle \epsilon_1 | \epsilon_2 \rangle = \mathbb{E}[\epsilon_1 \epsilon_2^T] \quad (\text{III.24})$$

Les relations de l'équation III.23 représentent dans le cadre de ce produit scalaire des « angles statistiques » (un angle plat signifiant une corrélation totale et un angle droit une décorrélation).

Lorsque l'on projette les vecteurs d'innovations \mathbf{d}_b^o , \mathbf{d}_a^o et \mathbf{d}_b^a sur les vecteurs propres de la matrices \mathbf{KH} , on obtient la figure III.7. Des applications élémentaires du théorème de Pythagore et d'Euclide à ce triangle permettent de retrouver les relations de l'équation III.23. Cela montre que si la matrice \mathbf{KH} et donc les matrices \mathbf{R} et \mathbf{B} , sont définies de manière optimale, alors les triangles $(\mathbf{Hx}^b, \mathbf{Hx}^t, \mathbf{y}^o)$ et $(\mathbf{Hx}^t, \mathbf{Hx}^a, \mathbf{y}^o)$ de la figure III.7 sont droits.

4.2.3 Le maximum de vraisemblance

Les critères du χ^2 et des cohérences de matrices de covariances sont essentiellement algébriques. Ils permettent de vérifier que les matrices de covariance \mathbf{R} et \mathbf{B} ont été spécifiées de manière mathématiquement cohérente. Néanmoins, ces tests ont potentiellement une infinité de matrices solutions et inversement un couple de matrices \mathbf{R} et \mathbf{B} peut être très proche des matrices optimales sans pour autant respecter les tests algébriques (Dee, 1995).

Ce manque d'informations précises pousse certains chercheurs à caractériser de manière statistique la confiance que l'on peut avoir en le couple de matrices \mathbf{R} et \mathbf{B} spécifié. Il s'agit du même raisonnement qui pousse à estimer de manière statistique les flux a posteriori.

Cette représentation statistique de la confiance en \mathbf{R} et \mathbf{B} est ancienne dans d'autres domaines des problèmes inverses (p. ex. Burg et al., 1982). En assimilation de données atmosphériques et météorologiques, cette formulation est plus récente (Dee, 1995; Chapnik et al., 2004; Michalak et al., 2005).

Elle se base sur une estimation de la maximisation de la vraisemblance des matrices \mathbf{R} et \mathbf{B} en connaissant les informations dont on dispose sur le système. Autrement dit, on cherche à estimer la densité de probabilité $p(\mathbf{R}, \mathbf{B} | \mathbf{y}^o, \mathbf{x}^b, \mathbf{H})$. En utilisant la règle de Bayes, cette densité de probabilité peut se réécrire :

$$p(\mathbf{R}, \mathbf{B} | \mathbf{y}^o, \mathbf{x}^b, \mathbf{H}) \propto p(\mathbf{y}^o | \mathbf{x}^b, \mathbf{H}, \mathbf{R}, \mathbf{B}) p(\mathbf{R}, \mathbf{B} | \mathbf{x}^b, \mathbf{H}) \quad (\text{III.25})$$

A priori, en l'absence d'information sur le système, nous ne disposons d'aucune information sur le couple (\mathbf{R}, \mathbf{B}) . Ainsi, la densité de probabilité $p(\mathbf{R}, \mathbf{B} | \mathbf{x}^b, \mathbf{H})$ est uniforme.

De plus, dans un modèle non biaisé gaussien, la densité de probabilité $p(\mathbf{y}^o | \mathbf{x}^b, \mathbf{H}, \mathbf{R}, \mathbf{B})$ est une loi normale centrée en $\mathbf{0}$ et de matrice de covariance $\mathbf{S}_{\mathbf{R}, \mathbf{B}}$:

$$p(\mathbf{y}^o | \mathbf{x}^b, \mathbf{H}, \mathbf{R}, \mathbf{B}) \propto \mathcal{N}(\mathbf{0}, \mathbf{S}_{\mathbf{R}, \mathbf{B}}) \quad (\text{III.26})$$

Il en découle que :

$$p(\mathbf{R}, \mathbf{B} | \mathbf{y}^o, \mathbf{x}^b, \mathbf{H}) \propto \mathcal{N}(\mathbf{0}, \mathbf{S}_{\mathbf{R}, \mathbf{B}}) \quad (\text{III.27})$$

Or, la matrice $\mathbf{S}_{\mathbf{R}, \mathbf{B}}$ se calcule explicitement :

$$\begin{aligned} \mathbf{S}_{\mathbf{R}, \mathbf{B}} &= \mathbb{E}[(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)^T] \\ &= \mathbb{E}[(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t + \mathbf{H}\mathbf{x}^t - \mathbf{H}\mathbf{x}^b)(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t + \mathbf{H}\mathbf{x}^t - \mathbf{H}\mathbf{x}^b)^T] \\ &= \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T \end{aligned} \quad (\text{III.28})$$

Ainsi, on en déduit que :

$$p(\mathbf{R}, \mathbf{B} | \mathbf{y}^o, \mathbf{x}^b, \mathbf{H}) \propto \frac{\exp\left\{-\frac{1}{2}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)\right\}}{\sqrt{(2\pi)^d |\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T|}} \quad (\text{III.29})$$

avec $|\cdot|$ l'opérateur déterminant.

Maximiser la vraisemblance revient donc à maximiser la fonction :

$$\ln p(\mathbf{R}, \mathbf{B} | \mathbf{y}^o, \mathbf{x}^b, \mathbf{H}) = -\frac{1}{2} \text{tr}(\mathbf{S}_{\mathbf{R}, \mathbf{B}}^{-1} \mathbf{S}) - \frac{1}{2} \ln |\mathbf{S}_{\mathbf{R}, \mathbf{B}}| \quad (\text{III.30})$$

avec $\mathbf{S} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^b$.

Cette fonction, appliquée à des sous-ensembles des espaces d'observation et d'état, revient à un diagnostic semblable au critère du χ^2 . Mais celle-ci est plus modulable que le χ^2 et permet d'estimer la fiabilité des matrices de covariances \mathbf{R} et \mathbf{B} en fonction de paramètres spécifiques.

Les ingrédients d'une inversion atmosphérique

Sommaire

1	Les observations	62
1.1	Les sites de surface	62
1.1.1	Les mesures flacons	63
1.1.2	Les mesures in situ	63
1.2	Les campagnes intensives et mesures ponctuelles	63
1.2.1	Les campagnes aéroportées YAK-AEROSIB	63
1.2.2	Le laboratoire mobile ferroviaire TROICA	64
1.3	Les satellites	64
2	Les flux a priori	65
2.1	Les inventaires anthropiques	65
2.2	Les modèles de processus biogéochimiques	67
2.3	Les bases de données de feux de biomasse	67
3	Les modèles de transport	67
3.1	Les modèles lagrangiens	67
3.2	Les modèles eulériens	68
3.2.1	Les eulériens directs et adjoints	68
3.2.2	Les eulériens mondiaux	68
3.2.3	Eulériens régionaux	69

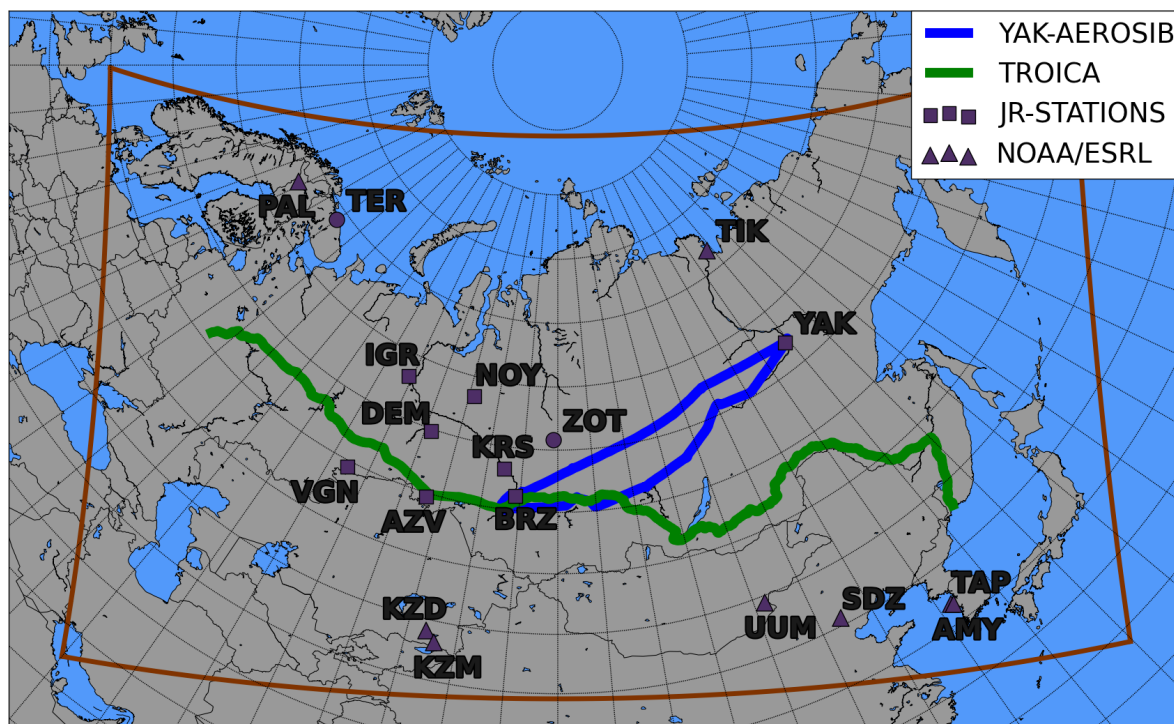


FIGURE IV.1: Distribution des sites de mesures de concentrations atmosphériques de méthane en Eurasie et des campagnes intensives de mesures. Les campagnes YAK-AEROSIB sont effectuées par le LSCE (voir sect. 1.2.1), les campagnes TROICA par l'Institut de physique de l'atmosphère Oboukhov (voir sect. 1.2.2). Le réseau NOAA/ESRL est décrit dans Dlugokencky et al. (1994), le réseau JR-STATIONS en sect. 1.1.2. Le site ZOT est maintenu par le Max Planck Institute (Winderlich, 2012) et le site TER par le Main Geophysical Observatory (Saint-Petersbourg). Le contour marron représente les frontières du domaine à aire limitée illustré en figure IV.4 page 69.

Comme nous l'avons vu au long du Chapitre III, l'inversion atmosphérique nécessite trois ingrédients principaux : des observations, des modèles de transport et des flux a priori. Nous présentons ici les éléments que nous utilisons par la suite pour nos inversions en Sibérie et en Europe.

1 Les observations

1.1 Les sites de surface

Les sites de surface permettent de mesurer des concentrations atmosphériques de manière pérenne. Une prise d'air est installée en un point donné, en général sur un mât ou une tour. Soit la prise d'air est reliée directement à un instrument de mesure, soit l'air est prélevé dans un flacon qui est analysé plus tard en laboratoire.

1.1.1 Les mesures flacons

Du fait de très faibles concentrations atmosphériques, les appareils d'analyse de la composition atmosphérique de méthane ont longtemps été très volumineux et nécessitaient des infrastructures disponibles uniquement dans les laboratoires de recherche. Des flacons hermétiques remplis d'air étaient donc prélevés en différents points du globe et acheminés jusqu'aux laboratoires de mesure.

Cette méthode est toujours utilisée car elle permet des mesures très précises et d'étudier la composition isotopique de l'air (ce que les instruments in situ ne font pas de manière systématique)

1.1.2 Les mesures in situ

Depuis les années 2000, les instruments ont été miniaturisés et surtout leur coût a diminué. Les infrastructures nécessaires pour les instruments de mesures atmosphériques continues sont maintenant réduites. Un abri et une alimentation électrique sont suffisants pour étalonner régulièrement l'instrument et garantir son fonctionnement.

Par conséquent, et malgré les étalonnages et opérations de maintenance régulières nécessaires, il est maintenant possible de mesurer à un coût abordable les concentrations de méthane à des intervalles de temps de l'ordre de la seconde directement au niveau du site de prise d'air. Cela a permis d'installer d'importants réseaux d'observations des concentrations de surface de méthane.

Par exemple, en Europe, le réseau ICOS (*Integrated Carbon Observing System*) vise à développer une infrastructure pérenne de 20–30 stations de surface mesurant en continu les concentrations de méthane, ainsi que d'autres gaz. Le réseau InGOS (*Integrated non-CO₂ greenhouse gas Observing System*) regroupe provisoirement plusieurs réseaux nationaux existant. En Sibérie, le *National Institute for Environmental Studies* (NIES, Tsukuba, Japon), en collaboration avec l'Institut d'Optiques de l'Atmosphère (IAO, Tomsk, Russie) et l'Institut du Permafrost (Yakutsk, Russie), entretient un réseau (*Japan–Russia Siberian Tall Tower Inland Observation Network*; JR-STATIONS) de mesures continues de concentrations (voir figure IV.1).

1.2 Les campagnes intensives et mesures ponctuelles

1.2.1 Les campagnes aéroportées YAK-AEROSIB

Le projet YAK-AEROSIB (*Airborne Extensive Regional Observations in SIBERIA*) est un projet de collaboration franco-russe démarré en 2004 visant à la base à établir des observations systématiques de CO, CO₂, O₃. L'objectif était de déterminer les sources et le transport de CO₂ en Eurasie, ainsi que les processus de chimie-transport amenant à la production d'O₃ au-dessus de la Sibérie (Paris et al., 2010a). Depuis 2010, des mesures de CH₄ sont aussi disponibles afin d'étudier le transport et les sources régionales.

Les mesures sont effectuées à l'aide d'un laboratoire embarqué dans un avion (Antokhin et al., 2011). L'avion suit un plan de vol en dents de scie de 800 à 8500 m (voir figure IV.2) entre Novossibirsk et Yakutsk (voir figure IV.1) afin d'échantillonner la basse et moyenne troposphère sibérienne.

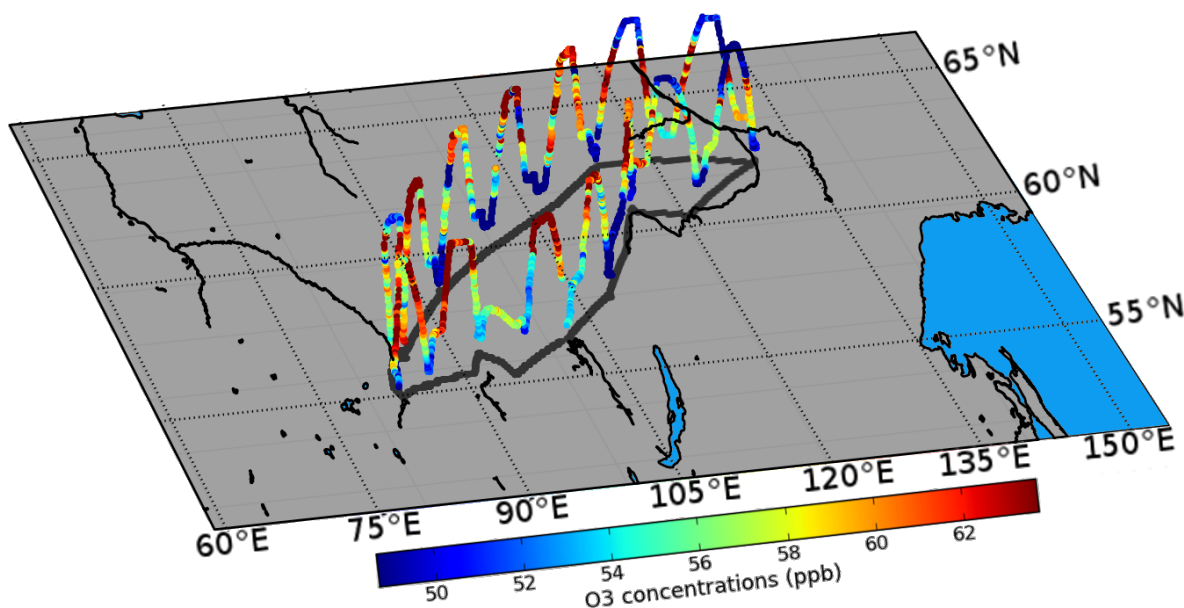


FIGURE IV.2: Plan de vol en 3 dimensions de la campagne YAK-AEROSIB d'avril 2010. Les mesures d'O₃ ont été superposées aux positions de l'avion.

Les instruments de mesures sont sensibles aux différences de pression et aux vibrations produites par l'avion en vol. Cette sensibilité peut détériorer la précision des appareils de mesure. Pour cette raison, de nombreux étalonnages en vol sont nécessaires et réduisent le nombre d'observation disponibles.

Le NIES effectue aussi régulièrement (en moyenne tous les mois) des mesures aéroportées au-dessus des sites NOY et IGR de la figure IV.1 (Tohjima et al., 1996; Umezawa et al., 2012).

1.2.2 Le laboratoire mobile ferroviaire TROICA

Depuis 1999, une voiture du transsibérien a été aménagée par l'Institut de physique de l'atmosphère Oboukhov (Moscou) en laboratoire de mesures de composition atmosphérique. Cette voiture peut être greffée à un convoi classique de voyageurs. Depuis 1999, une à deux campagnes de mesures sont effectuées chaque année sur le tracé Moscou-Vladivostok du transsibérien (Belikov et al., 2006).

1.3 Les satellites

Les sites d'observation par flacons ou in situ sont limités dans l'espace. De larges régions ne sont pas couvertes par ce type de mesures. De leur côté, les satellites sont en mesure de documenter les concentrations atmosphériques de méthane sur tout le globe depuis l'espace.

Pour cela, les missions satellitaires telles que l'instrument TANSO (Thermal And Near infrared Sensor for carbon Observation) embarqué sur le satellite GOSAT (*Greenhouse gases Observing SATellite*) et mesurant dans l'infra-rouge court ou IASI (Interféromètre Atmosphérique de Sondage Infrarouge) embarqué sur MetOp-A (satellite A de la série de trois satellites *Meteorological operational satellites* de l'ESA et EUMETSAT) utilisent des méthodes de télédétection passive consistant à analyser le spectre de la lumière solaire renvoyée par la surface

de la Terre. Les mesures de luminance permettent de déduire les concentrations de méthane intégrées sur des colonnes d'air (p. ex. Parker et al., 2011).

Les satellites des missions GOSAT et MetOp sont en orbite autour de la Terre et chaque orbite polaire se décale légèrement en longitude par rapport à la précédente. Ces satellites repassent donc au-dessus d'un même point au bout de quelques jours. De plus, l'instrument embarqué observe la surface survolée sur une largeur de quelques dizaines de km : c'est ce qu'on appelle la fauchée. De la sorte, la couverture spatiale des observations est mondiale, mais tous les points du globe ne sont pas couverts en permanence.

En dépit de leur supériorité en termes de couverture spatiale en comparaison des réseaux de surface, les satellites fournissent des informations incomplètes sur les concentrations atmosphériques de méthane. Tout d'abord, le paramètre mesuré par télédétection par le satellite est la concentration de méthane intégré sur une colonne d'air. On peut donc difficilement distinguer l'altitude des panaches détectés. De plus, la précision des satellite est inférieures à celle des sites de surface et les erreurs instrumentales sont de l'ordre de 1–2% contre moins de 0,1% pour les observations de surface. Enfin, certaines mesures satellitaires (TANSO pour le méthane) sont sensibles à l'ensoleillement et à la couverture nuageuse et neigeuse qui rendent les observations inexploitable. Par conséquent, en Sibérie, le nombre d'observations satellites est assez réduit du fait d'un ensoleillement réduit en hiver et d'une importante couverture nuageuse et neigeuse.

2 Les flux a priori

Nous avons déjà décrit en Sect. 2.1 du Chapitre II les principes de l'approche « montante » permettant de calculer les flux a priori nécessaires aux inversions. Nous donnons ici des détails sur les bases de données et modèles utilisés dans la suite.

2.1 Les inventaires anthropiques

Les flux anthropiques que nous utilisons sont déduits de la base de données EDGAR (*Emission Database for Global Atmospheric Research*) développée par le Centre de Recherche de la Commission Européenne d'Ispra (Italie) et la *Netherlands Environmental Assessment Agency* (PBL). La base de données EDGAR est entièrement décrite sur le site internet : <http://edgar.jrc.ec.europa.eu/>. Les facteurs d'émissions et les données statistiques d'activités sont agrégés depuis chaque institution locale et nationale au sein d'EDGAR.

Cette base de données fournit des cartes à une résolution spatiale de $0,1^\circ \times 0,1^\circ$ tous les ans (voir figure IV.3), ainsi que des bilans d'émissions par pays et par région géographique. Toutes ces informations sont disponibles pour une cinquantaine de sous-secteurs d'activité industrielle, agricole ou économique (par exemple, les émissions du secteur pétrolier, du transport, etc.).

Nous utilisons principalement les deux dernières versions de la base de données EDGAR : EDGARv4.2 et EDGARv4.2 Fast Track 2010 (FT2010). Dans la version EDGARv4.2 FT2010, les facteurs d'émissions sont issus d'études plus récentes qu'EDGARv4.2.

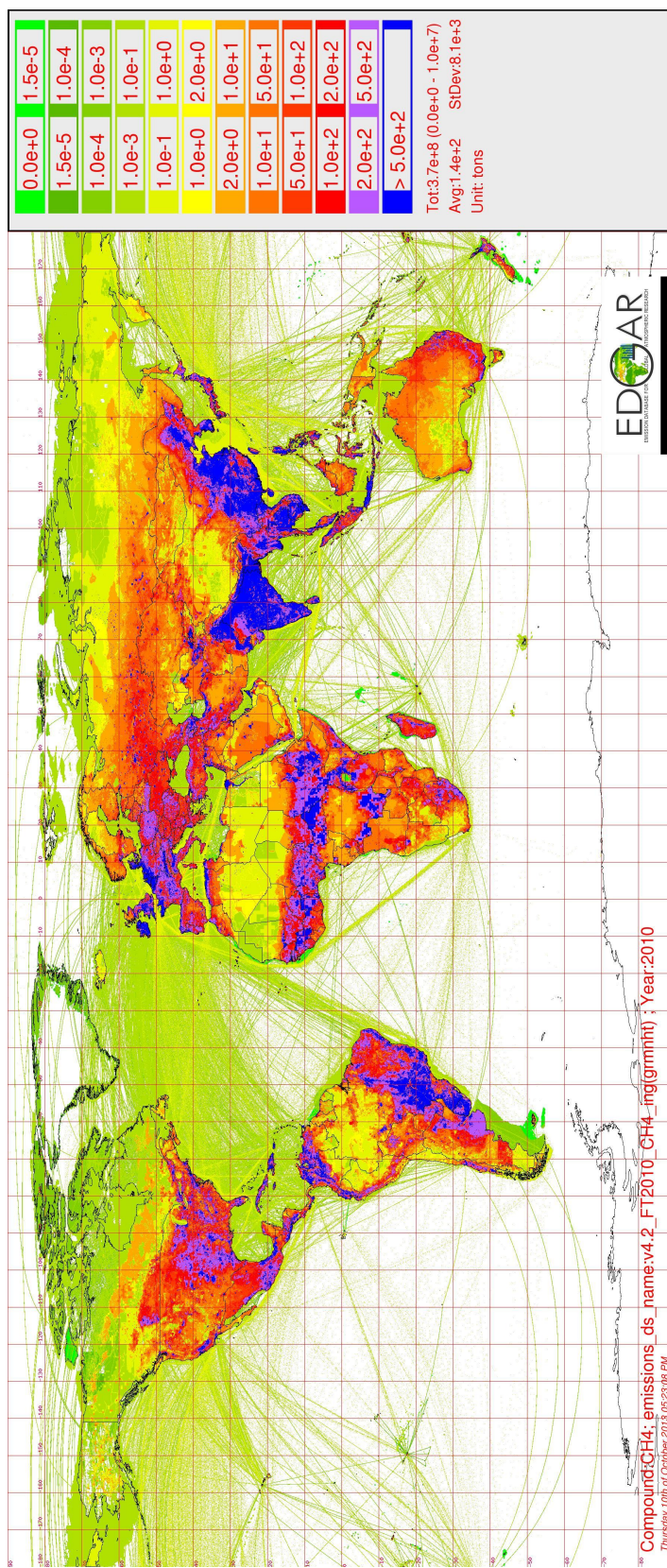


FIGURE IV.3: Répartition spatiale des émissions anthropiques de méthane en 2010 selon la base de donnée EDGARv4.2 FT2010. Toutes les émissions anthropiques sont sommées dans cette carte.

2.2 Les modèles de processus biogéochimiques

Nous utilisons le modèle LPJ (Lund-Postdam-Jena) et son extension LPX-Bern (*Land surface Processes and eXchanges*) pour la modélisation des flux de méthane biogéniques (Spahni et al., 2011). Ces deux modèles simulent de manière dynamique la répartition des zones inondées et l'évolution des tourbières (Spahni et al., 2013; Ringeval et al., 2014). Pour cela, le cycle de l'azote est aussi simulé afin d'estimer la fertilisation des plantes. Ces modèles utilisent des données météorologiques et hydrologiques (température, précipitations, couverture nuageuse, etc.) issues de la base de données CRU TS 3.21 (Harris et al., 2014) ainsi que d'observations de la concentration moyenne de CO₂ intervenant dans l'efficacité des plantes.

Ces deux modèles peuvent estimer les flux mensuels de méthane à une résolution de $0,5^\circ \times 0,5^\circ$.

2.3 Les bases de données de feux de biomasse

Les émissions a priori de feux sont calculées à partir de la version 4 de la base de données GFED (*Global Fire Emission Database*). La base de données GFED v4 est construite à partir de données satellitaires de l'instrument MODIS (*Moderate Resolution Imaging Spectroradiometer*). Cet instrument permet d'évaluer à une résolution de 500 m les surfaces brûlées à l'aide de l'algorithme de cartographie décrit dans Giglio et al. (2009). Les émissions de méthane sont ensuite déduites à des échelles mensuelles et journalières et à une résolution de $0,25^\circ \times 0,25^\circ$. Les émissions sont calculées à l'aide du modèle Carnegie-Ames-Stanford-Approach (CASA; Potter et al., 1993) calculant la quantité de matière organique disponible et de facteurs d'émissions par type de végétation (van der Werf et al., 2010).

3 Les modèles de transport

Les modèles de transport relient les émissions aux observations. Deux approches de modélisation de l'atmosphère existent : l'approche lagrangienne et l'approche eulérienne.

3.1 Les modèles lagrangiens

Dans l'approche lagrangienne, chaque parcelle d'air est suivie individuellement au cours du transport. En un point donné, l'air mesuré peut être considéré comme un assemblage de nombreuses particules virtuelles. Chaque particule virtuelle a été transportée au-dessus de différentes régions à différentes altitudes. Si la particule virtuelle est transportée à proximité du sol au-dessus d'une région d'émission, elle se chargera ponctuellement en méthane.

Les concentrations mesurées représentent la somme de toutes ces contributions de particules virtuelles (Lin et al., 2003). Il suffit donc de calculer la trajectoire des particules virtuelles atteignant le point de mesure à rebours dans le temps pour connaître la contribution de chaque région d'émission à la concentration mesurée, c'est-à-dire la matrice jacobienne \mathbf{H} .

Nous utilisons le modèle lagrangien de dispersion FLEXPART version 8.2.3 (Stohl et al., 2005) pour simuler les trajectoires des particules virtuelles. Afin de représenter la dispersion atmosphérique, ce modèle utilise un module de paramétrisation gaussienne. Selon les paramètres utilisés et les situations météorologiques, la dispersion peut être sous-évaluée ou exagérée.

Le modèle est contraint par des données météorologiques du Centre Européen pour le Prévisions Météorologiques à Moyen Terme (CEPMMT). Nous utilisons la base de données ERA INTERIM (*ECMWF Re-Analysis – Interim*) à $1^\circ \times 1^\circ$ de résolution spatiale, 60 niveaux verticaux et une fréquence de 3 h (Uppala et al., 2005).

Le calcul explicite de la contribution de chaque région d'émission est effectué en émettant l'hypothèse que la couche limite atmosphérique est parfaitement mélangée. Une particule dans la couche limite au-dessus d'une zone d'émission de méthane se chargera donc d'une quantité de méthane inversement proportionnelle à la hauteur de la couche limite.

Les modèles lagrangiens ont l'avantage d'être légers car des calculs sont effectués uniquement au niveau des sites d'observation. Ils permettent aussi de pouvoir très facilement changer la représentation ω du problème une fois les trajectoires virtuelles calculées.

3.2 Les modèles eulériens

3.2.1 Les eulériens directs et adjoints

Pour le transport de traceurs passifs aux échelles régionales, l'approche eulérienne est plus précise que l'approche lagrangienne. Dans cette approche, on simule explicitement tous les points de l'espace, ainsi que les flux de masse entre les différentes mailles de la représentation choisie. On dispose donc de l'évolution temporelle du champs de concentrations en 3 dimensions. Il suffit alors de comparer concentrations simulées de manière directe et concentrations mesurées. La matrice \mathbf{H} se calcule alors de manière directe colonne par colonne.

Certains modèles eulériens possèdent un modèle adjoint qui permet aussi de calculer la sensibilité de chaque observation aux émissions. Les modèles adjoints permettent de calculer la matrice \mathbf{H} ligne par ligne (Hourdin and Talagrand, 2006).

Suivant le nombre de lignes et de colonnes dans \mathbf{H} le modèle direct ou adjoint peut être choisi.

3.2.2 Les eulériens mondiaux

Une première classe de modèles eulériens simule l'atmosphère dans son ensemble. Ce sont des Modèles de Circulation Générale (MCG) ou *General Circulation Model* (GCM). Parmi ces modèles, nous utilisons le modèle LMDz du Laboratoire de Météorologie Dynamique dans la suite de ce rapport (Sadourny and Laval, 1984; Hourdin et al., 2006; Locatelli et al., 2014).

Modéliser l'atmosphère sur toute la colonne atmosphérique et sur toute la surface de la Terre est coûteux. De plus, une simulation de l'atmosphère complète doit être accomplie sur de longues périodes (typiquement supérieures à la décennies) afin que l'atmosphère soit complètement mélangée.

De ce fait, ces modèles utilisent en général une modélisation paramétrisée hydrostatique de l'atmosphère, c'est-à-dire que les gradients de pression sont uniquement contrebalancés par la force de gravitation. Ils opèrent sur des grilles à assez basse résolution. Par exemple, le modèle LMDz est en général utilisé avec une résolution spatiale de $3,75^\circ \times 2,25^\circ$.

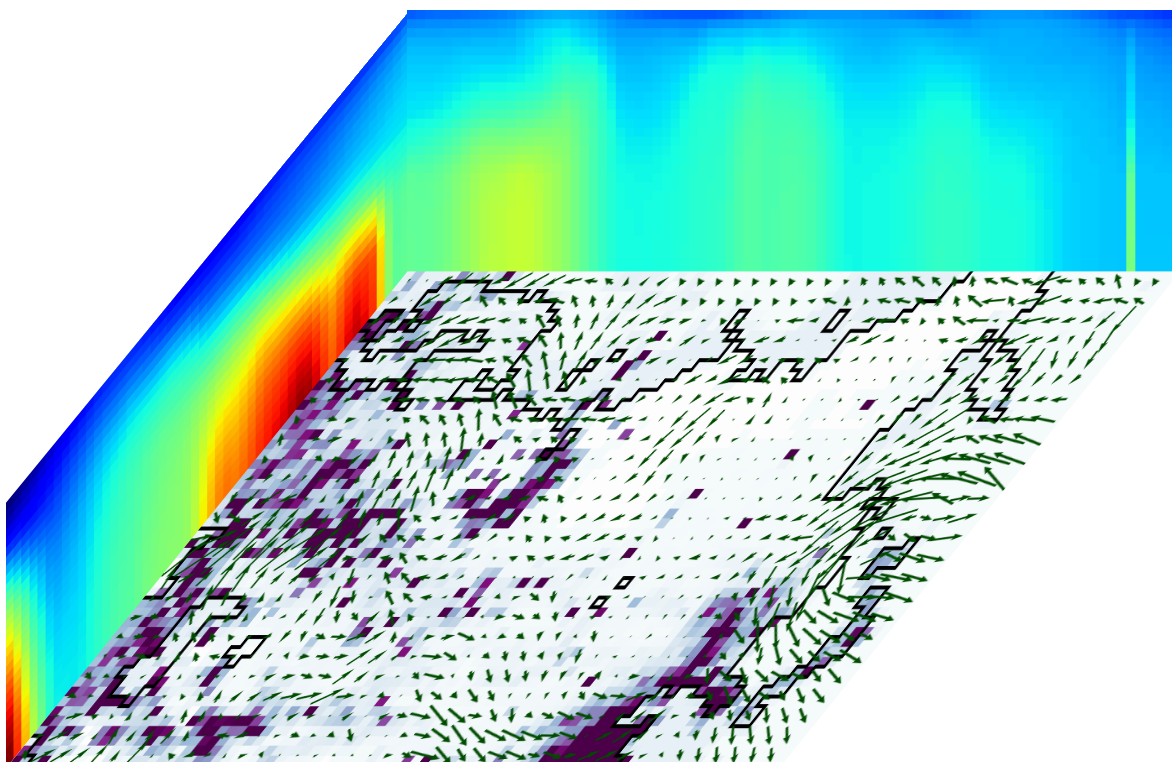


FIGURE IV.4: Représentation en trois dimensions d'un domaine à aire limitée pour des simulations eulériennes en Eurasie (voir contours du domaine sur la figure IV.1 page 62). Des émissions sont contraintes (nuances de violet) à l'interface atmosphère-surface et des concentrations au niveau des bords du domaine. Le modèle transporte ensuite les masses d'air à l'aide de champs de vents prescrits (flèches vertes). Le monde du modèle est discrétisé selon des mailles (par exemple, la limite des côtes en lignes noires).

3.2.3 Eulériens régionaux

Lorsque l'on s'intéresse à une région spécifique, il est possible d'utiliser un modèle eulérien régional sur un domaine limité, c'est-à-dire une boîte virtuelle fermée en longitude, latitude et altitude (voir figure IV.4). Tout comme le modèle mondial, un modèle de transport eulérien régional simule les flux de masse et les concentrations à l'intérieur du domaine d'intérêt. La taille limitée du domaine d'intérêt permet d'augmenter la résolution spatiale par rapport à un modèle mondial et de relâcher l'hypothèse d'hydrostatisme.

Des masses d'air entrent et sortent de la boîte virtuelle par les bords et le toit du domaine. Pour prendre en compte ces masses d'air, les modèles eulériens régionaux utilisent des champs de concentrations en 3 dimensions à basse résolution calculés par des modèles mondiaux.

Dans la suite, nous utilisons le modèle non-hydrostatique CHIMERE créé par le Laboratoire de Météorologie Dynamique (Vautard et al., 2001; Menut et al., 2013). Ce modèle développé dans le cadre d'études de qualité de l'air est aussi utilisé pour des simulations régionales de concentrations de gaz à effet de serre. Les champs météorologiques nécessaires au calcul des flux de masses sont extraits des bases de données du CEPMMT. Nous utilisons des données météorologiques à $0,5^\circ \times 0,5^\circ$ toutes les 3 h.

Troisième partie

**Applications et développements
scientifiques**

Facteurs influençant la composition atmosphérique en méthane en Sibérie

Sommaire

1	Étude du transport par une campagne aéroportée	74
1.1	Le transport : prérequis de l'approche atmosphérique	74
1.2	La campagne YAK-AEROSIB d'avril 2010	74
1.3	L'ozone comme traceur du transport atmosphérique	74
1.4	Une estimation lagrangienne des structures de transport	75
2	<i>Tropospheric ozone over Siberia in spring 2010</i>	75
2.1	<i>Introduction</i>	77
2.2	<i>Methods</i>	78
2.2.1	<i>Campaign description and sampling area</i>	78
2.2.2	<i>Instruments used and data processing</i>	79
2.2.3	<i>Air mass origin and back trajectories</i>	80
2.3	<i>Results and discussion</i>	81
2.3.1	<i>Average, clean and polluted air masses composition</i>	81
2.3.2	<i>Influence of combustion processes on ozone concentrations</i>	84
2.3.3	<i>Upper troposphere ozone excess and stratospheric intrusion</i>	87
2.3.4	<i>Widespread upper tropospheric low ozone concentrations</i>	90
2.4	<i>Conclusions</i>	91
3	Structures météorologiques et contributions distantes en Sibérie	92
3.1	Structures de transport caractérisées	92
3.2	Identification de sources ponctuelles	94
3.3	Estimation de l'influence des hydrates de méthane en mer des Laptev	94

1 Étude du transport en Sibérie à l'aide d'une campagne de mesures aéroportée

1.1 Le transport : prérequis de l'approche atmosphérique

Nous visons une quantification des flux de méthane en Sibérie par inversion atmosphérique. L'approche atmosphérique implique inévitablement des considérations statistiques, numériques et mathématiques. Néanmoins, il est nécessaire de commencer par estimer qualitativement le comportement de l'atmosphère et donc le transport atmosphérique dans la région d'intérêt. La connaissance des structures de transport dans la région permet d'évaluer les besoins en termes de précision dans le modèle de transport. Le second objectif de l'étude du transport est de déterminer quelles sources influencent la composition atmosphérique en méthane au-dessus de la Sibérie et donc doivent être prise en compte dans une inversion.

Ces sources peuvent être locales ou distantes. Le transport horizontal à longue distance peut suivre des trajectoires complexes induisant des influences et des contributions de sources contre-intuitives. Le mélange vertical des sources locales peut changer en fonction de la convection dans la couche limite et la convection profonde.

Nous nous appuyons sur l'analyse des mesures effectuées lors de la campagne aéroportée du projet YAK-AEROSIB en avril 2010 pour explorer ces différents points.

1.2 La campagne YAK-AEROSIB d'avril 2010

Les données des campagnes YAK-AEROSIB (voir sect. 1.2.1 page 63 du chapitre IV) fournissent un « instantané » en 3 dimensions de la composition atmosphérique sibérienne. Il est donc possible d'explorer les structures horizontales et verticales du transport dans la région.

Lors de la campagne YAK-AEROSIB d'avril 2010, des données ont été collectées sur les concentrations de CO, CO₂, CH₄ et O₃. Des spectres de tailles et des densités en aérosols ont aussi été mesurées, ainsi que des données météorologiques.

La co-émission de CH₄ et de CO et CO₂ peut fournir des informations sur le type de sources influençant la mesure. Cependant, les instruments de mesure de CH₄ et d'aérosols n'ont pas fonctionné correctement lors de cette campagne, rendant leurs données inexploitable.

Nous utilisons donc un moyen indirect d'étudier les sources distantes de méthane en Sibérie.

1.3 L'ozone comme traceur du transport atmosphérique

Dans l'article scientifique retranscrit en sect. 2, nous utilisons l'ozone pour étudier le transport en Sibérie.

Dans la sect. 1.3 page 10 du chapitre I, nous avons succinctement présenté l'ozone et son rôle dans le système chimique atmosphérique. L'ozone est produite en présence de NO_x et de COV. Il peut aussi être détruit dans certaines conditions par ces mêmes éléments, ainsi que par des aérosols et des composés halogénés. Ces réactions chimiques pendant le transport de masses polluées en NO_x ou en COV créent d'importants gradients de concentrations d'ozone.

Les gradients d'ozone peuvent donc être utilisés pour identifier le transport de masses d'air depuis des sources distantes de NO_x et de COV (Cooper et al., 2010).

Dans la troposphère sibérienne au printemps, les précurseurs de l'ozone sont principalement liés à des processus de combustion, c'est-à-dire à l'activité anthropique ou aux feux de biomasse. L'étude des concentrations d'ozone permet donc de déduire indirectement l'influence de sources distantes anthropiques de méthane par ces processus de combustion. À l'inverse, les sources distantes de méthane par des zones inondées ne peuvent être identifiées à l'aide de l'ozone. Néanmoins, les régions entourant la Sibérie sont des zones pour lesquelles l'activité anthropique domine : l'Europe et la Chine. L'ozone est donc un bon traceur du transport atmosphérique à grande distance du méthane.

Par ailleurs, la chimie de l'ozone est fortement non linéaire. L'étude du transport via cet indicateur est donc nécessairement qualitative et non quantitative.

1.4 Une estimation lagrangienne des structures de transport

Nous ne connaissons pas a priori les structures de transport en Sibérie. Une approche eulérienne nécessiterait donc de modéliser au moins toute l'atmosphère de l'hémisphère nord.

De ce fait, nous choisissons d'utiliser le modèle de dispersion lagrangien FLEXPART pour identifier l'historique et l'origine des masses d'air identifiées comme influencées par des sources distantes. Comme les panaches de pollution à l'ozone sont liés à des processus de combustion, nous identifions les masses d'air polluées à l'aide des concentrations en CO. Une masse d'air avec un fort excès en CO corrélé avec une nette augmentation ou diminution d' O_3 est identifiée comme une masse d'air polluée.

Des simulations du modèles FLEXPART sont systématiquement effectuées pour les masses d'air ainsi identifiées afin de caractériser le transport en Sibérie.

2 Tropospheric ozone over Siberia in spring 2010 : remote influences and stratospheric intrusion

Cet article est le fruit du traitement et de l'analyse des données brutes de la campagne de mesures intensives aéroportées YAK-AEROSIB d'avril 2010. Cette analyse a été effectuée en stage de master et poursuivie en thèse. Cet article a été publié le 4 juin 2013 dans le journal de l'*International Meteorological Institute* à Stockholm, *Tellus, Series B, Chemical and Physical Meteorology*.

Les auteurs ayant participé à ce travail sont : Antoine Berchet¹, Jean-Daniel Paris¹, Gérard Ancellet², Kathy S. Law², Andreas Stohl², Philippe Nédélec³, Michael Yu. Arshinov⁴, Boris D. Belan⁴, et Philippe Ciais¹.

¹Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CNRS-CEA-UVSQ, Orme des Merisiers, CEA Saclay, Gif-sur-Yvette, France.

²UPMC Univ. Paris 06, Université Versailles St-Quentin, CNRS/INSU, LATMOS-IPSL, Paris, France.

³Norwegian Institute for Air Research, Kjeller, Norway.

⁴Laboratoire d'Aérodynamique, CNRS-UPS, Toulouse, France.

⁵V. E. Zuev Institute of Atmospheric Optics, SB-RAS, Tomsk, Russia.

La version définitive de cet article est disponible à l'adresse internet suivante : <http://www.tellusb.net/index.php/tellusb/article/view/19688/0>. Nous le retranscrivons ci-dessous dans son intégralité.

Résumé

Nous avons identifié et caractérisé différents facteurs influençant l'ozone troposphérique au dessus de la Sibérie pendant le printemps 2010. Pour cela, nous avons analysé des mesures in situ de rapport de mélange d'ozone, de dioxyde de carbone, de monoxyde de carbone et de méthane. Ces mesures ont été acquises à l'aide d'analyseurs continus pendant la campagne de mesure intensive aéroportée du projet YAK-AEROSIB, effectuée entre le 15 et le 18 avril 2010, sur un volume s'étalant sur 3000 km entre 800 et 6700 m au dessus du sol. L'analyse et l'interprétation de ces observations se sont appuyées sur des simulations d'origine des masses d'air à l'aide du modèle lagrangien FLEXPART. L'analyse de la variabilité des rapports de mélange des gaz traces et de l'origine simulée des masses d'air a mis en évidence des panaches de pollution en provenance de l'est et de l'ouest de la plaine de Sibérie occidentale, ainsi que du nord-est de la Chine. Ces panaches ont été associés à des feux de biomasse et à de l'activité anthropique qui a augmenté les rapports de mélange d'ozone au cours du transport. Des rapports de mélange bas en ozone ont aussi été observés sur une large zone dans la haute troposphère, au-dessus de 5500 m. Cette masse d'air a été transportée depuis la couche limite marine de la mer de Norvège, au-dessus de laquelle les niveaux de fond en ozone sont bas au printemps. Le transport depuis la mer de Norvège jusqu'à la zone d'observation s'est fait sans mélange avec les masses d'air de la moyenne et haute troposphère riche en O₃. Enfin, la source stratosphérique d'ozone a été observée de manière directe avec une intrusion stratosphérique bien délimitée. L'analyse de cet événement a mis en évidence une contribution stratosphérique ponctuelle de $2,56 \pm 0,29 \times 10^7$ kg d'ozone associée à un flux régional descendant d'ozone de $9,75 \pm 2,9 \times 10^{10}$ molécules cm⁻² s⁻¹, en deçà des valeurs climatologiques hémisphériques.

Abstract

We have identified and characterised different factors influencing the tropospheric ozone over Siberia during spring 2010. This was done by analysing in-situ measurements of ozone, carbon dioxide, carbon monoxide, and methane mixing ratios measured by continuous analysers during an intensive airborne measurement campaign of the YAK-AEROSIB project, carried out between 15 and 18 April 2010. The analysis and interpretation of the observations, spanning 3000 km and stretching from 800 to 6700 m above ground level, were enhanced using the Lagrangian model FLEXPART to simulate backward air mass transport. The analysis of trace gas variability and simulated origin of air masses showed that plumes coming from east and west of the west Siberian plain and from north-eastern China related to biomass burning and anthropogenic activity had enhanced ozone mixing ratios during transport. In one case, low ozone mixing ratios were observed over a large region in the upper troposphere above 5500 m. The air mass was transported from the marine boundary layer over the Norwegian Sea where O₃ background concentrations are low in the spring. The transport was coherent over thousands of kilometres, with no significant mixing with mid-upper troposphere air masses rich in O₃. Finally, the stratospheric source of ozone to the troposphere was observed directly in a well-defined stratospheric intrusion. Analysis of this event suggests an input of

$2.56 \pm 0.29 \times 10^7$ kg of ozone associated with a regional downward flux of $9.75 \pm 2.9 \times 10^{10}$ molecules $\text{cm}^{-2} \text{s}^{-1}$, smaller than hemispheric climatology.

2.1 Introduction

Tropospheric ozone (O_3) is an atmospheric oxidant, a harmful pollutant and a greenhouse gas. It is created in situ from photochemical oxidation of precursors [hydrocarbons, carbon monoxide (CO) and methane (CH_4)] in the presence of nitrogen oxides (NOx). These species are mainly emitted during combustion and biogenic processes. Ozone is also transported from the stratosphere, where large photochemical production occurs.

The stratosphere contributes 375–677 TgO_3/yr to the global troposphere (e.g., Collins et al., 2000; Wild et al., 2004; Hauglustaine et al., 2005) or $\sim 9\%$ of the tropospheric O_3 source (Seinfeld J. H. and Pandis S. N., 2006). However, this input is not continuous in time. It mainly occurs regionally as stratospheric intrusions associated to midlatitude cyclones (Browning, 1997), where enhanced O_3 mixing ratios can subsequently mix with tropospheric background air (Ancellet et al., 1994; Stohl et al., 2003). Tropopause folds are common in the mid-latitudes, including Siberia. But compared to the rest of the Northern Hemisphere, the region is the smallest contributor to stratosphere-troposphere exchange (STE) in the North Hemisphere (James et al., 2003). Between 20 and 50 events occur per $5^\circ \times 5^\circ$ area each year (Beekmann et al., 1997) contributing roughly 10–50 TgO_3/yr . Stratosphere-troposphere exchange over Siberia is predicted to have its maximum intensity in summer and minimum in winter, in contrast to North America and the North Atlantic where the maximum in STE intensity is less pronounced (Stohl, 2001; James et al., 2003). However, few observations of stratospheric intrusions over Siberia have been reported in the literature (see the review by Stohl et al., 2003; aircraft observations in Ishijima et al., 2010; train-based observations in Shakina et al., 2001; model studies for example in Wernli and Sprenger, 2007).

Emissions of NOx, VOC (volatile organic compounds) from anthropogenic and biogenic sources is responsible for O_3 production and/or destruction downwind of source regions (review in HTAP, 2010). Observations in the Arctic and sub-Arctic areas have revealed production of O_3 in pollution and biomass burning plumes during transport (e.g., Oltmans et al., 2010; Jaffe and Wigder, 2012; Thomas et al., 2013). Ozone destruction was also reported in some plumes from biomass burning or anthropogenic pollution (e.g., Verma et al., 2009; Alvarado et al., 2010). Ozone is lost by photochemical destruction following photolysis and reaction with water vapour. Halogen oxidation in the Arctic lower troposphere (Gilman et al., 2010; Sommar et al., 2010) can also lead to significant O_3 destruction, but this is generally confined to the Arctic boundary layer in springtime. Another major sink for O_3 is dry deposition on leaves via stomatal exchanges, harming the vegetation. Ozone deposition is an important term of the tropospheric O_3 budget, with an estimated 530 TgO_3/yr sink in the Northern Hemisphere. Siberian ecosystems, with about 10% of global land surface, represent a vast sink region for O_3 . In central Siberia, previous studies suggest an overall large sink for O_3 as deduced from backward transport analysis (Paris et al., 2010b; Engvall-Stjernberg et al., 2012). However, the regional distribution of O_3 sources and sinks remains poorly understood.

Despite Siberia's vast dimensions and importance in the climate system, little is known about whether and how the regional O_3 budget differs from the rest of the Northern Hemisphere. For example, O_3 production in boreal wildfire plumes seems to be weaker or to turn

Table V.1: Campaign description. Coordinates and dates respectively for take-off and landing.

Flight n°	Date UTC	Time UTC	Local time	Coordinates		Vertical profiles
				Latitude	Longitude	
1	15 Apr. 2010	04:36-10:30	11:36-20:30	55-62°N	83-114°E	10
2	15-16 Apr. 2010	23:04-03:30	09:04-13:30	62-60°N	114-115°E	8
3	16 Apr. 2010	04:47-07:27	14:47-16:27	60-56°N	115-102°E	4
4	18 Apr. 2010	02:20-06:19	11:20-13:19	56-55°N	102-83°E	6

into net destruction, compared to fire plumes at lower latitudes (Jaffe and Wigder, 2012). This may be due to lower NO_x emissions and/or more sequestration of NO_x as PAN (peroxyacyl nitrates; although PAN can produce O₃ downwind). Also, given their importance for atmospheric environmental issues and the global greenhouse gas budget, more atmospheric measurements of O₃, its precursors and other pollutants over Siberia are needed (see Elansky, 2012). These data are particularly useful for the validation of atmospheric chemistry models and satellite products. For example, Pommier et al. (2012) used such data to validate IASI O₃ at high latitudes.

The Airborne Extensive Regional Observation over Siberia (YAK-AEROSIB) campaigns provide three-dimensional 'snapshots' of the Siberian tropospheric composition in spring and summer (review in Paris et al., 2010a). Six such campaigns have been performed between April 2006 and April 2010: four in summer and two in spring. In this paper, which describes for the first time the data collected in the 2010 campaign, we investigate O₃ and other trace gases (CO₂, CO and CH₄) in the troposphere over Siberia as measured in situ during several instrumented aircraft flights in April 2010. Through case studies, we focus on the respective contributions of various terms of the regional O₃ budget, such as STE, long-range transport, or O₃ atmospheric photochemistry. In Section 2, we describe the instrumentation used during the campaign and the FLEXPART Lagrangian model, which was used to investigate the origins of observed air masses using retro-plume calculations. In Section 3, we present results for different cases showing evidence for long-range pollution transport to Siberia, photochemical processing and intrusions of stratospheric air into the mid-troposphere over Siberia.

2.2 Methods

2.2.1 Campaign description and sampling area

The region overflown during the campaign corresponds to a latitude band from 55°N to 62°N (details in Tab. V.1; map shown in Fig. V.3). The area is mostly covered by forests, steppes and wetlands. The major Siberian cities flown over are Novosibirsk (55.02°N, 82.97°E), Tomsk (56.44°N, 85°E), Yakutsk (62.04°N, 129.79°E) and Krasnoyarsk (56.02°N, 92.88°E). All these cities and industrialised zones are located in the southern part of Siberia. To the north, most anthropogenic emissions occur in limited areas such as the Taymir peninsula and the nearby Norilsk area, associated with mining, and oil and gas extraction. Large amounts of SO₂, NO_x, aerosols and halogens have been measured in the boundary layer near these areas (Belan et al., 2007).

Table V.2: Instrument characteristics. ⁽¹⁾After 30-s slide averaging.

Species measured	Core technique /instrument	Acquisition frequency	Meas. uncertainties	Calibration frequency	Reference
CO ₂	NDIR LI-COR 6262	2s	0.15 ppm	30 min or each plateau	Paris et al. (2008)
CH ₄	OA-ICOS LGR FGGA	1s	6 ppb 0.27%	Pre and post flight	
CO	Gas filter correlation Thermo 48C	1s ⁽¹⁾	5 ppb 5% ⁽¹⁾	20 min	Nédélec et al. (2003)
O ₃	UV absorption Thermo 49	4s	2 ppb 2%		Paris et al. (2008)
Aerosols	Diffusional particle sizer	80s	10%	Pre and post flight	Reischl et al. (1991)
Relative humidity	HYCAL IH-3602-C Honeywell	1s	7%		Zuev et al. (1992)
Temperature		1s	0.5°C		Zuev et al. (1992)

Four flights were carried out from Novosibirsk to Yakutsk and back, with stop-overs in Mirnyy and Bratsk. The campaign took place from 15 to 18 April 2010. Flights were carried out during daytime and under all weather conditions, with the exception of very cloudy conditions. The climate being continental, temperatures in Siberia are still low in April. The mean surface temperatures from 10 to 20 April ranged from 0 to 5°C in southern Siberia to (10°C to the North according to NOAA/ESRL Physical Sciences Division reanalysis (available at <http://www.esrl.noaa.gov/psd/data/reanalysis/>). The relative humidity was high (>80% near surface during the period of interest). Most of the region where the flights took place was covered by snow.

Both the flight plan and the instrumental package are essentially those described in Paris et al. (2008, see their Fig. 1). A total of 14 ascents/descents were performed by the aircraft during the campaign with a see-saw pattern and with interposed horizontal legs (plateaus) of 10-30 min duration at three different levels: 6600, 2000 and 800 m above ground level. For each of the three levels, 14 horizontal sections were carried out. The Antonov-30 research aircraft used for the campaigns is described by Antokhin et al. (2011) and Paris et al. (2010a).

2.2.2 Instruments used and data processing

During the flights, concentrations of O₃, CO₂, CO, CH₄ and meteorological (wind speed, temperature, pressure, humidity) data are collected with in situ continuous analysers. The measurement equipment used for the YAK-AEROSIB campaigns is described in Paris et al. (2010a, consult Fig. 3 and Table A1 in the cited reference). Technical details of the instrumentation are summarised in Table V.2. The O₃ analyser is based on a commercial instrument (UV absorption; Thermo Instruments Model 49) modified to match aircraft requirements (pressure, temperature, vibrations). The instrument is calibrated in laboratory before and after the campaign with an O₃ generator. The instrument uncertainties are 2 ppb (2%) for an integration time of 4 seconds.

For the 2010 campaign, a CH₄ analyser was added to the instrumental package by the Institute of Atmospheric Optics of Tomsk. Methane was measured here for the first time in the YAK-AEROSIB campaigns, using the Los Gatos Research model FGGA (Fast Greenhouse Gas Analyser). No in-flight calibrations for this instrument were performed but calibration was carried out on the ground both before and after each flight with a standard gas mixture of methane (2308.37 ± 0.802 ppb). Measurements are reported on the Nippon Sanso Company scale maintained by Tohoku University, with an offset of 0.5 ppb relative to the NOAA CH₄ scale (Dlugokencky et al., 2005). Abnormal mixing ratios due to an identified instrument malfunction were filtered out, resulting in a precision of 10-30 ppb depending on flights. Standard deviation ($n = 90$) of the measured standard gas concentration during the pre- and post-flight calibration was 6.18 ppb. The limited precision prevents the detection of narrow plumes in CH₄ data.

Conditions in airborne campaigns (vibrations, thermal and pressure variation) impose recurrent calibrations for CO and CO₂ to compensate for drifts. This prevents the collection of continuous measurements during the entire flight duration. Carbon dioxide calibration gases are traceable to primary World Meteorological Organisation/NOAA standards and drifts are corrected by linear interpolation between the 'zero' reference gas injections.

2.2.3 Air mass origin and back trajectories with the Lagrangian model FLEXPART

The FLEXPART Lagrangian particle dispersion model (Stohl et al., 2005) was used for long-range transport analysis and for determining the origin of the sampled polluted air masses. Two different approaches were implemented. In the first method ('receptor-oriented'), designed for separating and quantifying remote contributions in a single plume, 20-d retroplumes were calculated using 60000 computational particles from short segments ($0.5 \times 0.5^\circ$) along the plume sections of the flight tracks. The model calculations were based on two alternative meteorological data sets to verify consistency of the simulated transport: operational analyses from the National Centres for Environmental Prediction's Global Forecast System (GFS) with a resolution of $0.5^\circ \times 0.5^\circ$ and 26 vertical levels and the European Centre for Medium-Range Weather Forecasts' (ECMWF) model with a resolution of $1^\circ \times 1^\circ$ and 91 vertical levels. All the figures were plotted using the ECMWF products. The FLEXPART model output (reported as ns.kg (1) is proportional to the residence time of the particles in a given volume of air and corresponds to potential emission sensitivity (PES). When convolved with the gridded emission fluxes from an emission inventory, maps of potential source contributions (PSC) are obtained. Integrating these maps over the globe yields a model-calculated mixing ratio of the emitted species at the location of the aircraft. EDGAR fast track 2000v3.2 (Emission Database for Global Atmospheric Research from PBL Netherlands Environmental Assessment Agency) was used for anthropogenic CO emissions other than biomass burning. Biomass burning CO emissions were based on fire locations detected with a confidence of more than 75% by the MODIS (Moderate Resolution Imaging Spectroradiometer) instrument onboard the Terra and Aqua satellites (Giglio et al., 2003, data imported from NASA's EOSDIS website; Earth Observing System and Information system; <http://earthdata.nasa.gov/data/near-real-time-data/firms>) and a land cover classification, as described by (Stohl et al., 2007). For every single plume, we estimate the characteristic residence time in the atmosphere using the modal age given as the age of the maximum contribution in the PSC.

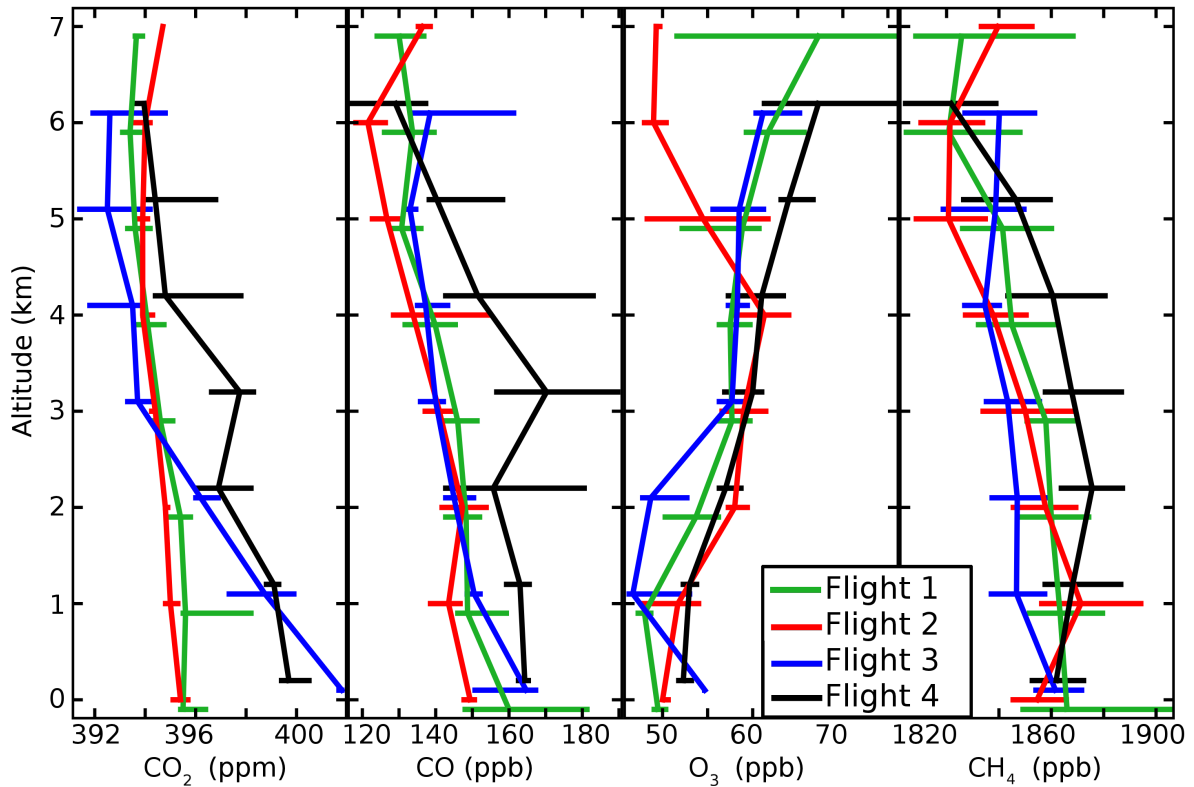


Figure V.1: Vertical profiles of CO_2 , CO , O_3 and CH_4 for each flight. Medians have been calculated for 1-km-height intervals (y-axis ticks mark the bottom of each layer). The horizontal error bars show the inter-quartile ranges over each altitude interval. In Flight 4, O_3 profile, the 3rd quartile at 6-7 km is ~ 100 ppb.

The second method ('flight curtains') is intended to study the spatial structure of air masses origin in the troposphere nearby the measurement route. Particles were released from a two-dimensional grid oriented as a vertical curtain along the flight track, with $0.5^\circ \times 0.5^\circ \times 500$ m boxes distributed in the height interval 1000-7000 m. Every release box was filled with 2000 virtual particles for which back trajectories are calculated. Back trajectories have been calculated from the sampling date back to 6 April 2010. For every time step, the fraction of particles originating from pre-selected regions of interest is calculated.

2.3 Results and discussion

2.3.1 Average, clean and polluted air masses composition

The overall tropospheric composition measured during the spring 2010 campaign is generally comparable to that observed during previous campaigns (Paris et al., 2008, 2010b,a) with a vertical distribution typical of late winter over Siberia. However, the spring 2012 campaign is on average less polluted than the previous spring campaign (2006). The mean CO mixing ratio was 145 ± 22 ppb in 2010, compared to 174 ± 28 ppb in 2006. In spring 2010, air masses came predominantly from clean Arctic regions, whereas in 2006, the main origins were Europe and north-eastern Asia.

Figure V.1 shows for each species the median mixing ratio vertical profiles for the four flights. Regarding O_3 , the mixing ratios generally increase with altitude (linear regression slope of $+1.4 \pm 0.03$ ppb/km; $r = 0.87$). This small positive O_3 mixing ratio gradient is usually attributed to a stratospheric source and to the surface deposition sink (supposedly limited to snow-free areas; e.g., Engvall-Stjernberg et al., 2012) as well as photochemical production, which is more efficient in the middle and upper troposphere (Reeves et al., 2002). In summer, previous campaigns have shown that the O_3 surface sink is more pronounced resulting in a steeper vertical gradient in the lower troposphere (Paris et al., 2010b). The CO_2 gradient of -0.6 ppm/km ($r = -0.90$) confirms the limited biogenic activity (confined to plant respiration while vegetation uptake has not started over Siberia in April) that causes the restricted spring O_3 sink. Indeed, the gradient is $+1$ ppm/km during the summer campaigns, due to a net sink of CO_2 through vegetation uptake (Paris et al., 2010a).

As for CO_2 , CO concentrations decrease with altitude (-4.1 ppb/km; $r = -0.87$), except in Flight 4 when pollution and fire plumes were observed (referred to as Plume A8 below in Section 3.2; plume details in Tab. V.3) in large sections of the flight in mid-troposphere. The same is true for CH_4 (-7.4 ppb/km; $r = -0.87$). Higher concentrations at lower altitudes are due to the proximity with the surface sources for CO (fossil fuel and biomass burning) and CH_4 (mainly anthropogenic emissions and to a limited extent wetlands in late winter). All these combustion-related emissions are expected to influence indirectly the regional O_3 budget.

Further in the paper, we focus on the influence of remote anthropogenic emissions and biomass burning on O_3 concentrations, from emitting regions to the observed field through transport. Despite the remoteness of most emissions, the observed plumes remained coherent and can be distinguished from the ambient background troposphere. Signatures of combustion-related emissions, traced by CO enhancements, are superimposed on background levels of CO resulting from CH_4 oxidation and the oxidation of biogenic hydrocarbons. Carbon monoxide deviations from this background are used for pollution plume identification. In this analysis of the YAK-AEROSIB campaign, the background is considered as uniform (see Pommier et al., 2010).

Before going further in plume description, we define, and describe, 'background' conditions using the following method. The quartile with the highest CO concentrations (> 153 ppb) was considered as polluted and screened out; we also excluded the Plume M (Table V.3) described in Section 3.4 which has very different characteristics from the troposphere observed during the rest of the campaign. In the remaining data, any 2-minute interval in the measurement time series with a CO standard deviation above 10 ppb was flagged out. For the remaining intervals, the first quartile of the concentration distribution was used as an estimate for the CO background concentration of the campaign. The CO background concentration is estimated to be 130 ppb during the campaign.

We consider that air masses with $\Delta CO < 5$ ppb (relative to the CO instrument uncertainties; see Table V.2) can be assimilated to background conditions. During the campaign, the air masses observed in background conditions account for $\sim 33\%$ of the whole data set (Fig. V.2, hatched area, excluding plume M in dark green in the figure and described in Section 3.4). In 'background' air, CO and CO_2 are correlated ($r = 0.78$, $p < 0.001$). This is explained by the global co-emission of these two species during combustion processes. Carbon monoxide

Table V.3: Selected single plumes and specific events. Identifiers begin with 'A' for anthropogenic (Section 3.2), 'M' for 'marine' plume (Section 3.4), 'S' for stratospheric intrusion (Section 3.3). Coordinates are the average position for the selected plumes. Correlations are indicated only if significant, i.e. $p < 0.01$.

Pl. ID	Time UTC	Coordinates			Concentrations (mean $\pm 1\sigma$ dev)				Correlation coef./ regression slope	
		Lat ($^{\circ}$ N)	Lon ($^{\circ}$ E)	Alt (m a.s.l)	CO ₂ (ppm)	CH ₄ (ppb)	O ₃ (ppb)	CO (ppb)	CO/CO ₂ (ppb/ppm)	O ₃ /CO (ppb/ppb)
A1	15 Apr. 04:45	56.3	84.9	6900	393.8 ± 0.3	1883 ± 13	52.5 ± 3.1	138.6 ± 2.4	-	-0.56 -2.26
A2	15 Apr. 05:20	56.4	85.7	1100	398.0 ± 2.7	1920 ± 12	61 ± 0.0	181.2 ± 6.8	0.97 6.1	0.68 0.3
A3	15 Apr. 23:15	62.6	115.3	4500	394.8 ± 0.6	1838 ± 11	63.2 ± 1.7	171.0 ± 13.7	0.94 27.5	-
A4	16 Apr. 05:15	59.5	112.5	6600	397.4 ± 1.3	1858 ± 19	60.7 ± 0.6	162.3 ± 1.7	-	- -2.26
A5	18 Apr. 07:25	56.4	101.8	1000	401.4 ± 0.6	1862 ± 14	54.4 ± 0.6	165.9 ± 4.4	0.74 12.2	-
A6	18 Apr. 04:30	55.7	91.5	5200	396.6 ± 1.1	1860 ± 21	67.5 ± 4.4	174.6 ± 17.5	0.81 12.3	0.56 0.45
A7	18 Apr. 05:15	55.3	88.2	3000	398.4 ± 0.5	1871 ± 17	59.6 ± 1.7	178.5 ± 19.4	0.85 11.0	0.73 0.13
A8	18 Apr. 05:30	55.3	86.3	1600	398.8 ± 0.5	1854 ± 17	52.3 ± 1.7	166.6 ± 2.9	-	0.6 0.81
S	18 Apr. 03:00	55.9	98.5	6700	388.7 ± 0.9	1801 ± 23	177 ± 35.3	90 ± 17	0.57 4.4	-0.92 -2.06
M	16 Apr. 01:00	61.8	121.7	6700	394.1 ± 1.6	1824 ± 22	49 ± 2.2	125 ± 9	0.94 11.1	-0.77 -3.00

and O₃ in background conditions have opposite vertical concentrations profile slopes, caused by (1) the background O₃ chemistry with low photochemical production or destruction in the low troposphere and transport; (2) the widespread surface CO sources; and (3) the occasional tropospheric propagation of CO-poor, O₃-rich stratospheric air.

In the following, Δ CO refers to the difference between a CO concentration and the background concentration. We estimate that 40% of the data is 'polluted' (defined here as Δ CO $>$ 5 ppb). Figure V.2 shows scatter plots of CO₂ and O₃ against CO, highlighting polluted air. In polluted air masses, the overall CO-CO₂ correlation is $r = 0.74$ ($p < 0.001$) with a regression slope of 5.5 ppb CO/ppm CO₂. For comparison, the biomass burning emission ratio is about 107 ppb CO/ppm CO₂ (Andreae & Merlet, 2001); the anthropogenic emission ratio is about 9 ppb CO/ppm CO₂ in the EDGARv4.2 inventory for the European Union and reaches 40 ppb CO/ppm CO₂ in the polluted region near Beijing (Wang et al., 2010). Our slope is somewhat smaller than the emission ratio of anthropogenic combustion sources, confirming that mixing with cleaner air during transport and oxidation of CO by OH has taken place.

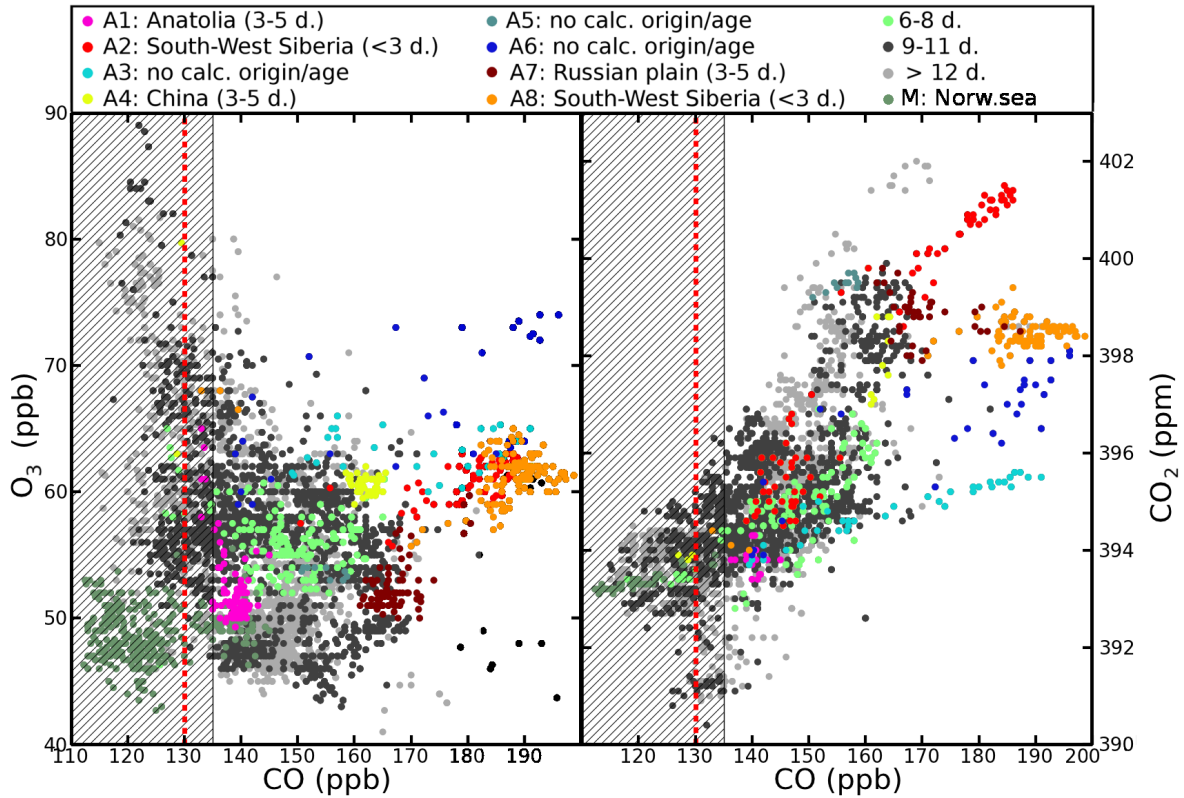


Figure V.2: Species-species scatter plots. Data are coloured according to FLEXPART air mass origin, with calculated modal transport age indicated (in days, between parentheses in the legend, see Section 2.3 for details). Dashed red line represents background CO concentration and observations in hatched area are considered as non-polluted (see Section 3.1). (Left) CO-O₃ scatter plot; (right) same for CO-CO₂.

2.3.2 Influence of combustion processes on O₃ concentrations

In Fig. V.2, plumes are coloured with respect to their origin, along with an estimate of their CO modal age (calculated with FLEXPART as explained in Section 2.3). The CO-O₃ correlations are determined by chemical reactions that can occur during the transport for these two species: CO can be oxidised by OH-radicals; O₃ is produced in the presence of NO_x; O₃ can also be destroyed with NO in the nascent plume, and subsequently produced or destructed in the plume. Negative CO-O₃ correlations are supposedly caused by significant destruction of O₃ during transport. Positive regression slopes significantly higher than for the background conditions can be associated with photochemical O₃ production. When positive correlations are close to the background ones, distinguishing photochemical production, destruction or surface deposition remains challenging. We find no significant CO-O₃ correlations for the whole set of selected observations of polluted air ($r^2 < 0.01$), making it difficult to infer any large-scale patterns of production or destruction of O₃ over Siberia during the period of interest. Correlations in individual plumes show high variability between plumes ($-0.95 < r < 0.9$). These observations are consistent with the large heterogeneity of possible photochemical patterns in combustion plumes in the free troposphere (for example trans-Atlantic pollution and wildfire plumes observed by Real et al., 2007, 2010). These

patterns are strongly dependent on the emission of precursors in the emission region and photochemistry along transport and mixing between air masses.

As we do not have measurements of NO_x and other O₃ precursors, our discussion is based on residence-time with plumes from similar regions having comparable precursor concentrations. We base our discussion of individual plumes on FLEXPART back trajectories and simulated CO enhancements; FLEXPART identified plumes aged of 3-5 d. In Fig.V.3, back trajectories for several selected plumes are plotted, superimposed on a map of a MODIS hot spot satellite detection during the period of interest. Every day backwards in time, the released virtual particles are grouped into five clusters, the position of which is indicated by the coloured symbols (the number within them corresponds to the backward time in days); the colour of the symbols are related to their mean altitude. Some plumes originate from southwestern Siberia (plumes A2 and A8 in Table V.3) and the western Russian plain (plume A7), two populated regions with heavy industry and numerous biomass burning hot spots (mainly agricultural in spring; Korontzi et al., 2006) according to MODIS hot spot satellite detection. In these plumes, we find statistically significant positive CO-O₃ correlations ($r > 0.60$, $p < 0.001$), with regression slopes of 0.1 to 0.8 ppb O₃/ppb CO. Regional anthropogenic activities are likely to emit precursors and hence enhance O₃ mixing ratios downwind in the free troposphere.

Another young plume (A4 in Table V.3 and Fig. V.3) with high CO mixing ratios ($\Delta\text{CO} > 32$ ppb) was calculated to be transported from the Korean peninsula and the eastern Sino-Russian border, where a conveyor belt quickly uplifted the plume. These two regions are industrialised regions with a contribution from biomass burning emissions (about 10% of the FLEXPART calculated CO enhancement is attributed to biomass burning). The plume was observed for about 5 minutes, but calibrations prevented sufficient simultaneous measurements to calculate reliable correlations between species. Nevertheless we observe an overall deficit of O₃ ($|\Delta\text{O}_3| > 15$ ppb) compared to the ambient air observed right before and after the selected plume, where $\Delta\text{CO} = +32$ ppb. The O₃ deficit and CO excess can be partly attributed to a non-negligible mixing with air from the planetary boundary layer (PBL), transported by the conveyor belt. Note that more than 15% of the sample resided in the PBL less than 3 d before measurement according to FLEXPART simulations. But if we use the vertical mean gradients calculated in Section 3.1, the O₃ deficit cannot be explained by only PBL influence. We then cannot exclude possible production with subsequent significant destruction in the plume (e.g., Real et al., 2010). An alternative explanation would be O₃ titration within the plume over the source region where there would be high NO_x concentrations. All these plumes well represented by the FLEXPART model are in their first days downwind of the emission, when photochemical regimes cause the steeper O₃ production and/or destruction.

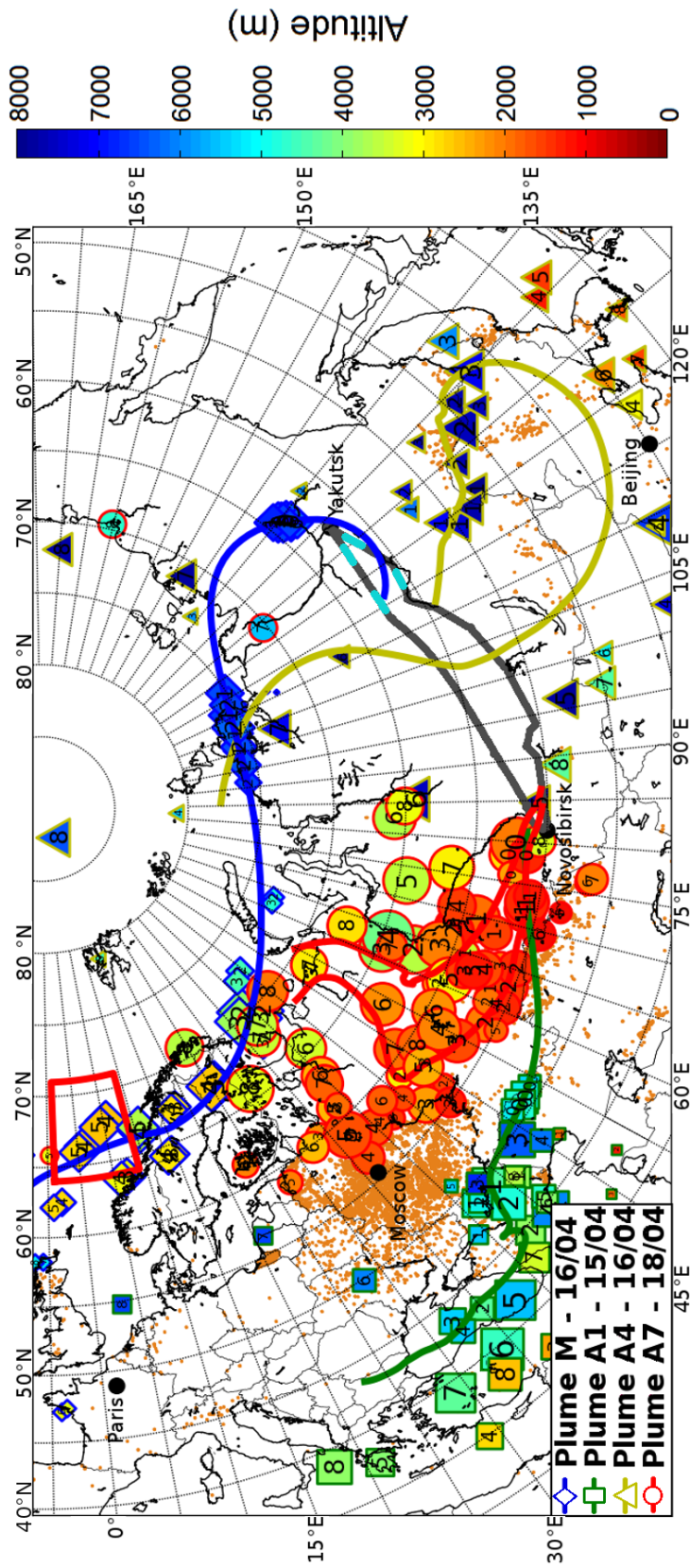


Figure V.3: Back trajectories for four specific plumes including pollution plumes related to biomass burnings and a marine plume, based on ECMWF wind fields. The lines show the mean back trajectory for each one of the four plumes. In addition, every 24 hours each particle cloud's position is divided into four clusters (see text for details), coloured according to their mean altitude. The numbers in each cluster's daily position symbol corresponds to the number of days prior to the particles' release. The aircraft pathway is highlighted by grey solid line; superimposed light blue lines figure plume M observations' positions. Orange dots are positions of fires detected by the MODIS instrument from 11 to 20 April (from the Fire Information for Resource Management System, available online at <http://earthdata.nasa.gov/data/near-real-time-data/firms>).

A few polluted plumes are not well simulated by FLEXPART. These plumes with observed high ΔCO and measured in narrow layers are in blue in Fig. V.2 (see also positions and average mixing ratios in Table V.3; there, plumes A3, A5 and A6). Another plume (A1 in Table V.3) with high modal age (> 6 d) and $\Delta\text{CO} = 10$ ppb was not well simulated by FLEXPART. Model limitations including driving wind fields' resolution, turbulence parameterisations and emission inventories, make the identification of some plumes difficult, especially the oldest ones which tend to be excessively diffused by the model and that we cannot analyse precisely. As a consequence, these inaccuracies in the model prevents a proper analysis of narrow plumes and old ones characterising long-range transport. This deficiency introduces a bias of sampling in our study. Nevertheless, our case studies remain consistent with previous studies with more numerous occurrences of O_3 enhancements than destruction in biomass burning and pollution plume.

2.3.3 Upper troposphere O_3 excess and stratospheric intrusion

Figure V.4a shows an observed decrease in CO and CO_2 mixing ratios and a steep increase in O_3 concentrations during Flight 4 at 7000 m (Plume S in Table V.3). In the sampled area (55.9°N , 98.5°E), observed mixing ratios for CH_4 , CO and CO_2 were much lower than elsewhere at this altitude ($\Delta\text{CO}_2 = -6.0 \pm 0.9$ ppm; $\Delta\text{CO} = -40.1 \pm 17.2$ ppb; $\Delta\text{CH}_4 = -17 \pm 23$ ppb compared to the median above 6000 m for the campaign), whereas O_3 ($\text{O}_3 = 176.8 \pm 35.3$ ppb; up to 215 ppb) was three times higher than in the other observations at these altitudes, characteristic of lower stratosphere air. Since the flight was well below the regional mean tropopause altitude during the campaign (300 hPa or ~ 9000 m in this area on 18 April 2010 according to NOAA/ESRL Physical Sciences Division reanalysis, available at <http://www.esrl.noaa.gov/psd/data/reanalysis/>), such high O_3 concentrations suggest that stratospheric air must have been mixed into the troposphere earlier, or that a tropopause fold was occurring simultaneously to the flight. The sampling of 'fresh' stratospheric air is confirmed by in-situ measured meteorological parameters: the wet bulb potential temperature was 300 K, about 20 K lower than the tropopause mean at this latitude (but almost 5 K higher than measured ambient air at 7000 m in the area), suggesting a quasi-adiabatic intrusion (2000-3000 m descent stretching over about 150 km) of stratospheric air, from the tropopause to mid/upper troposphere.

Over the area where the O_3 excess was measured (spanning roughly 5° by 5°), meteorological datasets (mainly ECMWF potential vorticity fields) suggest a frequency of 20-50 tropopause fold events per year (Beekmann et al., 1997). During frontal episodes, the dry intrusion air stream associated with extra-tropical cyclones generates a positive potential vorticity anomaly associated with the downward flux of stratospheric air. Under these particular circumstances, isentropic transport from stratosphere to troposphere can occur and the tropopause can drop to 320 K (i.e. 300-400 hPa or 6-8 km) for latitudes between 40 and 90°N (Holton et al., 1995).

During Flight 4, the occurrence of a stratospheric intrusion was confirmed by different independent data sets: in situ measured meteorological parameters, satellite imagery and ECMWF analysis fields. The intrusion is associated to a cut-off cyclone (about 150 km wide in its core, with an arm spanning over about 700 km; centred where we observe stratospheric air) with high wind speeds ($> 30 \text{ m.s}^{-1}$ and high potential vorticity (> 2 PVU; potential vorticity units; $1 \text{ PVU} = 10^{-6} \text{ K.m}^2.\text{kg}^{-1}.\text{s}^{-1}$) high wind speed (about 30 m.s^{-1} , downward

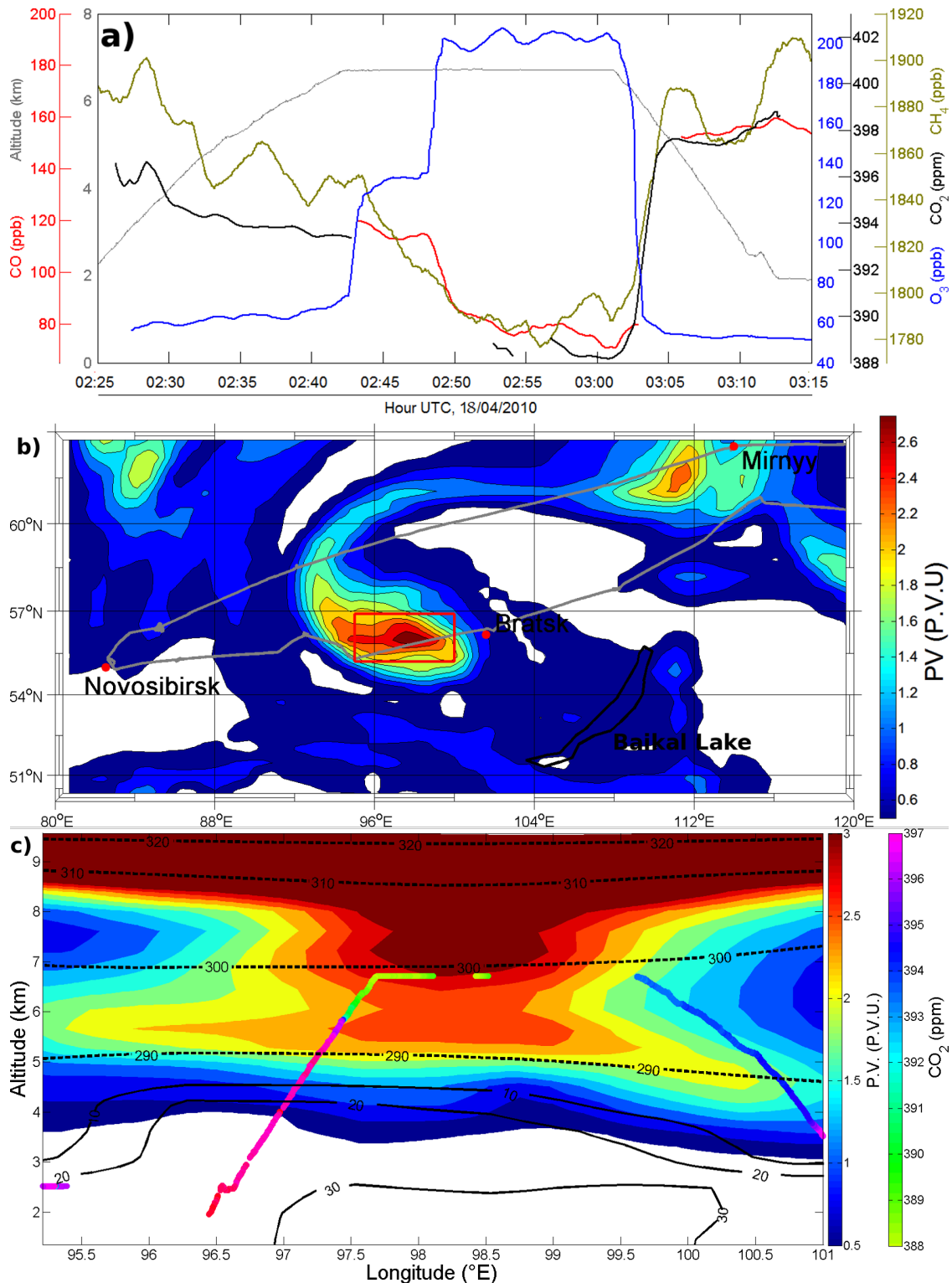


Figure V.4: Stratospheric intrusion. (a) Time series for all measured trace gases. CO_2 and CO data are not continuous because of concurrent in-flight calibrations and instrument correction; (b) ECMWF potential vorticity map at 450 hPa (~ 5700 m), 18 April 03:00; and (c) Potential vorticity section along aircraft trajectory (coloured regarding CO_2 concentrations). Dashed contours for potential temperature in K and solid contour for PBL origin in % calculated by FLEXPART 1 d before the snapshot.

flow (about $5 \text{ m}\cdot\text{s}^{-1}$, estimated with aircraft mean pitch angle during the intrusion) and high potential vorticity ($> 2 \text{ PVU}$). A significant downward flux was also estimated using the mean aircraft pitch during the event. We find evidence of this frontal situation in the measured mixing ratios of other trace gases: CO_2 mixing ratios were about 393.5 ppm east of the intrusion, whereas west of the intrusion they were 2.9 ppm higher (396.4 ppm) at the same altitude. These higher CO_2 concentrations coincide with a strong PBL influence as simulated by FLEXPART (Fig. V.4c, solid lines). These sharp gradients of O_3 and CO_2 mixing ratio across the intrusion's boundary are linked with the contrasted transport patterns across frontal regions, i.e. the warm conveyor belt lifting PBL air into the middle and upper troposphere, and dry intrusion of stratospheric air. Such sharp chemical composition gradients associated to fronts have been documented, e.g. by Esler et al. (2003) and Vaughan et al. (2003).

The FLEXPART backward simulations confirm the stratospheric intrusion, with a long residence time in the stratosphere (more than 7 d in the past 10 d for more than 75% of the sample) prior to detection. FLEXPART associates this stratospheric air mass with long-range transport in the lower stratosphere and upper troposphere. Most of the sampled stratospheric air is calculated to subsequently remain within the troposphere: 90% of the released particles are located in air masses with $\text{PV} < 1.5 \text{ PVU}$ and below 6 km after 8 d.

As the observed STE is very well defined both in the observation and in the ECMWF analysis, we use PV/O_3 correlation to estimate the O_3 influx into the troposphere. We take PV from the ECMWF analyses and O_3 from the measurements along the flight track. We find a PV/O_3 correlation with correlation coefficient $r = 0.90$ and a slope of $140 \text{ ppb } \text{O}_3/\text{PVU}$ along the flight track. Assuming that this correlation is typical for the entire dry air stream, we can use PV from ECMWF analyses to estimate the total amount of ozone carried by the dry intrusion. Hence, integrating inside a chosen domain in regards to PV fields (the red square area in Fig. V.4b), we obtain an O_3 stratospheric input of $2.56 \pm 0.29 \times 10^7 \text{ kg}$ ($9.05 \pm 0.9 \times 10^{13} \text{ kg}$ of air). FLEXPART forward trajectories suggest that 70% of the stratospheric intrusion is well-mixed in the troposphere after 3 d; hence the local flux associated to this event is $70.1 \pm 10.5 \text{ kg } \text{O}_3 \text{ s}^{-1}$. The evaluation of the uncertainties in chemistry-transport models (Stohl et al., 1998), especially in STE context (Meloan et al., 2003), is difficult because of, for example, numerical scheme errors, spatial or temporal mismatches and errors in meteorological forcings. The given uncertainties here are upper error limits calculated by convoluting errors at each step of our discussion. Three STE events are visible in Fig. V.4b over an area of $2.5 \times 10^6 \text{ km}^2$. Assuming that the flux are similar for these events, the corresponding regional flux of O_3 is therefore approximately $9.75 \pm 2.9 \times 10^{10} \text{ molecules } \text{O}_3 \text{ cm}^2 \text{ s}^{-1}$. If we consider that the region and time window that we investigated is large enough to be representative of the spring mean flux over Siberia, our estimation of downward O_3 flux can be compared to climatological studies. This estimation is in the same magnitude of the mean downward flux over Western Europe ($12 \pm 2.7 \times 10^{10} \text{ molecules } \text{O}_3 \text{ cm}^2 \text{ s}^{-1}$; Beekmann et al., 1997) and consistent with other studies values (Ancellet et al., 1994). Sprenger et al. (2007) carried out a climatological study of STE in the Northern Hemisphere. Their conclusions over Siberia estimated low downward flux in winter ($>$ two times lower than our figure) and downward flux of the same magnitude of our result during summer months. The upper troposphere appears to have been more perturbed than the winter climatology and was closer to its summer state rather during the 2010 late winter campaign.

Atmospheric chemistry transport models tend to poorly represent STE, notably because

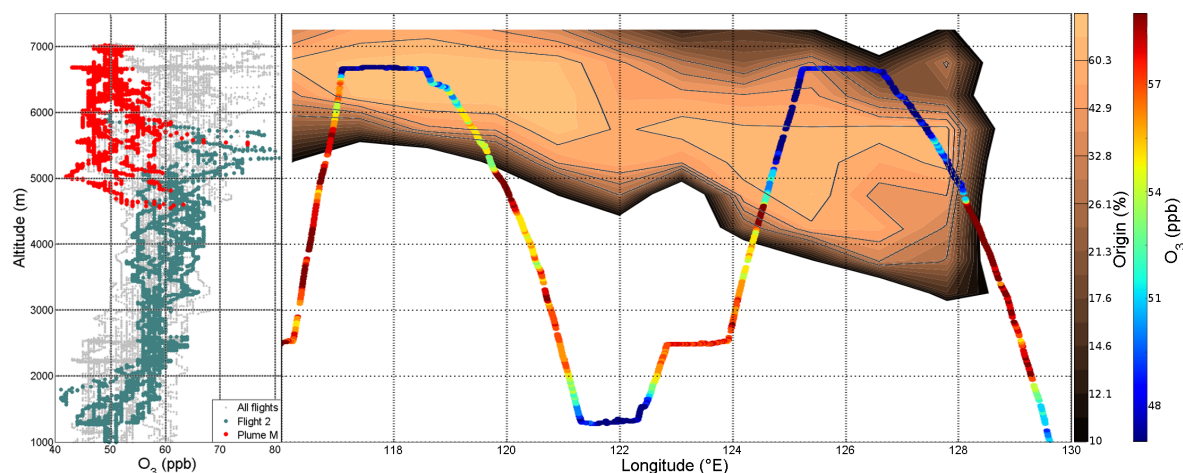


Figure V.5: Measured O_3 and air fraction originating from the North Atlantic lower troposphere for Flight 2 (16 April 2010). (Left) O_3 mixing ratios measured during Flight 2 (red and green symbols) are superimposed over the overall campaign data (gray symbols); the air mass originating from North Atlantic at high latitudes is indicated in red; (right) measured O_3 mixing ratios along aircraft pathway over FLEXPART-calculated North Atlantic origin, expressed in % of the original released particle cloud. The North Atlantic area is defined as: 64–71°N, 3°W–10°E, 0–2000 m (see red box in Fig. V.3 page 86).

the filamentary structure of the tropopause folds is lost by numerical diffusion in global circulation models (e.g., Cristofanelli et al., 2003; Roelofs et al., 2003). Detailed in-situ observations of STE, such as the case presented here, are important because they can contribute to model validation at fine spatial and temporal scales.

2.3.4 Widespread upper tropospheric low O_3 concentrations

Although O_3 concentrations normally increase with altitude in the troposphere, due to the stratospheric influx and upper troposphere photochemical production, we observed an air mass with low O_3 mixing ratios between 6000 and 7000 m, to the west of Yakutsk, sampled along four level-flight sections at 6700 m (plume M in Table V.3 spanning an area of roughly 800×200 km; see Fig. V.3). Observed O_3 concentrations in this air mass were 49.2 ± 2.2 ppb, with a $\Delta O_3 = -14$ ppb relative to the median for all other flight plateaus above 6000 m. There was a sharp gradient in O_3 concentrations between the surrounding mid-troposphere and the low-ozone air mass (see vertical profile in Fig. V.5). CO , CO_2 and CH_4 concentrations were not noticeably different from surrounding air. The concentrations relative to background values were $\Delta CO = -4.5 \pm 9.1$ ppb, $\Delta CO_2 = -0.6 \pm 1.6$ ppm and $\Delta CH_4 = +7 \pm 22$ ppb. As a result, the air mass does not appear to have been influenced by recent emissions from anthropogenic activity or biomass burning with enhanced levels of trace gases.

FLEXPART back trajectories point to similar pathways for all samples within this low- O_3 air mass (Fig. V.3). This air mass spent 3 d in the upper troposphere (> 5000 m) over the north Siberian coast, then 1-2 d in the middle troposphere (> 3000 m) over northern Scandinavia and, finally, 3 d over the Norwegian Sea in the lower troposphere and the PBL. In the right panel of Fig. V.5, we compute the 'flight curtains' approach with FLEXPART (as explained in

Sect. 2.3). For each altitude, we estimate the fraction of the air mass originating from the Norwegian Sea (see red box in Fig. V.3). In Fig. V.5, we also super-impose O_3 concentrations measured during the flight. For the air mass with low O_3 concentrations, at least 80% of FLEXPART particles are simulated to have travelled over the Norwegian Sea in the lower troposphere. Another plume with similar O_3 mixing ratios (47.3 ± 1.2 ppb) was observed at the same altitude during Flight 1, 15 hours earlier than Flight 2, and has similar back trajectories. Other air masses that have resided over the Arctic Ocean were also observed in the upper troposphere but they did not exhibit low O_3 concentrations (during other flights, $O_3 > 60$ ppb in plumes with dominant Arctic origin in the upper troposphere).

A likely explanation would be the transport of an air mass with already low O_3 concentrations over the Norwegian Sea. An air mass with originally low O_3 concentrations and no significant injection of O_3 precursors is likely to remain poor in O_3 during the transport due to reduced O_3 production in the troposphere. It has been observed that the conditions (high H_2O concentrations with low NO_x concentration) in the marine boundary layer tend to cause a net O_3 destruction in the lower troposphere over the North Atlantic (Reeves et al., 2002) and long term measurements in the marine boundary layer point to concentrations of 40 ppb O_3 during spring (Parrish et al., 1998). All the available pieces of information suggest that the observed air mass would have been influenced by marine photochemistry and then transported to East Siberia with a very stable photochemical regime (low O_3 precursor mixing ratios and limited solar radiation in spring in Arctic regions) and almost no mixing with the ambient air rich in O_3 in the upper troposphere. Though less likely, another explanation for the low O_3 concentrations could be the remote influence of the first phase of the Eyjafjallajökull eruption in spring 2010. Forward simulations point to volcanic plumes transported to Siberia with patterns similar to those observed in the upper troposphere. The first phase of the eruption was characterised by no significant emissions of ash and gas compared to the second phase (e.g., Stohl et al., 2011). However, the presence of volcanic fine ashes and/or halogens could possibly have accelerated the O_3 destruction in the observed plumes. This O_3 destruction could have increased the depletion caused by the marine boundary layer influence.

2.4 Conclusions

A new intensive airborne campaign was carried out in the framework of the YAK-AEROSIB project over Siberia in spring between 15 and 18 April 2010. We analysed backward transport calculated with the Lagrangian model FLEXPART to study processes influencing the O_3 budget in the Siberian troposphere.

The 28 vertical profiles of concentrations of a number of atmospheric species collected during the campaign exhibit slightly increasing CO_2 with decreasing altitude, indicating a net surface source of CO_2 (plant respiration, biomass burning and anthropogenic emissions). A corresponding gradient was also found for CO, suggesting the importance of combustion sources. The O_3 gradient, what is positive with increasing altitude, suggests limited deposition at the surface (especially on leaves and through stomatal activity) and net source in the lower stratosphere and upper troposphere.

We show cases of long-range transport of pollutants and biomass burning plumes that influence O_3 mixing ratios over Siberia by focusing on plumes with CO excess from background

conditions as a proxy of combustion processes. Plumes from Russian urban and industrial regions are typically associated with enhanced O_3 concentrations when observed 4-6 d after emission. But low O_3 was also observed in plumes transported from urban region.

An air mass, poor in O_3 ($\Delta O_3 = -20.5$ ppb) but with no significant difference in CO and CO_2 concentrations relative to the background, was observed at 6700 m over a large area. The backward transport simulations indicate that the air mass resided for a long time in the lower troposphere over the Norwegian Sea and the Arctic/North Atlantic Ocean 5 d before airborne measurements. This suggests O_3 destruction in the marine boundary layer with no significant recovery due to low O_3 precursor concentrations and little significant mixing with background air masses richer in O_3 than the original marine air mass until the observation.

Our data also indicate a significant contribution of the stratosphere to the tropospheric O_3 budget over Siberia. We have quantified the input of a large stratospheric intrusion over Siberia. A downward flux was observed in a frontal system with huge enhancement in O_3 mixing ratios (+200%, $\Delta O_3 = +150$ ppb) and depletion in other trace gases. Observations combined with satellite imagery and ECMWF analysed meteorological fields were precise enough to infer an approximate estimation of transported air in this single STE ($2.56 \pm 0.29 \times 10^7$ kg O_3). Global estimates and model representation of processes leading to STE would benefit from comparisons with detailed in-situ measurements such as the one described here.

Acknowledgements

We thank the scientific and flight crew for their expertise and for carrying out successfully the campaigns since 2006. The measurement campaigns were funded under the project LIA YAK-AEROSIB by the ANR Blanc POLARCAT, CLIMSLIP-LEFE, CNRS, the French Ministry of Foreign Affairs, CEA (in France), and RAS and RFBR (in Russia). A. Stohl was supported by the Norwegian Research Council in the framework of the CLIMSLIP project. Funding support from the French ANR Blanc SIMI 5-6 021 01 Climate Impact of Short-lived Pollutants and Methane in the Arctic (CLIMSLIP) is acknowledged.

3 Structures météorologiques et contributions distantes en Sibérie

3.1 Structures de transport caractérisées

Les gradients verticaux de composition atmosphérique ont été analysés lors de la campagne YAK-AEROSIB. Ces gradients en CO_2 , CO et CH_4 , positifs de la surface à la haute troposphère, révèlent une influence significative des sources régionales de surface. Ils sont majoritairement dus à des sources anthropiques et aux feux de biomasses. Néanmoins, on peut aussi s'attendre à une influence significative sur la composition atmosphérique sibérienne des zones inondées en été (juin-septembre).

Des transports à longue distance de masses d'air polluées en CO, CO_2 et O_3 venant de Chine, du Moyen-Orient, de Russie Occidentale et d'Europe ont été identifiés. Par conséquent, ces influences distantes doivent être prises en compte dans une inversion atmosphérique dans la région.

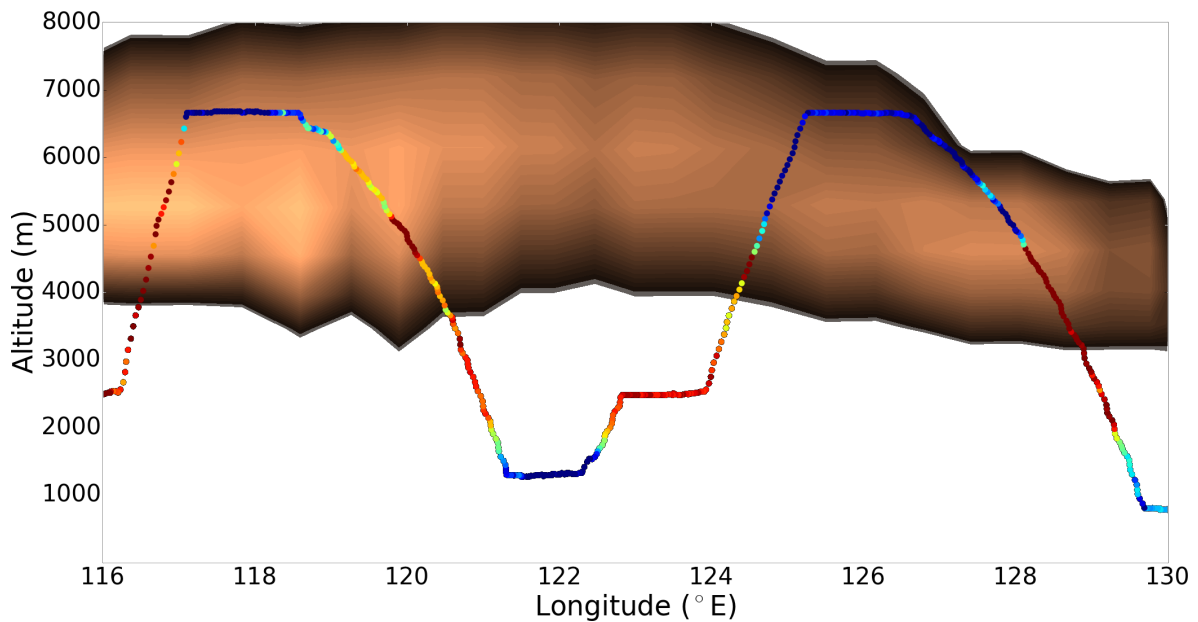
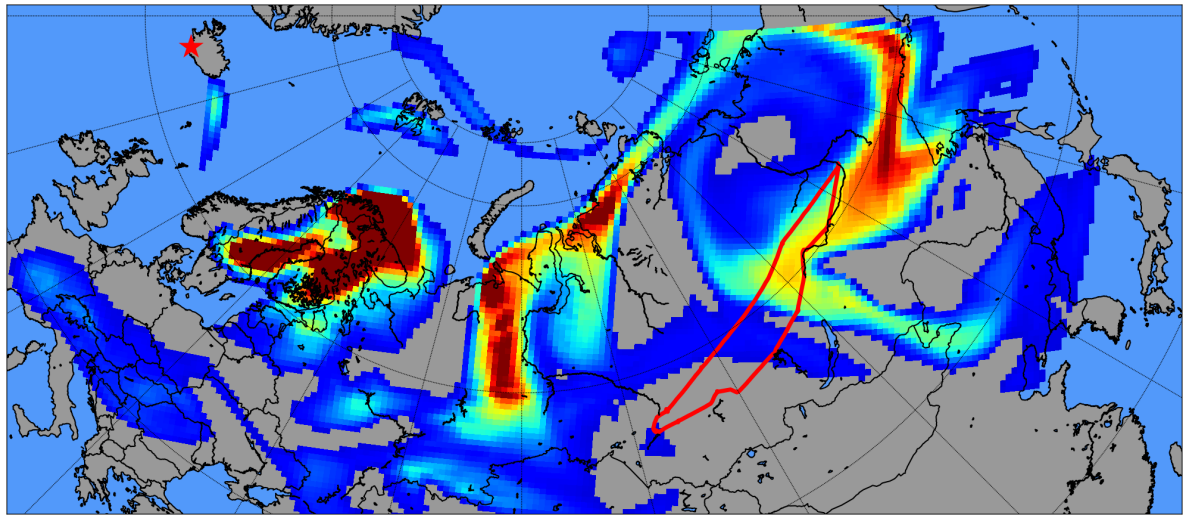


FIGURE V.6: Influence à longue distance du volcan Eyjafjöll sur le Vol 2 de la campagne YAK-AEROSIB d'avril 2010. (haut) Cartes de concentrations relatives simulées par le modèle CHIMERE d'un traceur passif émis par le volcan Eyjafjöll. Les concentrations ont été extraites à un niveau du modèle à ~7000 m d'altitude. (bas) Coupe verticale des concentrations simulées le long de la trajectoire de l'avion. La trajectoire de l'avion et les concentrations mesurées sont superposées à la coupe de la même manière que dans la figure V.5 page 90.

Par ailleurs, l'étude a permis de confirmer qu'un modèle de transport tel que FLEXPART était à même de reproduire de manière cohérente les structures de transport atmosphérique en Eurasie, malgré des imprécisions liées aux paramétrisations et aux résolutions spatiale et temporelle.

3.2 Identification de sources ponctuelles

Le modèle de dispersion lagrangien FLEXPART identifie difficilement des sources ponctuelles distantes. En effet, le nombre de particules virtuelles est fini. On ne dispose donc que d'un maillage discret de la zone influençant potentiellement la masse d'air que l'on étudie. Par conséquent, des émissions ponctuelles peuvent échapper à ce maillage. Pour cette raison, nous avons choisi dans l'article de nous intéresser uniquement à de grandes régions d'émissions potentielles.

Cette identification par régions n'a pas permis de caractériser clairement la cause des faibles concentrations d'ozone mesurées à haute altitude lors du deuxième vol de la campagne. La cause la plus probable était le transport de masses d'air appauvries en ozone venant de la couche limite marine au large de la Norvège. Mais l'éruption de l'Eyjafjöll en Islande pouvait aussi avoir engendré des destructions photochimiques d'ozone dans les panaches de gaz volcaniques.

Pour confirmer ou infirmer cette hypothèse, nous avons procédé à des simulations directes à l'aide du modèle eulérien CHIMERE avec une source ponctuelle située aux coordonnées du volcan islandais durant la première phase de l'éruption de 2010. La figure V.6 résume ces simulations. Les émissions du volcan sont transportées jusqu'à la zone échantillonnée par la campagne YAK-AEROSIB au moment du deuxième vol. La répartition verticale de l'influence du volcan Eyjafjöll est compatible avec les altitudes auxquelles ont été observées les masses d'air appauvries en ozone. Néanmoins, la répartition verticale de l'influence de la mer de Norvège de la figure V.5 page 90 semble à première vue plus cohérente avec les concentrations mesurées.

Sans information supplémentaire, il est difficile de conclure car chaque modèle peut sembler se tromper ou au contraire bien reproduire la réalité pour de mauvaises raisons.

3.3 Estimation de l'influence des hydrates de méthane en mer des Laptev

La mer des Laptev peut émettre une quantité de méthane non négligeable à l'échelle régionale (voir sect. 2.1.4 page 30 du chapitre II). Il est nécessaire d'estimer si des masses d'air venant de la mer des Laptev sont transportées jusqu'à des sites qui seront ensuite utilisés dans l'inversion atmosphérique. L'analyse des données collectées pendant la campagne YAK-AEROSIB d'avril 2010 n'a pas permis de trancher cette question.

Par conséquent, nous avons estimé la possible influence de la mer des Laptev (voir figure V.7) à l'aide du modèle CHIMERE centré sur le pôle Nord (modifié à partir de la version distribuée de CHIMERE ; voir Annexe C) pour l'année 2010 complète. Grâce à ce domaine, nous simulons la trajectoire des masses d'air tout autour du cercle polaire et pouvons vérifier si des panaches riches en méthane venant de la mer des Laptev atteignent les sites d'observations sibériens.

La circulation moyenne autour du pôle Nord se fait d'Ouest en Est. Les émissions de la mer des Laptev sont donc préférentiellement exportées vers l'Alaska et l'extrême nord du Canada. Les rares masses d'air circulant jusqu'à la Sibérie suivent donc un chemin détourné et sont très diluées lorsqu'elles survolent la Sibérie. Par conséquent, l'influence d'une mer des Laptev émettant $8 \text{ TgCH}_4 \cdot \text{a}^{-1}$ (Shakhova et al., 2010) ne dépasse pas les quelques ppb en Sibérie, ce qui est négligeable (quelques %) par rapport aux contributions simulées des émissions

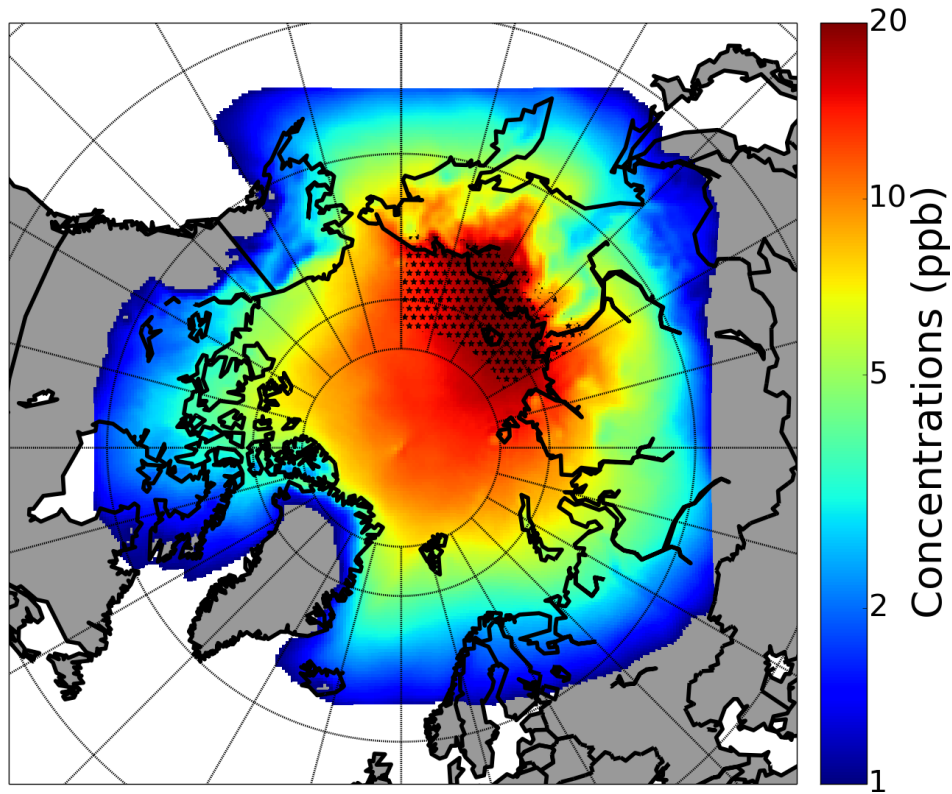


FIGURE V.7: Transport des émissions des hydrates de méthane en mer de Laptev. Concentrations moyennes simulées sur l'année 2010 pour des émissions de $8 \text{ TgCH}_4 \cdot \text{a}^{-1}$ par des hydrates de méthane en mer de Laptev (zone étoilée).

régionales anthropiques et des zones inondées.

Dans la suite, nous négligerons la source des hydrates de méthane dans nos inversions. Une estimation des émissions de méthane en mer de Laptev a été effectuée à l'aide d'une analyse statistique simple des simulations de la configuration polaire de CHIMERE comparées à des observations. Cette estimation a donné lieu à la rédaction d'un article soumis à *Geophysical Research Letters* et dont le manuscrit est retranscrit en Annexe D.

Adaptation de méthodes de quantification des statistiques d'erreurs

Sommaire

1	Choix des observations et erreurs associées	99
1.1	Types d'observations pertinentes pour contraindre les flux en Sibérie	99
1.2	Traduction dynamique des informations des observations de surface	100
2	Critères mathématiques objectifs d'ajustement des erreurs	101
2.1	Mauvaise connaissance identifiée des matrices de covariance	101
2.2	Ajustement objectif des matrices	102
2.3	Application à un cas d'étude	102
3	<i>Error statistics for atmospheric inversion</i>	102
3.1	<i>Introduction</i>	104
3.2	<i>Methods</i>	106
3.2.1	<i>Inversion system</i>	106
3.2.2	<i>Error configuration : description of the algorithms</i>	110
3.3	<i>Results</i>	114
3.3.1	<i>Patterns in the error variances for the 3 methods</i>	115
3.3.2	<i>Temporal variability and diurnal cycle of the observation errors</i>	116
3.3.3	<i>Correlations in the observation errors</i>	119
3.3.4	<i>Correlations in the background errors</i>	120
3.4	<i>Flux inversion : the Elgin leak case</i>	123
3.4.1	<i>Physical relevance of the inverted fluxes</i>	123
3.4.2	<i>Using the optimised tuples (\mathbf{R}_{ND}, \mathbf{B}_{ND}) for flux inversion</i>	123
3.5	<i>Discussion</i>	125
3.5.1	<i>Limitations and hypothesis probation</i>	125
3.5.2	<i>Implications for data selection</i>	126
3.6	<i>Conclusions</i>	129
4	Des erreurs objectivées mais encore incertaines	130

4.1	Choix de la méthode d'ajustement de matrices	130
4.2	Incertitudes de la méthode	130
4.3	Acquis sur le système d'inversion	131

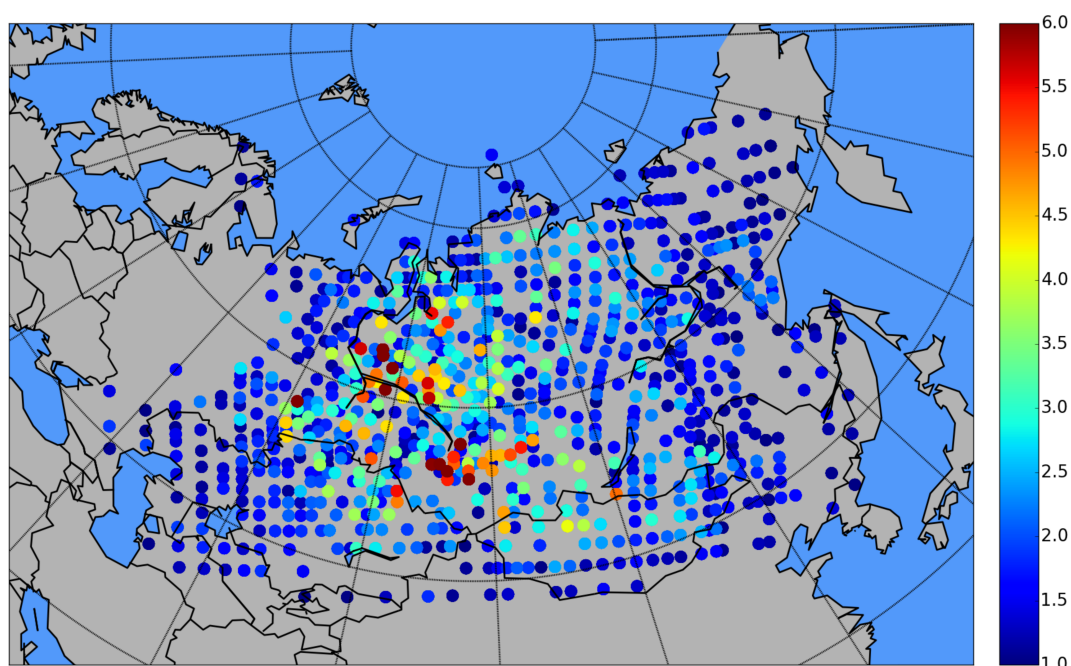


FIGURE VI.1: Contributions régionales sibériennes aux colonnes GOSAT en ppb. Les émissions sibériennes ont été transportées à l'aide du modèle CHIMERE et projetées sur les colonnes d'air GOSAT.

Nous souhaitons évaluer de la manière la plus objective possible les statistiques d'erreur pour une inversion (voir sect. 3.3 page 54 du chapitre III). Pour cela, nous commençons par sélectionner le type d'observations pertinentes pour une inversion en Sibérie et identifions les erreurs dans la comparaison modèle-mesure pour ces observations (sect. 1). Ensuite, nous choisissons des critères mathématiques pouvant nous permettre d'ajuster les matrices d'erreurs de manière objective (sect. 2). Nous présentons dans la sect. 3 un cas d'étude dans lequel nous testons plusieurs méthodes possibles d'ajustement de matrices. Nous faisons enfin le bilan des enseignements tirés de ce cas d'étude (sect. 4).

1 Choix des observations et erreurs associées

1.1 Types d'observations pertinentes pour contraindre les flux en Sibérie

La majorité des flux sibériens de méthane se situent au niveau de la plaine de Sibérie occidentale. Dans la figure V.7 page 95, nous avons vu que l'influence d'une zone d'émission devient faible après quelques centaines de km, même pour un flux conséquent de $8 \text{ TgCH}_4 \cdot \text{a}^{-1}$. Bousquet et al. (2011) ont fait le même type d'expérience et montré qu'une émission ponctuelle est entièrement diffusée dans le signal atmosphérique de fond en quelques centaines de km. De plus, la structure du transport ne permet pas toujours de contraindre des flux avec des observations, même lorsque celles-ci sont à proximité des zones d'émissions. De ce fait, nous devons disposer d'observations à proximité et en aval des émissions suivant la circulation atmosphérique.

Les observations satellitaires proposent une très bonne couverture spatiale d'observation.

Néanmoins, les instruments embarqués sur des satellites estiment des concentrations d'air intégrées sur des colonnes d'air. Le signal atmosphérique est donc moyenné, ce qui rend difficile la détection d'un signal régional. En effet, comme on le voit dans la figure VI.1 page 99, les émissions régionales de méthane contribuent pour moins de 6 ppb aux colonnes totales de méthane GOSAT, alors que le signal mesuré par GOSAT varie entre 1750 et 1850 ppb, avec une erreur instrumentale de 1–2%, soit 15–30 ppb. Les mesures satellitaires ne sont donc pas appropriées pour une implémentation dans une inversion régionale des émissions de méthane en Sibérie.

Les mesures aéroportées, elles, sont précises (en l'absence de dysfonctionnement instrumental), mais les campagnes de mesures sont très ponctuelles en temps et en espace. Or, nous avons vu dans le chapitre V que les masses d'air survolent la Sibérie en quelques jours. Une campagne ponctuelle pourra donc contraindre les flux de méthane uniquement sur de courtes périodes. Elle ne permettra par conséquent pas d'estimer la variabilité temporelle des flux régionaux de méthane.

Pour ces raisons, nous utilisons dans la suite les données d'observations de surface in situ et en flacons (voir sect. 1.1 page 62 du chapitre IV). Ce choix impose d'identifier les erreurs possibles faites par les modèles atmosphériques à proximité de la surface.

1.2 Traduction dynamique des informations des observations de surface

Les sites d'observations choisis fournissent des mesures de concentration atmosphérique de méthane à des hauteurs inférieures à 300 m du sol. Les mesures sont de plus quasi-continues pour la plupart. Les mesures quasi-continues sont intégrées à une échelle horaire afin d'améliorer la qualité des données.

Ces caractéristiques des mesures de surface imposent des contraintes sur le modèle utilisé pour simuler le transport atmosphérique.

Premièrement, la faible hauteur par rapport au sol des points de mesure situe les sites de surface dans la couche limite atmosphérique la grande majorité du temps. Par conséquent, le mélange vertical atmosphérique local doit être bien représenté dans le modèle. La manière dont nous relierons les émissions aux concentrations grâce au modèle lagrangien FLEXPART néglige la structure du mélange vertical dans la couche limite. De ce fait, l'utilisation de FLEXPART pour une inversion est exclue.

Nous utiliserons donc un modèle eulérien.

La basse résolution d'un modèle eulérien mondial détériore la représentation synoptique du transport des masses d'air. Les zones frontales de forts gradients de concentrations sont mal représentées par un modèle mondial tel que LMDz. Une mauvaise représentation de la variabilité horaire du transport rend critique l'utilisation d'observations horaires. Pour ces raisons, nous choisissons d'utiliser le modèle eulérien régional CHIMERE.

De surcroît, les mesures à proximité de la surface doivent être reproduites par un modèle s'appuyant sur une orographie précise. Un modèle à basse résolution moyenne l'orographie sur de grandes zones. L'orographie choisie peut alors être imprécise (voir figure VI.2). De la sorte, il sera difficile de choisir le niveau du modèle auquel il faut comparer l'observation. Les exigences en termes d'orographie confirment donc le choix d'un modèle régional tel que

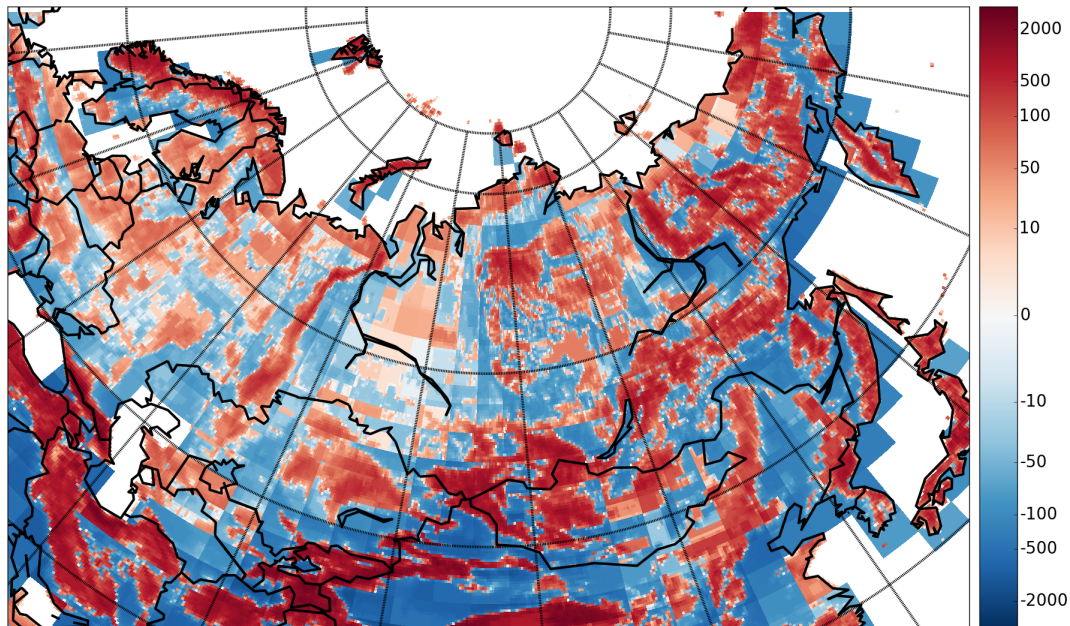


FIGURE VI.2: Comparaison de l'orographie (en m) dans ETOPO (voir figure II.1 page 22) et dans LMDz au-dessus de la Sibérie.

CHIMERE.

2 Critères mathématiques objectifs pour l'ajustement des statistiques d'erreurs

2.1 Mauvaise connaissance identifiée des matrices de covariance

Les considérations physiques ci-dessus permettent d'anticiper un certain nombre d'erreurs possibles dans l'inversion et de potentiellement garantir de meilleurs résultats d'inversion par rapport à d'autres choix physiques. Nous avons focalisé notre réflexion sur les observations, mais des considérations physiques peuvent aussi améliorer notre connaissance des incertitudes sur les flux a priori

Néanmoins, à ce point de notre étude, il est impossible d'estimer explicitement la forme des matrices de covariance d'erreurs \mathbf{R} et \mathbf{B} , ainsi que la valeur des variances d'erreurs.

Nous nous attendons en effet à des erreurs de mélange vertical supérieures la nuit par rapport au jour, lorsque la couche limite est bien développée. Les zones frontales sont par ailleurs un problème connu dans les simulations du transport atmosphérique de traceurs. Les processus internes aux mailles tels que la turbulence sont simulés en moyenne via des paramétrisations, ce qui contribue aussi encore aux erreurs. On peut anticiper de surcroît que les erreurs liées à la hauteur de couche limite seront corrélées spatialement puisque, pour une hauteur de couche limite donnée, le modèle réagit d'une manière prescrite.

La liste de réflexions possibles sur les erreurs est longue, mais il est difficile d'attribuer une valeur précise à toutes les erreurs repérées. Il est de plus techniquement hasardeux de

déterminer quelles erreurs dominent, avec quelle localisation et à quelles périodes. Enfin, le « jugement expert » dont nous disposons sur les erreurs est nécessairement subjectif et certaines erreurs peuvent être oubliées.

Nous avons vu dans la sect. 3.3.2 page 54 du chapitre III que des informations précises sur les matrices d'erreurs \mathbf{R} et \mathbf{B} étaient requises pour garantir des résultats d'inversion précis. Le « jugement expert » n'apporte pas les informations nécessaires. En conséquence, il nous faut déterminer une manière objective de prescrire les matrices d'erreurs avant de procéder à une inversion.

2.2 Ajustement objectif des matrices

Nous avons vu dans la sect. 4.2 page 56 du chapitre III que les matrices de covariance d'erreurs possèdent un ensemble de propriétés mathématiques que l'on peut utiliser après une inversion pour vérifier la cohérence mathématique de celle-ci. Généralement, ces propriétés sont difficiles à vérifier à cause de calculs lourds.

Nous prenons le parti de simplifier le problème en réduisant la dimension des espaces d'état et d'observation afin de rendre les calculs de ces propriétés mathématiques simples. De la sorte, nous pouvons ajuster de manière itérative les matrices de covariance \mathbf{R} et \mathbf{B} jusqu'à converger vers un point respectant la propriété mathématique choisie.

En cela, nous suivons ce qui a été fait précédemment dans d'autres domaines que les inversions de flux de méthane (Dee, 1995; Chapnik et al., 2004; Michalak et al., 2005; Winiarek et al., 2012). Nous adaptons à notre contexte les études citées et les appliquons à notre cas d'étude d'inversion atmosphérique de méthane.

2.3 Application à un cas d'étude

La Sibérie est une région peu explorée en termes d'inversion atmosphérique de méthane. Les flux y sont mal connus et les observations sont dispersées. Par conséquent, nous commençons par adapter les méthodes d'ajustement de matrice sur un cas d'étude européen pour lequel les émissions sont bien répertoriées et les observations assez denses.

En mars 2012, un incident a causé une importante fuite de gaz naturel au niveau d'une plate-forme pétrolière au large de l'Écosse. Cette fuite a été colmatée deux mois plus tard. Pendant cette période, des émissions correspondant à environ 1% des émissions régionales de méthane ont été estimées.

Dans ce cadre, notre objectif est d'évaluer si un réseau de mesures de surface (voir figure VI.3 page 107) autour de la source ponctuelle est en mesure de détecter une telle émission. Pour ce faire, du fait de la faible contribution de la fuite à l'échelle régionale, les matrices de covariance d'erreurs doivent être définies de manière optimale.

3 Towards better error statistics for atmospheric inversions of methane surface fluxes

Cet article se base sur une adaptation à l'inversion atmosphérique de méthodes développées dans le domaine de l'assimilation de données en météorologie. Cet article a été publié le

29 juillet 2013 dans le journal de l'éditeur Copernicus de l'*European Geosciences Union, Atmospheric Chemistry and Physics*. Il est disponible dans sa version définitive à l'adresse internet suivante : <http://www.atmos-chem-phys.net/13/7115/2013/acp-13-7115-2013.html>. Nous le transcrivons ci-dessous dans son intégralité.

Les auteurs ayant participé à ce travail sont : Antoine Berchet¹, Isabelle Pison¹, Frédéric Chevallier¹, Philippe Bousquet¹, Sébastien Conil², Michael Geever³, Tuomas Laurila⁴, Jost Lavric⁵, Morgan Lopez¹, John Moncrieff⁶, Joan Necki⁷, Michel Ramonet¹, Martina Schmidt¹, Martin Steinbacher⁸, et Jérôme Tarniewicz¹.

Résumé

Nous adaptons des méthodes statistiques générales afin d'estimer des matrices de covariance d'erreurs optimales dans un système d'inversion régional des émissions de surface de méthane à partir de concentrations atmosphériques. En nous appuyant sur un nombre minimal d'hypothèses physiques sur la forme des matrices d'erreurs, nous calculons des statistiques d'erreurs optimales au regard de critères statistiques objectifs. Dans le cadre de cette approche générale appliquée à un cas réaliste, nous reproduisons des erreurs dans les observations et le vecteur d'état qui sont compatibles avec le « jugement expert » alors que seuls des critères statistiques ont été spécifiés. Avec un nombre réduit d'hypothèses sur la forme des erreurs, nos résultats rendent compte de la variabilité et de l'interdépendance des erreurs. Ces variabilité et interdépendance sont causées par la mauvaise représentation des observations dans le modèle de transport, ou par l'incapacité du modèle à reproduire des situations de forts gradients de concentrations. Les situations sujettes à de forts biais (p. ex. la nuit quand le mélange vertical est mal reproduit par le modèle de transport) peuvent aussi être diagnostiquées à l'aide de nos méthodes. Ce diagnostic peut permettre de déterminer les priorités d'amélioration d'un modèle de transport pour augmenter la précision des inversions. Nous avons aussi analysé la sensibilité de l'inversion aux observations. Cette analyse permet d'optimiser la sélection des données implémentées dans l'inversion. Nous avons appliqué notre méthode à un cas d'étude réel. Un accident dans une plate-forme pétrolière en mer de Nord a généré de fortes émissions de méthane dans l'atmosphère au printemps 2012. Nos estimations des émissions sont cohérentes avec les déclarations officielles.

Abstract

We adapt general statistical methods to estimate the optimal error covariance matrices in a regional inversion system inferring methane surface emissions from atmospheric concentrations. Using a minimal set of physical hypotheses on the patterns of errors, we compute a guess of the error statistics that is optimal in regard to objective statistical criteria for the specific inversion system. With this very general approach applied to a real-data case, we recover

¹Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CNRS-CEA-UVSQ, Orme des Merisiers, CEA Saclay, Gif-sur-Yvette, France.

²Andra, DRD/OS, Observatoire Pérenne de l'Environnement, France.

³National University of Ireland, Galway, Ireland.

⁴FMI, Finnish Meteorological Institute, Helsinki, Finland

⁵Max Planck Institute for Biogeochemistry, Jena, Germany

⁶University of Edinburgh, Edinburgh, United Kingdom

⁷AGH University of Science and Technology, Krakow, Poland

⁸Empa, Laboratory for Air Pollution/Environmental Technology, Duebendorf, Switzerland

sources of errors in the observations and in the prior state of the system that are consistent with expert knowledge while inferred from objective criteria and with affordable computation costs. By not assuming any specific error patterns, our results depict the variability and the inter-dependency of errors induced by complex factors such as the misrepresentation of the observations in the transport model or the inability of the model to reproduce well the situations of steep gradients of concentrations. Situations with probable significant biases (e.g., during the night when vertical mixing is ill-represented by the transport model) can also be diagnosed by our methods in order to point at necessary improvement in a model. By additionally analysing the sensitivity of the inversion to each observation, guidelines to enhance data selection in regional inversions are also proposed. We applied our method to a recent significant accidental methane release from an offshore platform in the North Sea and found methane fluxes of the same magnitude than what was officially declared.

3.1 Introduction

Quantifying the methane (CH_4) fluxes between the surface and the atmosphere, establishing their temporal variability and spatial distribution, and estimating the anthropogenic and natural contributions to these fluxes is critical for closing the present-day methane budget. One of the approaches used for this purpose, called the atmospheric inversion, assimilates information about atmospheric composition to infer surface fluxes. This type of top-down estimation relies on the assimilation of in-situ observations of atmospheric concentrations (Houweling et al., 1999, 2006; Hein et al., 1997; Pison et al., 2009; Bousquet et al., 2011, 2006; Bergamaschi et al., 2005) and/or of remote-sensing data from satellite-based instruments (e.g., Bergamaschi et al., 2009). Using observations for inversions at the global scale reduces the uncertainties on the mean CH_4 flux balances on large regions (typically a few millions of km^2 large). At the regional and mesoscales, high-resolution inversions potentially provide the spatial distribution of the fluxes, so that the characterisation of the processes involved can be improved (Bergamaschi et al., 2010). Inversions at any scale depend on simulations of the atmospheric mixing and advection by Chemistry-Transport Models (CTMs) to estimate the influence of emissions and sinks on the atmospheric concentrations where they are measured. Whether they are based on coarse (Chen and Prinn, 2006; Hein et al., 1997), varying (Peylin et al., 2005) or high (Lauvaux et al., 2008; Sarrat et al., 2007) resolutions, all the CTMs suffer to a certain extent from uncertainties in reproducing the atmospheric concentrations. The uncertainties are due to the transport errors (Baker et al., 2006; Geels et al., 2007; Peylin et al., 2002; Ahmadov et al., 2007; Prather et al., 2008), to the assumption that a point observation can be compared to the mean simulated concentration on the corresponding grid box, i.e., the representation errors (Gerbig et al., 2003a; Tolk et al., 2008), or to the errors from aggregating the fluxes on large regions (Kaminski et al., 2001).

In the framework of Bayesian atmospheric inversion (Enting et al., 1993; Tarantola, 1987), the implementation of a system requires obtaining an advanced understanding of the statistics of the observational and instrumental errors, the transport errors, the representation errors, and the errors of the prior distribution and magnitude of the fluxes prescribed in the system. Most of the cited works empirically assigned these error statistics. Objective methods of tuning the errors in the system also exist (Wahba et al., 1994; Dee, 1995; Desroziers and Ivanov, 2001) and have been applied to get the general structure of the errors (Michalak et al., 2005; Winiarek et al., 2012). But these methods rely on subjective prior knowledge on the error

structure (e.g., isotropic spatial correlation or temporal decay in the correlations), which can limit the generality of the results.

In this study, we apply three different methods based on the statistical and algebraic properties of the errors, but with a minimum of additional physical assumptions on the error patterns. In inversion systems typically solving fluxes at the model resolution (e.g., $\sim 0.5^\circ \times 0.5^\circ$ each week during a season or a year in regional scale studies), this approach would require the handling of matrices of error covariances the total size of which exceeds billions or even trillions of components. To embrace memory limitations and reduce the computation costs, we have chosen a short window of inversion and have aggregated the surface fluxes on synoptic-scale regions. This simplification allows applying powerful generic methods, but induces limitations (Kaminski et al., 2001; Bocquet et al., 2011) that have to be taken into account when moving to a full-resolution inversion system.

Our study exploits a recent unexpected release of CH_4 in the North Sea in spring 2012 to apply this statistical approach and test the ability of a European network of atmospheric observations to detect the leak. On 25 March 2012, an offshore oil platform on the Elgin field, located 200 km east of Scotland shores (57° N , 1.53° E), was evacuated due to a gas leak. The company operating the platform gave a rough evaluation of the flux reaching $200\,000 \text{ m}^3 \text{ d}^{-1}$ or 140 metric tons per day (td^{-1}) for CH_4 , which accounts for less than 1% of the daily regional emissions (within a radius of $\sim 750 \text{ km}$ around the leak point) according to the Emission Database for Global Atmospheric Research (EDGAR v4.2; <http://edgar.jrc.ec.europa.eu>) for the year 2008. The leak was stopped two months after. The methane plume emitted by this point source is difficult to extract from the observation noise and from the variability of the other sources, which makes the assignment of error statistics particularly critical (Winiarek et al., 2012). We develop and apply a regional inversion framework based on CHIMERE CTM simulations (Vautard et al., 2001) on a domain covering the European continent (Fig. VI.3). Relying on objective statistical criteria, we optimise the covariance matrices of the errors of the observations and of the prior state vector (surface fluxes, initial and lateral boundary conditions) for two independent time windows of inversion: the 2 weeks before the beginning of the leak and the 2 weeks after. The three methods of optimisation are implemented with acceptable computation times. They managed to produce error covariance matrices, which are specifically suited to the system and the inversion window, and that are a best guess of the optimum in regard to the objective criteria of each method. Complex error structures are then retrieved. And within the framework of the underlying assumptions, every piece of information provided by the observations and the prescribed fluxes is entirely used. We then use the computed matrices to invert the European fluxes before and after the leak start and test whether or not the atmospheric network detected this methane plume.

In Sect. 3.2, we describe the inversion methods and the dataset used in the study. We also present the algorithms that we implemented following the literature to specify the inversion system configuration. In Sect. 3.3, the results of these tuning methods are presented. The inversion results from these sets are analysed in Sect. 3.4 and their limitations and possible adaptation to larger systems are discussed in Sect. 3.5.

3.2 Methods

3.2.1 Inversion system

Theory: analytical framework

We apply classical data assimilation methods based on the Bayesian formalism (Courtier et al., 1994; Enting et al., 1993, 1995; Tarantola, 1987). In the following we use the unified notation by Ide et al. (1997). Assuming a Gaussian nature for all the errors, the method basically relies on the minimisation of the cost function:

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}(\mathbf{y}^o - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x})) \\ &\quad + \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) \\ &= J^o(\mathbf{x}) + J^b(\mathbf{x}) \end{aligned} \quad (\text{VI.1})$$

J^o (resp. J^b) is the contribution of the observations (resp. the background) to the total cost function. \mathbf{y}^o accounts for the observation vector (dim $\mathbf{y}^o \sim 5000$; description in Sect. 3.2.1); \mathbf{x} is the state vector (i.e., the series of the variables to optimise), \mathbf{x}^b the background vector including the prior knowledge on the state of the system (dim $\mathbf{x} \sim 100$; description in Sect. 3.2.1).

H is the observation operator converting the information in the state vector to the observation space. The atmosphere advects and mixes the emissions and the initial and lateral boundary conditions. Since the time of residence of air masses within our domain is of typically two weeks, to be compared to the mean CH_4 lifetime in the atmosphere of about nine years (e.g., Dentener et al., 2003), the chemistry along the transport is neglected in all the following. For each observation $(\mathbf{y}^o)_i$ at a specified place and time, an equivalent $H(\mathbf{x})_i$ is induced from the state vector \mathbf{x} with a CTM (description in Sect. 3.2.1). Throughout the study, H is supposed linear: the operator is represented by its Jacobian matrix \mathbf{H} and $H(\mathbf{x})$ is identified to $\mathbf{H}\mathbf{x}$. With our relatively low-dimensional system, it is possible to explicitly estimate the observation operator with so-called “response functions”. Response functions (Bousquet et al., 1999b) are calculated for each component of the state vector by running the CTM in forward mode and extracting the simulated concentration at each point where a corresponding observation is available.

The covariance matrix \mathbf{R} describes the errors $\epsilon = \mathbf{y}^o - \mathbf{H}\mathbf{x}$ between the observations and their reconstruction from the state vector with the model. We assume that the errors are unbiased, i.e., $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. \mathbf{R} then encompasses the errors directly related to the measurement process, but also to the transport model (Ahmadov et al., 2007; Peylin et al., 2002), to the model representation, i.e., its inability to represent the local variability within the grid cells (Tolk et al., 2008; Geels et al., 2007) and to the aggregation process (Kaminski et al., 2001; Bocquet et al., 2011). \mathbf{B} is the covariance matrix of the errors on the background vector \mathbf{x}^b (details in Sect. 3.2.1).

For this linear problem, the cost function J admits a global minimum reached for the optimum state vector $\hat{\mathbf{x}}^a$ such that:

$$\hat{\mathbf{x}}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \quad (\text{VI.2})$$

where $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$ is the Kalman gain matrix.

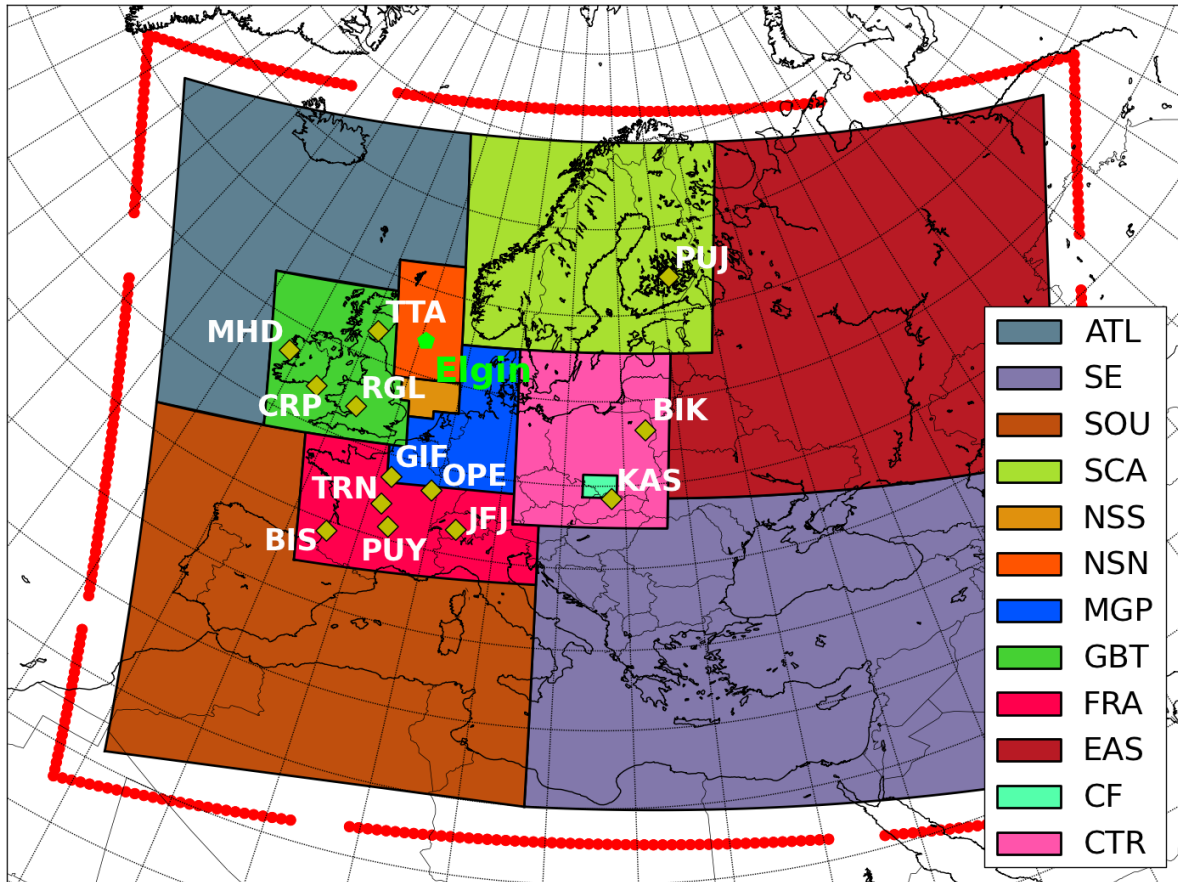


Figure VI.3: Spatial extension of the simulations with the CTM CHIMERE: observation sites marked by yellow diamonds and white legends; the leak position in the northern part of the North Sea is pointed by the green pentagon. The fluxes, the spatial distribution of which is interpolated from EDGAR v4.2 inventory, are aggregated by regions figured by the coloured area. The red edges of the domain denote the spatial distribution of the lateral boundary condition components.

The associated covariance matrix of errors, representing the a posteriori uncertainties in Gaussian assumptions, is given by:

$$\mathbf{P}^a = \mathbf{B} - \mathbf{KHB}. \quad (\text{VI.3})$$

The state vector

In Eq. (VI.1), \mathbf{x} stands for the state vector which is optimised by the inversion and \mathbf{x}^b figures the assumed state (called the “background”) before the assimilation of the observations. The vector \mathbf{x} contains all the information on all the degrees of freedom of the system, e.g., on the emissions (spatial and temporal distribution), boundary conditions (all CH_4 concentration fields at the edges of the domain) and initial conditions (3-dimensional distribution of the CH_4 concentrations at the first step of the period of interest). In most realistic systems, dealing with the complete state vector implies prohibitive computational costs and/or unaffordable memory needs. Solving a high dimension system needs the implementation of

variational algorithms (Chevallier et al., 2005). An alternative is to reduce the size of the inverse problem.

We drastically simplify the state and background vectors in order to allow the manipulation of the matrix \mathbf{B} . For each of the two windows of inversion (the two weeks before and the two after the leak start), computed independently, the simplified vector has a dimension of 99 and contains coefficients of linear corrections on: (1) the aggregated emissions on 12 regions (see Fig. VI.3); the spatial and temporal distributions of the emissions in each region are interpolated from the EDGARv4.2 database (<http://edgar.jrc.ec.europa.eu>); (2) the concentrations on the boundaries of the domain; boundaries are divided into 17 sub-parts: one for the top side, 16 for the lateral sides (8 horizontal parts \times 2 vertical components per part; spatial horizontal distribution in Fig. VI.3; the vertical partition is situated at the half of the domain in pressure coordinates, i.e., at ~ 700 hPa); one coefficient is attributed per sub-part per period of 3 days; for each sub-part, boundary concentrations are supposed constant and uniform; the lateral boundary conditions (LBC) are then fixed by $(8 \times 2 + 1) \times 5$ periods = 85 coefficients; (3) the initial condition 3-dimensional concentration field (IC); the model is first vented with background boundary concentrations (extracted from global simulations with the CTM LMDz; Bousquet et al., 2011) and EDGAR emissions during 10 days before the period of inversion; and (4) an offset, constant and uniform along the whole domain; the prior offset was calculated from the available observations as an estimation of the background concentrations during the window of inversion; the initial and boundary concentrations are expressed as perturbations from this offset.

This simplification implicitly implies the hypothesis of pure correlation of the information within each aggregated region of the state vector (see Sect. 3.5.1). We chose an extended domain compared to the network coverage in order to cope with the spatial and temporal ill-representation of the LBC (Lauvaux et al., 2012).

Atmospheric transport model

We use the Eulerian mesoscale non-hydrostatic chemistry transport model CHIMERE for this study (Vautard et al., 2001). This model was developed in a framework of pollution simulations (Schmidt et al., 2001a; Pison et al., 2007), but is also used for greenhouse gas studies (Broquet et al., 2011). We use here a regular horizontal grid of 50 km-side cells with 25 layers geometrically spaced from the surface to 450 hPa (~ 6000 m). The model time step varies dynamically from 4 to 6 min depending on the maximum wind speed in the domain. The model is an off-line model which needs meteorological fields as forcing. The forcing fields are deduced from interpolated meteorological fields from the European Centre for Medium-range Weather Forecast (ECMWF) with a horizontal resolution of $0.5^\circ \times 0.5^\circ$ every 3 h. The model is operated in a domain of limited area spanning over the whole continental Europe (roughly 24×10^6 km²; see Fig. VI.3).

Observations

The study is based on the assimilation of measurements of the atmospheric composition. Concentrations of CH₄ are measured in-situ in 13 European sites (see Fig. VI.3; details in Table VI.1) at altitudes from sea level up to 3580 m a.s.l. and with different instrumentation and time resolution. Some stations are equipped with CRDS analysers (frequency of up to 1 Hz

Table VI.1: Site characteristics. The altitudes of the sites are given as m above sea level (a.s.l.) and the inlet height is in m above ground level (a.g.l.). The sites are grouped into three categories relatively to the topography and their close environment: rural (R), mountain (M), coastal (C). * These sites are recent and still do not have related publications. ¹ Observatoire Pérenne de l'Environnement.

Station	ID	Location			Inlet height (m)	Site type	Reference
		Lon (° E)	Lat (° N)	Alt (m)			
Bialystok	BIK	23.0	53.2	183	300	R	Popa et al. (2010)
Biscarosse	BIS	-1.2	44.4	120	47	C	Ahmadov et al. (2009)
Carnsore Point	CRP	-6.4	52.2	9	4	C	*
Gif-sur-Yvette	GIF	2.2	48.7	160	7	R	Lopez et al. (2012)
Jungfraujoch	JFJ	8.0	46.6	3580	5	M	Reimann et al. (2008)
Kasprowy	KAS	19.6	49.1	1989	0	M	Necki et al. (2003)
Mace Head	MHD	-9.9	53.3	25	15	C	Ramonet et al. (2010)
Obs. Pér. Environ. ¹	OPE	5.5	48.6	390	10 50 120	R	Ramonet et al. (2011)
Puijo	PUJ	27.0	62.0	232	84	R	Ramonet et al. (2011)
Puy-de-Dôme	PUY	3.0	45.8	1465	10	M	Lopez (2012)
Ridge Hill	RGL	-2.5	52.0	199	45 90	R	*
Trainou	TRN	2.11	48.0	131	5 50 100 180	R	Lopez et al. (2012)
Angus	TTA	-3.0	56.6	254	222	R	Vermeulen et al. (2005)

magnitude), whereas others are Gas Chromatographs. Hourly aggregates were used as input in y^0 for the inversion. The observation sites can be split into three categories: (1) mountain sites which monitor almost all the time the free troposphere; they are scarcely influenced by the local emissions and are representative of the continental and global budget; (2) coastal sites with primary influence from the ocean when the air flows towards lands; as mountain sites, they are representative of global patterns; and (3) rural sites, inland but remote from anthropogenic emissions hot spots. All instruments are calibrated by tanks traceable to the NOAA 2004 CH₄ scale (Dlugokencky et al., 2005) with a calibration precision of ± 2 ppb.

3.2.2 Error configuration: description of the algorithms

In order to apply the Bayesian inversion framework, a perfect knowledge of the background and observation error statistics is needed. The tuple of covariance matrices (\mathbf{R} , \mathbf{B}) must then be established. Tuning and calculating optimal covariance matrices has long been of interest in data assimilation (e.g., Talagrand, 1998; Desroziers and Ivanov, 2001; Chapnik et al., 2004). Statistical studies on large sets of data are required to reach a sufficient threshold of information to get a reliable approximation of \mathbf{R} and \mathbf{B} . In most cases, the sets of data are not available and the covariance matrices are built relying on physical considerations and an expertise on the observation and model behaviours (Bergamaschi et al., 2010). In this subsection, we describe different objective methods to infer the best tuple of \mathbf{R} and \mathbf{B} matrices: first, the Desroziers' scheme, second, the maximisation of the likelihood, third, observation space diagnostics. The Desroziers' scheme and the maximisation of the likelihood are computed on the sub-space of the diagonal matrices for both \mathbf{R} and \mathbf{B} , while the observation space diagnostics allow the recovery of full matrices.

Before further discussion in Sects. 3.3 and 3.4, the relevance of the 3 proposed methods is validated by a χ^2 test (see Sect. 3.2.2), used in many studies (e.g., Lauvaux et al., 2012; Winiarek et al., 2012; Peylin et al., 2002; Rayner et al., 1999).

Validation test: χ^2 distribution

It can be shown, within Gaussian assumptions, that for the state vector $\hat{\mathbf{x}}^a$ minimising the cost function J , $J(\hat{\mathbf{x}}^a) = J^o(\hat{\mathbf{x}}^a) + J^b(\hat{\mathbf{x}}^a)$ has the statistics of a χ^2 distribution with a mean equal to $d/2$, d being the total number of available observations. We then define a χ^2 index $\frac{2J(\hat{\mathbf{x}}^a)}{d}$ that shall be close to 1.

The index can be written:

$$\chi^2(\tilde{\mathbf{R}}, \tilde{\mathbf{B}}) = \frac{1}{d} \left[\begin{array}{l} (\mathbf{y}^o - H(\hat{\mathbf{x}}^a))^T \tilde{\mathbf{R}}^{-1} (\mathbf{y}^o - H(\hat{\mathbf{x}}^a)) \\ + (\hat{\mathbf{x}}^a - \mathbf{x}^b)^T \tilde{\mathbf{B}}^{-1} (\hat{\mathbf{x}}^a - \mathbf{x}^b) \end{array} \right] \quad (\text{VI.4})$$

Though insufficient, this test provides a low-cost insurance that we got a well-defined tuple of covariance matrices. The three methods presented below are tested and validated in regard to this test. A deeper analysis of the algorithms' results is presented in Sect. 3.3.

Desroziers' scheme: subsets application

We describe here a method to roughly infer the shapes of \mathbf{R} and \mathbf{B} covariance matrices with a very low computation cost. It has been shown by Talagrand (1998) and Desroziers and Ivanov (2001) that for a given $(\mathbf{x}^b, \mathbf{y}^o)$, for any subspace j independent of the complementary space in the observation or state space, the optimal tuple of matrices (\mathbf{R} , \mathbf{B}) follows the expression below:

$$\begin{aligned} \mathbb{E}[J_j^o] &= \frac{1}{2} \left[p_j - \text{tr}(\mathbf{P}_j(\mathbf{H}\mathbf{K})\mathbf{P}_j^T) \right] \\ \mathbb{E}[J_j^b] &= \frac{1}{2} \left[n_j - \text{tr}(\mathbf{P}_j(\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{P}_j^T) \right] \end{aligned} \quad (\text{VI.5})$$

where $\text{tr}(\cdot)$ stands for the trace operator, $\mathbb{E}[J_j^o]$ (resp. $\mathbb{E}[J_j^b]$) is the expectation of the contribution at the maximum $\hat{\mathbf{x}}^a$ to the cost function J of the independent subspace j of the

observation (resp. state) space, \mathbf{P}_j is a projector from the whole observation (resp. state) space to the subspace j and p_j (resp. n_j) the dimension of the subspace j . \mathbf{I}_n stands for the identity matrix in the background space.

One cannot ascertain statistical independence between subspaces of the observation or of the state space before running the algorithm. To apply the method, we then make the following assumptions to divide the observation and the state spaces into 41 independent subspaces: (1) each measurement site is independent from all others, (2) at each site, all the observations during the afternoon (12:00–17:00) are gathered in one subspace (named “day” period in the following) and all the remaining observations during the morning and the night in a second subspace (named “night”); the planetary boundary layer (PBL) during the night is ill-represented by models; ensuing erroneous vertical mixing is expected to deteriorate the ability of the CTM to simulate realistic concentrations during the night; the shared cause of the enhanced errors justifies the assumed dependency of night-time observations; for similar reasons, we group the remaining observations during the afternoon with well-mixed PBL. (3) the LBC are independent from other state dimensions, (4) same with the IC, (5) same with the offset, (6) every aggregated region of emissions is independent from the others.

Doing so, we have 26 independent subspaces within the observation space (13 sites \times day/night) and 15 for the state space (12 regions + LBC + IC + offset). Desroziers and Ivanov (2001) proposed an iterative tuning procedure that converges to a tuple that satisfies Eq. (VI.5); we refer to this procedure as Desroziers’ scheme (DS). Let us rewrite the cost function:

$$J_k(\mathbf{x}) = \sum_j \frac{1}{(s_{j,k}^b)^2} J_{j,k}^b(\mathbf{x}) + \frac{1}{(s_{j,k}^o)^2} J_{j,k}^o(\mathbf{x}) \quad (\text{VI.6})$$

where $s_{j,k}^b$ and $s_{j,k}^o$ denote the adapted weights for the subspace j at step k of the iterative procedure to balance the observations and the background in the cost function.

Desroziers’ scheme is described by the following system of equations for every step k :

$$\begin{cases} (s_{j,k+1}^b)^2 &= \frac{2J_{j,k}^b(\hat{\mathbf{x}}^a)}{n_j - \text{tr}(\mathbf{P}_j(\mathbf{I}_n - \mathbf{K}_k \mathbf{H})\mathbf{P}_j^T)} \\ (s_{j,k+1}^o)^2 &= \frac{2J_{j,k}^o(\hat{\mathbf{x}}^a)}{p_j - \text{tr}(\mathbf{P}_j(\mathbf{H}\mathbf{K}_k)\mathbf{P}_j^T)} \\ \mathbf{B}_{j,k+1} &= s_{j,k+1}^b \mathbf{B}_{j,0} \\ \mathbf{R}_{j,k+1} &= s_{j,k+1}^o \mathbf{R}_{j,0} \end{cases} \quad (\text{VI.7})$$

\mathbf{B}_j and \mathbf{R}_j stand for the diagonal sub-matrices associated to the subset j . \mathbf{K}_k is the Kalman gain matrix calculated with $(\mathbf{R}_k, \mathbf{B}_k)$ (see Eq. VI.2). The method implicitly relies on the χ^2 distribution of the cost function. The computed tuples then converges to a tuple filling the χ^2 criterion.

We start this algorithm from a tuple $(\mathbf{R}_0, \mathbf{B}_0)$ with no physical assumption, i.e., $\mathbf{R}_0 = \mathbf{I}_d$ and $\mathbf{B}_0 = \mathbf{I}_n$. We then stop the iterative scheme when every subset contribution to the cost function $E[J_j^o]$ or $E[J_j^b]$ is less than 1% away from its theoretical expected value (right member of Eq. VI.5); this is equivalent to get a χ^2 test greater than 99%. The algorithm converges in no more than 15 steps (i.e., a couple of minutes on a standard office computer).

Maximum of likelihood

Desroziers' scheme (DS) relies on coarse approximations and, for example, cannot extract the variability of the observational errors day-by-day and hour-by-hour. The following method allows the computation of a tuple (\mathbf{R}, \mathbf{B}) which is tuned not by block but by component individually. This improvement implies computational cost drastically higher than for DS method, but still affordable (less than a day) in our case study. In Gaussian assumptions, the likelihood of the observations \mathbf{y}^o for given \mathbf{R} and \mathbf{B} can be written as follows (Michalak et al., 2005):

$$p(\mathbf{y}^o | \mathbf{R}, \mathbf{B}) = \frac{e^{-\frac{1}{2}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)}}{\sqrt{(2\pi)^d |\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T|}}$$

The function diverges to infinity when $\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T \sim (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)^T$. But when supposing that \mathbf{R} and \mathbf{B} are diagonal definite positive, \mathbf{S} cannot be written as a matrix $\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$. Hence, in these assumptions, one can prove that the function is bounded and admits a computable maximum (Burg et al., 1982). A proper (\mathbf{R}, \mathbf{B}) tuple for the inversion system is necessarily a maximum of the function (Dee, 1995).

For memory limitation reasons, we do not maximise the function itself, but equivalently its logarithm:

$$\ln p(\mathbf{y}^o | \mathbf{R}, \mathbf{B}) = -\frac{1}{2} \text{tr}(\mathbf{S}_{\mathbf{R}, \mathbf{B}}^{-1} \mathbf{S}) - \frac{1}{2} \ln |\mathbf{S}_{\mathbf{R}, \mathbf{B}}| + C$$

with $\mathbf{S}_{\mathbf{R}, \mathbf{B}} = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$, $\mathbf{S} = (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)^T$ and C a constant not relevant for computing the maximum of the function. $|\cdot|$ stands for the determinant operator.

The function maximum cannot be easily computed analytically in general. Hence, we use an ascending pseudo-Newtonian method based on the calculation of the gradient of the log-likelihood. The algorithm converges to a local maximum (Chapnik et al., 2004), but we have no insurance of converging to the global maximum. The result of the algorithm can be very dependent from the (\mathbf{R}, \mathbf{B}) tuple chosen as a starting point. To ensure the robustness of the result, we test this method with 2 different starting tuples: (1) one constructed relying on expert knowledge (e.g., diagonal background matrix with variances consistent with inventories specification), (2) the other with uniform errors of 50 ppb for observations, the LBC and the offset and 10% for the emissions. The results are very similar: the difference between the two errors related to an observation j does not exceed 5% and is less than 1% in average.

Additionally, to accelerate the convergence of the algorithm, one can notice (details in Burg et al., 1982) that the log-likelihood is maximum only if (\mathbf{R}, \mathbf{B}) satisfies:

$$\alpha = \frac{\text{tr}(\mathbf{S}_{\mathbf{R}, \mathbf{B}}^{-1} \mathbf{S})}{d} = 1$$

Consequently, we force α to stay close to 1 at each step of the algorithm by rescaling (\mathbf{R}, \mathbf{B}) by α . This normalisation is equivalent to the χ^2 test; hence the constrained maximum of likelihood algorithm necessarily fulfills the χ^2 test.

This method makes the observation and background error variances statistically consistent with the prior difference between \mathbf{y}^o and $\mathbf{H}\mathbf{x}^b$. When assuming that \mathbf{R} and \mathbf{B} are diagonal, a unique set of variances matches this criterion on $\mathbf{y}^o - \mathbf{H}\mathbf{x}^b$. Hence, a maximum total amount of

d pieces of information is used in the algorithm, while more are available in the background and the observation operator for the subsequent computation of $\hat{\mathbf{x}}^a$. A more precise and complete quantification of the balance between the information used for the optimisation of the matrices and for the inversion itself is difficult in a real case study. A dedicated OSSE (Observing System Simulation Experiment) could improve our knowledge in this direction.

Observation space diagnostics

With the two previous methods, \mathbf{R} and \mathbf{B} are confined to the sub-space of diagonal matrices. But the errors on the observations are known to be correlated through the \mathbf{H} operator errors amongst others. Errors on the background are also correlated, mainly because of shared errors in the inventory methods and in flux process modelling. We carry out in this subsection an algorithm to produce a setup of non-diagonal matrices \mathbf{R} and \mathbf{B} .

Inquiring into error correlations requires a huge amount of information. The available information is not sufficient to characterise deterministically the full non-diagonal matrices \mathbf{R} and \mathbf{B} . This subsection must then be seen as a way to infer guidelines for covariance building.

Design on \mathbf{R} Desroziers et al. (2005) showed that the innovation vectors $\hat{\mathbf{d}}_a^o = \mathbf{y}^o - \mathbf{H}\hat{\mathbf{x}}^a$ and $\hat{\mathbf{d}}_b^o = \mathbf{y}^o - \mathbf{H}\hat{\mathbf{x}}^b$ should fulfill the following equation:

$$\mathbb{E}[\hat{\mathbf{d}}_a^o(\hat{\mathbf{d}}_b^o)^T] = \mathbf{R} \quad (\text{VI.8})$$

Equation (VI.8) is valid if and only if (1) the observation operator is linear, (2) the errors are Gaussian and unbiased, and (3) the matrix $\mathbf{H}\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$ is consistent with the statistics of the background and observation errors.

In order to build a matrix \mathbf{R} consistent with Eq. (VI.8), we test an iterative algorithm similar to Desroziers' scheme using the following instruction for every k :

$$\begin{cases} \mathbf{R}'_{k+1} &= \mathbb{E}[\hat{\mathbf{d}}_a^o(\hat{\mathbf{d}}_b^o)^T] \\ \mathbf{R}_k + 1 &= \chi^2(\mathbf{R}'_{k+1}, \mathbf{B}_k) \cdot \mathbf{R}'_{k+1} \\ \mathbf{B}_k + 1 &= \chi^2(\mathbf{R}'_{k+1}, \mathbf{B}_k) \cdot \mathbf{B}_k \end{cases} \quad (\text{VI.9})$$

The tuple is normalised by the associated χ^2 test in order to constrain better the algorithm.

The expectation could be calculated explicitly as a combination of \mathbf{R}_k , \mathbf{B}_k and \mathbf{H} . However, as suggested by Desroziers et al. (2005), the exact iterative scheme is not expected to converge. In our case, we indeed found no convergence. Therefore, we use a Monte Carlo evaluation of the expectation with 50 000 perturbations of \mathbf{y}^o and \mathbf{x}^b with Gaussian distribution of covariances \mathbf{R}_k and \mathbf{B}_k . The tuples of covariance matrices generated with the Monte-Carlo algorithm have likelihood values bigger (hence closer to the maximum of likelihood) than the one inferred from the explicit algorithm. Their likelihood (-480) is also significantly higher than with the diagonal tuple calculated in Sect. 3.2.2 ($\sim -15\,000$). The non-diagonal tuple is then more 'likely' in the sense of Sect. 3.2.2 compared with the diagonal one.

Any inconsistency in the requirements of Eq. (VI.8) will make the expectation non-symmetric positive definite. In particular, biases in the observations carry information that is considered by the algorithm as random errors. The built matrices are then partly influenced by the intrinsic biases in the system. However, a simple diagnostic on the spectrum shows that after 50 000 perturbations, only less than 0.1 % of the Eigen values are negative. The biases that create asymmetry in the matrix then do not seem to significantly impact our case. This could not always apply to other frameworks. The limitation of the algorithm is to build up statistics from a single occurrence of $\hat{\mathbf{y}}^o$ and $\hat{\mathbf{x}}^b$ and therefore is not fully generic.

But since asymmetry does not seem to be critical in our case, we rebuild a symmetric semi-definite positive matrix by correcting \mathbf{R}'_{k+1} spectrum. We start the algorithm from two different tuples of matrices: the one which maximises the log-likelihood in the diagonal assumption, and another with the same \mathbf{B} but with $\mathbf{R} = \mathbf{I}_d$. The convergence is slower with the second starting tuple but the two optimum tuples are similar.

Calculating a matrix expectation of dimension 5000×5000 with only 50 000 Monte-Carlo perturbations is very unstable. Additionally, the perturbations on the observations are generated from \mathbf{y}^o , whereas it should be computed from the unknown perfect unperturbed observation vector $\tilde{\mathbf{y}}$. Hence, the closer \mathbf{R}_k gets to the optimal value, the weaker is the approximation based on \mathbf{y}^o . For this reason, after a few steps of improvement of the log-likelihood value, the algorithm quickly diverges. We then keep the last step before the beginning of the divergence as a guess for the true covariance matrices tuple.

Design on \mathbf{B} Desroziers et al. (2005) also reported an equation constraining \mathbf{B} in the observation space:

$$\mathbb{E}[\hat{\mathbf{d}}_b^a(\hat{\mathbf{d}}_b^o)^T] = \mathbf{H}\mathbf{B}\mathbf{H}^T \quad (\text{VI.10})$$

with the innovation vector $\hat{\mathbf{d}}_b^a = \mathbf{H}(\hat{\mathbf{x}}^a - \hat{\mathbf{x}}^b)$. The expectation is again based on a Monte-Carlo estimation. We compute an iterative scheme similar to that developed in Eq. (VI.9) with also a normalisation by the χ^2 test. We use the tuple calculated above as a starting point.

Equation (VI.10) relies on the same assumptions as Eq. (VI.8). In particular, bias and other mismatches between $\mathbf{H}\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$ and the correct statistics can induce inconsistent asymmetries that must be corrected. Another critical point in building \mathbf{B} from constraints in the observation space is that the computed expectation could be outside the ensemble $\{\mathbf{H}\mathbf{B}\mathbf{H}^T / \mathbf{B} \text{ symmetric definite positive}\}$. We then project the expectation onto this structure to recover a compatible \mathbf{B} matrix.

Because of the Monte Carlo and the indirect estimation of \mathbf{B} , the instability is even sharper than for the computation of \mathbf{R} . We then do a unique iteration of the algorithm to get an evaluation of the potential optimal correlation coefficients in \mathbf{B} .

3.3 Results

We run the three algorithms described in Sect. 3.2.2 to infer a best guess for the optimum tuple (\mathbf{R}, \mathbf{B}) of the error covariance matrices. In the following subsection, we describe the shape of the calculated matrices in regard to known physical patterns of errors. We will refer to the diagonal tuple computed from the Desroziers' Scheme (resp. the Maximum of log-

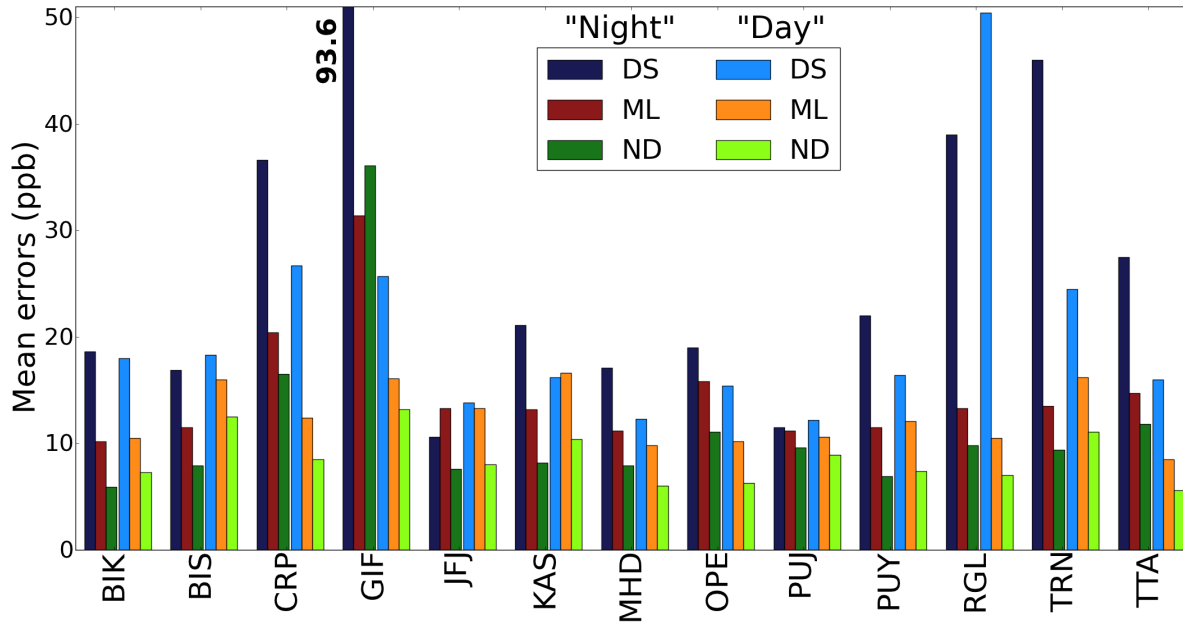


Figure VI.4: Comparison of the observation errors calculated by the 3 methods presented in Sect. 3.2.2 for the inversion window after the leak start (25 March 2012). The errors are averaged on two periods each day: “day” (12:00 to 17:00) and “night” (the remaining hours of the day) for comparison to DS method. Values out of the graph window are printed on the matching bar.

likelihood) as $(\mathbf{R}_{\text{DS}}, \mathbf{B}_{\text{DS}})$ (resp. $(\mathbf{R}_{\text{ML}}, \mathbf{B}_{\text{ML}})$); the non-Diagonal tuple from the observation space algorithm will be referred to as $(\mathbf{R}_{\text{ND}}, \mathbf{B}_{\text{ND}})$.

3.3.1 Patterns in the error variances for the 3 methods

In Fig. VI.4, the variances of the observation errors in the 3 methods are compared for the period after the leak start. For the ML and ND algorithms, the variances of the observation errors are averaged along the same subspaces as in DS method (see Sect. 3.2.2) to be comparable. Most of the observation errors remain within the same interval (5–20 ppb) with the 3 methods. Their magnitudes are comparable with other studies which build the errors from physical considerations (e.g., Bergamaschi et al., 2010). At most sites (GIF, JFJ and KAS excepted), the 3 methods attribute averaged observation errors that follow the same order: $\text{DS} > \text{ML} > \text{ND}$. The errors from \mathbf{R}_{ML} are in average 34% smaller than the errors in \mathbf{R}_{DS} . The error variances in \mathbf{R}_{ND} are calculated to be even smaller (54% less than in \mathbf{R}_{DS} in average). In the Bayesian unbiased framework, the inversion with the tuple from the maximum of likelihood is then expected to be more constrained by the observations than the one from Desroziers’ scheme. The non-diagonal tuple seems more constrained by the observations than the other two, but the covariances make it difficult to precisely compare only the variances. The 3 methods share the same day/night patterns at all sites apart from the 3 mountain sites (JFJ, KAS and PUY) and the two sites BIS and BIK: compared to the errors during the “night” (17:00–00:00 plus 00:00–12:00), the errors during the “day” (12:00–17:00) are 25% (resp. 23% and 31%) smaller for the DS (resp. ML and ND) method. The errors are consistently smaller when the PBL is well developed, i.e., when the local emissions are quickly mixed in

the atmosphere, and hence when the CTM more realistically simulates the atmospheric concentrations. At the mountain sites (JFJ, KAS and PUY; see Table VI.1), the rough DS method does not calculate the same patterns as the other two; this primarily suggests that, for the sites mostly located in the free troposphere in spring (characterised by synoptic variability), the averaging on “day” and “night” intervals is less relevant than for the sites influenced by the PBL. Additionally, in mountain sites, the low-precision DS method disagrees with the other two because it cannot compute the errors that occur when the PBL height is close to the site altitude and when polluted air masses can be locally uplifted to the site. This phenomenon occurs at time scales that are smaller than and not synchronized with the partitioning made in Desroziers’ scheme, that makes the DS method unable to accurately detect this source of error. On the opposite side, the two other methods which are handled individually each observation can detect this phenomenon and take it into account in the error modelling, giving averaged errors that are consistent with each other.

Regarding the background, the attributed errors partly balance the confidence attributed to the observations. For the two comparable diagonal algorithms (ML and DS), the ML method better optimises the use of the observations than DS does on large subspaces. To avoid corrections dominated by the observations, background errors 7 times smaller than for the DS algorithm are then computed by the ML algorithm. On the other hand, no clear general pattern appears with the tuple $(\mathbf{R}_{\text{ND}}, \mathbf{B}_{\text{ND}})$, compared with the other two. While the errors on the observations are on average smaller than the ones calculated with the ML method, the background errors with the ND method are of the same magnitude as the DS ones, i.e., higher than the ML ones. With non-diagonal covariance matrices, the variances by themselves are insufficient to identify all the properties of the errors. The non-diagonal elements in \mathbf{R}_{ND} and \mathbf{B}_{ND} , characterising correlations of errors, must be taken into account to understand the error patterns. Strong negative (resp. positive) correlations are expected to increase (resp. decrease) the average confidence in the background. We analyse the effects of these correlations more precisely in Sect. 3.3.3.

3.3.2 Temporal variability and diurnal cycle of the observation errors

We focus here on the hour-by-hour variability of the observation errors; as a consequence, the DS method is not commented in this subsection. The variability in the variances is comparable in \mathbf{R}_{ML} and \mathbf{R}_{ND} ($r = 0.89$; linear regression coefficient: $\lambda = 0.97$). While no physical assumptions have been added to the algorithms, we notice that the observation errors follow known physical patterns for both algorithms. One of these is the source of errors related to the PBL height mis-estimation and the vertical mixing parameterization in CTMs. In Table VI.2, we show for each station the linear correlations between the modelled PBL heights and the calculated error variances for ML method. We also compute a logarithm transformation that shows that some PBL errors follow an exponential decay: $\exp(-h_{\text{PBL}}/h_0)$, with h_{PBL} the modelled PBL height and h_0 a reference height. The magnitude of these calculated reference heights is 1000 m for the sites with significant correlations; above this reference threshold, the PBL can be approximately considered as well mixed. Logarithmic correlations are of the same magnitude in ND method at the sites with small p values ($p < 10^{-2}$; e.g., logarithm correlation: at CRP $r = -0.64$, at GIF $r = -0.57$, at MHD $r = -0.37$). The logarithm correlations are stronger than the linear ones in the sites where the p value is small ($< 10^{-2}$), except at RGL where the linear fit is better. For these sites (including RGL), we approximate

Table VI.2: Linear and logarithmic correlations (r) between calculated error standard deviation and the simulated PBL height (given by ECMWF). At each site, the calculation of r and the p value are carried out with \mathbf{R} from ML method. Some p values are very small because calculated on large samples of observations. Error contribution (μ) for log-regression is the mean ratio between the fit and the computed errors, when the correlation is significant.

Stat	Linear regression		Log-regression		
	r	p	r	p	μ
BIK	0.10	0.06	0.8	0.16	–
BIS	0.13	0.01	0.11	0.04	–
CRP	–0.33	2.6×10^{-10}	–0.60	1.7×10^{-35}	60 %
GIF	–0.34	1.3×10^{-8}	–0.50	4.3×10^{-18}	50 %
JFJ	–0.04	0.5	–0.05	0.36	–
KAS	–0.01	0.8	–0.10	0.08	–
MHD	–0.41	2.9×10^{-15}	–0.47	2.1×10^{-19}	52 %
OPE	–0.31	5.6×10^{-22}	–0.36	1.5×10^{-28}	54 %
PUJ	–0.04	0.5	0.05	0.33	–
PUY	0.06	0.5	0.01	0.84	–
RGL	–0.17	9.1×10^{-6}	–0.12	0.002	58 %
TRN	0.21	7.9×10^{-11}	0.02	0.59	–
TTA	–0.64	2.1×10^{-22}	–0.72	2.3×10^{-29}	41 %

the errors by the exponential fit and estimate the relative contribution of the PBL error as the ratio between the fit error and the total error. The PBL error then seems to account for about half of the total error.

The strong correlations between the PBL height and the errors on the observations can also be partly related to systematic errors (especially during the night) in the CTM, which are difficult to distinguish from random ones based on the available data. The systematic nature of some errors can conflict with the hypothesis of random unbiased errors in the inversion framework: $\epsilon = \mathbf{y}^o - \mathbf{H}\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ in the unbiased framework would become $\epsilon \sim \mathcal{N}(\eta, \mathbf{R}')$ with a systematic bias η . Our methods will then try to find \mathbf{R} such as $\mathbf{R} \sim \mathbf{R}' + \eta\eta^T$. Though statistically consistent with all the available pieces of information, such a biased formulation of the problems will have critical effect on the inversion results (e.g., Dee, 2005). Therefore, the biases have to be taken into account and our method seems to provide an efficient tool to detect them (e.g., high redundant errors associated to nighttime vertical mixing) in order to potentially fix them before the inversion.

In Fig. VI.5, the statistics (median and interquartile range) of the errors are displayed by layers of 250 m for h_{PBL} for the sites with significant correlations as calculated in Table VI.2 on the one side and for the mountain sites on the other side. The anti-correlation between h_{PBL} and the errors is confirmed and a more complex behaviour appears for mountain sites. For example, the relation to the PBL is inverted for the Kasprowy mountain site (KAS; see Fig. VI.5): the site is set on the summit of a mountain ridge in vicinity of a region with high CH_4 emissions due to coal mining. The resolution of the model makes the simulation of the upward winds from the valley impossible. Then, an observation can be noticeably influenced by polluted air masses from the PBL while the model considers the observation to be in the

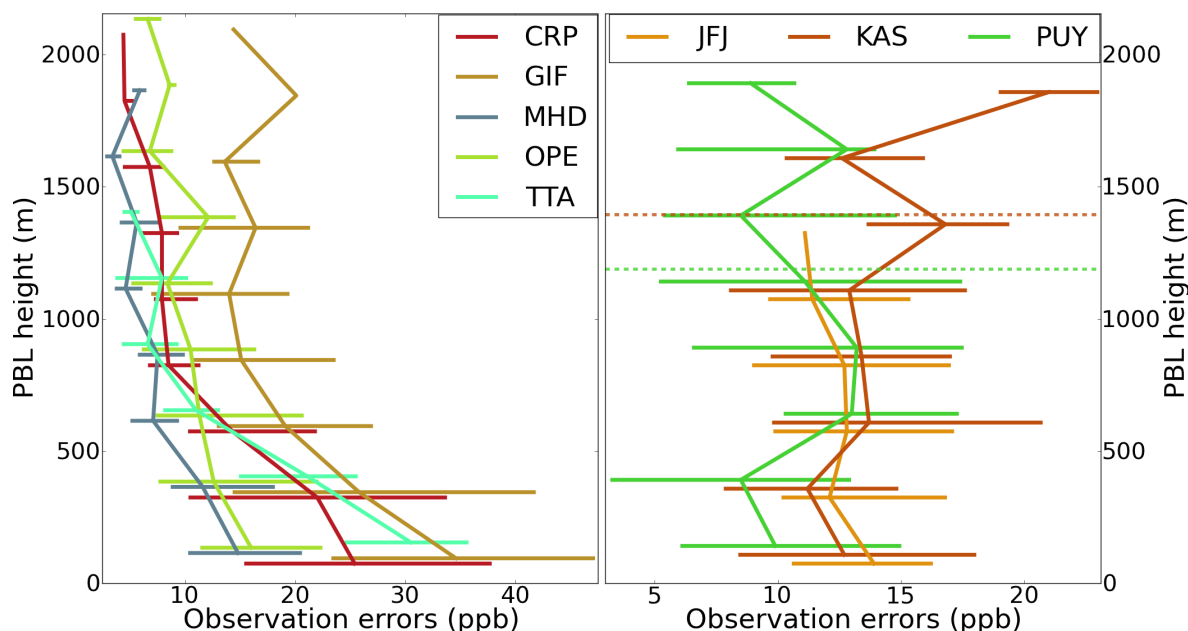


Figure VI.5: Statistics of the errors projected along ECMWF-simulated PBL height for ML method: median and inter-quartile gap per 250 m-high layer. (Left) sites with strong correlations as calculated in Table VI.2; not displayed site RGL exhibits the same patterns but with higher errors. (Right) Mountain sites with influence from the PBL less prevailing. Dashed lines refer to the station altitude in the model; JFJ is above the maximum simulated PBL height.

free troposphere, and inversely; that explains the peak in errors when the modelled h_{PBL} is close to the altitude of the site in the CHIMERE model: ~ 1400 m above the local pixel orography level. The mountain site at Puy-de-Dôme (PUY) appears to pose the same problem in other studies (Broquet et al., 2011); but in the particular meteorological situation of our window of inversion (2 weeks in spring), the issue does not clearly appear.

Another identified source of errors (not shown in the figures) is the temporal and spatial mismatches which can occur in situations of steep gradients of concentrations when the air masses are changing, like in frontal systems. Air mass changes occur 0 to 3 times during each inversion window of two weeks at the sites. The quantification of these errors is uncertain during night-time since it is difficult to separate the errors related to the PBL and the errors due to the gradients (e.g., mismatches or numerical diffusion). So, we focus on air mass changes occurring during daytime. In these cases, the computed errors exceed 4 times the mean error during daytime; the increased errors are then directly attributable to the inability of the model to simulate high-resolution phenomena. The relationship between the observation errors and the temporal evolution of the 3-D meteorological fields can be complex; our method allows a relevant estimation of the errors specific to the meteorological conditions at each site.

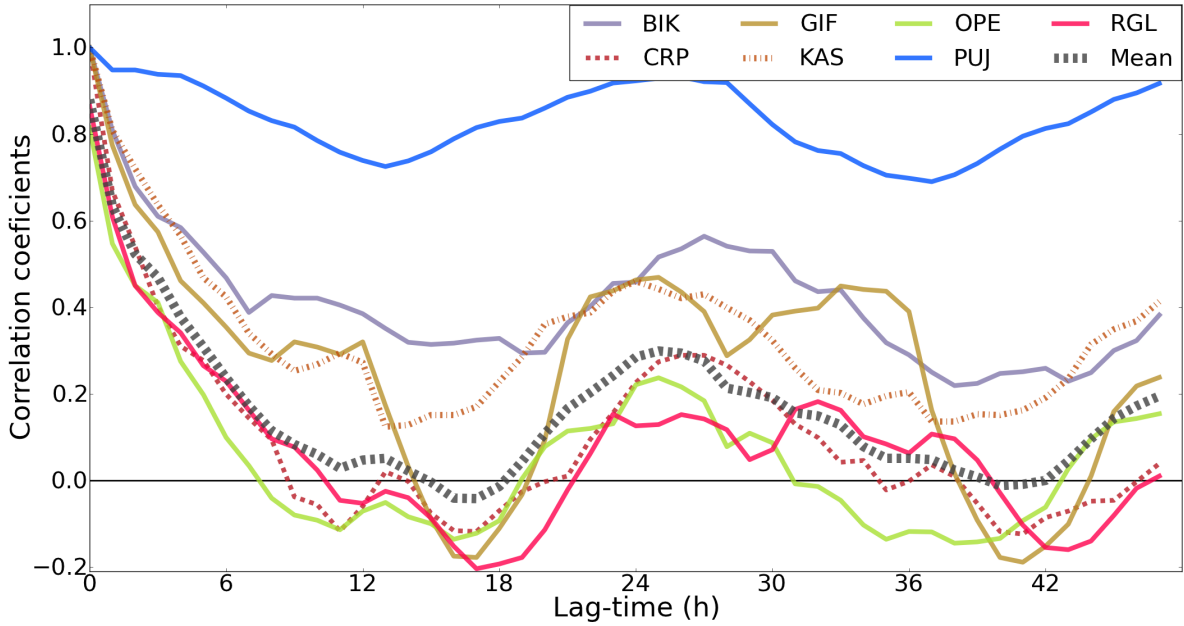


Figure VI.6: Mean temporal auto-correlation. For each site, are figured the average correlations of all the observations from 12 to 7 p.m. with the following 48 h. Only the sites with a local maximum at 24 h are figured: the 6 sites with strong correlation with the PBL (in Table VI.2) and KAS mountain site. Solid lines for rural sites; dashed (resp. dotted) lines for coastal (resp. mountain) sites.

3.3.3 Correlations in the observation errors

If we focus on the non-diagonal terms of \mathbf{R}_{ND} , a large part (64%) of the correlation coefficients are very large ($|r| > 0.9$ with even positive and negative repartition). Strong positive and negative correlations mark a redundancy of the information provided by the observations. We then group the available observations into classes of redundant information. Amongst each class, we chose $|r_{i,j}| > 0.9$ between every tuple (i, j) of observations. Following this criterion, we divide the set of observations into 625 balanced classes, figuring 625 independent pieces of information given by the observations. This figure can be compared to other studies which filter out part of the data before the assimilation. For example, Bergamaschi et al. (2010) proposed to keep one observation per day and per site only, to avoid an over-constraining due to spatial and temporal correlations; in our case, it would have meant keeping 195 observations (i.e., 3 times less than the result of ND algorithm). Our method suggests keeping more independent pieces of information but the amount remains low compared to the total set of observations ($\sim 15\%$ of the ~ 5000 observations). Then, in our system with a drastically reduced state space dimension, the network over-constrains the fluxes; but this result is very dependent of the analytical framework with aggregated regions and is unlikely to apply to full-resolution configurations. In systems with state vectors larger than the observation vector, one would expect that the operator H will not reduce the amount of independent pieces of information that can be assimilated so much.

The temporal structure of the error correlations is shown in Fig. VI.6. At each site, the mean time auto-correlations of the errors at lags 0 to 48 h are calculated. The average auto-

correlations are computed with starting hours from 12:00 to 19:00 only; the patterns for the other hours of the day are the same, but with mean correlations that are closer to 0. It appears that the absolute correlations quickly decrease below $|r| < 0.25$ in about 5 h at each site. In Fig. VI.6, we display the sites with a day-to-day significant correlation. For these sites, we observe a maximum of correlation at 24 h, related to processes with a diurnal cycle. But this is not necessarily the PBL diurnal cycle since every site with strong correlations between the PBL height and the errors does not exhibit the 24 h peak. The 24 h periodic correlation could be related to the surface temperature diurnal cycle for instance or any other diurnal cycle in the atmospheric state.

A spatial structure of the observation covariances could also have been expected (Lauvaux et al., 2009) at sites close to each other such as GIF and TRN, which are about 100 km distant. But the calculated \mathbf{R}_{ND} covariance matrix does not exhibit any global spatial patterns. Distant observations can be strongly correlated, but the dominant underlying process is likely related to the PBL height.

3.3.4 Correlations in the background errors

The lateral boundary conditions (LBC)

In \mathbf{B}_{ND} , the components related to the LBC are found to be independent ($|r| < 0.1$) from the ones of the aggregated fluxes. Amongst the LBC, two independent groups of regions appear. Within each group, the components are very strongly correlated or anti-correlated ($|r| > 0.9$) with the other components of the group and are independent from the ones not in the group. About 10% of the LBC cannot be attributed to any of the two groups. In Fig. VI.7, the boundary regions have been sorted accordingly to these groups. The colours in the figure denotes this classification: blue and red for the two groups and green for the few remaining boundaries with mild correlations with other regions. The boundaries have also been sorted according to the influence of the assimilation of the observations on them. We quantified this influence by using the diagonal elements of the matrix \mathbf{KH} (see Eqs. VI.2 and VI.3 in Sect. 3.2.1) which are necessarily in the interval $[0, 1]$. We highlight the regions strongly influenced ($\mathbf{KH}_{i,i} > 0.9$; “+” sign) against the others (“o” sign). Then it comes that the regions unseen (resp. constrained) by the inversion are strongly correlated with each other and not correlated with the constrained (resp. unseen) regions.

Drawing conclusions about the signs of the correlations between the boundary components is more difficult. Within the “unseen” group, none of the three algorithms can retrieve information about these components of the state vector since the corresponding elements in \mathbf{H} are negligible compared to the others. Thus, the positive and negative correlations are likely to be numerical artifacts in the algorithms. For the constrained group, a large negative correlation between two regions means that the prior mean contribution of the two regions is well constrained. Equivalently, the two LBC components are well known on average but the two individual contributions are not separated from each other; this may indicate that the contributions are similar in magnitude with simultaneous transitions observed at the sites, hence not discernible. On the opposite, a positive correlation reveals a constraint on the difference of the contributions, hence on the spatial and/or temporal gradients in the LBC. For example, for two LBC regions positively correlated and upwind the observations, if a clear transition between the contribution of one region and the other is observable from the

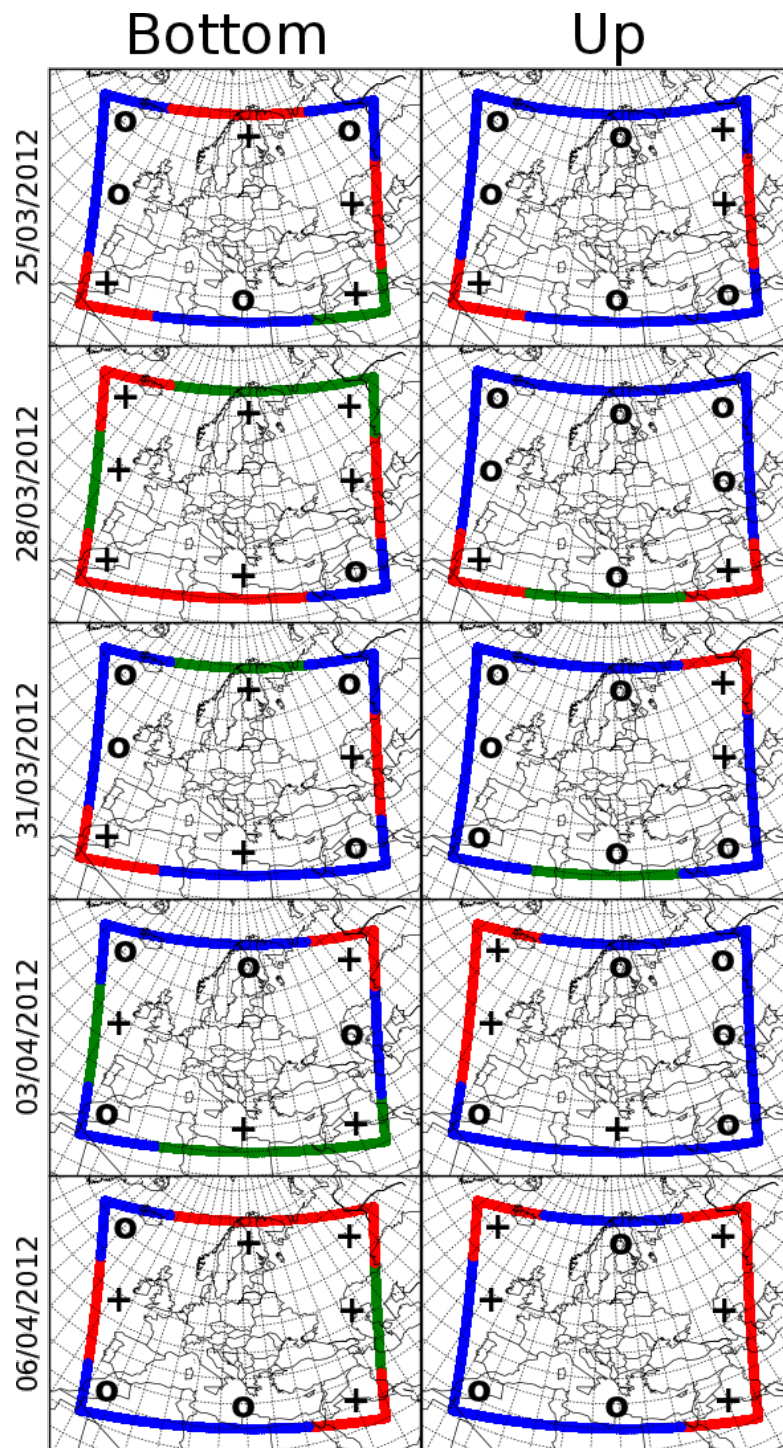


Figure VI.7: Correlations between the errors on the LBC elements of the background vector for the two weeks after the leak start, calculated from \mathbf{B}_{ND} . Dates are the starting dates of every 3 day window of uniform constant boundary concentrations. The bottom boundaries span from the surface to ~ 700 hPa; up is the remaining part of the vertical direction. Red and blue highlight 2 classes of boundaries very strongly correlated ($|r| > 0.9$). Green boundaries are not significantly correlated to any other regions. “+” (resp. “o”) signs corresponds to boundaries strongly (resp. lightly) influenced by the inversion (criterion of selection described in Sect. 3.3.4).

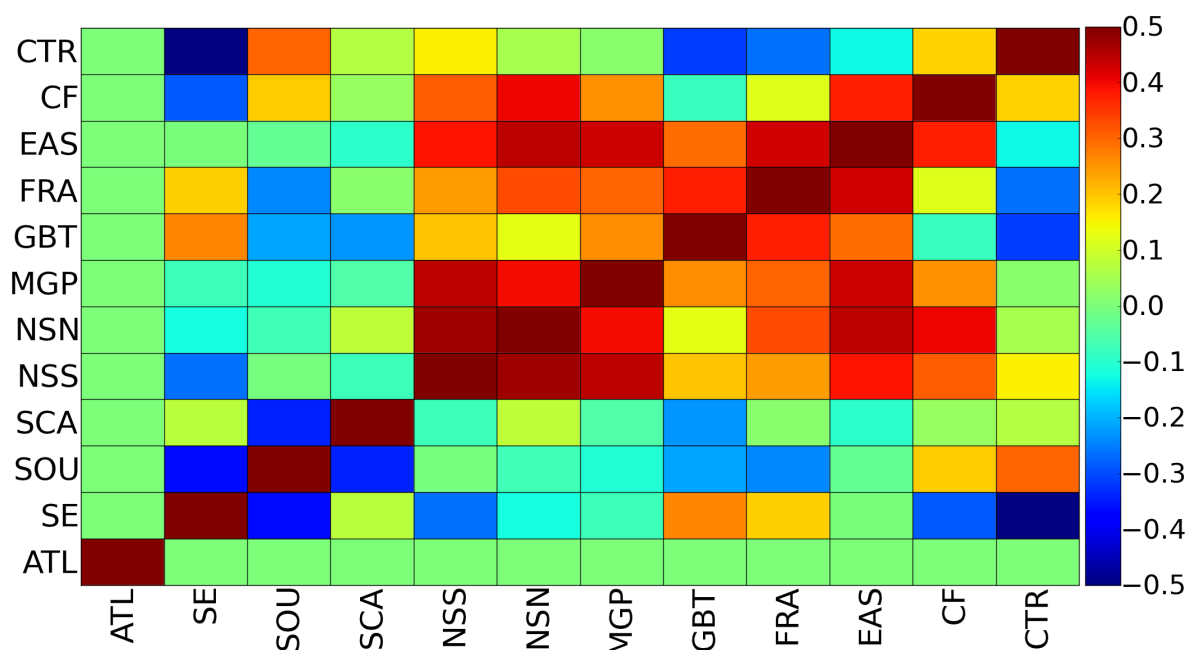


Figure VI.8: Correlations between the errors of the background aggregated regions for the non-diagonal matrix B_{ND} for the two weeks after the leak start. Refer to Fig. VI.3 for the region names, location and extension.

network, the gradient between the two will be well fixed, whereas the total balance will be unclear and biased by all the other components of the state vector. The sign of the correlation then mainly depends on the meteorological conditions at the site and of the air mass history when it is observed. It can be very variable and virtually unpredictable with a general formula. Our method gives an objective estimation of the issue.

The aggregated emissions

The error correlations for the aggregated regions of emissions are displayed in Fig. VI.8. Negative correlations occur only between regions that are not upwind any sites (i.e., in Fig. VI.8, for the period after the leak start: “ATL”, “SCA”, “SE”, and “SOU”; see Fig. VI.3; e.g., $r = -0.36$ between “SOU” and “SE”). As for LBC, interpreting the correlations between unconstrained regions is hazardous. On the opposite, the positive error correlations between the regions close and upwind the sites (for example, $r = 0.45$ between the neighbouring regions NSS and MGP) denote the confidence in the background flux gradients amongst these regions but not necessarily in the total flux balance. The confidence in the prescribed gradients in emissions is consistent with the methodology used to build the inventory maps. Activity maps by sectors are convolved with emission factors. In Europe, the declared activity is considered reliable. Then, the emission factors, hence the overall magnitude of the emission, is a more critical source of uncertainties than the spatial distribution.

3.4 Flux inversion: the Elgin leak case

From here, we use the optimised tuples of matrices to actually compute the inversion with Eqs. (VI.2) and (VI.3).

3.4.1 Physical relevance of the inverted fluxes

In this study we have selected a domain for the simulations which spans over a region much bigger than the network coverage area. This choice is expected to decrease the errors due to the coarse approximations made on the LBC. But, in return, we do not expect the results to be accurate on the outer regions because of the inability of the inversion to either constrain unobserved regions, or distinguish the contributions from these regions and from the LBC when they are all upwind the observations.

The three methods rely on the assumption of the Gaussianity of all the errors. In particular, this assumption theoretically allows surface fluxes to be negative. With the optimised state vector \mathbf{x}^a calculated with $(\mathbf{R}_{DS}, \mathbf{B}_{DS})$, increments (defined as $\frac{(\hat{\mathbf{x}}^a)_j - (\mathbf{x}^b)_j}{(\mathbf{x}^b)_j}$) of -150% and below appear on the emissions in regions close to the sides of the domain. Hence the inversion generates strongly negative surface fluxes, due to inaccurate separations between emissions and boundary condition signals. Net surface uptake is physically not acceptable for CH_4 in Europe where anthropogenic emissions are largely prevailing (Bergamaschi et al., 2010) compared to soil uptake. Then, despite its very low-computation cost, a simple DS can not be applied unless one uses it on a larger number of subsets and with additional physical constraints.

In regard of the LBC issue, the other two algorithms (ML and ND) seem reliable. The ML algorithm does not compute absolute increments over 35% for these regions and keeps posterior uncertainties compatible with a 0% increment (i.e., no change from the prior flux). The non-diagonal tuple $(\mathbf{R}_{ND}, \mathbf{B}_{ND})$ leads to increments incompatible with the positivity of the fluxes in some of the outer regions but in an acceptable range ($> -10\%$) considering the posterior errors in the matrix \mathbf{P}^a for these regions.

Moreover, the diagonal algorithm DS and ML do not compute any significant posterior error correlation ($r^2 < 0.1$) between outer regions and the LBC; inversely, the ND tuple explicitly estimates strong posterior error correlations for these components (numerous correlation coefficients $r^2 > 0.5$). Hence, ND method can account for the erroneous separation from the outer emissions and the LBC.

3.4.2 Using the optimised tuples $(\mathbf{R}_{ND}, \mathbf{B}_{ND})$ for flux inversion

The non-diagonal tuple reliably takes the LBC ill-separation into account. We then use the inversion carried out with this tuple to analyse the posterior fluxes on the regions close to the network that are better constrained than the outer ones. The increments for the two periods (15-day long each) before and after the leak start are shown in Fig. VI.9. The posterior errors are not displayed because they do not exceed 1% for most regions (maximum of 1.1% for “NSN” region before the leak start).

These surprisingly small posterior uncertainties can seem physically inconsistent. But the figures apply to very large regions. Hence, uncertainties on pixels are expected to be

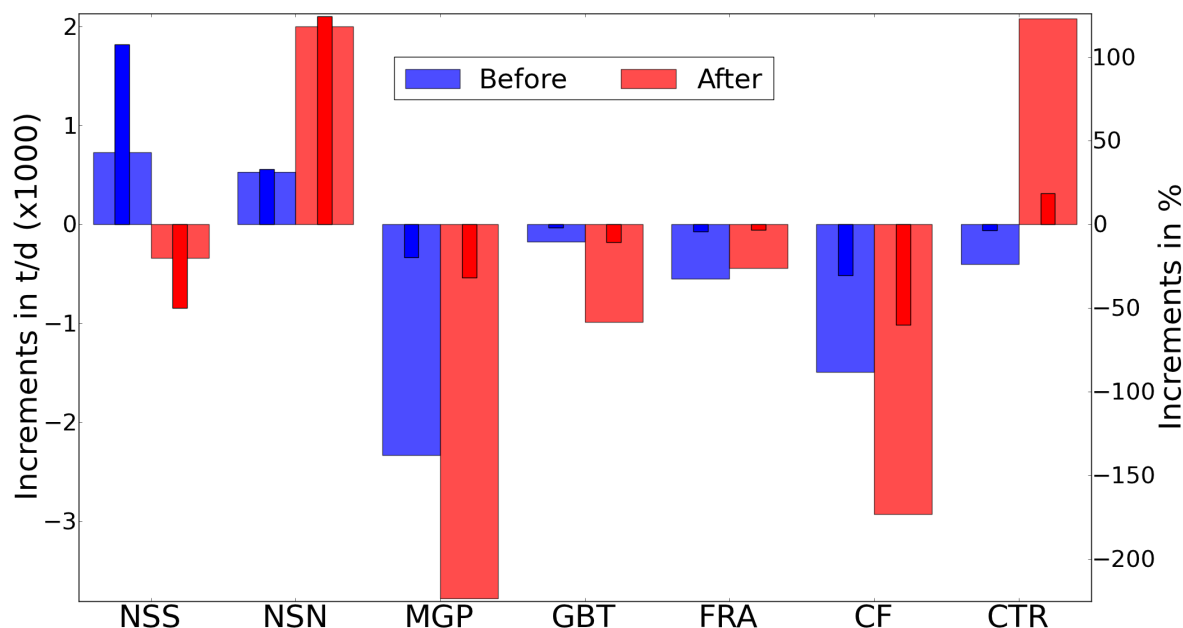


Figure VI.9: Increments on the emissions for the regions not contiguous to the boundaries in t d^{-1} (large bars) and in % (thin bars) of the initial total with the non-diagonal tuple $(\mathbf{R}_{\text{ND}}, \mathbf{B}_{\text{ND}})$. The two temporal windows of inversion are the two weeks before (blue) the leak start and the two weeks after (red). Acronyms in Fig. VI.3.

smoothed by the aggregation. Moreover, as CH_4 emissions in spring in Europe are mainly anthropogenic and well documented, even the pixel errors could be relatively low.

Before the leak start, the total emissions from these inner regions (“NSS”, “NSN”, “MGP”, “FRA” and “CTR”) are corrected from $52\,455 \pm 16\,838 \text{ t d}^{-1}$ to $51\,262 \pm 1999 \text{ t d}^{-1}$, that is to say an increment of $-2.3 \pm 3.9\%$. After the leak start, the correction to the inner budget is from $52\,455 \pm 12\,431 \text{ t d}^{-1}$ to $48\,062 \pm 48 \text{ t d}^{-1}$, i.e. an increment of $-8.4 \pm 0.1\%$. The inversion of the CH_4 emissions over the inner regions suggests an over-estimation in the inventories which actually were not designed for year 2012 but for 2008. The decrease in the emissions by $3200 \pm 2000 \text{ t d}^{-1}$ (i.e., $-6.1 \pm 3.8\%$ reported to the prior total balance) between the two inversion windows remains consistent with the uncertainties of the inventories and can be explained by the typical variability of the emissions. But one should recall that the results are averaged on aggregated regions, whereas the areas of influence of the sites do not necessary overlay the whole region (see Sect. 3.5.1). The inversion corrects simulated concentrations, not considering the implications in regards to absolute emissions. The increments can be amplified and suffer from aggregation errors and sampling heterogeneity (Kaminski et al., 2001). More critically, the aggregation errors on the areas that are not within the footprint of the network cannot be recovered from our methods since the operator H and the covariance matrix \mathbf{B} are aggregated before the algorithms are run.

Focusing on the estimation of the CH_4 release from the Elgin platform, the inversion suggests an increase in the regional emission where the leak occurred (“NSS” region) of $+1472 \pm 30 \text{ t d}^{-1}$ (the errors on the difference is calculated assuming that the two inversion windows are independent). But the two parts of the North Sea are expected to be

ill-distinguished by the system, as confirmed by a posterior correlation coefficient r of -0.78 between the errors in the two regions after the leak start. Then the flux that can be attributed to the leak is defined as the difference of the emissions after and before the leak start over the whole North Sea area. Our inversion computes a flux of $+406 \pm 33 \text{ t d}^{-1}$. The inversion detects an increase of emissions from the direction of the leak, but fails to unambiguously affect the increased flux to the proper region. The figure we compute is of the same magnitude (3 times higher) than the estimation given by the operator and does not exceed 18% of the background emissions related to oil and gas extraction in the North Sea.

Our results also reveal a high dependency to the meteorological situation during an inversion window. The ratio between the increment and the posterior errors on the emission budget in the inner regions is very different for the two periods (~ 1 before and $\gg 1$ after the leak start). The reconstructed error on the total budget largely depends on the correlation coefficients in \mathbf{P}^a . For the period before the leak start, most posterior correlations are large and positive ($r > 0.9$). The gradients are then well constrained while the total budget stays uncertain. On the opposite, after the leak start, neighbouring regions exhibit negative correlations by pair (e.g., “NSN” – “NSS”, “MGP” – “GBT”). The assimilation of the observations cannot separate the contribution from these close regions, but it leads to a good reduction of the error on the total balance. These two different behaviours may be related to different synoptic regimes during each inversion window: before the leak start, an anticyclone was laying on central and western Europe; after the leak start, air masses coming from North Europe vented the domain.

3.5 Discussion

3.5.1 Limitations and hypothesis probation

All the results depend on strong statistical and physical hypotheses, which may not all be robust. First, we show in Sect. 3.3.1 that the assumed Gaussian errors of the background can produce physically inconsistent inverted fluxes. Adding Lagrangian correcting factors (e.g., Göckede et al., 2010) to the cost function (Eq. VI.1) can ensure physically consistent fluxes. But that would alter the algebraic properties of the problem and make the implementation of our methods more complex. With regions that act as buffers against the uncertainties on the LBC, the ND methods proved to acceptably deal with the issue.

Second, all CTMs suffer from weaknesses and errors in their parameterizations and numerical scheme. The induced errors can be systematic and not only random, as suggested in Sect. 3.3.2. They should then be considered as a bias η in the observational errors $\epsilon = \mathbf{y}^o - \mathbf{H}\mathbf{x} \sim \mathcal{N}(\eta, \mathbf{R}')$. Further investigations on the effects of the parameterizations, the resolution and the inputs to the CTM shall be carried out to quantify and fix as much as possible the bias η . More specifically, in Sect. 3.3.2, we showed that very high diagnosed error variances during the night could be related to systematic state-dependent biases in the CTM vertical mixing and in PBL modelling.

Third, aggregating fluxes within bigger regions implicitly implies full correlations of the errors on the background in each region. Despite this strong assumption, our methods are supposed to diagnose the error on the aggregated fluxes that are within the footprint of the network. On the opposite, the aggregated regions that are partly within and partly outside the network footprint will exhibit strongly biased diagnosed errors and increments.

Kaminski et al. (2001) studied the issue and found potential errors of the same magnitude as the fluxes themselves. A better choice of the resolution and of aggregated regions considering the prior fluxes and the transport patterns (e.g., Wu et al., 2011) during the window of inversion should significantly improve the results of the methods.

Despite these weaknesses in our methods, the optimal tuple of covariance matrices gives better results than a tuple built on expert considerations: either these expert-built tuples, which are most of the time diagonal, are similar to $(\mathbf{R}_{DS}, \mathbf{B}_{DS})$ that causes inconsistent negative CH_4 fluxes, or the observation errors are enhanced to reduce their impact on the inversion; but in this latter configuration applied to our inversion windows, the corresponding inverted fluxes remain close to the prior ones and the flux uncertainties are not noticeably reduced. Our objectively calculated tuple gives better inversion results, with reduced posterior uncertainties. Moreover, some of the computed error patterns are generic and are transferable to other larger systems. In this study, we chose a particular representation of the complete full-resolution state vector. Most errors represented by the covariance matrix \mathbf{R} are independent of this representation (Bocquet et al., 2011). As a consequence most results on \mathbf{R} will remain valid in the framework of a full-resolution state vector. The recovery of the errors of the non-aggregated background vector are more ambiguous and only large patterns could be inferred for finer resolutions. Additional hypotheses must be made on the shape of the full-resolution background errors to deduce their values from the aggregated matrix. Our methods can then be seen as a way to simplify and project problems with large state vectors in order to infer the patterns of the errors with relatively small computation costs.

3.5.2 Implications for data selection

The framework we chose allows the explicit computation of the sensitivity of the inversion to each observation. We follow Cardinali et al. (2004) and calculate the influence matrix, which gives the effect of a small change of \mathbf{y}^o on $\mathbf{H}\hat{\mathbf{x}}^a$: $\mathbf{S} = (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T$. For every observation \mathbf{y}_j^o , high observation errors reduce the contribution of the observation j to the total inversion. On the opposite, high local contributions tend to enhance the amount of information the inversion can extract from a single observation. The sensitivity matrix \mathbf{S} encompasses both these compensating effects. In Fig. VI.10, each site is colored according to its contribution to the inversion, calculated by adding all the diagonal terms of \mathbf{S} associated to the site. KAS and JFJ have contributions 2.5 times higher than the average contribution, related to their situations in the free troposphere and the constraints they give on the LBC. As a site filling a gap in the observations, PUJ also has a strong contribution (1.5 times the average). On the other hand, GIF, TTA and PUY are very close to the core of the network. The algorithm attributes negligible contributions to these sites in favour of the other sites. Totalling the influence of the whole set of observations, we get the figure of about 50% of the posterior state vector fixed by the observation; the other half comes from the background.

Usually, inversion systems assimilate only a few hours of observations per day, while efforts are made to monitor the atmospheric composition continuously. For example, Bergamaschi et al. (2010) chose to average 3 h of observation per day and per site (bands in Fig. VI.11a). This choice is justified by the confidence given to the model during the afternoon when the vertical mixing in the PBL is strong. Flagging out the other data may be acceptable when inquiring into patterns at the continental scale, but one needs more information at the local and regional scales. In Sect. 3.3.1, we show that the observation errors

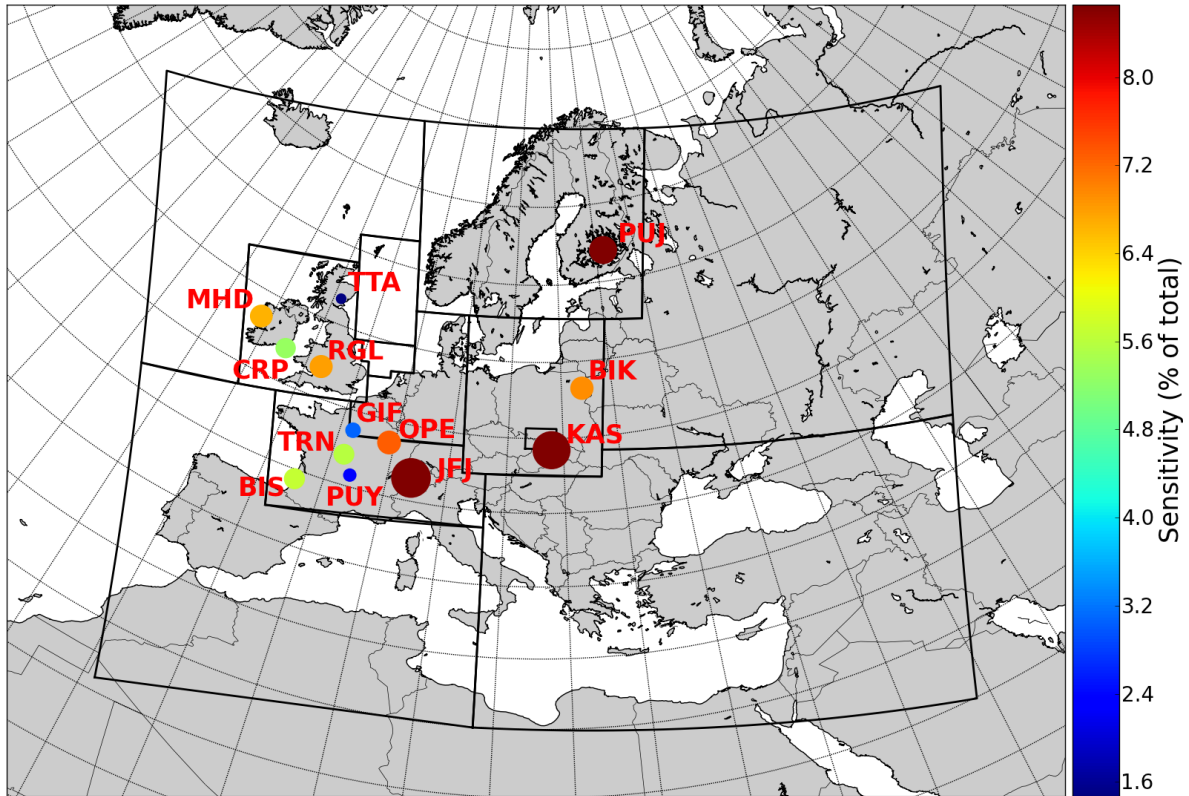


Figure VI.10: Sensitivity of the inversion to each site as the sum of the diagonal elements of the sensitivity matrix S (details in Sect. 3.5.2) associated to the site. The figures are normalised by the total influence of the observations, i.e., the trace of S .

during the day are only 30 % lower than during the night, though this figure is mitigated by non-negligible systematic biases during the night. This small difference suggests that night observations could reasonably be assimilated if the biases had a smaller impact.

In Fig. VI.11a, the sensitivity by site of each hour of the day is shown. Apart from MHD, the usual intervals of selected data (green band for the plain site, blue for the mountain ones) do not exhibit significantly higher sensitivity in the inversion system than the other hours of the day. This selection then leads to a global loss of more than 85 % of information: the sum of the diagonal elements of the sensitivity matrix S related to the observations in the selection band reaches 15 % of the total trace. The mountain stations are known to suffer from the issue of ascendant polluted streams from the PBL as developed in Sect. 3.3.2. Our method implicitly automatically filters out these air masses. Then, the usable observations are not confined to the middle of the night. Moreover, concerning the plain sites (see Fig. VI.11b), the selected band corresponds to the minimum of sensitivity, that is to say the data that least constrain the local and regional fluxes we inquire into. The most influential observations are situated just before the beginning of the day, when the high errors of the night start decreasing and the local contributions are still significant. The end of the afternoon (at about 18:00) is also more influential than the middle of the day, but less than early morning. One should expect a better confidence in late afternoon results from the CTM, when the vertical mixing is still active (though reduced), than early morning.

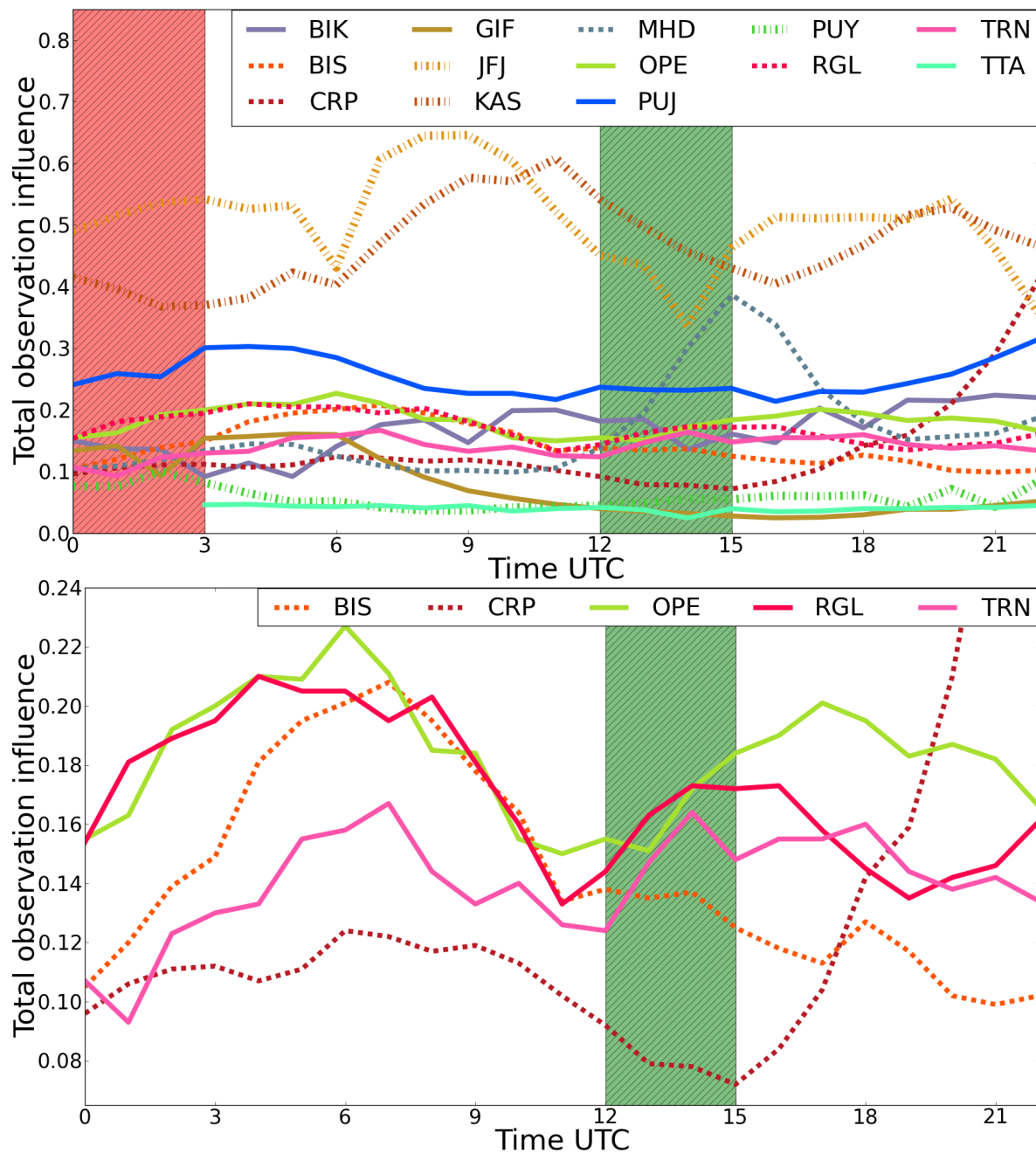


Figure VI.11: Total sensitivity per site computed for each hour of the day (time UTC). Details on the sensitivity computation are described in Sect. 3.5.2. A sensitivity of 1 roughly corresponds to the constraint on 1 degree of freedom of the system. The green (resp. red) band highlights the interval of data selection generally used in most global inversion systems for the plain (resp. mountain) observation sites. Solid lines for rural sites; dashed (resp. dotted) lines for coastal (resp. mountain) sites.

We emphasised in Sect. 3.3.2 the diurnal patterns of the errors on the observations. Further efforts have to be made in modelling the PBL height and the vertical mixing to ensure better quality and reduced bias in the simulations at the end of the night. Our method would then allow a better use of the observations for local and regional inversions.

3.6 Conclusions

We inquired into the possibility of precisely and objectively estimating the covariance matrices of the errors on the observations and the background (\mathbf{R} and \mathbf{B}) that best fit inversion system requirements. A best guess of these matrices with regard to objective criteria is needed in the Bayesian inversion framework, especially for regional studies. To do so, we used algorithms developed in a theoretical framework, but too complex to be tested in full-resolution systems. The translation to a regional configuration was carried out by simplifying the system and reducing the total size of the covariance matrices to allow an algorithmic tuning of \mathbf{R} and \mathbf{B} that estimates optimum tuples (\mathbf{R} , \mathbf{B}) in terms of statistical properties.

We tested 3 algorithms of growing complexity (and computational costs) to estimate the optimal tuple (\mathbf{R} , \mathbf{B}). Unlike other studies, which make strong physical assumptions, such as isotropic spatial correlations in the observation errors or temporal decay of the correlations, we minimised the number of assumptions to keep a better objectivity in our results. In principle, all the patterns of errors can then be recovered specifically to the system during the window of the inversion we focused on.

Amongst other noteworthy patterns, our algorithms retrieved the errors due to the misestimation of the planetary boundary layer height in global circulation models, i.e., large errors during the night and lower during the day when the CTM reproduces the atmospheric transport better. This source of errors contribute in a large part (50% in most plain sites) to the diurnal variability of the observation errors and also causes significant temporal correlations within a 24 h period.

Additionally, our approach does not require a prior filtering of the observations we could consider as ill-simulated by the model. In theory, all the available observations can then be assimilated in the inversion system and not only the ones during early afternoon. However, our study points at probable significant systematic bias in the CTM during the night. The night observations should then be excluded. But the algorithms give objective tools to diagnose the need for efforts to better simulate the atmospheric behaviour during the late night when the observations seem to have the biggest impact on the inversion results. Late afternoon observations were also computed to have a significant influence on the inversion results. A cautious implementation of these observations into an inversion system is expected to enhance the efficiency of the system.

The prospects from this work will be to quantify the uncertainties in our methods and their impact on the optimised fluxes. A dedicated Observing System Simulation Experiment could be carried out in that sense. The computational costs should also be reduced by running our scripts in parallel. In the framework of an inversion system with full temporal and spatial resolution, when variational algorithms are necessary to compute the optimal fluxes, our general method may overburden the computer and memory capacity. Indeed, the limiting factor in our algorithms comes from the diagonal maximisation of the log-likelihood needed to compute the non-diagonal optimal tuple. The maximising algorithm induces computational costs

limited by the size of the ‘background’ vector \mathbf{x}^b . The computational complexity is at least $O([\dim(\mathbf{x}^b)]^3)$ while full-resolution state space dimension is of several orders of magnitude larger than our reduced state space. Nevertheless, the method could be tested in systems of intermediate complexity to infer additional knowledge on the statistics of the errors.

Acknowledgments

We thank all the PIs and supporting staff from the sites we used for making available the validated observations quickly. We wish to thank Bertrand Bonan (INRIA), Lin Wu (Laboratoire des Sciences du Climat et de l’Environnement) and Victor Winiarek (Université Paris-Est, CERE) for the fruitful discussions and advices. We are grateful to the referees for their valuable comments which led to a substantial improvement in the presentation of the paper. This research was supported by the Commissariat à l’Énergie Atomique et aux Énergies Renouvelables.

4 Des erreurs objectivées mais encore incertaines

4.1 Choix de la méthode d’ajustement de matrices

Nous avons testé trois méthodes sur notre cas d’étude : l’une se basant sur le critère du χ^2 , la deuxième sur une maximisation de la vraisemblance et la troisième sur la cohérence de matrices d’erreurs avec des covariations de vecteurs d’innovation. Ces trois méthodes sont de complexité variable et possèdent chacune des contraintes techniques et des approximations sous-jacentes.

Alors que la méthode du critère du χ^2 repose sur de fortes approximations et que la méthode consistant à vérifier des cohérences de matrices de covariance sature les capacités numériques en termes de calcul et de mémoire, la maximisation de la vraisemblance de matrices diagonales présente le meilleur compromis entre réalisme et coût de calculs. Cette méthode permet de reproduire la variabilité temporelle et spatiale des erreurs tout en étant implémentable en un temps acceptable (bien inférieur à la fenêtre d’inversion choisie).

4.2 Incertitudes de la méthode

La fonction de vraisemblance a une forme difficile à caractériser. De ce fait, nous avons besoin de garanties supplémentaires pour assurer que le point de convergence n’est pas un maximum local. De plus, malgré la rigueur mathématique sous-jacente de la méthode choisie, les calculs numériques sont sujets à des approximations et des erreurs d’arrondis.

Une manière d’évaluer ces erreurs est le calcul de la hessienne de la vraisemblance au point de convergence (voir sect. 4.2.3 page 58 du chapitre III). Ce calcul n’est pas trivial et dépasse parfois les capacités de calcul disponibles. Or, développer une méthode objective de quantification des covariances d’erreurs est d’une utilité limitée sans une évaluation des incertitudes intrinsèques à la méthode.

En conséquence, la suite du travail doit se concentrer sur une manière d’estimer ces incertitudes. Pour cela, la taille de l’espace d’état et d’observation doit être astucieusement choisie afin de ne pas dépasser les capacités informatiques accessibles.

4.3 Acquis sur le système d'inversion

Malgré la nécessité d'évaluer les incertitudes des méthodes utilisées, le cas d'étude nous a permis d'acquérir des certitudes sur les systèmes d'inversion régionaux de flux de méthane.

Les corrélations d'erreurs ont montré une importante redondance des informations contenues dans les observations horaires. Des données horaires consécutives portent des informations communes sur la circulation à la méso-échelle et à l'échelle synoptique. Les informations indépendantes restantes concernent les processus locaux que le modèle reproduit mal. En conséquence, les données horaires auxquelles nous avons accès grâce aux sites de mesures in situ en continu sont en excédant par rapport à notre capacité d'analyse par un système d'inversion atmosphérique à la méso-échelle.

Par ailleurs, la variabilité temporelle des erreurs et les corrélations avec la hauteur de la couche limite révèle le rôle prépondérant du mélange vertical local dans les erreurs. Nos calculs ont aussi mis en évidence de très probables biais dans les concentrations simulées de nuit. La figure VI.5 page 118 nous suggère de plus de n'utiliser que les données uniquement mesurées sous une couche limite épaisse d'au moins 500 m.

Ces critères nous amènent à n'introduire qu'une donnée d'observation par jour (l'après-midi) et par site afin d'éviter les biais et les redondances. Ce choix est habituellement fait dans les inversions atmosphériques, mais notre étude a permis de le mettre en évidence objectivement. Le calcul objectif des matrices de covariances a de plus révélé que les données que l'on peut sélectionner sont celles qui apportent le moins d'information au système d'inversion. Les biais et redondance d'information condamnent donc notre système d'inversion atmosphérique à n'utiliser qu'une minorité des informations contenues dans les mesures de concentrations atmosphériques de méthane.

Développement d'un système d'inversion bayésienne par marginalisation

Sommaire

1	Construction heuristique de la méthode générale	135
1.1	Approche pour l'estimation des incertitudes sur les matrices d'erreurs	135
1.2	Définition d'un problème adapté à la théorie	135
1.3	Cas d'étude sibérien	136
2	<i>Objectified uncertainty quantification in Bayesian inversions</i>	137
2.1	<i>Introduction</i>	139
2.2	<i>Marginalized Bayesian inversion</i>	141
2.2.1	<i>Context and motivation for the marginalization</i>	141
2.2.2	<i>Marginalization of the inversion</i>	144
2.3	<i>Informed definition of the problem</i>	146
2.3.1	<i>Principle for problem reduction</i>	148
2.3.2	<i>Representation choice</i>	149
2.3.3	<i>Numerical artefacts</i>	151
2.4	<i>Validation experiments</i>	152
2.4.1	<i>Required tests</i>	152
2.4.2	<i>OSSE evaluation</i>	154
2.5	<i>Set up of the OSSEs</i>	155
2.5.1	<i>Virtual true state \mathbf{x}^t</i>	156
2.5.2	<i>Simulation of the observation operator \mathbf{H}</i>	158
2.5.3	<i>Synthetic observations \mathbf{y}^o</i>	159
2.6	<i>Results and discussion</i>	160
2.6.1	<i>Robustness of the method</i>	160
2.6.2	<i>Spatial evaluation</i>	163
2.6.3	<i>Limitations and benefits</i>	165
2.7	<i>Conclusions</i>	166
3	Une méthode générale et cohérente	167

3.1	Vers un système opérationnel	167
3.2	Améliorations à prévoir	167

Dans la sect. 3.3 page 54 du chapitre III et 4.2 page 130 du chapitre VI, nous avons défini et mis en évidence l'erreur statistique faite par les systèmes d'inversion bayésienne. Dans ce chapitre, nous développons une méthode capable d'estimer cette erreur statistique. Nous commençons par décrire de manière heuristique l'approche choisie pour développer cette méthode et par donner des pistes pour appliquer cette méthode (sect. 1). Nous présentons ensuite en détail la méthode et la soumettons à des tests systématiques sur un cas d'étude d'inversion atmosphérique en Sibérie (sect. 2).

1 Construction heuristique de la méthode générale

1.1 Approche pour l'estimation des incertitudes sur les matrices d'erreurs

Le cadre théorique de l'inversion atmosphérique décrit dans le chapitre III repose sur une nécessaire estimation des statistiques d'erreurs sans prendre en compte l'incertitude statistique. Une manière de prendre en compte cette incertitude consiste à réécrire la densité de probabilité cherchée $p(\mathbf{x}|\mathbf{y}^o, \mathbf{x}^b)$ au moyen d'une marginalisation sur les statistiques d'erreurs. La marginalisation consiste à obtenir la probabilité voulue en sommant les différentes probabilités correspondant à tous les couples (\mathbf{R}, \mathbf{B}) possibles, chacun de ces termes étant pondéré par la probabilité du couple en question (voir figure VII.1).

Mathématiquement, cela s'écrit :

$$p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) = \int p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b, \mathbf{R}, \mathbf{B}) \times p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) d(\mathbf{R}, \mathbf{B}) \quad (\text{VII.1})$$

Cette expression générale inclut le cas classique où on ne prend qu'un seul couple (\mathbf{R}, \mathbf{B}) , considéré comme connu. Il s'agit en effet du cas où $p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ est une distribution en Dirac. Néanmoins, dans le cas général que nous considérons dans la suite, $p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ n'est pas aussi bien connue.

Dans la théorie bayésienne gaussienne, la densité de probabilité a posteriori peut être explicitée. Dans notre approche plus générale, cette explicitation n'est pas triviale. L'approche de marginalisation consistera donc à échantillonner par un algorithme de Monte-Carlo la densité de probabilité des matrices de covariance et à calculer les résultats de l'inversion pour chaque membre de l'échantillon.

On voit donc que l'approche par marginalisation Monte-Carlo s'appuiera sur de nombreuses inversions explicites. Or, une inversion atmosphérique complète peut coûter jusqu'à quelques jours, voire quelques semaines de calcul dans notre cas. Il est par conséquent indispensable de réduire la taille du problème inverse que l'on se pose afin que la marginalisation ne soit pas trop gourmande en ressources de mémoire et de puissance de calcul et ne reste pas théorique.

1.2 Définition d'un problème adapté à la théorie

Les ressources nécessaires pour chaque inversion et la taille de l'échantillonnage Monte-Carlo de la marginalisation dépendent de la taille des espaces d'état et d'observation. Une définition brutale du problème avec une haute résolution temporelle et spatiale pour les flux et les mesures pousserait très vite les serveurs de calculs à leurs limites. Nous tâchons donc de définir le problème (c'est-à-dire la représentation ω) aussi petit que possible. Pour cela,

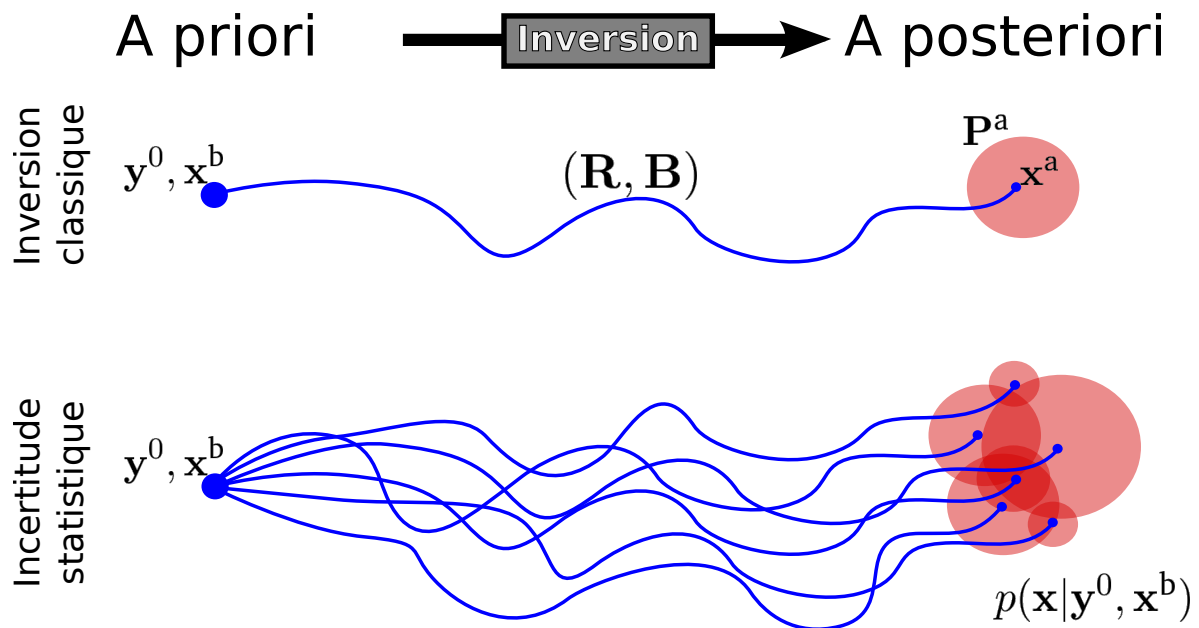


FIGURE VII.1: Incertitude statistique dans l'inversion bayésienne. L'inversion déduit l'état a posteriori \mathbf{x}^a en fonction de $(\mathbf{y}^0, \mathbf{x}^b)$. Dans l'inversion bayésienne classique, \mathbf{x}^a est déduit avec une incertitude \mathbf{P}^a en fonction des statistiques d'erreurs (\mathbf{R}, \mathbf{B}) . Un ensemble de couples (\mathbf{R}, \mathbf{B}) peuvent être choisis : c'est ce qu'on appelle l'incertitude statistique. On a donc un ensemble de $(\mathbf{x}^a, \mathbf{P}^a)$ qui forment $p(\mathbf{x}|\mathbf{y}^0, \mathbf{x}^b)$.

nous agrégeons les observations par après-midi comme spécifié dans la sect. 4.3 page 131 du chapitre VI et les flux par régions d'émissions et par plages de temps.

Agréger des flux par régions engendre d'inévitables incertitudes (Kaminski et al., 2001; Bocquet et al., 2011). Les deux sources majeures d'erreurs d'agrégation sont : la mauvaise distribution spatiale a priori des flux au sein d'une région d'agrégation et la tendance de l'inversion à corriger par bloc les régions agrégées alors même que seule une partie de la région est contrainte par les observations.

Wu et al. (2011) proposent une manière optimale de choisir la représentation ω pour chaque système d'inversion. Néanmoins, leurs calculs sont très coûteux. Par conséquent, nous n'implémentons pas leur calculs et suivons plutôt l'heuristique de leur travail avec des régions plus petites à mesure que l'on se rapproche des sites d'émissions. Plus précisément, nous estimons la zone d'influence de chaque site d'observation à l'aide de calcul de rétro-trajectoires FLEXPART et sélectionnons à la main les régions en fonction de l'influence du réseau de mesure complet (voir figure VII.2).

1.3 Cas d'étude sibérien

Nous testons la pertinence et l'efficacité de la marginalisation sur un cas d'étude académique en Sibérie basé sur des observations (voir figure VII.5 page 156) et un transport atmosphérique réalistes. Nous définissons une « réalité virtuelle » pour les flux et les observations et procédons à des inversions marginalisées avec plusieurs configurations possibles (flux a priori et modèle de transport).

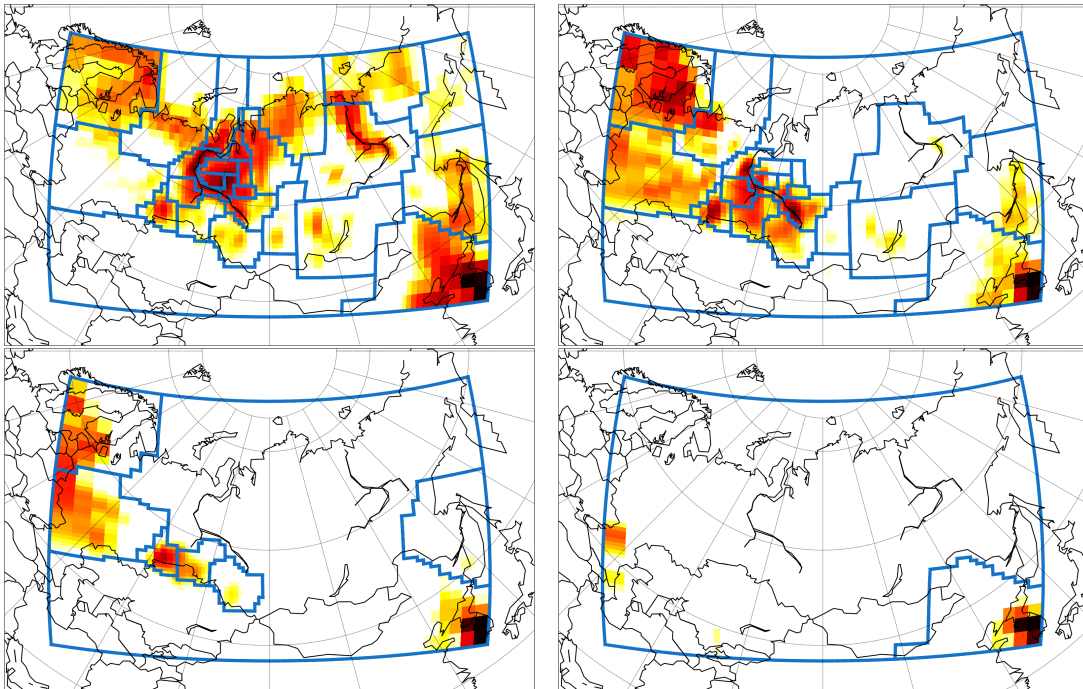


FIGURE VII.2: Choix des régions d'agrégation en fonction de l'influence du réseau de mesures à quatre périodes de l'année. La zone d'influence du réseau (matérialisée par les nuances de rouge sur la carte) dépend du transport et des émissions. Un ensemble de région d'agrégation cohérentes physiquement est spécifié (en haut à gauche). Pour une influence significative par toutes les régions du motif de base, l'ensemble des régions est pris en compte dans l'inversion. Les régions contiguës avec une faible influence sont fusionnées pour réduire le nombre de régions dans l'inversion (en haut à droit et bas à gauche). En bas à droite, un minimum d'influence impose la fusion de presque toutes les régions.

Pour définir les régions d'agrégation, nous définissons des ensembles physiquement cohérents de zones inondées (par écosystème) et d'émissions anthropiques (par sous-région), puis, en fonction de l'influence du réseau d'observations utilisées, et donc du transport, nous regroupons ou non ces régions pré-établies (voir figure VII.2). Une fois la structure d'agrégation choisie, nous calculons l'opérateur d'observation \mathbf{H} à l'aide de simulations directes de transport avec CHIMERE pour chaque région d'agrégation, c'est-à-dire chaque colonne de \mathbf{H} .

Le cas d'étude est décrit plus en détails dans la sect. 2.

2 Objectified quantification of uncertainties in Bayesian atmospheric inversions

Cet article a été soumis le 28 mai 2014 dans le journal de l'éditeur Copernicus de l'*European Geosciences Union*, *Geoscientific Model Development*. Il est depuis le 29 juillet 2014 en discussion dans le même journal (voir à l'adresse <http://www.geosci-model-dev-discuss.net/7/4777/2014/gmdd-7-4777-2014.html>). Une version revue a été soumise en avril 2015 et est retranscrite ci-dessous.

Les auteurs ayant participé à ce travail sont : Antoine Berchet¹, Isabelle Pison¹, Frédéric Chevallier¹, Philippe Bousquet¹, Jean-Louis Bonne¹, et Jean-Daniel Paris¹.

Les résultats préliminaires de cette étude ont été présentés lors d'une conférence internationale et d'un atelier international (voir Annexe B page 245) : présentation orale à l'*European Geosciences Union General Assembly* (Avril 2013 ; Vienne, Autriche), présentation orale à l'*International ZOTTO Workshop* (Septembre 2013 ; Krasnoyarsk, Russie).

Résumé

Les inversions atmosphériques bayésiennes classiques traitent des observations atmosphériques et des émissions a priori. Observations et émissions sont connectées par un opérateur d'observation représentant principalement le transport et le mélange atmosphériques. Ces inversions classiques s'appuient sur la définition d'erreurs sur les observations, les émissions et l'opérateur d'observation. À la méso-échelle, les résultats des inversions sont très sensibles aux distributions d'erreurs prescrites, alors que celles-ci sont mal connues. Le cadre théorique bayésien classique quantifie mal l'impact sur les flux optimisés des distributions d'erreurs mal spécifiées. Dans le but de lutter contre cette défaillance du cadre bayésien, nous nous appuyons sur les résultats de travaux de recherche récents et modifions la théorie bayésienne. Pour cela, nous appliquons une marginalisation sur toutes les distributions d'erreurs qui peuvent être prescrites dans le système d'inversion. La marginalisation consiste à procéder à des inversions pour toutes les erreurs possibles, les résultats des inversions étant pondérés par la probabilité d'occurrence de chaque distribution d'erreur. La distribution a posteriori des flux ainsi calculées ne peut être explicitée. De ce fait, nous procédons à un échantillonnage Monte-Carlo sur une approximation de la densité de probabilité des distributions d'erreurs. Cette approximation est déduite d'un algorithme déjà testé de maximum de vraisemblance. De la sorte, l'inversion marginalisée repose sur un diagnostic objectif des statistiques d'erreurs sans ajout préliminaire d'information sur les matrices d'erreurs. Elle inclut de manière robuste les incertitudes sur les distributions des erreurs, à l'inverse des inversions classiques avec des matrices d'erreurs figées. La pertinence et la robustesse de la méthode sont testées sur un cas d'étude : l'inversion des flux de méthane à la méso-échelle en Eurasie avec des sites d'observations réalistes. Des expériences de simulations de systèmes d'observation sont effectuées avec différentes structures de transports, distributions de flux, et émissions totales. La méthode reproduit de manière statistiquement cohérente la « réalité » prescrite dans la plupart des cas. De plus, notre méthode nous procure des scores d'influence et des matrices de corrélation a posteriori. Une interprétation exhaustive des résultats de l'inversion est alors possible. Nous sommes ainsi en mesure d'évaluer l'impact du réseau d'observation sur la caractérisation des flux de surface. Les corrélations a posteriori nous révèlent les régions qui sont mal séparées. Nous pouvons en déduire les échelles temporelles spatiales et temporelles que le système d'inversion peut typiquement contraindre. Ces échelles se sont révélées cohérentes avec les schémas d'agrégation choisis heuristiquement.

¹Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CNRS-CEA-UVSQ, Orme des Merisiers, CEA Saclay, Gif-sur-Yvette, France.

Abstract

Classical Bayesian atmospheric inversions process atmospheric observations and prior emissions, the two being connected by an observation operator picturing mainly the atmospheric transport. These inversions rely on prescribed errors in the observations, the prior emissions and the observation operator. When data pieces are sparse, inversion results are very sensitive to the prescribed error distributions, which are not accurately known. The classical Bayesian framework experiences difficulties in quantifying the impact of mis-specified error distributions on the optimized fluxes. In order to cope with this issue, we rely on recent research results to enhance the classical Bayesian inversion framework through a marginalization on a large set of plausible errors that can be prescribed in the system. The marginalization consists in computing inversions for all possible error distributions weighted by the probability of occurrence of the error distributions. The posterior distribution of the fluxes calculated by the marginalization is not explicitly describable. As a consequence, we carry out a Monte Carlo sampling based on an approximation of the probability of occurrence of the error distributions. This approximation is deduced from the well-tested method of the maximum likelihood estimation. Thus, the marginalized inversion relies on an automatic objectified diagnosis of the error statistics, without any prior knowledge about the matrices. It robustly accounts for the uncertainties on the error distributions, contrary to what is classically done with frozen expert-knowledge error statistics. Some expert knowledge is still used in the method for the choice of an emission aggregation pattern and of a sampling protocol in order to reduce the computation cost. The relevance and the robustness of the method is tested on a case study: the inversion of methane surface fluxes at the meso-scale with virtual observations on a realistic network in Eurasia. Observing System Simulation Experiments are carried out with different transport patterns, flux distributions and total prior amounts of emitted methane. The method proves to consistently reproduce the known 'truth' in most cases, with satisfactory tolerance intervals. Additionally, the method explicitly provides influence scores and posterior correlation matrices. An in-depth interpretation of the inversion results is then possible. The more objective quantification of the influence of the observations on the fluxes proposed here allows us to evaluate the impact of the observation network on the characterization of the surface fluxes. The explicit correlations between emission aggregates reveal the mis-separated regions, hence the typical temporal and spatial scales the inversion can analyse. These scales are consistent with the chosen aggregation patterns.

2.1 Introduction

Characterizing the global biogeochemical cycles of greenhouse gases requires a reliable understanding of the exchanges at the surface-atmosphere interface. The description of these exchanges must encompass the absolute amounts of gas released to and removed from the atmosphere at the surface interface, the spatial distribution and the temporal variability of the fluxes, and the determination of the underlying physical processes of emissions and sinks. Such an integral depiction is still missing for most greenhouse gases (Ciais et al., 2013). One of the possible approaches to inquire into the surface fluxes is the analysis of the atmospheric signal. The drivers of the spatial and temporal variability of the atmospheric composition are atmospheric transport, chemistry and surface fluxes. Therefore, monitoring the atmospheric composition and using a representation of the atmospheric transport and chemistry with Global Circulation Models (GCMs) or Chemistry-Transport Models (CTMs) can help in

inferring information on the fluxes (Bousquet et al., 2006; Bergamaschi et al., 2010). This approach, called atmospheric inversion, suffers from two practical issues in its implementation. First, the atmospheric composition is still laconically documented, though the number of global monitoring projects with extensive surface observation networks and satellite platforms has been increasing in the last decades (e.g., Dlugokencky et al., 1994, 2009). Indeed, the satellite platforms have a global coverage but the observed atmospheric composition is integrated over the vertical column, while the surface sites can provide continuous observations but only at fixed point locations. Second, the atmosphere behaves as an integrator and the air masses are mixed ambivalently through the transport (Enting et al., 1993). Thus, the inverse problem of tracking back the fluxes from the variability of the atmospheric composition cannot be solved univocally. The Bayesian formalism allows statistical analyses of the atmospheric signal, so that one can identify confidence intervals of fluxes compatible with the atmospheric composition (Tarantola, 1987).

Bayesian inversions have been extensively used at the global scale, providing insights on the greenhouse gas budgets (e.g., Gurney et al., 2002; Kirschke et al., 2013; Bergamaschi et al., 2013). However, non compatible discrepancies in the results appear between the possible configurations of atmospheric inversion systems (Peylin et al., 2013). The various configurations include the choice of the atmospheric transport, its spatial and temporal resolutions, the meteorological driving fields, the type and density of the observations, etc. In the Bayesian formalism, some assumptions also have to be made on the transport model error statistics, on the errors made when comparing a discretized model to observations (Geels et al., 2007) and on the confidence we have on the prior maps and time profiles of emissions (Enting, 2002). All these choices are based on technical considerations and on the expert perception of the problem to solve. Comparing results based on different choices that are physically adequate, but subjective, is difficult, especially to track inconsistencies, which enlarge the range of flux estimates.

In the following, we focus on the development of an enhanced Bayesian method that objectifies the assumptions on the statistics of the errors and that takes the unavoidable uncertainties generated by our lack of knowledge on these error statistics into account. In this approach, the confidence ranges of the optimized surface fluxes are computed by a Monte Carlo marginalization on all the possible error statistics, which is more general than the usual Bayesian approach deducing posterior uncertainties from a single error statistic combination only. The weight function for the marginalization is inferred from an already-tested maximum likelihood approach (e.g., Dee, 1995; Michalak et al., 2005), processing the pieces of information carried by the differences between the measurements and the prior simulated concentrations. The potential and consistency of the method is tested through Observing System Simulation Experiments (OSSEs) on a realistic configuration of atmospheric inversion.

The case study is the quantification of methane fluxes in the Siberian Lowlands with a network of surface observation sites that have been operated for a few years by the Japanese National Institute for Environmental Studies (Sasakawa et al., 2010) and the German Max Planck Institute (Winderlich et al., 2010). The characterization of the region is challenging, with co-located massive methane emissions from anthropogenic activity (oil and gas extraction) and from wetlands in summer. Moreover, the wetland emissions have a very high temporal variability (due to their sensitivity to the water table depth and to the temperature; e.g., Macdonald et al., 1998; Hargreaves and Fowler, 1998). Their quantification is then

difficult. In order to catch the influence of the sampling bias due to non-regularly distributed observation sites and non-continuous measurements, we produce virtual observations from a known 'truth' at locations where real observations are carried out and at dates when the logistical issues do not prevent the acquisition of measurements. We then check the capability of our method to reproduce consistent flux variability and distribution with seven degraded inversion configurations (perturbed transport, flat flux distributions, etc.).

In Sect. 2.2, we describe the theoretical framework of our method of marginalization. The enhancements on the general theoretical framework need a cautious definition of the problem to be implementable in term of computational costs and memory limits. In Sect. 2.3, guidelines for a suitable definition of the problem are developed. The whole structure of the method is summarized in Sect. 2.4.1. In Sect. 2.4, we present the particular set up of the OSSE carried out for proving the robustness of the method. The specific Siberian configuration we test our method on is detailed in Sect. 2.5. The OSSE are evaluated along defined objective statistical scores in Sect. 2.6.

2.2 Marginalized Bayesian inversion

We first describe the motivations for using a marginalized inversion in Sect. 2.2.1. In Sect. 2.2.2, we describe the marginalization itself and the Monte Carlo approach chosen in order to compute it.

2.2.1 Context and motivation for the marginalization

2.2.1.1 Bayesian inversion framework The surface-atmosphere fluxes, through transport, cause a variability in the atmospheric mixing ratios of the species we are interested in. The atmospheric inversion relies on the processing of the atmospheric variability in order to infer the surface-atmosphere fluxes. Since the atmosphere is diffusive and irreversibly mixes air masses from different origins, it is physically impossible to infer univocal information on the fluxes from the integrated atmospheric signal alone (Tarantola, 1987; Enting, 2002). We then pursue a thorough characterization of the *pdf* of the state of the system \mathbf{x} (e.g., the spatial and temporal distribution of the surface fluxes, but also background concentrations and baselines in some cases), assuming some prior knowledge on the system and having some observations of the atmospheric physical variables related to our problem. That is to say, we want to calculate the *pdf* $p(\mathbf{x}|\mathbf{y}^o - H(\mathbf{x}^b), \mathbf{x}^b)$ for all possible states \mathbf{x} ; \mathbf{y}^o is a vector gathering all the available observations, \mathbf{x}^b is the background vector including the prior knowledge on the state of the system and H is the observation operator converting the information in the state vector to the observation space. Typically, H embraces the atmospheric transport and the discretization of the physical problem. In the scope of applications of the atmospheric inversions, the observation vector \mathbf{y}^o gathers measurements of dry air mole fraction. As for the observation operator, it is computed with a model which simulates mixing ratios. As we are interested in trace gases, we will consider that the dry air mole fractions can be treated as mixing ratios. In all the following, we also consider that H is linear; hence, H is represented by its Jacobian matrix \mathbf{H} and $H(\mathbf{x}^b) = \mathbf{H}\mathbf{x}^b$. This approximation is valid for all non reactive atmospheric species at scales large enough, so that the treatment of the local scale turbulence by the model does not generate numerical non-linearity. When the atmospheric chemistry must be taken into account (for instance with methane), either the window of inversion must be short compared with the typical lifetime in the atmosphere for the linear assumption to be

valid, or the concentration fields of the reactant species (e.g., OH radicals for methane) must be accurately known.

In general, the characterization of the *pdf* is built within the Bayesian formalism with the assumption that all the involved *pdfs* are normal distributions (Enting et al., 1993). The *pdfs* are then explicitly described through their mode and their matrix of covariance. In this case, the *pdf* $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \propto \mathcal{N}(\mathbf{x}^a, \mathbf{P}^a)$ is defined by its mode, \mathbf{x}^a , the posterior state, and its matrix of covariance, \mathbf{P}^a . In addition to the linear assumption, we also consider that the uncertainties are unbiased. That is to say: $p(\mathbf{x} - \mathbf{x}^b) \propto \mathcal{N}(\mathbf{0}, \mathbf{B})$ and $p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t) \propto \mathcal{N}(\mathbf{0}, \mathbf{R})$ where \mathbf{x}^t is the true state of the system. The uncertainty matrix \mathbf{B} (resp. \mathbf{R}) encompasses the uncertainties on the background \mathbf{x}^b (resp. on the measurements and on the model, including representation errors, i.e. the errors made when approximating the real world by a numerical gridded model). Under these assumptions, we can explicitly write the posterior vector and the posterior matrix of covariance:

$$p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \propto \mathcal{N}(\mathbf{x}^a, \mathbf{P}^a) : \begin{cases} \mathbf{x}^a &= \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \\ \mathbf{P}^a &= \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B} \end{cases} \quad (\text{VII.2})$$

with $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$ the Kalman gain matrix.

2.2.1.2 Ambivalent uncertainty set-up Atmospheric inversion is straightforward (apart from technical issues in the numerical implementation of the theory) as long as the uncertainty matrices \mathbf{R} and \mathbf{B} are defined.

Some of their components can be calculated unambiguously, such as measurement errors in matrix \mathbf{R} . Other errors are derived, in most cases, following expert knowledge on, e.g., the behaviour of the atmospheric transport and of the surface fluxes. This expert knowledge is acquired, for example, through extensive studies on the sensitivity of the transport model to its parametrization and forcing inputs (e.g., Denning et al., 1999; Ahmadov et al., 2007; Lauvaux et al., 2009; Locatelli et al., 2013), or by comparing prior fluxes to measured local fluxes (e.g., Chevallier et al., 2006). Some studies also rely on pure physical considerations (e.g., Bergamaschi et al., 2005, 2010).

However, the complex and unpredictable structure of the uncertainties is hard to reproduce accurately from the expert knowledge alone and an ill-designed couple of uncertainty matrices (\mathbf{R}, \mathbf{B}) can have a dramatic impact on the inversion results (e.g., Berchet et al., 2013b; Cressot et al., 2014). The discrepancies between the possible configurations of inversion can also reveal some biases, η , in the models: in that case $p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t) \propto \mathcal{N}(\eta, \mathbf{R})$ instead of $p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t) \propto \mathcal{N}(\mathbf{0}, \mathbf{R})$, which would require a different handling of Eq. VII.2. For example, the horizontal wind fields can be biased or the vertical mixing in the planetary boundary layer systematically erroneous. That makes it difficult to compare simulated concentrations in the boundary layer to measurements (e.g., Peylin et al., 2002; Dee, 2005; Geels et al., 2007; Williams et al., 2013; Lauvaux and Davis, 2014). Biases can have critical impacts on inversion results and must be inquired into independently (e.g., Bocquet, 2011). Nevertheless, for our study, we decide to neglect the biases in the inversion. We discuss in Sect. 2.6.3 the potential impacts of biases that are not significant in our specific application. We then focus only on the mis-specification of the uncertainty matrices \mathbf{R} and \mathbf{B} .

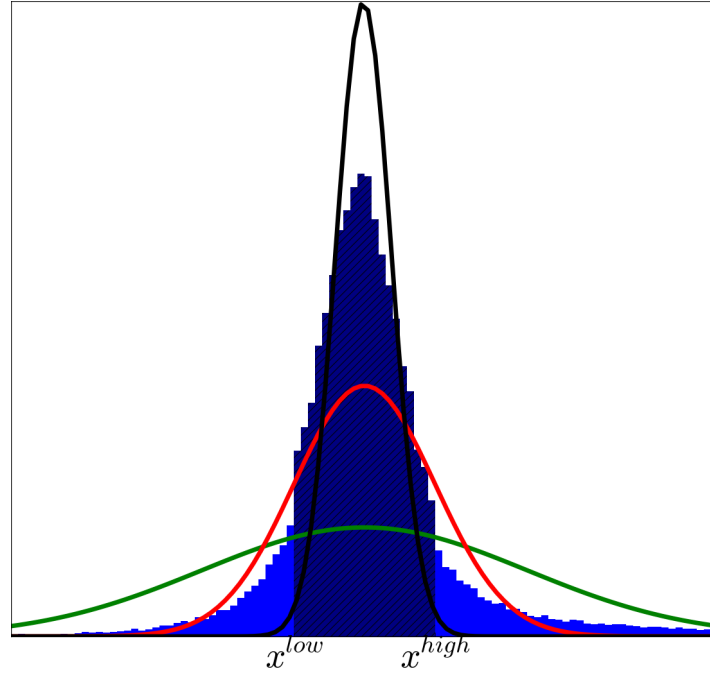


Figure VII.3: Distribution of one component of the Monte Carlo posterior ensemble. The histogram displays the raw posterior distribution. The dark hatched part of the histogram depicts the proportion of the ensemble within the tolerance interval TI_{68} , $[x^{\text{low}}, x^{\text{high}}]$ (as defined in Sect. 2.2.2). The red curve represents the Normal distribution with the same mode and tolerance interval; the green one stands for a Normal distribution with the same mode and the same standard deviation; the black one is the posterior distribution computed with the maximum likelihood couple of uncertainty matrices, presenting under-estimated skewness compared with the Monte Carlo distribution.

2.2.1.3 Possible uncertainty handling In order to address the uncertainty issue in atmospheric inversions, efforts are carried out towards objectifying the way the error statistics are chosen (e.g., Schwinger and Elbern, 2010; Winiarek et al., 2012; Berchet et al., 2013b). These efforts focus on specific algebraic properties of the uncertainty matrices (e.g., Desroziers and Ivanov, 2001; Desroziers et al., 2005) or more generally on understanding the likelihood of the prior innovation vector, $\mathbf{y}^o - \mathbf{H}\mathbf{x}^b$, as a function of the uncertainty matrices (Dee, 1995). Under Gaussian assumptions, the likelihood of the innovation vector can be written:

$$p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b | \mathbf{R}, \mathbf{B}, \mathbf{x}^b) = \frac{e^{-\frac{1}{2}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)^\top (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^\top)^{-1} (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)}}{\sqrt{(2\pi)^d |\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^\top|}} \quad (\text{VII.3})$$

with d the dimension of the observation space and $|\cdot|$ the determinant operator.

In the likelihood framework, the couple of uncertainty matrices (\mathbf{R}, \mathbf{B}) that maximizes Eq. VII.3 is considered as optimal and will be hereafter referred to as the maximum likelihood. This maximum likelihood optimally balances the observation and prior state error variances and covariances according to the prior innovation vector $\mathbf{y}^o - \mathbf{H}\mathbf{x}^b$ (Chapnik et al., 2004). A direct algorithm computing the maximum likelihood (applied to atmospheric inversion

in, e.g., Winiarek et al., 2012; Berchet et al., 2013b) is then supposed to provide a good approximation of the couple of optimal matrices $(\mathbf{R}_{\max}, \mathbf{B}_{\max})$ which can be used forward in the inversion (Eq. VII.2). In order to dampen the computation cost of the maximum likelihood estimation, most studies just maximize the likelihood on hyperparameters (e.g., correlation lengths), describing the couple of matrices (\mathbf{R}, \mathbf{B}) in a more simple way.

Though general, the estimation of the innovation vector maximum likelihood relies on strong assumptions, it can suffer from strong numerical errors and it is not necessarily univocal. More explicitly, as showed by previous works, the *pdf* of the uncertainty matrices $p(\mathbf{R}, \mathbf{B})$ behaves as a χ^2 distribution with d degrees of freedom, d being the dimension of the observation space. Thus, the likelihood is highly dominated by the mode of $p(\mathbf{R}, \mathbf{B})$, co-located with the maximum likelihood. However, the peaked likelihood argument may be too rough in some cases. As the number of observations decreases compared to the number of state dimensions, this optimal case becomes less univocal. In the frameworks where observations are too scarce, the maximum likelihood may lead to flawed results. To assess the validity of the peak assumption, estimations of the Hessian matrix of the likelihood at its maximum have been used (e.g., Michalak et al., 2005; Wu et al., 2013). Hessian matrices give the magnitude of the uncertainties on the computation of the uncertainty matrices. Nevertheless, to our knowledge, no atmospheric inversion account for the impact of the Hessian matrix of the likelihood on the inversion results.

In addition, even when the *pdf* $p(\mathbf{R}, \mathbf{B})$ is intensely peaked at its maximum, the inferred inversion results from Eq. VII.2 with a direct maximum likelihood algorithm would erroneously under-estimated uncertainties on the result (see Fig. VII.3 and, e.g., Berchet et al., 2013b). Indeed, at the maximum likelihood, all the pieces of information in the system are considered perfectly usable by the inversion which then gives too optimistic posterior uncertainties in this case.

2.2.2 Marginalization of the inversion

2.2.2.1 Theoretical formulation Here, we compute the *pdf* $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ by a marginalization on the uncertainty matrices to comprehensively account for the uncertainties in the characterization of the uncertainties and to quantify the impact of ill-specified uncertainty matrices. In statistics, marginalizing a probability density function (*pdf*) $p(\mathbf{x})$ consists in rewriting it as a sum of conditional probabilities $p(\mathbf{x}|\mathbf{z})$ weighted by $p(\mathbf{z})$.

Thus, the complete *pdf* $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ classically described by Eq. VII.2 is separated into a sum of the contribution of each possible couple of covariance matrices (\mathbf{R}, \mathbf{B}) weighted by the probability of occurrence of the couple (\mathbf{R}, \mathbf{B}) :

$$\begin{aligned} & p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \\ &= \int_{(\mathbf{R}, \mathbf{B})} p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b, \mathbf{R}, \mathbf{B}) \\ & \quad \times p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \quad (\mathbf{R}, \mathbf{B}) \\ & \propto \int_{(\mathbf{R}, \mathbf{B})} \mathcal{N}(\widetilde{\mathbf{x}}^a, \widetilde{\mathbf{P}}^a) \\ & \quad \times p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \quad (\mathbf{R}, \mathbf{B}) \end{aligned} \quad (\text{VII.4})$$

In Eq. VII.4, $(\widetilde{\cdot})$ depicts a dependency to the couple (\mathbf{R}, \mathbf{B}) . The complete *pdf* $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ then has the shape of an infinite sum of weighted normal distributions. This infinite

sum could be described as a multi-variate T-distribution (Bocquet, 2011).

The general expression of Eq. VII.4 encompasses the classical case with only one couple of matrices (\mathbf{R}, \mathbf{B}) which considers $p(\mathbf{R}, \mathbf{B} | \mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ as a Dirac-like distribution (centered at the maximum likelihood or at any expert-based couple of uncertainty matrices). More generally, $p(\mathbf{R}, \mathbf{B} | \mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ is not so well known as discussed in Sect. 2.2.1.3 above.

2.2.2.2 Monte Carlo sampling Hereafter, a direct Monte Carlo characterization of Eq. VII.4 is carried out to deduce $p(\mathbf{x} | \mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$.

The Monte Carlo ensemble is to be defined along the *pdf* $p(\mathbf{R}, \mathbf{B})$, but the exact distribution of the error statistics is intricate. In all the following, we then approximate the *pdf* $p(\mathbf{R}, \mathbf{B})$ by a multi-variate χ^2 distribution with d (the number of observations) degrees of freedom, centered at the maximum likelihood of the prior innovation vector (following Dee, 1995). The Monte Carlo marginalization is consequently a direct extension of the maximum likelihood estimation now classically used in the atmospheric inversion framework.

The maximum likelihood can be estimated first by a quasi-Newtonian descent method. However, descent methods have high computation costs and thus require a reduced number of hyperparameters (variances, correlation lengths, etc.) to describe the full uncertainty matrices. From here, we decide to reduce the distribution of the matrices (\mathbf{R}, \mathbf{B}) to the subspace of the diagonal positive matrices. Using a subspace of the possible error statistics can dampen the generality of the method. In particular, error correlations will be excluded with diagonal uncertainty matrices. Correlations can be used in some frameworks to detect the biases in the system (Berchet et al., 2013b). But, more importantly, correlations of observation or background errors can indicate redundant pieces of information in the inversion system. For instance, an inversion computed with no observation correlation tries to use too much information and is expected to give too optimistic a reduction of uncertainties on the fluxes. Nevertheless, in Sect. 2.3, we reduce the observation and state spaces in order to numerically compute the Monte Carlo marginalization. The reduction of the observation and state spaces indirectly depicts correlations in the full-resolution system. In this configuration, the correlation issue is then attenuated and the diagonal assumption is valid.

At the end, for each diagonal term of the uncertainty matrices (\mathbf{R}, \mathbf{B}) , we prescribe a χ^2 distribution with d (i.e. the dimension of the observation space) degrees of freedom, rescaled so that its average equals the associated term in the computed maximum likelihood couple $(\mathbf{R}_{\max}, \mathbf{B}_{\max})$. That is to say, for each diagonal element $\mathbf{r}^{i,i}$ of the matrix \mathbf{R} (equivalently of the matrix \mathbf{B}):

$$p\left(\frac{\mathbf{r}^{i,i}}{\mathbf{r}_{\max}^{i,i}} \times d\right) \propto \chi^2(d) \quad (\text{VII.5})$$

as the mean of the χ^2 distribution with d degrees of freedom, $\chi^2(d)$, is d .

The χ^2 distributions are then sampled on a large ensemble – the Monte Carlo approach stabilizes after a few tens of thousands draws in our case study – to characterize the final output *pdf* $p(\mathbf{x} | \mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$. Each samples of the ensemble must take into account the spread of $\mathcal{N}(\tilde{\mathbf{x}}^a, \tilde{\mathbf{P}}^a)$ in Eq. VII.4. To do so, we describe the *pdf* $p(\mathbf{x} | \mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ not from the ensemble of posterior fluxes $(\tilde{\mathbf{x}}^a)$, but from a perturbed ensemble of $(\tilde{\mathbf{x}})$, with each $\tilde{\mathbf{x}}$ a random sample of $\mathcal{N}(\tilde{\mathbf{x}}^a, \tilde{\mathbf{P}}^a)$.

2.2.2.3 Processing the Monte Carlo posterior ensemble In Fig. VII.3, we draw an example of the distribution of the Monte Carlo posterior vector ensemble along one component of the state space. The black curve depicts the posterior distribution inferred from the maximum likelihood, with under-estimated spread compared to the Monte Carlo distribution. On the opposite, as illustrated by the green curve, a Normal distribution with the same mode and the same standard deviation gives a misleading flat shape. As for a Gaussian, we then define the symmetric tolerance interval, so that 68.27% of the samples are in the range (the hatched portion of the histogram in Fig. VII.3). This interval is equivalent to the Gaussian $\pm\sigma$ interval, with σ the standard deviation. One shall remind that the computed tolerance interval does not depict a Normal distribution. A Normal distribution with the same tolerance interval (the red curve in Fig. VII.3) is still misleadingly flat. In all the following, we will write the tolerance interval $\text{TI}_{68}, [\mathbf{x}^{\text{low}}, \mathbf{x}^{\text{high}}]$.

To summarize (as represented in the block diagram of Fig. VII.4), the maximum likelihood is first estimated using a quasi-Newtonian algorithm, similarly to what has been done in the literature (e.g., Winiarek et al., 2012; Berchet et al., 2013b). We deduce from this maximum likelihood a plausible distribution of the uncertainty matrices (\mathbf{R}, \mathbf{B}) . Through a Monte Carlo sampling of uncertainty matrices (\mathbf{R}, \mathbf{B}) along the deduced distribution, we compute an ensemble of possible posterior vectors $(\mathbf{x}^{\text{a}}_{(\mathbf{R}, \mathbf{B})})$. We can then define the tolerance intervals TI_{68} and a posterior covariance matrix filled by the covariances of the ensembles of state components with each other.

Posterior covariance matrices are not always easy to compute in the atmospheric inversion framework. Here, the posterior covariance matrix is computed explicitly and objectively. The explicit definition of this matrix can give valuable information on the ability of the inversion to separate co-located emissions and emissions at different periods and locations. This capacity is used for the evaluation of the OSSEs in Sect. 2.4.2 and 2.6.

2.3 Informed definition of the problem

The general approach defined in Sect. 2.2 applies a Monte Carlo method on tens of thousands individual inversions after the completion of a maximum likelihood algorithm. This procedure requires extensive amounts of memory and computation power that can't be afforded in most real cases. For instance, the explicit computation of \mathbf{H} with a Chemistry-Transport Model (CTM) closely depends on the dimension of the state space: every column of the observation operator needs one model simulation (Bousquet et al., 1999a). Additionally, each step of the algorithm to compute the maximum likelihood of the prior innovation vector and each step of the Monte Carlo method relies on matrix products, matrix determinants and matrix inverses. At first sight, all these operations are as many technical issues in high dimension problems.

As a consequence, the application of the theoretically simple framework developed in Sect. 2.2 relies closely on an informed definition of the problem. The dimensions of the observation and state spaces should be reduced to dampen the numerical obstacles, but one shall keep resolutions physically relevant for the system we are analysing. By synthesising the recent literature on the subject, we show in the following that approximations can be reasonably applied to the full-resolution problem while not impacting the quality of the marginalized inversion results. Applying the Monte Carlo marginalized inversion is then technically feasible

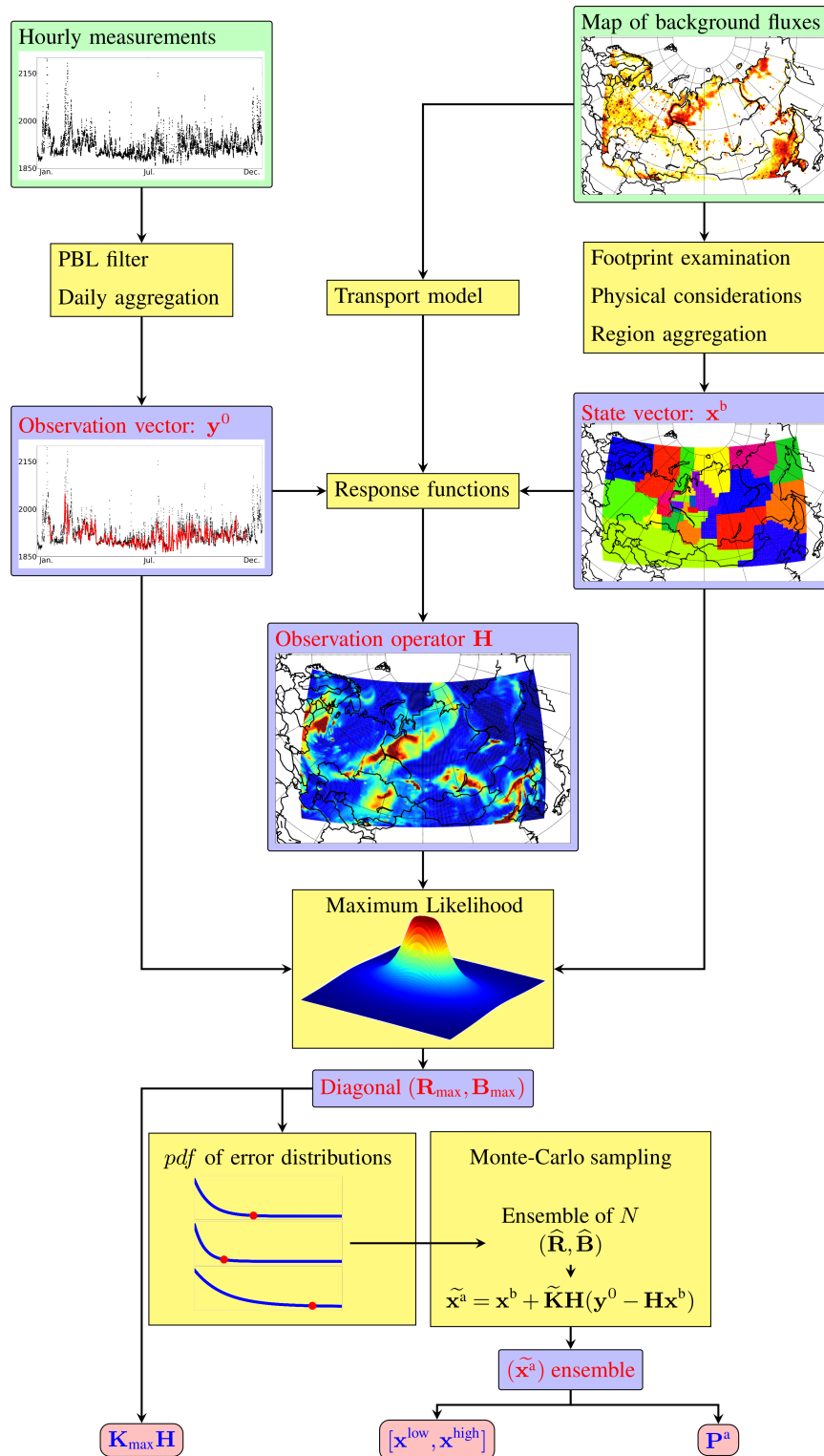


Figure VII.4: Block diagram of the method. Green boxes represent the raw inputs of the system. The blue ones are intermediary results and red ones the outputs to be interpreted. The yellow ones depict the algorithms to compute. Details in Sect. 2.2 and 2.3. Insights for output analyses are given in Sect. 2.4.2.

in a problem defined with a reduced dimension from the full-dimension problem.

2.3.1 Principle for problem reduction

2.3.1.1 Motivations and definition In the observation space, more and more surface observation sites nowadays provide quasi-continuous measurements (at least a few data points per minute in the data set we use; Sasakawa et al., 2010; Winderlich et al., 2010). For long windows of inversion at the regional scale (of a few weeks or months), such a frequency of acquisition generates a number of data points technically impossible to assimilate all together in our framework. Concerning the fluxes, one shall aim at a characterization of the fluxes on very fine pixels and at a high temporal resolution. As the window of inversion lengthens and the domain widens, the number of flux unknowns grows dramatically.

In the inversion framework, the most straightforward way of minimizing the dimension of a problem is to reduce the dimensions of the observation and state spaces. Aggregating components of the state space and sampling observations are classically used for this purpose. In most studies, the reduction of the problem is carried out arbitrarily. However, aggregation can generate large errors (Kaminski et al., 2001; Bocquet et al., 2011), which would mitigate the benefits of the Monte Carlo marginalized approach compared to more classical ones applied in other atmospheric inversion studies with no aggregation (e.g., variational inversions; Courtier et al., 1994; Bergamaschi et al., 2005; Pison et al., 2009). Here, we propose a more objective way to do so following recent literature.

Using the formalism from Bocquet et al. (2011), we aim at defining a representation ω that encompasses the horizontal and temporal resolution of the fluxes, the choice of the regions of aggregation and the temporal sampling of the observations. The representation ω is defined through two operators Γ_ω and Λ_ω , which projects respectively the full-resolution state and observation space to smaller ones. After the state space “projection” with Γ_ω , the inversion applies corrections on regions of aggregation with fixed emission distributions, instead of on single pixels. The adjoint of this operator, Γ_ω^T , then redistributes total emissions on finer scales with the same fixed emission distribution. The choice of Γ_ω impacts both the state vector \mathbf{x} and the observation operator \mathbf{H} . The observation sampling Λ_ω can consist in averaging or picking one value per time step (chosen accordingly to the physical resolution inquired into). For instance, one can decide to average the observations by day in order to study the synoptic variability of the atmosphere, related to the fluxes at the meso-scale. The observation sampling applies to both the observation vector \mathbf{y}^o and the observation operator \mathbf{H} . The observation operator \mathbf{H} computes the contribution from single sources to single observations. The adjoint of the observation sampling, Λ_ω^T , will then redistribute an average or a sample for each chosen time steps along this time step. The redistribution will follow the raw observed temporal profile within the processed time step.

2.3.1.2 Mathematical formulation At first glance, choosing the aggregation pattern and the sampling protocol can be considered as two independent physical problems. However, as they both influence the dimension of the observation operator \mathbf{H} , they cannot be fixed separately. More explicitly, we can derive a formula, which links Γ_ω and Λ_ω . Indeed, our final objective is to compute total posterior fluxes for each aggregated region that are as close as possible to the posterior fluxes from a full-resolution inversion aggregated afterwards. That is to say, we want to confine the norm of $\mathbf{x}_\omega^a - \Gamma_\omega \mathbf{x}_t^a$ with \mathbf{x}_ω^a the posterior state vector resolved

in the representation ω and \mathbf{x}_t^a the posterior state vector computed with a full-resolution representation of the problem. Algebraic manipulations lead to:

$$\mathbf{x}_\omega^a - \Gamma_\omega \mathbf{x}_t^a = \Gamma_\omega \mathbf{B} \mathcal{E}_\omega (\mathbf{y}^o - \mathbf{H} \mathbf{x}^b) \quad (\text{VII.6})$$

where:

$$\mathcal{E}_\omega = \mathcal{P}_\omega \mathbf{H}^T \Lambda_\omega^T \mathcal{S}_\omega^{-1} \Lambda_\omega - \mathbf{H}^T \mathcal{S}^{-1}, \quad (\text{VII.7a})$$

$$\mathcal{S} = \mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T, \quad (\text{VII.7b})$$

$$\mathcal{S}_\omega = \Lambda_\omega \left\{ \mathbf{R} + \mathbf{H} (\mathcal{A}_\omega + \mathcal{P}_\omega \mathbf{B} \mathcal{P}_\omega) \mathbf{H}^T \right\} \Lambda_\omega^T, \quad (\text{VII.7c})$$

$$\mathcal{P}_\omega = (\Gamma_\omega)^T \Gamma_\omega, \quad (\text{VII.7d})$$

$$\mathcal{A}_\omega = (\mathbf{I} - \mathcal{P}_\omega) \mathbf{x}_t \mathbf{x}_t^T (\mathbf{I} - \mathcal{P}_\omega), \quad (\text{VII.7e})$$

$$\mathbf{x}_t = \text{the true state of the system}, \quad (\text{VII.7f})$$

$$\mathbf{I} = \text{the identity matrix}. \quad (\text{VII.7g})$$

In Eq. VII.6, \mathbf{R} and \mathbf{B} are the full-resolution matrices of the error statistics.

For the aggregation errors to be limited, \mathcal{E}_ω (Eq. VII.7a) must tend towards $\mathbf{0}$. Then, \mathcal{S} (Eq. VII.7b) and \mathcal{S}_ω (Eq. VII.7c) must be as close as possible to each other and the impact of \mathcal{P}_ω (Eq. VII.7d) and of the sandwich product with Λ_ω , $\Lambda_\omega^T(\cdot)\Lambda_\omega$, must be as small as possible. Γ_ω^T extrapolates the fluxes from the aggregated regions to a finer resolution following an a priori repartition. The matrix \mathcal{P}_ω then redistributes the fluxes over a region with respect to the prior repartition, but keeping the same total emissions on the region.

In Sect. 2.3.2 below, we explain how to reduce these terms. The exact estimation of Eq. VII.6 is complicated and requires extensive numerical resources (e.g., Wu et al., 2011). In the following, we use physical considerations towards minimizing Eq. VII.6. The errors that are intrinsic to the aggregation process and that are unavoidable are controlled so that the benefit from the general marginalization is not wasted. We show in Sect. 2.6.3 that the physical considerations for choosing the representation ω in our case do not depreciate the inversion results compared to what would have been obtained with the exact resolution of Eq. VII.6.

Considering the computer resources we use, all the principles we define are applied in order to limit the size of the observation space (resp. the state space) to a dimension of roughly 2000 (resp. 1500). For instance, in the meso-scale Eurasian case study described in Sect. 2.5, these considerations lead to the aggregation patterns displayed in Fig. VII.4 and Fig. VII.8. With this problem dimensions, the ensemble used in the Monte Carlo sampling consists of 60000 draws.

When the observation and the state space aggregation are chosen, the operator \mathbf{H} can be computed with the so-called 'response functions', based on forward simulations of the transport for each state component (Bousquet et al., 1999a).

2.3.2 Representation choice

2.3.2.1 Observation space sampling The sandwich product with Λ_ω , $\Lambda_\omega^T(\cdot)\Lambda_\omega$, aggregates the errors in the observation space and diffuses them back within each aggregate along

a prescribed temporal profile. For example, it can typically compute the average error per day; then it allocates for each sub-daily dimension an error proportional to the contribution of the related component of y^o to the daily mean. However, a daily averaging would be dominated by the outliers (e.g, singular spikes or night-time observations when the emissions remain confined close to the surface due to weak vertical mixing) that are generally associated to very high observation errors (due to fine scale mis-representations of the transport and erroneous night vertical mixing). For this reason, we decide to define Λ_ω as the sampling operator, which, for each day and observation site, picks the component of the observation vector when the daily minimum of concentrations within a planetary boundary layer higher than 500 m is observed. Below this threshold, the vertical mixing by the model is known to be possibly erroneous (e.g., Berchet et al., 2013b). The daily resolution is chosen in order to keep a representation of the transport relevant to the meso-scale expectations on flux characterization. Higher time resolution would not improve the inversion efficiency due to strong within-day temporal correlations of errors (Berchet et al., 2013b).

2.3.2.2 Observational constraints One can notice that far from the observational constraints, the atmospheric dispersion (depicted by the sandwich product with \mathbf{H} , $\mathbf{H}(\cdot)\mathbf{H}^T$) makes the potential errors negligible compared to the errors generated in the areas close to the stations. Indeed, gathering two close hot spots of emissions thousands of km away from the observation sites is not problematic since the air masses coming from the two spots will be well mixed. On the opposite, two hot spots that are as distant from each other as the first two, but close to an observation site, will generate plume-like air masses with a very high sensitivity to the errors of mixing and transport in the model. We use an estimation of the observation network footprints (approximating \mathbf{H}^T) in order to fix the typical regions constrained by the network and avoid unfortunate grouping. At this step, approximate footprints are preferred to the heavy computation of the complete \mathbf{H}^T and are sufficient for our physical considerations. Within the constrained regions, we use a small spatial resolution for the fluxes and the transport and fine aggregation patterns; outside of them, we choose a coarse resolution and large aggregation patterns. These guidelines for using footprints prior to an inversion can be applied more systematically, as what is done in Thompson and Stohl (2014). An illustration of aggregation patterns in our case study can be looked at in Fig. VII.8.

2.3.2.3 Flux aggregation Some terms in Eq. VII.6 are directly related to the aggregation of the fluxes. The term $\mathbf{H}\Lambda_\omega\mathbf{H}^T$ in Eq. VII.7c depicts the aggregation errors coming from the uncertain distribution and temporal profile of the fluxes within each aggregation region, then transported to the observation sites. It must be close to $\mathbf{0}$. In our application below, this is particularly important for hot spots of emissions the location of which is poorly known. The term $\mathbf{H}\mathcal{P}_\omega\mathbf{B}\mathcal{P}_\omega\mathbf{H}^T$ in Eq. VII.7c must be as close as possible to $\mathbf{H}\mathbf{B}\mathbf{H}^T$. The factors of divergence between these two terms come from the areas that are not well constrained by the observations. If, within a region of aggregation, a part is upwind the observation sites, while the other is not seen, then the aggregation errors will scatter on the unseen part of the region. The main sources of errors can then be separated into two different types: 1) the resolution/representation type, and 2) the constraint type.

The type-1 errors are mainly related to the resolution of the observation operator. The models consider that the fluxes and the simulated atmospheric mixing ratios are uniform on a sub-grid basis and neglects sub-grid processes. This discretization contributes to type-

1 errors, as 'representation' errors (Tolk et al., 2008). Additionally, the distribution within each aggregation region is fixed and sub-region re-scaling are forbidden. The fine resolution close to the observation network is bound to confine type-1 errors (e.g., Wu et al., 2011). Additionally, the representation error is critical for co-located emissions, especially when the typical temporal and spatial scales of these emissions are different. For instance, grouping hot spots from oil extraction emissions with widespread wetland emissions that quickly vary in time is hazardous. We then aggregate the emissions along their typical time and space scale, hence according to the underlying physical process. An in-depth analysis of the footprints and the small patterns of aggregation close to the observation sites should limit the type-2 constraint errors. Area under high observational constraints should not be grouped with under-constrained areas.

The resolution and aggregation choices can be computed objectively, but at a very high cost and only within a framework of prescribed frozen error matrices (Bocquet, 2009; Wu et al., 2011; Koohkan et al., 2013). For our purpose, we cannot afford such computation costs and rely on heuristic choices: small resolution and aggregation patterns close to the observation sites, aggregation by type of emission, separation of constrained/under-constrained areas by analysing the footprints. These non-optimal subjective choices may damp the efficiency of our method and must be carried out cautiously. Nevertheless, in our case, afterwards checking shows that our choices did not have critical impact on the inversion results in our case.

2.3.3 Numerical artefacts

In addition to the need of defining a well-sized problem, smart adaptations can be applied to the computation of the method in order to enhance the efficiency of the algorithm. We face several sources of numerical artefacts in the computation of the method. In the quasi-Newtonian maximum likelihood algorithm, numerical artefacts are generated by the under-constrained regions. After a few steps, the computed gradient of the likelihood is dominated by these regions and the algorithm stays stationary. This issue could be partly related to the under-optimality of the chosen representation ω as suggested by the optimality criteria described in Bocquet et al. (2011). The stagnation of the maximum likelihood algorithm could then be used to detect too small regions of aggregation.

The under-constrained regions perturbing the maximum likelihood algorithm can be diagnosed using the diagonal terms of the influence matrix \mathbf{KH} (with \mathbf{K} defined in Eq. VII.2 and following Cardinali et al., 2004). This matrix represents the sensitivity of the inversion to elementary changes in the observations. Strong observation constraints are related to high sensitivity. After stagnation, the regions with a diagnosed $\mathbf{KH} < 0.5$ are flagged out and the algorithm is carried on. This way, only the sufficiently constrained components of the state vector are processed until the algorithm converges. A third to half of the regions are flagged out this way in our case study.

The detection of the mis-representation of hot-spot plumes should also be enhanced. Despite the minimum daily sampling and the fine resolution close to the observation network, the plume issue can still generate strong temporal and spatial mismatches. For example, a point source can influence a station in the real world, but not in the model because it has been mis-located, and conversely. This creates significant differences between the simulated

and the observed concentrations. The maximum likelihood algorithm attributes such mismatches to prior errors and/or observation errors. High diagnosed errors in the maximum likelihood algorithm are then a criterion for plausible mismatches. We know such plumes must be flagged out from the inversion to avoid irrelevant high influence from very local sources hard to represent. Since we notice that the observation and prior computed errors seems to follow a Fischer-Snedecor distribution, we choose to flag out the observations that are within the 95% tail of the distribution.

2.4 Validation experiments

In Sect. 2.2, we described our modified atmospheric inversion by marginalization. In Sect. 2.3, we proposed some essential rules to follow in order to properly define the problem, so that the rather simple theoretical framework is not hindered by finite numerical resources. The marginalization method has to be validated along objective criteria. In the following, we summarize the general structure of the method in order to identify the critical points to test in the method (Sect. 2.4.1). We deduce from these points some OSSEs to carry out. In Sect. 2.4.2, we define the scores according to which the method will be evaluated.

2.4.1 Required tests

2.4.1.1 Method summary The method described in Sect. 2.2 and 2.3 is condensed in the block diagram in Fig. VII.4. The marginalized inversion takes the same input as any other atmospheric inversion: some atmospheric measurements and prior maps of fluxes with specified resolution and temporal profiles. In Sect. 2.3, we gave recommendations on the processing of the 'raw' inputs, so we get an observation vector \mathbf{y}^o , a prior state vector \mathbf{x}^b and an observation operator \mathbf{H} that are small enough to be computable by the method. These highlights are mainly the sampling of the observations per day (in accordance with our objective of characterizing meso-scale fluxes in our case study) and the aggregation of the fluxes by regions (based on physical considerations and footprint analysis). The maximum likelihood algorithm processes \mathbf{y}^o , \mathbf{x}^b and \mathbf{H} in order to find a couple of optimal diagonal error matrices $(\mathbf{R}_{\max}, \mathbf{B}_{\max})$. This maximum likelihood is found by a quasi-Newtonian descent method. We then infer from $(\mathbf{R}_{\max}, \mathbf{B}_{\max})$ the approximate χ^2 shape of the distribution of all the possible error matrices (\mathbf{R}, \mathbf{B}) . We carry out a Monte Carlo sampling on these distributions of errors and get an ensemble of posterior state vectors $(\widehat{\mathbf{x}}^a)$. The processing of this ensemble provides the final output of the method: a tolerance interval TI_{68} of the posterior state and the posterior correlations between the components of the state space. The method also allows the explicit computation of the influence matrix $\mathbf{K}_{\max}\mathbf{H}$ in order to analyse the constrained regions of emissions only.

To summarize, the marginalized inversion processes two vectors and one operator: \mathbf{y}^o , \mathbf{x}^b , and \mathbf{H} , as any other atmospheric inversion. The main difference with most other atmospheric inversions resides into the objective and automatic computation of the influence of ill-specified error statistics, in contrast with the traditional assigning of frozen error matrices based on expert knowledge and with the more recent online computations of error hyperparameters. Thus, we do not have to inquire into the sensitivity of our method to the prescribed spatial correlations of flux errors, or to the error variances. Such a sensitivity is transposed to the choice of the aggregation patterns and the sampling protocol, as defined in Sect. 2.3.1. The chosen configuration of aggregation and the sampling protocol are checked afterwards

Table VII.1: OSSEs summary. Three parameters of the inversion (sub-total masses emitted per regions, emission distribution and transport) can be perturbed compared with the 'truth'. The seven possible combinations are depicted by = and \neq signs for each parameter and each OSSE. Every OSSE is evaluated along the scores defined in Sect. 2.4.2. The scores are given in % for the best correlation threshold for grouping the state space components as presented in Sect. 2.4.2. The influence score must be as close to 100% as possible. The other two scores must be as small as possible. The regions are grouped along a correlation criterion r_{\max} (see Sect. 2.4.2); we present the scores only for r_{\max} with the best results. For OSSE 7, the scores are zeros for the fossil fuel regions because most of these regions were filtered out. The few remaining ones are very well constrained.

	OSSE 1		OSSE 2		OSSE 3		OSSE 4		OSSE 5		OSSE 6		OSSE 7	
Inversion inputs:														
x sub-totals	\neq		=		=		\neq		\neq		=		\neq	
x distributions	=		\neq		=		\neq		=		\neq		\neq	
H	=		=		\neq		=		\neq		\neq		\neq	
Optimal r_{\max}	0.5		0.5		0.5		0.5		0.6		0.5		0.4	
Scores:	ff	wet	ff	wet	ff	wet	ff	wet	ff	wet	ff	wet	ff	wet
Relative score	79	94	16	27	40	84	3	66	30	117	20	93	0	112
Absolute score	9	16	2	11	36	24	1	27	18	40	37	30	0	15
Influence	63	56	39	37	45	30	37	28	46	58	32	32	13	33

to be relevant in our case study. OSSEs are then to be carried out to evaluate the sensitivity of the method to y^o , x^b , **H**.

2.4.1.2 Test strategy We assume that, in our case, the method is not sensitive to errors in y^o . Indeed, in all the following, we consider that the measurement errors are negligible compared to transport errors; this is true for surface sites that fulfil the World Meteorological Organisation strict recommendations for accuracy and precision (WMO/GAW, 2011). This approximation does not hold for satellite total columns measurements, for which the transport errors are smoothed over the vertical atmospheric column and the instrument errors are larger. In addition, representativeness errors may also impact y^o . OSSEs should account for these errors. However, OSSEs may face difficulties in explicitly highlighting these errors. Therefore, we do not perturb y^o in order to represent the instrumental uncertainties and representativeness errors in the OSSEs.

The OSSEs are then based on perturbations of x^b and **H**. The discrepancies between the background x^b and the 'truth' x^t are of two types: 1) the erroneous distribution and temporal profile of the fluxes within aggregation regions, and 2) incorrect total emissions by region. For example, in Eurasia, the maps of the distribution of the wetlands differ drastically from one database to another (Frey and Smith, 2007). Apart from the distribution, the amount of gas emitted by each process is uncertain, due to mis-parametrizations or, for anthropogenic emissions, mis-specified activity maps (e.g., Rypdal and Winiwarter, 2001). The transport **H** differs from the 'true' transport mainly because of the resolution of the model, the parametrization of sub-grid processes (such as vertical turbulent mixing in the planetary boundary layer or deep convection), and the meteorological forcing fields (which are not necessarily optimized for transport applications).

The main sources of errors in the inversion are then: 1) a wrong representation of the transport (highly dependent of the transport model used, its resolution, its parametrization and the exactitude of forcing wind fields), 2) an erroneous distribution of the fluxes within aggregation regions (each inventory and database has different statistical methods and parameters to reproduce surface fluxes), and 3) incorrect total emissions by regions. In order to evaluate the impact of each point on the inversion result, we carry out OSSEs with perfect synthetic observations from a nature run (i.e. with 'true' emissions and 'true' transport, as defined in the set-up in Sect. 2.5). We test the ability of the marginalized inversion to reproduce the 'true' fluxes or, at least, to consistently include the 'truth' within the tolerance intervals. There are eight possible combination of correct or perturbed phases of the 3 parameters. The 'all true' combination is not relevant: then $\mathbf{y}^o - \mathbf{H}\mathbf{x}^b = 0$ and the maximum likelihood algorithm is stationary. Seven combinations remain, detailed in Tab. VII.1. We run the marginalized inversion for the seven OSSEs and evaluate them along the scores defined in Sect. 2.4.2 below.

2.4.2 OSSE evaluation

2.4.2.1 Scoring system We expect an atmospheric inversion to provide reliable ranges of uncertainties for surface fluxes. That is to say, for as many components of the state vector \mathbf{x}_i as possible, the 'truth' \mathbf{x}_i^t should be within the tolerance interval TI_{68} , $[\mathbf{x}_i^{\text{low}}, \mathbf{x}_i^{\text{high}}]$ (defined in Sect. 2.2). In order to evaluate the ability of producing consistent fluxes, we define a relative score \mathbf{z}_{rel} for each component of the state vector: $(\mathbf{z}_{\text{rel}})_i = 2 \frac{|\mathbf{x}_i^a - \mathbf{x}_i^t|}{\mathbf{x}_i^{\text{high}} - \mathbf{x}_i^{\text{low}}}$. Hereafter, all the scores will be expressed in % for better readability. Statistically, \mathbf{z}_{rel} has no upper bound. Relative scores bigger than 100% are not statistically inconsistent, but, for the method to be validated, we expect that the proportion of state components with $\mathbf{z}_{\text{rel}} < 100\%$ is dominant.

Furthermore, the atmospheric inversion is supposed to reveal pieces of information to the understanding of the system. Then, we also expect that a correct relative score below 100% is not reached by specifying huge tolerance intervals. To evaluate the ability of the marginalization of getting close to the reality, i.e. providing valuable information on the state of the system, we define an absolute score \mathbf{z}_{abs} : $(\mathbf{z}_{\text{abs}})_i = \left| \frac{\mathbf{x}_i^a}{\mathbf{x}_i^t} - 1 \right|$. The smaller the absolute score, the more accurate the marginalized inversion.

An inversion also must be able to evaluate the observation constraints on the regions. An objective estimator of the constraints on the regions is the influence matrix \mathbf{KH} defined in Sect. 2.3. The Kalman gain matrix depends on the couple (\mathbf{R}, \mathbf{B}) . Amongst all the Monte Carlo draws, we compute the influence matrix $\mathbf{K}_{\text{max}}\mathbf{H}$ for the couple associated to the maximum likelihood. The diagonal terms of this matrix range from 0 to 1. They give for all components of the state space the constraint given by the observations. We then define the influence score: $(\mathbf{z}_{\text{infl}})_i = (\mathbf{K}_{\text{max}}\mathbf{H})_i$. The closest to 100% these terms, the more constraints the inversion provides. We can then deduce the typical range of influence of the observation sites and detect the blind spots of the used network.

For each component i of the state space, we then have defined 3 indicators:

$$\begin{cases} (\mathbf{z}_{\text{rel}})_i &= 2 \frac{|\mathbf{x}_i^a - \mathbf{x}_i^t|}{\mathbf{x}_i^{\text{high}} - \mathbf{x}_i^{\text{low}}} \\ (\mathbf{z}_{\text{abs}})_i &= \left| \frac{\mathbf{x}_i^a}{\mathbf{x}_i^t} - 1 \right| \\ (\mathbf{z}_{\text{infl}})_i &= (\mathbf{K}_{\text{max}} \mathbf{H})_i \end{cases} \quad (\text{VII.8})$$

2.4.2.2 Posterior correlation processing Another point most inversions do not compute explicitly and objectively is the typical temporal and spatial scales the inversion can differentiate in the fluxes, considering the atmospheric transport and the density of the observations. Our marginalized inversion gives access to an explicit matrix of correlations as defined in Sect. 2.3. Strong positive and negative correlations between two components of the state space indicate that the inversion cannot separate the contributions from the two components. For example, air masses observed at a station and coming from two regions upwind the station will have a mixed atmospheric signal difficult to analyse. Co-located emissions are also not necessarily differentiated in the atmospheric signal. Moreover, in a regional framework, when a model of limited area is coupled to lateral boundary conditions (LBC), the inversion must explicitly alert on the regions that cannot be separated from the boundary conditions, i.e. from the baseline signal.

In the case of strong correlations between two components of the state space in the posterior covariance matrix, we consider that it is not relevant to try to infer specific information for the two separate components. Then, we group the state space components according to their posterior correlations. We define a threshold of correlation r_{max} and associate couples of regions (i, j) within groups such that $|r_{i,j}| > r_{\text{max}}$. If we prescribe $r_{\text{max}} = 0$, all the regions will be grouped; on the opposite, if $r_{\text{max}} = 1$, no group will be formed. The optimal correlation threshold is not evident. We test the grouping for all possible r_{max} . We flag out from the processing of the results all the groups, which include some contributions from the LBC. Thus, from this post processing, we only keep the regions that are clearly constrained by the observation sites, with no interference from the LBC. Moreover, we can infer the spatial and temporal scale that the inversion can resolve from the grouping patterns.

In Tab. VII.1, the three scores defined in Eq. VII.8 are averaged on the whole domain of interest for the optimal correlation threshold r_{max} (as discussed in Sect. 2.6.1).

2.5 Set up of the OSSEs

We compute the OSSEs that we described in Sect. 2.4 in a realistic meso-scale case. We focus on a domain spanning over Eurasia, from Scandinavia to Korea. At this scale, the air masses residence time is typically of days to a few weeks. This time scale is small compared to the lifetime of methane of 8-10 years in the atmosphere (mainly due to oxidation by OH radicals; Dentener et al., 2003). Hence, the observation operator can be considered linear. We apply the method on a region characterized by significant fluxes, with collocation of different sources with different emission time-scales: Siberia. We describe the region of interest and the chosen 'truth' for the experiments in Sect. 2.5.1. We use two transport models in order to simulate atmospheric transport. The technical details on these models are summarized in Sect. 2.5.2. In Sect. 2.5.3, we explain how we choose and compute the synthetic observations for our experiments.

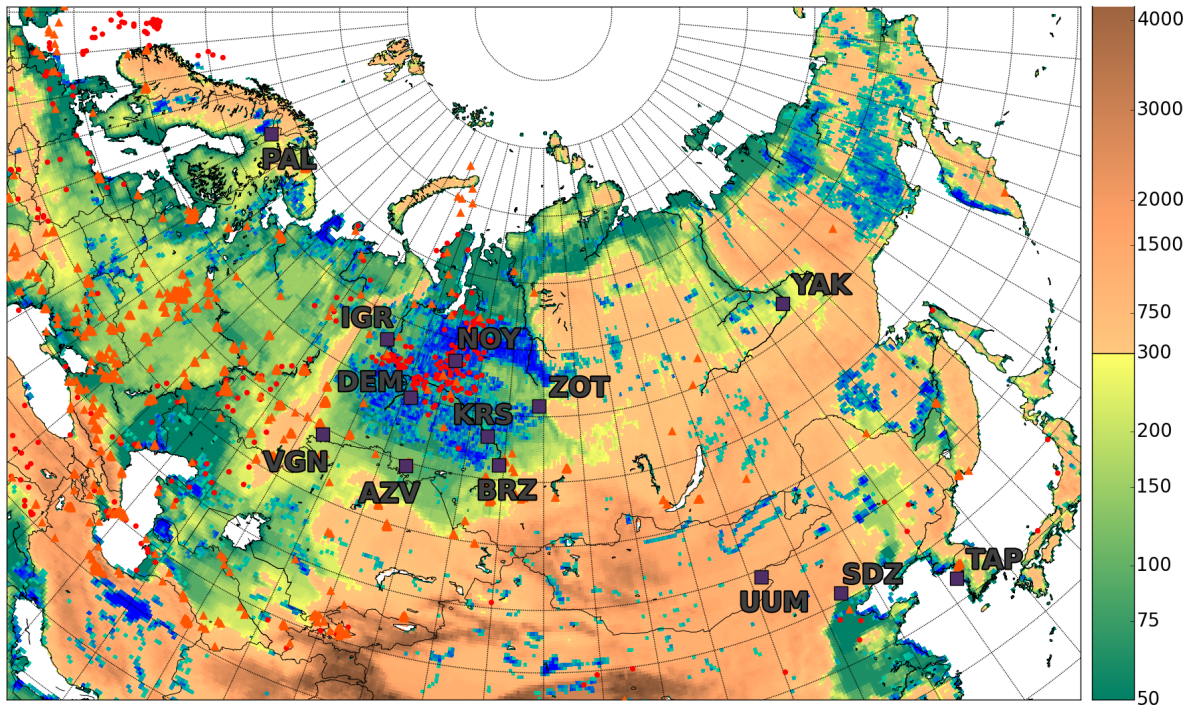


Figure VII.5: Topographic map of the domain of interest. The colorbar shows the altitude above sea level (from ETOPO1 database; Amante and Eakins, 2009). Red dots (resp. orange triangle) depicts hot spots of CH_4 emissions (based on EDGAR v4.2 inventory; see Sect. 2.5.1) related to oil welling and refineries (resp. gas extraction and leaks during distribution in population centers). Purple squares highlight the observation site localization. Blueish shaded areas represent average inundated regions, wetlands and peatlands (from the Global Lakes and Wetlands Database; Lehner and Döll, 2004)

2.5.1 Virtual true state x^t

2.5.1.1 State space components In the region of interest, the emissions of methane are dominated by wetland, anthropogenic (here, mainly related to the oil and gas industry) and wildfire emissions. In Fig. VII.5, the distributions of the wetlands and of the oil and gas industry in the region are displayed. Anthropogenic emissions of methane in the region are mainly hot spots related to the intense oil and gas industry in the Siberian Lowlands and to the leaks in the distribution system in population centres in the South part of Siberia. Wetland emissions are mainly confined in the lower part of Siberia in the West Siberian plain, half of which is lower than 100 meters above sea level.

The spatial distribution of the associated fluxes is deduced from: 1) EDGAR database v4.2 (<http://edgar.jrc.ec.europa.eu>) for year 2008 for anthropogenic emissions, 2) LPX-Bern v1.2 process model at a monthly scale for wetland emissions (Spahni et al., 2011), 3) GFED database at daily scale for wildfires (Giglio et al., 2009). The EDGAR inventory uses economic activity maps by sectors and convolves them with emission factors estimated in laboratories or with statistical studies (Olivier et al., 2005). LPX-Bern is an update of process model LPJ-Bern (Spahni et al., 2011). It includes a dynamical simulation of inundated wetland areas, dynamic nitrogen cycle, and dynamic evolution of peatlands (Spahni et al., 2013; Ringeval

et al., 2014). The model uses CRU TS 3.21 input data (temperature, precipitation rates, cloud cover, wetdays) and observed atmospheric CO₂ for each year for plant fertilization. GFED v4 is built from burnt area satellite product (MCD64A1). CH₄ emissions at monthly and daily scales are deduced from the burnt areas using the Carnegie-Ames-Stanford-Approach (CASA model; Potter et al., 1993) and emission factors (van der Werf et al., 2010). Wildfire emissions can be very strong and are punctual in time and space; they are then difficult to analyse by the inversion. Wildfires are included as inputs to the marginalized inversion, but are automatically filtered out during the computation. In all the following, we evaluate the OSSEs only in terms of anthropogenic and wetland emissions.

In addition, at the meso-scale, we use a CTM (see Sect. 2.5.2.2) with a limited area domain. Initial and lateral boundary conditions (IC and LBC) are then also to be optimized in the system to correct the atmospheric inflow in the domain. Lateral concentrations are deduced from simulations at the global scale by the general circulation model LMDz with the assimilation of surface observations outside the domain of interest (Bousquet et al., 2006). We aggregate the LBC along 4 horizontal components and 2 vertical ones (1013-600 hPa and 600-300 hPa).

2.5.1.2 Generation of a perturbed reference state \mathbf{x}^t The EDGAR fluxes are given at the yearly scale and the LPX fluxes are calculated at a monthly scale. Additionally, LPX monthly fluxes exhibit smoothed patterns while wetland emissions can vary drastically from a point to another. We want the nature run for OSSEs to reproduce the potential spatial and temporal variability of the emissions. To do so, we intensify the spatial and temporal contrasts from the databases to the nature run. We then compute the 'true' state vector \mathbf{x}^t by perturbing EDGAR emissions on a monthly basis and LPX on a 10-day basis. That is to say: $\mathbf{x}^t = \alpha \otimes \mathbf{x}^{\text{data}}$, with the vector α depicting the scaling factors by state space component, \otimes the point-wise multiplication operator and \mathbf{x}^{data} the emissions from the databases. The perturbations in α from original EDGAR and LPX databases applied to get the 'truth' are scaling factors up to 10. These scaling factors could have been chosen randomly, but we prefer inferring them with a raw expert-knowledge-based inversion using real data. The purpose of using real data for computing \mathbf{x}^t is to generate potentially realistic variations within the state space.

For both anthropogenic and wetland emissions, the scaling factors can significantly differ from a period of inversion to another. We can then evaluate the ability of the marginalized inversion to catch quick variations. The distribution of the scaling factors α is shown in Fig. VII.6. These factors are plausible, knowing the uncertainties on the wetland emissions and gas leakage (e.g., Reshetnikov et al., 2000). Such target scaling factors are at the edge of the validity of the Gaussian assumption and of the positivity of methane fluxes. The ability of the marginalization to recover such correction factors will prove its robustness.

As for anthropogenic and wetland emissions, we apply the scaling factors α on the components of \mathbf{x}^t related to LBC by periods of 10 days.

The OSSEs rely on \mathbf{x}^b perturbed from \mathbf{x}^t , or not. We apply two types of perturbations as summarized in Tab. VII.1. In OSSE 1, 4, 5 and 7, we only implement prior fluxes with different total emissions on the regions of aggregation. We take the emissions of the raw inventories as background to test the ability of recovering 'true' fluxes from realistic background fluxes without assigning frozen prior errors. In OSSE 2, 4, 6 and 7, the distribution of the prior

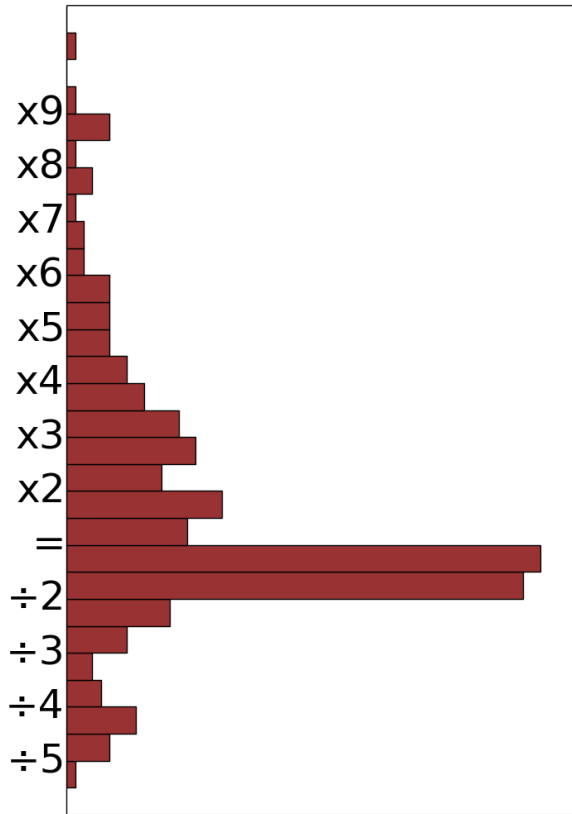


Figure VII.6: Distribution of the scaling factors applied to the emission databases in order to compute the 'truth'. All the emission components of the state vector have been included in the histogram. The selection of the scaling factor distribution is detailed in Sect. 2.5.1.

fluxes is modified from the 'truth'. We choose all flat flux distributions for each region of aggregation as prior fluxes.

2.5.2 Simulation of the observation operator \mathbf{H}

The observation operator \mathbf{H} is deduced from simulations of atmospheric transport. We use two different transport models in order to evaluate the impact of the transport on the inversion. We define $\mathbf{H}_{\text{FLEXPART}}$ with the Lagrangian dispersion model FLEXPART and $\mathbf{H}_{\text{CHIMERE}}$ with the Eulerian Chemistry-Transport Model CHIMERE. Any transport model can be considered at some point biased compared with the reality. Confronting the results from FLEXPART to those from CHIMERE will allow us to test the robustness of our method to the biases.

2.5.2.1 The Lagrangian model: FLEXPART

With the Lagrangian dispersion model FLEXPART (Stohl et al., 2005), we can compute the footprints of the observations, hence $\mathbf{H}_{\text{FLEXPART}}^T$. We use FLEXPART version 8.2.3 to compute numerous back-trajectories of virtual particles from the observation sites. The model is forced by the European Centre for Medium-range Weather Forecast (ECMWF) ERA-Interim data at an horizontal resolution of $1^\circ \times 1^\circ$, with 60 vertical levels and 3 hour temporal resolution (Uppala et al., 2005). Virtual particles are released in a 3D box centred around each observation site with a 10-day lifetime backwards

in time. The footprints are computed on a $0.5^\circ \times 0.5^\circ$ horizontal grid, following the method of Lin et al. (2003), taking the boundary layer height at each particle location into account. The footprints only have to be convolved with the emission fields in order to get simulated concentrations at the observation sites. The method for computing the footprints considers that only the particles within the boundary layer are influenced by surface emissions and that the boundary layer is well-enough mixed to be considered as uniform. The uniform vertical mixing of the mixing layer can generate a bias on the surface simulated concentrations. Such a bias is critical in the classical inversion framework and consequently in the one we describe since all the uncertainties are considered unbiased.

FLEXPART can easily compute an estimation of the adjoint of the full-resolution observation operator before choosing the representation ω . Hence, despite the expectable biases, we use this model to estimate the footprints of the network and deduce the aggregation patterns needed to compute $\mathbf{H}_{\text{CHIMERE}}$. This same model FLEXPART may also be used to compute explicitly and rigorously the representation ω according to objective criteria (Koohkan and Bocquet, 2012).

2.5.2.2 The Eulerian model: CHIMERE Using the Eulerian meso-scale chemistry transport model CHIMERE (Vautard et al., 2001; Menut et al., 2013) constrained by non-hydrostatic meteorological fields, we explicitly define the observation operator $\mathbf{H}_{\text{CHIMERE}}$ by computing the forward atmospheric transport from the emission aggregated regions (defined according to Sect. 2.3 criteria) to the observation sites. This model was developed in a framework of air quality simulations (Schmidt et al., 2001a; Pison et al., 2007), but is also used for greenhouse gas studies (Broquet et al., 2011; Berchet et al., 2013b). We use a quasi-regular horizontal grid zoomed near the observation sites after the considerations of Sect. 2.3. The domain of interest is of limited area and spans over the mainland of the Eurasian continent (see Fig. VII.5). The average side length of the grid cells near the stations is 25 km, while it spans over 150 km away from the observation sites. The 3D-domain roughly embraces all the troposphere, from the surface to 300 hPa (~ 9000 m), with 29 layers geometrically spaced. The model time step varies dynamically from 4 to 6 min depending on the maximum wind speed in the domain. The model is an off-line model which needs meteorological fields as forcing. The forcing fields are deduced from interpolated meteorological fields from ECMWF with a horizontal resolution of $0.5^\circ \times 0.5^\circ$ every 3 h.

2.5.3 Synthetic observations y^o

We compute the nature run, i.e., the synthetic observations, from the 'true' state vector, with the CTM CHIMERE. That is to say, in all the following, we consider that: $y^o = \mathbf{H}_{\text{CHIMERE}}\mathbf{x}^t$. The site and date of available observations are chosen according to the operated observation sites in the region. Thirteen Eurasian surface sites have been selected. These sites are maintained by NIES (Tsukuba, Japan; Sasakawa et al., 2010), IAO (Tomsk, Russian Federation), MPI (Iena, Germany; Winderlich et al., 2010), NOAA-ESRL (Boulder, United States of America; Dlugokencky et al., 2009), and KMA (Seoul, Korea). The description of the sites is given in Tab. VIII.1. The observation sites do not carry out measurements all the year round due to logistical issues and instrument dysfunctions. In order to reproduce this sampling bias, we generate synthetic observations only when real measurements are available from January to December 2010.

Table VII.2: Eurasian site characteristics (Sect. 2.5.3). The altitudes of the sites are given as m above sea level (a.s.l.) and the inlet height is in m above ground level (a.g.l.).

Station	ID	Location			Inlet height (m a.g.l.)
		Lon (° E)	Lat (° N)	Alt (m a.s.l)	
Azovo	AZV	73.03	54.71	100	50
Berezorechka	BRZ	84.33	56.15	150	80
Demyanskoe	DEM	70.87	59.79	75	63
Igrim	IGR	64.42	63.19	25	47
Karasevoe	KRS	82.42	58.25	50	67
Noyabrsk	NOY	75.78	63.43	100	43
Pallas	PAL	24.12	67.97	560	5
Shangdianzi	SDZ	117.12	40.65	287	0
Tae-ahn Peninsula	TAP	126.12	36.72	20	0
Ulaan Uul	UUM	11.08	44.45	914	0
Vaganovo	VGN	62.32	54.50	200	85
Yakutsk	YAK	129.36	62.09	210	77
Zotino	ZOT	89.35	60.80	104	301

2.6 Results and discussion

After the description of the set-up in Sect. 2.5, we now have a 'true' state \mathbf{x}^t and some reference observations \mathbf{y}^o . We also have two observation operators $\mathbf{H}_{\text{CHIMERE}}$ and $\mathbf{H}_{\text{FLEXPART}}$ and several possible prior fluxes \mathbf{x}^b as inputs for the marginalized inversion developed in Sect. 2.2. In order to evaluate the method, we now carry out the OSSEs described in Tab. VII.1 following the complete procedure in Fig. VII.4. In Sect. 2.6.1, we examine the average robustness of the method. Then, in Sect. 2.6.2, we explore the spatial efficiency of the marginalized inversion in our case study. In Sect. 2.6.3, we discuss the enhancement provided by our method compared to the classical Bayesian framework, despite some limitations.

2.6.1 Robustness of the method

2.6.1.1 Impact of the correlation processing The marginalization should consistently reproduce the nature run in the OSSEs or, at least, it should detect its inability in characterizing the fluxes from the given atmospheric constraints. As detailed in Sect. 2.4.2, the aggregation regions may have strong posterior correlations after the marginalized inversions. This denotes the difficulties that the inversion encounters in separating some emissions. The aggregation regions can be grouped along correlation thresholds r_{max} arbitrarily chosen in order to explicitly take the emission dipoles into account. In Fig. VII.7, we plot the profiles of the scores defined in Sect. 2.4.2 along the possible correlation thresholds r_{max} for grouping the regions. Specifying a correlation threshold r_{max} allows identifying the typical temporal and spatial scales that the inversion can separate. In the case of a limited domain CTM, the influence of the LBC and of the inside fluxes can be mis-separated. The correlations take this issue into account and the correlation threshold specifies the tolerance to such mis-separations.

For all OSSEs, the influence score \mathbf{z}_{infl} increases with r_{max} . In the correlation processing after the computation of the marginalized inversion, the threshold r_{max} depicts the degree

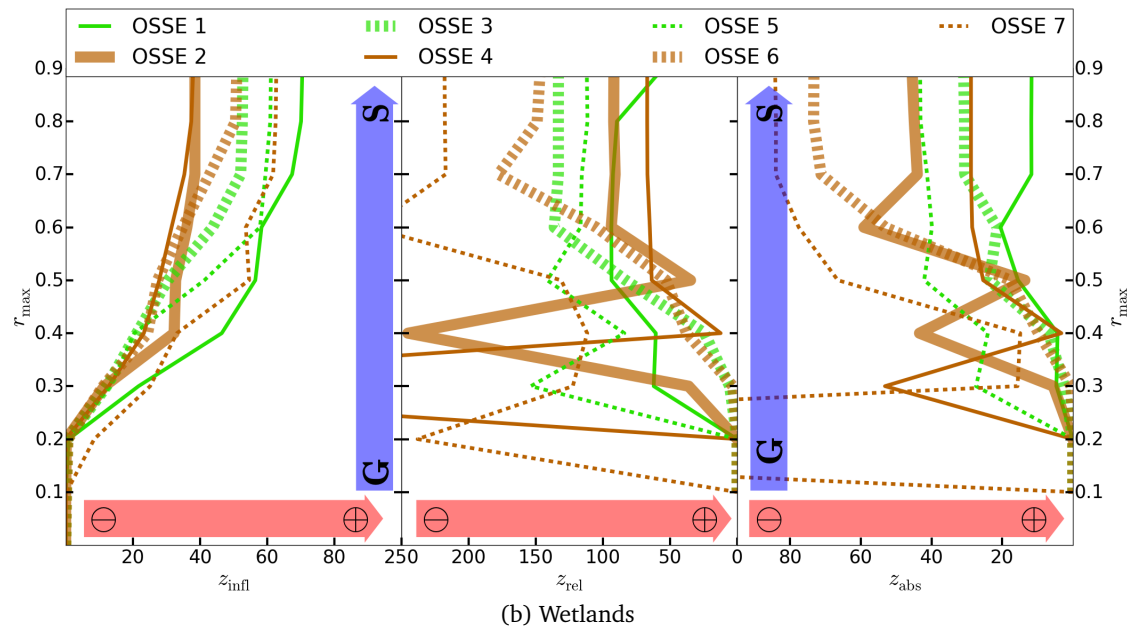
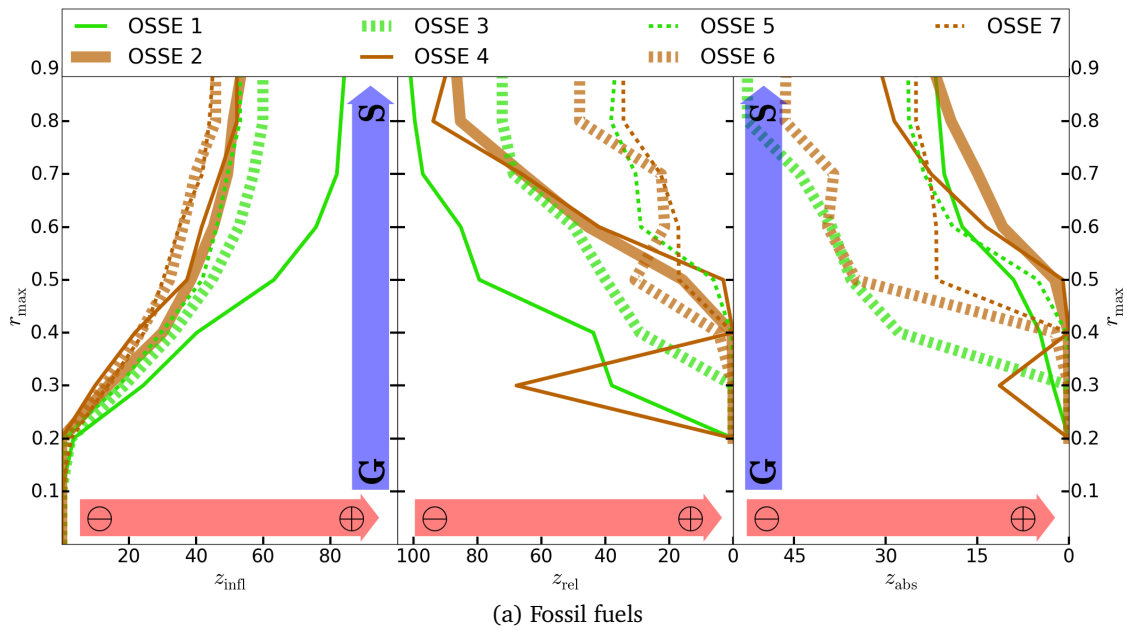


Figure VII.7: Score comparison on fossil fuel (up) and wetland (bottom) regions for all OSSEs along correlation thresholds r_{\max} of region grouping (see details in Sect. 2.4.2). (left) Influence correlation z_{infl} profile. (center) Relative score z_{rel} correlation profile. (right) Absolute score z_{abs} correlation profile. The red arrows depict the direction from lowest scores to best ones. The blue arrows denote the direction of grouping, from all grouped ('G', $r_{\max} = 0$) to all separated ('S', $r_{\max} = 1$). The OSSEs are indexed along Tab. VII.1 numerotation. Thin (resp. thick) lines stand for correct (resp. perturbed) sub-total emissions. Green (resp. brown) lines depict correct (resp. perturbed) emission distributions. Solid (resp. dotted) lines represent correct (resp. perturbed) transport. As in Sect. 2.4.2, the scores are noted in %.

of tolerance to mis-separation between inside fluxes and LBC. The higher the threshold of tolerance r_{\max} , the fewer inside fluxes are considered unseparable from the LBC. Hence, fewer aggregation regions are eliminated from the inversion and more fluxes are corrected by the inversion. As the number of constraints increases, we notice that the absolute and relative scores, \mathbf{z}_{abs} and \mathbf{z}_{rel} , also tend to increase with r_{\max} . That is to say, if we only try to get average information on big under-resolved regions, the posterior fluxes can be expected to be closer to the 'truth'. On the opposite, if we try to process too much spatial information from the inversion, we must expect more discrepancies with the 'truth'.

In particular, in Fig. VII.7, one can notice some outlier peaks for low r_{\max} . For low r_{\max} , very few regions are computed in the inversion. The peaks are created by the regions that are not any more considered as mis-separated from the LBC when r_{\max} increases. For some OSSEs, these newly computed regions have very wrong scores and dominate upon the other few computed regions. For this reason, one should be very careful in the chosen correlation threshold. In order to avoid the score instability, the optimum threshold should be chosen higher than 0.4. Above 0.5, in our meso-scale case study, as described above, the inversion is limited by the temporal and spacial variability of the fluxes to optimize and by the transport biases. Then, it can't reach the requirement of consistent reproduced fluxes.

One should find a balance between the physical scales one want to separate and the consistency of the results. In Tab. VII.1, we summarize the scores of every OSSE for a chosen correlation threshold with respect to result consistency.

2.6.1.2 Hot-spots and large-area emissions Both in Tab. VII.1 and Fig. VII.7, looking at a given correlation threshold r_{\max} , one would expect influence, relative and absolute scores that get more wrong when the inversion conditions degrades.

The fossil fuel influence score follows this trend: the more perturbed the transport and the prior fluxes are, the more state space components are considered un-invertible. The hot-spot regions of emissions are broadly filtered out and the remaining regions can be well characterized by the inversion even with wrong distribution and transport patterns. Some effects in the degrading conditions of the inversion can though compensate each other. For example, the absolute scores of OSSEs 5 and 7 are better than the one of OSSEs 3 and 6.

The situation for wetland emissions is different. These emissions are smoother than oil and gas emissions and are then not excluded because of wrong transport or distributions. For this reason, the influence score does not exhibit a clear trend with degrading inversion conditions. For wetland regions, transport seems to be the predominant factor of errors. OSSEs 3, 5, 6 and 7 do not consistently reproduce the 'truth' with relative scores higher than 100% when $r_{\max} \geq 0.4$. These discrepancies can be attributed to the very high variability prescribed in the 'true' wetland emissions. An erroneous transport will fail in detecting brutal changes of emissions at the synoptic scale. The wetland emissions should then be grouped temporally and spatially in order to average the point releases of methane.

The erroneous tolerance intervals can also be attributed to the biased transport in FLEXPART compared with CHIMERE. Since we filtered out most of the plumes with spatial and temporal mismatches with the observations, the horizontal biases in the transport are confined. Concerning the vertical bias, a wrong simulated vertical mixing in the planetary boundary will affect all the fluxes. This bias will then have an impact on the atmospheric concen-

trations that is relatively smoothed, uniform and constant. Therefore, an accurate detection of such a bias is very difficult. Any inversion relies on the unbiased assumption of the errors. The inversion will attribute the biases to the flux for wetland regions, impacting the result of the inversion. As other inversions, despite the marginalization, it appears that the results on wetland regions may be sensitive to vertical transport biases in the models (see discussion in Sect. 2.6.3.2).

Thus, the marginalized inversion seems to be sensitive to transport biases and to fluxes varying too quickly, as any other inversions. Nevertheless, a post-processing is made possible by the explicit and objective computation of the posterior covariances and of the influence matrix. This post-processing proves that the atmospheric inversion is not able to inquire into very fine scales in our case study. The correlation grouping of indifferentiable regions allows an accurate analysis of the best possible signal detectable by the inversion. In the following, we take a correlation threshold of 0.5 as a good balance between sufficient constraints on the system and consistent posterior fluxes.

2.6.2 Spatial evaluation

We have chosen a threshold of correlation grouping the regions so that the averaged scores on the whole domain of interest are optimal. The scores are not uniformly distributed. In Fig. VII.8, the distributions of the three scores are displayed for fossil fuel regions and wetlands for OSSE 1 (transport and distribution of the fluxes same as the 'truth', perturbed masses by regions; see Tab. VII.1). We choose the 'easiest' OSSE configuration in order to evaluate the behaviour of the marginalized inversion in the best configuration possible, thus getting the upper bound for the expectable quality of the results. Any more realistic set-up likely gives less good results. In the figure, the scores are projected on the aggregation grid built on the considerations in Sect. 2.3. Most of the observation sites are located in the centre of the domain (see Fig. VII.5). Then, the influence score is on average better close to the core of the network for the wetlands. For the fossil fuel regions, the influence score is relatively high also upwind the monitoring network (dominant winds blow West to East in the region). Additionally to the network density, the inversion suffers from mis-separation of side regions and LBC. For this reason, side regions tend to be less constrained than centre ones. However, one can notice in both wetland and fossil fuel maps that some centre regions are significantly less constrained than the core of the domain on average. These are regions of very high and dense emissions close to the observation sites (< 500 km). The air masses coming from these regions to the observation sites are plume shaped air masses. The inversion has troubles in assimilating single plumes. In Sect. 2.3, filters have been implemented in order to detect these problematic regions. The marginalized inversion effectively filtered out these regions.

The absolute and relative scores also show unexpected patterns. Scandinavia and China regions own some of the best absolute and relative scores. These two side regions are filtered out most of the time because of strong correlations with the LBC components of the state space (confirmed by their low influence score). Consequently, when not filtered out, these regions are very well and unambiguously constrained, so the good relative and absolute scores. For the rest of the domain, the scores are mostly the better, the closer to the observation network.

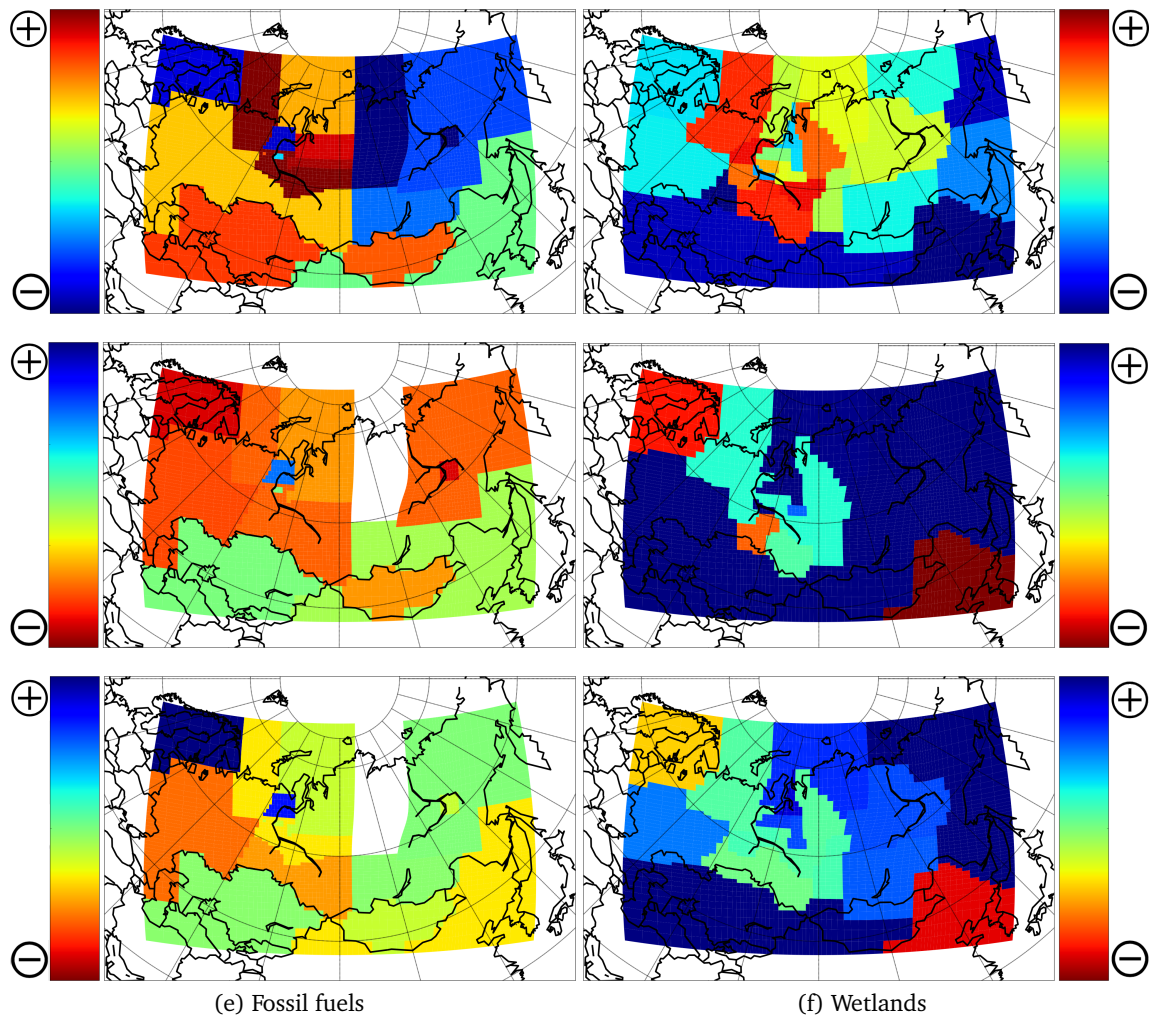


Figure VII.8: Map of the average scores as defined in Sect. 2.4.2 for OSSE 1 (see Tab. VII.1) projected on the aggregation grid defined in Sect. 2.3. (up) Influence score z_{infl} . (middle) Relative score z_{rel} . (bottom) Absolute score z_{abs} . The colour maps have been chosen so that redder regions correspond to better scores (denoted by \ominus and \oplus symbols). The zoom and map physical projection are the same as in Fig. VII.5.

2.6.3 Limitations and benefits

2.6.3.1 Promising computation of the uncertainties The marginalized inversion provides an objectified quantification of the errors in the inversion system. With the Monte Carlo approach we implemented, we are able to consistently take the sources of uncertainties in the inversion process into account, especially those from the prescribed error covariance matrices. As evaluated through OSSEs, the method proved to consistently catch 'true' fluxes on average in the particular Siberian set-up. Moreover, the Siberian set-up is a difficult case study for atmospheric inversions, with co-located intense fluxes that vary at temporal and spatial scales smaller than the meso-scale. The processing of hot-spots, critical in most inversion configurations, is consistently managed, through filters on the plume-shaped air masses. An in-depth analysis of the temporal variability of the fluxes is carried out in a sister publication with the Siberian set-up and real observations (Berchet et al., 2014b). Additionally, as a comparison, we carried out the same OSSEs on the same particular Siberian set-up, but with expert-knowledge frozen error matrices (diagonal matrices with the same representation ω as for the other OSSEs). The correlation profiles and the spatial structures of the scores with the expert-knowledge matrices are not shown because the general patterns are very similar to what is described for the marginalized inversion. Though similar in patterns, the values of the scores are significantly depreciated from the marginalized inversion to the expert-knowledge one. The expert-knowledge relative and absolute scores are several times bigger than the ones from the marginalized inversion, thus statistically incompatible with the 'truth'.

The marginalized inversion explicitly and objectively computes the posterior covariance matrix and the influence matrix. The physical interpretation of the inversion results are then enhanced by a clear analysis of the observation constraints to the fluxes. The processing of the posterior correlations makes the detection of the dipoles and of indistinguishable regions possible. The influence of the lateral boundary conditions, specific to the meso-scale and to the use of limited area CTMs, is estimated. Thus, the regions upwind the observation sites and mixed with lateral air masses can be excluded from the inversion. From the correlations, the grouping of regions gives an estimate of the typical spatial and temporal scale the method can compute. In our case, with few and distant observation sites, the groups of regions cover very large areas. Thus, a grid-point high resolution inversion would not have given deep insights into the fluxes we are looking at. The reduced problem approach described in Sect. 2.3 is then relevant when computed cautiously.

2.6.3.2 Subjective choices and biases Despite all these benefits compared with the classical Bayesian framework, our method still has limitations. The technical implementation of the method needs extensive computation power and memory requirements. For this reason, we have to drastically reduce the size of the problem to solve. The size reduction relies on rigorous considerations that are difficult to formulate analytically. Therefore we applied heuristic principles in order to choose the aggregation patterns of the observations and the fluxes. This subjective procedure can modify the results of the inversion and must be carried out very cautiously. The way we group the regions after the marginalized inversion in order to physically interpret the results is also subjective. We choose a correlation threshold of 0.5 in order to counter-balance the need of useful constraints from the inversion and the requirements of consistently reproducing the 'true' fluxes. Other thresholds could have been chosen and the typical distinguishable temporal and spatial scales would slightly differ from

one threshold to another. But, in any chosen correlation threshold, we notice that most aggregation regions are grouped within bigger ensembles, suggesting that the chosen aggregation patterns are small enough to have reduced impact on the inversion post-processed results.

The marginalized inversion suffers from transport biases as any other inversion. However, the maximum likelihood algorithm considers the biases as random errors and includes them into the error matrix \mathbf{R}_{\max} . The biases are then taken into account in the marginalized inversion, though as random errors. Biases can be represented, or at least detected, with non-diagonal matrices as suggested by Berchet et al. (2013b), but a non-diagonal framework would make the computation of the marginalized inversion critically complicated. Despite the implicit inclusion of the biases as random error in \mathbf{R}_{\max} , we reduced the impact of the horizontal transport biases through filters on the plume-shaped air masses. The vertical biases are smoother and more difficult to detect. This issue must be inquired into in further works. Biases can be studied through marginalizations on the input vectors (e.g., Bocquet, 2011). Coupled marginalizations on the input vectors and on the error statistics would provide a more complete view on atmospheric inversion uncertainties.

2.7 Conclusion

At the meso-scale, inconsistencies between inversion configurations appear in the classical Bayesian framework. One of the main sources of inconsistencies is the specification of the error matrices and the non inclusion of the tenacious uncertainties on these matrices. Synthesizing the recent literature, we developed an updated Bayesian method of inversion from the classical Bayesian framework based on a marginalization on the error matrices and on an objectified specification of the probability density function of the error matrices. This new method makes the comprehensive inclusion of the impact of ill-specified uncertainty matrices possible for the first time in atmospheric inversion to our knowledge. In principle, this method needs very high computation power and memory resources. To avoid technical limitations, we reduce the size of the problem by aggregating the fluxes by region, following objective principles for reducing aggregation errors. We test this method through OSSEs on methane in a domain of interest spanning over Eurasia with significant emissions of different types and different time and space scales. The OSSEs are based on synthetic observations generated from a nature run. We evaluate the consistency and robustness of the method on OSSEs with inversion configurations from the more favourable to the most disadvantageous one (perturbed atmospheric transport, flat flux distribution and wrong total masses). The method produces very consistent and satisfactory results. In most cases, the tolerance intervals given by the inversion include the 'true' fluxes and the results remain close to the 'truth'. The method also provides an explicit computation of the constraints on the regions and allows flagging out regions critically mis-separated from the lateral boundary condition. We hence have developed a robust and objectified method able to consistently catch 'true' greenhouse gas emissions at the meso-scale and to explicitly group the regions that are physically un-distinguishable with the atmospheric signal only. In addition, we developed a method that explicitly produces posterior tolerance intervals on the optimal distinguishable time and space flux scales and that computes the observation network influence on the fluxes.

The robustness of our method on the Siberian case with a biased transport proves that it can be generically applied to other meso-scale frameworks. The high spatial and temporal variability of the fluxes in Siberia ensures the possibility of using the system in 'easier' inver-

sion set-up. Actual observations from the sites we used for the validation of the method are exploited in further steps of our work in order to quantify the 'real' methane fluxes in the Siberian Lowlands (Berchet et al., 2014b).

Acknowledgements

We thank all the PIs from the sites we used for providing us with information on their data. We especially thank Jost Lavrič and Jan Winderlich (Max Planck Institute, Jena, Germany), Motoki Sasakawa (Center for Global Environmental Research, NIES, Tsukuba, Japan), and Michael Yu. Arshinov (V. E. Zuev Institute of Atmospheric Optics, SB-RAS, Tomsk, Russia) for the information on the Siberian sites. We are thankful to the two reviewers, including Marc Bocquet (CEREA, Champs-sur-Marne, France), for their constructive criticism of the first version of this manuscript and for their fruitful comments. We thank François Marabelle (LSCE) IT support team for the maintenance of computing resources. This study extensively relies on the meteorological data provided by the European Centre for Medium-range Weather Forecast. This research was supported by the Commissariat à l'Énergie Atomique et aux Énergies Alternatives.

3 Une méthode générale et cohérente

3.1 Vers un système opérationnel

Nous avons décrit une méthode générale susceptible de quantifier les incertitudes dans les systèmes d'inversions atmosphériques via une marginalisation sur les statistiques d'erreurs. En particulier, l'incertitude statistique est prise en compte.

L'inversion marginalisée testée avec différentes perturbations et biais se comporte de façon très cohérente et reproduit statistiquement la réalité. La variation des flux spatiale et temporelle est identifiée, ainsi que les corrections à appliquer aux flux. De plus, dans cette inversion marginalisée, nous disposons d'informations explicites sur l'influence des observations et sur les régions mal-séparées.

À ce point du travail de thèse, nous sommes en mesure d'implémenter une méthode et un système d'inversion satisfaisants et opérationnels. De plus, les ressources de calcul disponibles pour nos inversions permettent d'exécuter ce système en environ deux semaines pour une fenêtre d'inversion d'une année. Ces temps de calculs sont tout à fait accessibles et comparables à la durée d'une inversion classique.

Il est par conséquent possible d'envisager une quantification par inversion marginalisée des flux de méthane en Eurasie à la méso-échelle à partir de mesures réelles.

3.2 Améliorations à prévoir

Pour l'objet de notre étude, les performances du système marginalisé tel que nous l'avons décrit sont satisfaisantes. Néanmoins, certaines améliorations peuvent être trouvées.

Les points faibles de notre méthode sont : l'agrégation des émissions, l'hypothèse de diagonalité des matrices d'erreur et la distribution de type χ^2 pour les matrices de covariance.

L'agrégation des émissions se fait de manière heuristique dans notre étude. Des pistes plus objectives telles que celles développées par Wu et al. (2011) pourraient être explorées à condition de ne pas augmenter considérablement les temps de calcul. Les hypothèses sur la forme des matrices et leur distribution sont fortes et nécessitent un travail théorique et expérimental supplémentaire afin d'être confirmées ou corrigées.

Application du système d'inversion marginalisée à la Sibérie en 2010

Sommaire

1	Mise en place de l'inversion marginalisée en Sibérie	171
1.1	Choix d'une fenêtre d'inversion	171
1.2	Une étude unique	172
2	<i>Natural and anthropogenic methane fluxes in Eurasia</i>	172
2.1	<i>Introduction</i>	174
2.2	<i>Marginalized inversion framework</i>	176
2.2.1	<i>Motivations towards marginalizing</i>	176
2.2.2	<i>Method outline</i>	178
2.2.3	<i>Output analysis</i>	179
2.2.4	<i>Size reduction and filters</i>	180
2.3	<i>Set up for an Eurasian domain</i>	182
2.3.1	<i>The observation network : y^o</i>	182
2.3.2	<i>Estimates of the network footprints</i>	183
2.3.3	<i>Prior fluxes and state vector : x</i>	184
2.3.4	<i>The observation operator : H</i>	186
2.3.5	<i>Independent observations for evaluation</i>	187
2.4	<i>Diagnostics of the marginalized inversion</i>	188
2.4.1	<i>Observation weight in the inversion</i>	188
2.4.2	<i>Constrained regions</i>	191
2.4.3	<i>Solved time and space resolution</i>	192
2.5	<i>Results of the marginalized inversion</i>	194
2.5.1	<i>Inverted fluxes</i>	194
2.5.2	<i>Siberian Lowland CH_4 budget</i>	196
2.6	<i>Evaluation of the inversion</i>	198
2.6.1	<i>Performance on filtered out data</i>	198
2.6.2	<i>GOSAT evaluation</i>	198

	2.6.3	<i>Toward using satellite measurements in regional frameworks</i>	200
	2.7	<i>Conclusions</i>	202
3		Des contraintes objectivées sur le budget sibérien	203
	3.1	Des observations insuffisantes	203
	3.2	Incertitude réduite sur le budget annuel de méthane en Sibérie	204

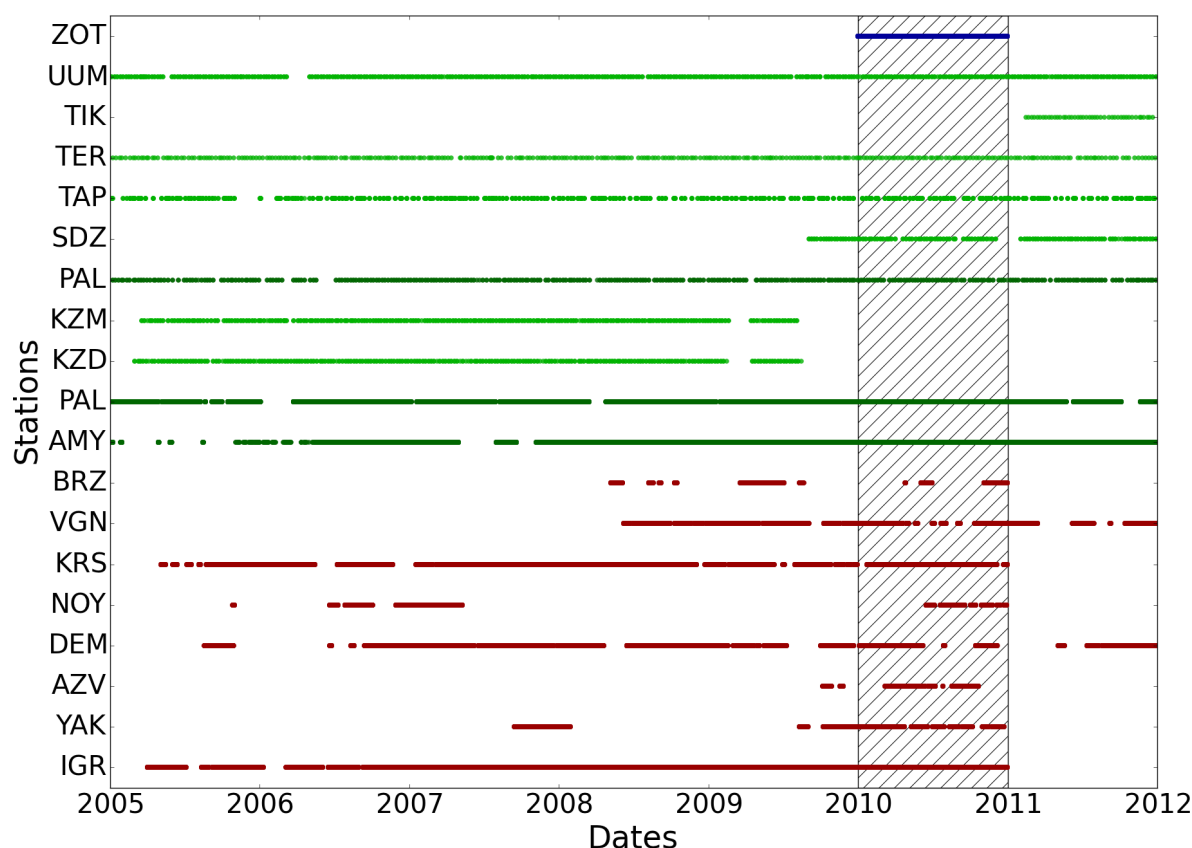


FIGURE VIII.1: Disponibilité des données en Eurasie entre 2005 et 2012 (voir figure IV.1 page 62 pour l'emplacement des stations). En bleu, les données fournies par J. Lavrič du Max Planck Institute ; en vert clair (respectivement foncé), les données flacons (respectivement continues) disponibles sur le *World Data Center for Greenhouses Gases* (<http://ds.data.jma.go.jp/gmd/wdcgg/>) ; en rouge, les données transmises par M. Sasakawa du NIES. En hachuré, la période choisie pour effectuer l'inversion des flux en Sibérie entre janvier et décembre 2010.

1 Mise en place de l'inversion marginalisée en Sibérie

1.1 Choix d'une fenêtre d'inversion

Nous poursuivons deux objectifs principaux en procédant à des inversions de méthane à la méso-échelle en Sibérie. Nous souhaitons d'abord réduire les incertitudes sur le budget annuel sibérien, qui atteignent 10% du budget mondial. Nous voulons aussi explorer la capacité des observations atmosphériques à améliorer notre connaissance sur la distribution spatiale et la variabilité temporelle des émissions (en particulier le début et la fin de la saison d'émission de méthane des zones inondées sibériennes).

Afin de remplir ces objectifs, nous devons choisir une fenêtre d'inversion d'au moins une année avec le plus d'observations possibles. Dans la figure VIII.1, nous référençons les observations de surface disponibles en Eurasie entre 2008 et 2012. Nous voyons que les contraintes logistiques limitent la continuité des observations. De plus, l'échange de données concernant

la Sibérie peut être lent. Par conséquent, seule l'année 2010 présente suffisamment de sites d'observation disponibles avec un échantillonnage temporel important.

En outre, une inversion régionale demande de prescrire des conditions aux bords les plus précises possible. Nous extrayons ces concentrations aux limites du domaine de simulations mondiales LMDz optimisées avec des observations de surface (Bousquet et al., 2006). Lorsque le travail d'inversion à proprement parler a débuté sur la Sibérie, les concentrations mondiales disponibles les plus récentes étaient celles de l'année 2010. Il en était de même pour les bases de données et modèles utilisés pour construire les flux a priori.

En conséquence, nous choisissons d'appliquer notre système d'inversion marginalisé à l'année 2010 complète afin de contraindre un cycle saisonnier complet.

1.2 Une étude unique

La quantification à la méso-échelle des flux en Sibérie que nous faisons est inédite sur deux points.

Premièrement, la méthode d'inversion marginalisée nous fournit des informations critiques dans une région telle que la Sibérie où les observations sont rares. L'évaluation précise de l'influence du réseau d'observation utilisé et l'estimation des échelles spatio-temporelles contraintes sont primordiales pour améliorer la couverture des observations sur le continent eurasien.

Deuxièmement, des inversions atmosphériques ont déjà été faites en Sibérie (comme détaillé en Sect. 2.3.3 page 32 du Chapitre II). Mais ces études ne s'appuyaient que sur un petit nombre de sites d'observations et des modèles mondiaux de transport. On s'attend à ce que l'augmentation du nombre de contraintes atmosphériques et le gain de précision dans la simulation du transport atmosphérique améliore la pertinence des résultats de l'inversion.

2 Natural and anthropogenic methane fluxes in Eurasia : a meso-scale quantification by generalized atmospheric inversion

Ce travail a été permis par le partage des données collectées en Sibérie par le NIES et le MPI. Cet article a été soumis au journal de l'éditeur Copernicus de l'*European Geosciences Union, Biogeosciences*. Il a été publié sous forme de discussion le 13 octobre 2014 et corrigé après revue en avril 2015.

Les auteurs ayant participé à ce travail sont : Antoine Berchet¹, Isabelle Pison¹, Frédéric Chevallier¹, Jean-Daniel Paris¹, Philippe Bousquet¹, Jean-Louis Bonne¹, Michael Yu. Arshinov², Belan B.D.², Cindy Cressot¹, Davydov D.K.², Ed Dlugokencky³, Fofonov A.V.²,

¹Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CNRS-CEA-UVSQ, Orme des Merisiers, CEA Saclay, Gif-sur-Yvette, France.

²V. E. Zuev Institute of Atmospheric Optics, SB-RAS, Tomsk, Russia.

³NOAA Earth System Research Laboratory, Boulder, Colorado, USA.

Galanin A.⁴, Jost Lavrič⁵, Toshinobu Machida⁶, Robert Parker⁷, Motoki Sasakawa⁶, Renato Spahni⁸, Benjamin Stocker⁸, et Jan Winderlich⁹.

Les résultats préliminaires de ce travail ont été présentés lors d'une conférence internationale (voir Annexe B 245) : présentation orale à l'*International ZOTTO Workshop* (Septembre 2013 ; Krasnoyarsk, Russie).

Résumé

Huit sites d'observations de surface mesurant de manière continue la composition atmosphérique en méthane fonctionnent en Sibérie depuis le milieu des années 2000. Pour la première fois dans une unique étude, nous assimilons l'ensemble de ces données in situ dans une inversion atmosphérique. Notre objectif est de quantifier les flux de surface de méthane d'origine anthropique et des zones inondées à la méso-échelle sur la plaine de Sibérie occidentale pour l'année 2010. Nous nous intéressons aussi à la quantification de l'utilisation des observations dans l'inversion et des contraintes apportées aux flux. Les inversions atmosphériques à la méso-échelle sont sensibles aux sources d'erreurs mal spécifiées. Pour cette raison, nous nous appuyons sur de récentes innovations théoriques et utilisons une approche nouvelle d'inversion capable de quantifier les incertitudes plus objectivement que les inversions classiques. Les incertitudes causées par la mauvaise spécification des statistiques d'erreurs dans l'inversion sont calculées de manière systématique. En raison des erreurs de représentation dans le transport atmosphérique et de la redondance des informations fournies par les observations, nous n'utilisons dans l'inversion qu'une seule observation par site et par jour. Le signal à haute résolution restant dans ces observations est caractéristique de structures d'émissions plus locales que la méso-échelle. L'analyse de l'utilisation par l'inversion des observations révèle que chaque site d'observation contraint les émissions dans un rayon de 500 km environ. Par conséquent, des sites d'observations supplémentaires seraient nécessaire pour contraindre l'ensemble de la plaine de Sibérie occidentale. Néanmoins, les flux contraints par les sites d'observations le sont de manière objective. Ainsi, nous trouvons des intervalles de confiance pour les flux a posteriori de l'ordre de 20% et 50% pour les flux respectivement anthropiques et des zones inondées. Si nous extrapolons les chiffres que nous donne l'inversion sur les zones contraintes à l'ensemble de la plaine de Sibérie occidentale, nous obtenons un budget annuel régional de 5–28 TgCH₄ pour l'année 2010, soit 1–5% des émissions totales de méthane. Malgré le caractère objectif de l'approche d'inversion utilisée, des observations indépendantes sont nécessaires afin d'évaluer les résultats de l'inversion. Les observations in situ étant rares dans la région, nous utilisons les données du satellite GOSAT pour évaluer notre inversion. Les colonnes totales mesurées ne révèlent cependant pas une sensibilité significative aux flux régionaux. Elles sont donc plutôt adaptées pour contraindre des flux à l'échelle mondiale.

⁴P. E. Melnikov Permafrost Institute SB-RAS, Yakutsk, Russia.

⁵Max Planck Institute for Biogeochemistry, Jena, Germany.

⁶Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan.

⁷University of Leicester, Leicester, United Kingdom

⁸Climate and Environmental Physics, Physics Institute, and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.

⁹Max Planck Institute for Chemistry, Mainz, Germany.

Abstract

Eight surface observation sites providing quasi-continuous measurements of atmospheric methane mixing ratios have been operated since the mid-2000's in Siberia. For the first time in a single work, we assimilate all of these in-situ data in an atmospheric inversion. Our objective is to quantify methane surface fluxes from anthropogenic and wetland sources at the meso-scale in the Siberian Lowlands for the year 2010. To do so, we first inquire into the way the inversion uses the observations and the way the fluxes are constrained by the observation sites. As atmospheric inversions at the meso-scale suffer from mis-quantified sources of uncertainties, we follow recent innovations in inversion techniques and use a new inversion approach which quantifies the uncertainties more objectively than the previous inversions. We find that, due to errors in the representation of the atmospheric transport and redundant pieces of information, only one observation every few days is found valuable by the inversion. The remaining high-resolution quasi-continuous signal is representative of very local emission patterns. An analysis of the use of information by the inversion also reveals that the observation sites constrain methane emissions within a radius of 500 km. More observation sites than the ones currently in operation are then necessary to constrain the whole Siberian Lowlands. Still, the fluxes within the constrained areas are quantified with objectified uncertainties. At the end, the tolerance intervals for posterior methane fluxes are of roughly 20% (resp. 50%) of the fluxes for anthropogenic (resp. wetland) sources. About 50–70% of emissions are constrained by the inversion on average on an annual basis. Extrapolating the figures on the constrained areas to the whole Siberian Lowlands, we find a regional methane budget of 5–28 TgCH₄ for the year 2010, i.e. 1–5% of the global methane emissions. As very few in-situ observations are available in the region of interest, observations of methane total columns from the Greenhouse Gas Observing SATellite (GOSAT) are used for the evaluation of the inversion results, but they exhibit marginal signal from the fluxes within the region of interest.

2.1 Introduction

Methane (CH₄) in the atmosphere contributes to climate forcing as a greenhouse gas and is involved in the atmospheric oxidizing capacity (Forster et al., 2007). Characterizing the variability of the atmospheric CH₄ composition requires accurate understanding of the methane biogeochemical cycle, in particular of the surface-atmosphere fluxes, of their spatial distribution and of their temporal variability. The quantification of these contributions to the methane cycle still experiences high uncertainties (Kirschke et al., 2013). The global surface to atmosphere CH₄ fluxes range between 500 and 600 TgCH₄.y⁻¹ (1 Tg = 10¹² g). Two of the main contributors to the global CH₄ budget are natural emissions from inundated areas and anthropogenic sources from coal, oil and gas extraction and distribution. Inundated areas are responsible for 145–260 TgCH₄.y⁻¹, i.e. 25–50% of total emissions with a very high heterogeneous spatial distribution and year-to-year variability (e.g., Bousquet et al., 2006; Dlugokencky et al., 2009; Bergamaschi et al., 2009). The anthropogenic sources from fossil fuel burning and leaks account for 100–150 TgCH₄.y⁻¹, i.e. 20–30% of total emissions, according to the EDGAR inventory (depending on the year and the version of the inventory; <http://edgar.jrc.ec.europa.eu>).

The West Siberian plain concentrates significant sources of CH₄ of both wetland and anthropogenic types (Lechtenböhmer et al., 2005; Spahni et al., 2011). On one side, with 50–

70% of its area covered by peatlands (Peregon et al., 2009), about 13% of global wetlands are located in the West Siberian plain. On the other side, Russia produces 20% of the natural gas in the world, mostly extracted in Siberia; and 0.1 to 10% (i.e. $0.5\text{--}40\text{ TgCH}_4\cdot\text{y}^{-1}$) of this gas are estimated to leak into the atmosphere (e.g., Hayhoe et al., 2002). Large amounts of methane are also released during the oil welling, of which Russia is also a major producer ($\sim 10\text{--}15\%$ of the global production), with 1-2% of the oil production leaked into the atmosphere as methane (e.g., E.P.A., 2011). Documenting the emissions of methane in the West Siberian plain is thus critical to reduce the uncertainties on the global methane budget.

However, accurately quantifying the wetland and anthropogenic emissions in the West Siberian plain is challenging. On the one hand, wetland emissions at high latitudes like in Siberia exhibit a clear year-to-year variability (Bergamaschi et al., 2013) and a distinct seasonality (e.g., Pickett-Heaps et al., 2011) due to high sensitivity to the soil temperature and humidity, to the water table depth and to the total inundated surface which can vary up to 25% from year to year (Ringeval et al., 2010). This high sensitivity to the local meteorological parameters could cause still unobserved drastic increases of CH_4 emissions from boreal wetlands with climate change (Bohn et al., 2007). On the other hand, anthropogenic sources are mainly related to uncontrolled leakage which is difficult to estimate. Quantifying these leaks raises many issues: pipelines with tiny leaks from chemical permeability span over thousands of kilometres, single leaks range on a spectrum of several orders of magnitude (from the lower with pipeline permeability leaks, to the higher during the welling), and leaks can appear (and disappear when detected and repaired) very quickly.

Despite the importance of quantifying and understanding the contribution of the West Siberian plain to the global CH_4 budget, few studies have been dedicated to this region. Glagolev et al. (2011) carried out extensive field measurements of local wetland CH_4 emissions in order to characterize the emission patterns of each different environment. They upscaled their results to the whole region using wetland distribution maps. However, considering the discrepancies on the wetland distribution (Peregon et al., 2009; Frey and Smith, 2007) and their extension variability (Ringeval et al., 2010), numerous hard-to-quantify uncertainty sources are expected from this approach. For the quantification of the anthropogenic sources, for instance, Dedikov et al. (1999) measured mixing ratios of methane close to gas lines and gas facilities to deduce emission factors. They upscaled their figures to the Russian territory and got an emission factor for CH_4 lower than 1% of the total production of natural gas. Reshetnikov et al. (2000) reviewed the existing literature about Siberia and found emission factors from 0.4 to 14% of the total production. Another approach is the analysis at different places and dates of the local variations of the atmospheric composition (mixing ratios and isotopic fraction) in CH_4 (and related species). The variations of the atmospheric composition provides information on the relative contribution of the different local processes in the Siberian budget. Such analysis has been carried out using observations from mobile platforms, such as aircraft profiles (Yamada et al., 2005; Umezawa et al., 2012) or train and ship measurements (Tarasova et al., 2006), and fixed stations (Sasakawa et al., 2012). Nevertheless, these studies are not systematic and comprehensive: they give local instantaneous information and only knowledge about the relative contributions of particular processes to the total regional fluxes. High spatial and temporal resolutions, absolute and consistent quantification and separation of the different main sources is essential to better characterizing and estimating the contribution of Siberian emissions to the global methane budget.

A first step towards this goal was done by Winderlich (2012). In this work, a systematic analysis of the variability of the atmospheric composition at the ZOTTO tall tower (described in Winderlich et al., 2010) and at a small set of auxiliary sites was carried out. Their approach relied on atmospheric inversion techniques based on the data assimilation Bayesian theory (Enting et al., 2012; Tarantola, 1987) and in principle allows an objective use of the information (here the variability of the atmospheric composition). This provided insights on, e.g., the likely under-estimation of the local anthropogenic sources from gas leak in some inventories (EDGAR v4.1 in Winderlich, 2012). However, regional atmospheric inversions with only a few sites give only local constraints on the surface fluxes. More critically, a regional inversion with a small number of observations experiences difficulties in identifying and separating the different contributions to CH₄ emissions. This mis-separation can be related to mis-quantified sources of errors in the atmospheric inversion in addition to the lack of information. This issue is dominant in the West-Siberian plain because of the co-located wetland and anthropogenic emissions. The objective of quantifying and separating the regional Siberian sources requires an inversion based on numerous observation sites, with a comprehensive approach of quantification of the uncertainties.

Recent literature highlights the need for precisely and objectively quantifying all errors in the inversion (transport, representation, flux distribution, etc.) in order to evaluate their impact on the inversion results (Lauvaux et al., 2009; Winiarek et al., 2012; Berchet et al., 2013b; Ganesan et al., 2014). In Berchet et al. (2014a), we proposed a general method in order to objectively quantify most of the critical sources of errors in the inversion. This improved algorithm is based on a Monte Carlo approach superimposed to maximum likelihood estimators (Chapnik et al., 2004; Michalak and Kitanidis, 2005).

For the first time in Eurasia, we use this improved algorithm on a network of 8 surface sites (Sasakawa et al., 2010; Winderlich et al., 2010) covering a large part of the Siberian lowlands and of 5 remote sites that constrain the air masses coming into the domain and getting out of it. These sites, which have been operated since the mid-2000's, are implemented into the inversion system with objectified uncertainty quantification from Berchet et al. (2014a). Here, our goal is to deduce an accurate quantification of the fluxes at the meso-scale with a temporal resolution of a few days from the variability of the atmospheric CH₄ composition at the 8+5 observation sites.

We explain the theoretical background used in our study in Sect. 2.2. In Sect. 2.3, the datasets and models used in the inversion are introduced. We then present the results on the fluxes and the limitations of the inversion in Sect. 2.4 and Sect. 2.5. Our inversion is then evaluated in Sect. 2.6 by using in-situ measurements and independent satellite observations, as the very few available surface observations do not and do not allow to keep some for evaluation.

2.2 Marginalized inversion framework

2.2.1 Motivations towards marginalizing

As the atmosphere mixes irreversibly air masses from different CH₄ sources, using the atmospheric signal as in an atmospheric inversion cannot lead to a deterministic characterization of the surface fluxes. In the classical Bayesian framework (Tarantola, 1987), the objective of the inversion is to inquire into the probability density function (*pdf*) of the sur-

face fluxes, or more generally of the state of the system, with some knowledge about the atmospheric composition and on the prior state distribution. The sought *pdf* can be written: $p(\mathbf{x}|\mathbf{y}^o - H(\mathbf{x}^b), \mathbf{x}^b)$.

In this formula, the vector \mathbf{y}^o gathers all the available observations; H is the observation operator converting the information in the state vector to the observation space; the vector \mathbf{x} depicts the state of the system (mainly CH_4 surface fluxes in our case) and \mathbf{x}^b is the background vector including the prior knowledge on the state \mathbf{x} of the system. H typically represents the discretization of the problem to be computed numerically and the atmospheric transport from the emission areas to the observation sites. In the following, we consider H linear and associate it to its Jacobian matrix \mathbf{H} . In our case, we simulate the atmospheric transport only on a domain of limited area as detailed in Sect. 2.3.3 and 2.3.4. Additional information about the atmospheric composition at the boundaries of the domain of interest is then necessary to compare observed with simulated atmospheric composition. Therefore, the state vector \mathbf{x} encompasses the surface fluxes, but also the lateral boundary conditions related to the observed baselines at each observation site.

Thus, the inversion computes the pieces of information contained within the observations, the prior state and the representation of the transport with their associated uncertainties (e.g., measurement errors, uncertainties in the flux inventories, etc.). With the usual Gaussian assumption, all the uncertainties are considered as normal *pdfs* and can be described with modes and uncertainty covariance matrices. The inversion then deduces an optimal posterior state vector $\widehat{\mathbf{x}}^a$ and posterior uncertainties \mathbf{P}^a . Within the Gaussian assumption, the posterior state vector and uncertainty matrix can thus be explicitly defined:

$$\begin{cases} \widehat{\mathbf{x}}^a &= \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \\ \mathbf{P}^a &= \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B} \end{cases} \quad (\text{VIII.1})$$

In Eq. VIII.1, the matrix $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$ is the Kalman gain matrix. The matrices \mathbf{R} and \mathbf{B} are the covariance matrices describing the observation and background uncertainties. Observation uncertainties encompass measurement, discretization and transport errors. Background uncertainties include the uncertainties in the spatial distribution of the fluxes, in their temporal variability and in their absolute value.

As long as these uncertainty matrices are known, the inversion only faces technical issues (e.g., matrix inverses and products in large dimension problems) for numerical implementation. However, only the uncertainties in the measurements are objectively quantified during the calibration process. The errors in the transport or in the prior fluxes are not perfectly known and, in most case, they are built relying on some expert knowledge about the system. But this subjective knowledge can lead to ill-specified matrices, which have a dramatic impact on the inversion results (e.g., Cressot et al., 2014). Recent studies inquired into objectified ways of specifying these matrices (e.g., Michalak and Kitanidis, 2005; Winiarek et al., 2012; Berchet et al., 2013b). The approach in these papers was to find optimal uncertainty matrices \mathbf{R} and \mathbf{B} along an objective statistical criterion: the maximum likelihood. The implementation of the method gave encouraging results, but the impact of the uncertainties within the maximum likelihood computation were complicated to evaluate.

Berchet et al. (2014a) used a general marginalization approach in order to quantify objec-

tively all the uncertainties impacting on the posterior fluxes. In the following, we summarize this approach.

2.2.2 Method outline

The marginalization consists in computing the complete *pdf* $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ as a weighted sum of the Gaussian *pdfs* $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b, \mathbf{R}, \mathbf{B})$ calculated for each possible uncertainty matrices \mathbf{R} and \mathbf{B} . This can be written as follows:

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \\ = \int_{(\mathbf{R}, \mathbf{B})} p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b, \mathbf{R}, \mathbf{B}) \\ \times p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \quad (\mathbf{R}, \mathbf{B}) \end{aligned} \quad (\text{VIII.2})$$

To compute the marginalized integral in Eq. VIII.2, a large number (60 000 in our case) of posterior vectors $\widehat{\mathbf{x}}^a$ and posterior uncertainty matrices $\widehat{\mathbf{P}}^a$ is computed through individual inversions as in Eq. VIII.1 with different uncertainty matrices $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{B}}$. This Monte Carlo sampling is carried out based on an estimate of the *pdfs* of the uncertainty matrices, $p(\mathbf{R})$ and $p(\mathbf{B})$. This estimate is deduced from the objectified maximum likelihood approach.

Posterior uncertainties and correlations are in the end computed from the ensemble of 60 000 computed individual inversion results as follows:

$$\begin{cases} \mathbf{x}^a &= \frac{1}{N} \sum_{i=1..N} \widehat{\mathbf{x}}^{a^i} \\ \mathbf{P}^a &= \frac{1}{N-1} \sum_{i=1..N} (\widehat{\mathbf{x}}^{a^i} - \mathbf{x}^a) (\widehat{\mathbf{x}}^{a^i} - \mathbf{x}^a)^T \end{cases} \quad (\text{VIII.3})$$

with N the number of Monte Carlo draws.

As detailed in Sect. 2.2.1, the main motivation for marginalizing the classical inversion framework is to use the available information (in-situ observations, flux inventory, transport model) in a way which is as objective as possible. The marginalized inversion gives an explicit and objectified access to pieces of information (described in details in Sect. 2.2.3) required:

1. to evaluate the efficiency of the observation network for constraining regional emissions and to give guidelines for improving monitoring deployment;
2. to inquire into atmospheric inversion skills in terms of resolved temporal and spatial resolution and of emission process separation, and to deduce observation and modeling requirements for future better inversions; and
3. to assess the robustness of emission inventories and process-based surface-atmosphere exchange models

These three key points are further discussed in Sect. 2.4.

One of the main drawbacks in our method we use is its numerical cost. Computing 60 000 individual inversions is cumbersome and requires extensive amounts of memory. Smartly chosen filtering criteria, observation sampling and flux aggregation patterns must be carried out before the marginalized inversion in order to reduce its complexity without degrading

the inversion optimality. As recent efforts have been made to reduce the subjectivity in the inversion, we rely on objective criteria (though not computed in an exact manner due to computational limitations) to complete the observation sampling, flux aggregation and filtering. These criteria are shortly developed in Sect. 2.2.4 and detailed in Berchet et al. (2014a).

2.2.3 Output analysis

As described above, the marginalized regional inversion as we compute it answers three main questions. Below are given details on how these questions are treated.

2.2.3.1 Network efficiency The evaluation of the network coverage is carried out through the explicit computation at the maximum likelihood of the influence $\mathbf{K}_{\max}\mathbf{H}$ and sensitivity matrices $\mathbf{H}\mathbf{K}_{\max}$ following Cardinali et al. (2004). With these two matrices, we can compute the weight ω of each observation site in the inversion and the unconstrained emissions ϵ as follows:

$$\omega_i = \sum_{j \text{ at } i} (\mathbf{H}\mathbf{K}_{\max})_{j,j} \text{ for each site } i \quad (\text{VIII.4a})$$

$$\epsilon_i = (1 - \mathbf{K}_{\max}\mathbf{H}) \times \phi_i \text{ for each emission pixel } i \quad (\text{VIII.4b})$$

with ϕ_i the emissions at the pixel i .

In Eq. VIII.4a, the higher the score ω_i the more the inversion uses the site i . Observation sites downwind emissions will have a strong impact on the inversion, but observations constraining air masses coming into the domain of interest are also key stones for regional inversions. Eq. VIII.4b depicts emitted CH_4 that is not seen or constrained by the inversion. Regions with the highest ϵ_i are unseen areas with strong emissions (as considered by inventories), where additional observation sites would be required.

2.2.3.2 Solved spatial and temporal scales The spatial and temporal resolutions that the inversion can solve are described by posterior error covariances. Mis-separated regions are usually detected through so-called flux dipoles (e.g., Rödenbeck et al., 2003). In our case, as we explicitly and objectively compute the posterior matrix \mathbf{P}^a from the Monte Carlo ensemble $(\widehat{\mathbf{x}}^a)$, posterior correlations are used for post-processing groups of ill-separated regions (these regions are defined through a dynamically-chosen aggregation pattern prior to the inversion; see Sect. 2.2.4.2). Both strongly positive and negative correlations $r_{i,j}$ point to ill-separated regions i and j . Following Observing System Simulation Experiments (OSSEs) carried out in Berchet et al. (2014a), we group two posterior flux regions i and j when:

$$|r_{i,j}| = |\mathbf{P}_{i,j}^a| > 0.5 \quad (\text{VIII.5})$$

For instance, some neighbouring emissions or successive ones could be mixed in the atmospheric signal and then be not separated from the inversion point of view. This post-processing makes it possible to group posterior fluxes in order to avoid dipoles and reduce uncertainties on bigger regions. Thus, it is also possible to filter out regional fluxes that are not separated from the boundary conditions. Every group of correlated regions including some contribution

from the boundary conditions are excluded from the analysis of inversion results. Indeed, errors on the lateral boundary conditions can mis-lead the inversion about the regional fluxes.

Through this post-grouping, we can assess the typical size of aggregated regions that the inversion can constrain with our set-up (observations and transport model resolution). The typical time-scale that the inversion can detect for a change in emission can also be assessed with this post-processing. As we want to separate contributions from anthropogenic and wetland emissions, the post-grouping may group or not co-located emissions from different processes and then give insights into the separation ability of our regional inversion.

2.2.3.3 Posterior flux analysis Ideally, an atmospheric inversion provides insights about emissions. From the correlation grouping applied to the Monte Carlo ensemble of posterior fluxes, we can compute tolerance intervals of posterior fluxes so that 68.27% of the ensemble is within the interval. The number 68.27% makes the tolerance intervals equivalent to the $\pm\sigma$ interval in the Gaussian framework, as the Monte Carlo posterior ensemble does not necessarily a Gaussian distribution. The inversion thus indicates that the fluxes we are inquiring into are very likely in the defined posterior tolerance interval. Deviating posterior tolerance intervals compared to prior fluxes point to required updates in the used prior database. Below we present our visualization approach to control posterior fluxes.

Figure VIII.6 synthesizes the inverted methane fluxes for Siberian lowlands (hatched domain in Fig. VIII.2). As detailed in Sect. 2.2.3.2, anthropogenic emissions (inverted at a monthly scale) can be grouped with wetland emissions (considered at the weekly scale). So, the lowest common multiple on which the fluxes can be analysed is the monthly scale. Then, for the Siberian Lowlands, for each month, we define the proportion of the fluxes that are within regions constrained through the inversion (written on the left of each pie chart in Fig. VIII.6). The proportions of anthropogenic and wetland emissions that are constrained, mixed with another type of emission or unseen by the observation network are represented in the pie charts in Fig. VIII.6. Finally, within the proportion of constrained regions, we analyse the inversion correction on anthropogenic, wetland and mixed emissions (bar plots in Fig. VIII.6). For each type of emission (anthropogenic, wetland and mixed), we present in the bar diagrams of Fig. VIII.6 the tolerance intervals of posterior fluxes and the prior uncertainties as calculated by the maximum likelihood algorithm (that is to say \mathbf{B}_{\max} ; see Sect. 2.2.2). For each month, the tolerance intervals on the total prior and posterior methane budget in the Siberian Lowlands is also given in TgCH_4 .

2.2.4 Size reduction and filters

As suggested in Sect. 2.2.2, the marginalized inversion requires some filtering, observation sampling and flux aggregation, so it can be numerically computed. Below, we explain how we carry out these pre-processing in a way that do not dampen the advantages of the marginalization.

2.2.4.1 Observation sampling Here we try to reduce the dimension of the observation space. At the regional scale, considering the spatial resolution of our transport model, only the synoptic variability of the observed signal is relevant. We then decide to keep only one piece of information per site and per day as it is commonly done at the global and continental scales. In addition, simulated vertical mixing close to the surface where observations are carried

out is known to be flawed when the planetary boundary layer (PBL) is shallow (typically at night and in winter in Siberia). We then sample the observations during the afternoon when the PBL is higher than 500 m as suggested by prior studies (e.g., Berchet et al., 2013b) and we pick the observed and simulated mixing ratios at the time when the observations are minimum.

As surface emissions dominate on surface sinks for CH₄, keeping the minimum observed mixing ratio by afternoon is equivalent to detecting the time when the PBL is at its maximum, hence when the atmospheric model is the more accurate. This criterion filters out outliers generated by local influences which cannot be reproduced by an atmospheric transport model with a resolution larger than 25 km, and which only add noise to the system.

For our case study, out of 127000 hourly measurements available in 2010, 30000 pass through the PBL height and night filters (see black dots in Fig. VIII.3). Out of these 30000 data points, about 2000 daily aggregates are selected. Details by observation site are given in Tab. VIII.2.

2.2.4.2 Flux aggregation and constraints The following procedures are bound to decrease the state space dimension. To define aggregation pattern, we divide our domain into 35 physical regions for each emission process (according to vegetation types, demography, industrial activity, etc.) as a basic pattern. This basic pattern is chosen so that the mesh gets finer closer to the observation network. The resolution of the transport representation is also chosen finer close to the observation network. To further reduce the number of aggregated regions, under-constrained regions are grouped together. This is carried out by analysing the observation network footprints estimated with a Lagrangian model (Sect. 2.3.2) which offers an efficient way to compute them.

In addition to the footprint aggregation, the influence matrix $\mathbf{K}_{\max}\mathbf{H}$ makes it possible to quantify observational constraints on the fluxes. Below a given threshold of constraint for a flux (related to the flux contribution to the atmospheric signal), the inversion cannot deduce any valid information on the flux. For this reason, we also filter out a region i with very low constraints from the marginalization if:

$$(\mathbf{K}_{\max}\mathbf{H})_i^a < 0.5 \quad (\text{VIII.6})$$

2.2.4.3 Plume filtering It is known that atmospheric transport models suffer from temporal and spatial mismatches when simulating air masses with strong mixing ratio gradients (e.g., for plumes well-delimited from the background air masses). When a plume is transported to the wrong place and time, as we sample air masses at a given point location, the differences between simulated and observed mixing ratios can reach unrealistic values. Such strong model-observation differences have a significant impact on inversion corrections on fluxes.

To dampen such undesirable effects, we introduce a procedure to filter out plume-like air masses from the inversion input. As explained in Sect. 2.2.2, a maximum likelihood estimation is computed prior to the marginalization. We take advantage of the maximum likelihood computation to detect air masses critically ill-reproduced by the transport model. As the maximum likelihood estimation computes optimal uncertainty matrices \mathbf{R}_{\max} and \mathbf{B}_{\max} , we filter out observations with a too high computed uncertainty. That is to say, for each observation i ,

the data point is excluded if:

$$(\mathbf{R})_{\max}^{i,i} > 20 \text{ ppb} \quad (\text{VIII.7})$$

This criterion does not necessarily exclude only plumes generating a big signal, but also the ones that are very poorly reproduced by the transport model.

One should notice that this criterion is computed in association with the low constraint criterion of Eq. VIII.5. That is to say, a region which always illuminates the observation network through plume-like air masses will have all its constraining observations filtered out. As a consequence, it will be considered poorly constrained and then be itself eliminated from the inversion.

Taken all together, the criteria defined above filter out a large part of the available observations and regional fluxes. In particular, hot spot emissions are largely removed from the inversion (except for some exception when the observations are far enough, so that the transport manages to well reproduce the plume generated by the emissions).

2.3 Set up for an Eurasian domain

We are interested in the Eurasian surface-atmosphere budget of methane. As developed in Sect. 2.2, the computation of the marginalized inversion needs measurements (\mathbf{y}^o ; Sect. 2.3.1), prior fluxes (\mathbf{x}^b ; Sect. 2.3.3) and an observation operator computed with an atmospheric transport model (\mathbf{H} ; Sect. 2.3.4). The footprints of the observation network required for choosing the aggregation patterns (necessary to the computation of \mathbf{H}) are estimated with a Lagrangian model (Sect. 2.3.2). The observations are compared to the fluxes through the observation operator \mathbf{H} . The operator \mathbf{H} is computed with a transport model, which simulates mixing ratios, whereas the observation sites provide dry air mole fractions. For trace gases such as CH_4 , dry air mole fractions and mixing ratios are equivalent. In all of the following, we affiliate dry air mole fractions to mixing ratios.

2.3.1 The observation network: \mathbf{y}^o

The inversion is based on measurements of atmospheric CH_4 dry air mole fractions at 14 Eurasian surface sites. The Eurasian sites are situated all over the Eurasian continent, from South Korea to Scandinavia (see Fig. VIII.2). They are maintained by the National Institute for Environmental Studies (NIES, Tsukuba, Japan), the Institute for Atmospheric Optics (IAO, Tomsk, Russian Federation), the Max Planck Institute (MPI, Iena, Germany), the P. E. Melnikov Permafrost Institute (Yakutsk, Russian Federation), the Finnish Meteorological Institute (FMI, Helsinki, Finland) and NOAA Earth System Research Laboratories (Boulder, United States of America). Most stations provide quasi-continuous measurements of methane atmospheric composition. A few stations only collect flasks at a weekly scale (see Tab. VIII.1). The location, measurement type, maintaining institute and 3-letter symbol of each observation site are described in Tab. VIII.1.

Flasks samples are collected at SDZ, TAP and UUM sites as partners of the NOAA global cooperative air sampling network (Dlugokencky et al., 1994). These flask sampling sites are designed to monitor large-scale and global variability. They are then placed at locations,

which are most of the time not illuminated by local and regional sources. Thus, the observations from these sites generally pass the filtering criteria of Sect. 2.2.4. Flasks samples are sent to and analysed at NOAA-ESRL by gas chromatography with flame ionization detection against the NOAA 2004 CH₄ standard scale (Dlugokencky et al., 2005).

NIES sites (AZV, BRZ, DEM, IGR, KRS, NOY, VGN and YAK) are part of the Japan–Russia Siberian Tall Tower Inland Observation Network (JR-STATION; Sasakawa et al., 2010) and are equipped with CH₄ semiconductor sensors based on a tin dioxide natural gas leak detector developed by Suto and Inoue (2010). The instruments are calibrated on tanks traceable to NIES 94 CH₄ scale. The NIES 94 CH₄ scale is higher than the NOAA 04 scale by 3.5–4.6 ppb in the range between 1750 and 1840 ppb (Zhou et al., 2009). ZOT site is operated by MPI since April 2009 (Winderlich et al., 2010) and air is analysed by an EnviroSense 3000i analyser (Picarro Inc., USA, CFADS-17) based on the cavity ring-down spectroscopy technique (CRDS Crosson, 2008). The calibration system uses tanks traceable to NOAA 04 scale. All the mixing ratios are reported to the NOAA 04 scale before being implemented into the inversion system.

JR-STATION and ZOT sites are located in the vicinity of anthropogenic and wetland sources. These local sources strongly influence the Siberian network, which was designed to monitor regional emissions. As a consequence, numerous observations from the Siberian network are ill-reproduced by our transport model and then are filtered out from the inversion as “plume” observations (following criteria in Sect. 2.2.4; see also Fig. VIII.3 and Tab. VIII.2 in Sect. 2.6.1). In particular, measured mixing ratios at BRZ site are not well simulated by our model, possibly due to missing local emissions in the prior. Most BRZ observations are thus filtered out from the inversion.

Due to logistical issues or instrument dysfunctions, observation sites do not provide measurements all year round. Fig. VIII.3, described in Sect. 2.4.1, details the temporal availability of the observations. The sampling bias is known to impact the inversion results (Villani et al., 2010). The issue is discussed in Sect. 2.4.1, but the general method developed in Sect. 2.2 consistently takes into account such a bias into increased posterior tolerance intervals and constraints on the emissions.

The observation vector y^o is defined after Sect. 2.2.4.1 sampling method. The final size of y^o implemented in the inversion is 2000. On average, 0.4 observations per station per day are validated for the inversion.

2.3.2 Estimates of the network footprints

As the observations that will be implemented in the system are known, the observation network footprints, necessary to choose the aggregation patterns in order to define the prior state vector x^b and the observation operator \mathbf{H} (as detailed in Sect. 2.2.4.2), can be computed. As we do not carry out a quantitative analysis of the footprints, we only need a rough estimation of the footprint patterns. Thus, we compute simulations with the Lagrangian dispersion model FLEXPART version 8.2.3 (Stohl et al., 2005) to get such an estimation. To build the footprints, we compute numerous back-trajectories of virtual particles from the observation sites at the times when measurements are available and valid for the inversion.

The model is forced by ECMWF ERA-Interim data at an horizontal resolution of $1^\circ \times 1^\circ$, with 60 vertical levels and 3 hours temporal resolution (Uppala et al., 2005). Virtual particles

Table VIII.1: Sites characteristics. The altitudes of the sites are given as m above sea level (a.s.l.) and the inlet height is in m above ground level (a.g.l.). The frequency column depict the type of instrument in the site: C=quasi-continuous, F=flask sampling.

Station	ID	Location			Inlet height (m a.g.l.)	Frequency	Network / Institute
		Lon (° E)	Lat (° N)	Alt (m a.s.l)			
Azovo	AZV	73.03	54.71	100	50	C	JR-STATION
Berezorechka	BRZ	84.33	56.15	150	80	C	JR-STATION
Demyanskoe	DEM	70.87	59.79	75	63	C	JR-STATION
Igrim	IGR	64.42	63.19	25	47	C	JR-STATION
Karasevoe	KRS	82.42	58.25	50	67	C	JR-STATION
Noyabrsk	NOY	75.78	63.43	100	43	C	JR-STATION
Pallas	PAL	24.12	67.97	560	7	C	FMI
Shangdianzi	SDZ	117.12	40.65	287	0	F	NOAA/ESRL
Tae-ahn Peninsula	TAP	126.12	36.72	20	0	F	NOAA/ESRL
Ulaan Uul	UUM	11.08	44.45	914	0	F	NOAA/ESRL
Vaganovo	VGN	62.32	54.50	200	85	C	JR-STATION
Yakutsk	YAK	129.36	62.09	210	77	C	JR-STATION
Zotino	ZOT	89.35	60.80	104	301	C	MPI

are released in a 3D box (10 km per side and 1000 m high) centered around each observation site with 10-day lifetime backwards in time. The footprints are computed on a $0.5^\circ \times 0.5^\circ$ horizontal grid, following the method of Lin et al. (2003), taking into account the boundary layer height at each particle location. This method considers that only the particles within the boundary layer are influenced by surface emissions and that the boundary layer is well-enough mixed to be considered as uniform.

2.3.3 Prior fluxes and state vector: x

The inversion system optimizes prior fluxes grouped in the regions aggregated by the pre-processing procedure (see Sect. 2.2.2). The prior spatial distribution and temporal variability of the fluxes are deduced from: 1) EDGAR database v4.2 FT2010 (<http://edgar.jrc.ec.europa.eu>) for year 2010 for anthropogenic emissions, 2) LPX-Bern v1.2 process model (Stocker et al., 2014) at a monthly scale for wetland emissions 3) GFED v4 database at a daily scale for wildfires. In Fig. VIII.2, the distributions of the anthropogenic hot spots of emissions and of the wetland regions are represented, superimposed over the regional topography. Anthropogenic emissions of methane in the region are mainly hot spots related to the intense oil and gas industry in the Siberian Lowlands and to the leaks in the distribution system in population centres in the vicinity of the Trans-Siberian Railway in the Southern part of the Siberian plain. Wetland emissions are mainly confined to the lower part of Siberia in the West Siberian plain, half of which is lower than 100 meters above sea level. Wildfires occur mainly in spring and summer in the Eurasian forest-covered areas; they emit CH_4 as intense hot spots.

The EDGAR inventory uses economic activity maps by sectors and convolved with emission factors calculated in laboratories or with statistical studies (Olivier et al., 2005). The Bern based Land surface Processes and eXchanges (LPX-Bern v1.2) model is an update of the

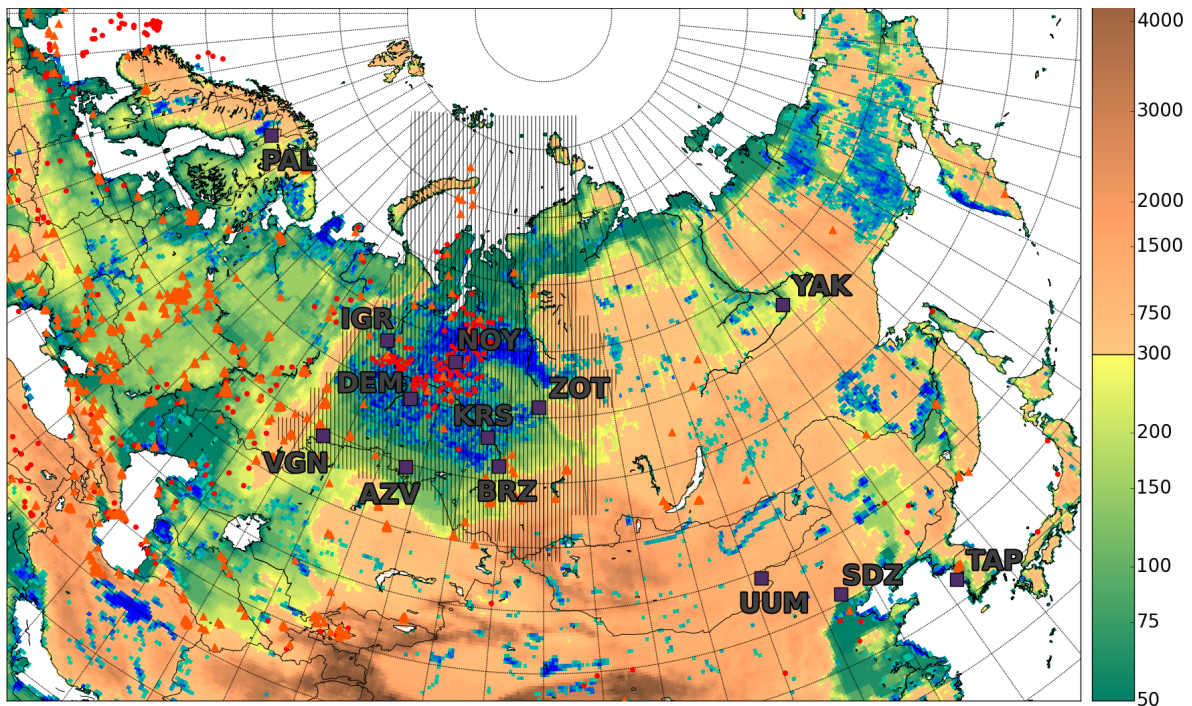


Figure VIII.2: Topographic map of the domain of interest. The colour bar paints the altitude above sea level (from ETOPO1 database; Amante and Eakins, 2009). Red dots (resp. orange triangle) depict hot spots of CH_4 emissions (based on EDGAR v4.2 inventory; see Sect. 2.3.3) related to oil welling and refineries (resp. gas extraction and leaks during distribution in population centres). Purple squares highlight the observation site localizations. Blueish shaded areas represent inundated regions, wetlands and peatlands (from the Global Lakes and Wetlands Database; Lehner and Döll, 2004). The Siberian budget in Sect. 2.5.2 is calculated on the hatched area.

dynamic global vegetation model LPJ (Spahni et al., 2011). It includes a dynamical simulation of inundated wetland areas (Stocker et al., 2014), dynamic nitrogen cycle (Stocker et al., 2013), and dynamic evolution of peatlands (Spahni et al., 2013; Stocker et al., 2014). The model uses CRU TS 3.21 input data (temperature, precipitation rate, cloud cover, wet days), observed atmospheric CO_2 and prescribed nitrogen deposition (Lamarque et al., 2011) for each year for the simulation of dynamic forest and peatland vegetation growth. The GFED v4 database is built from the 500 m Collection 5.1 MODIS DB burned-area mapping algorithm (Giglio et al., 2009). CH_4 emissions at monthly and daily scales are deduced from the burnt areas using Carnegie-Ames-Stanford-Approach (CASA model; Potter et al., 1993) and emission factors (van der Werf et al., 2010).

We are aiming at a separation of the types of emissions at the meso-scale. We therefore aggregate the emissions along the three different types of sources, with specific spatial patterns and temporal profiles for each type of emissions. Anthropogenic sources are hot spots emitting all year round. Wetlands are responsible for diffuse fluxes on large areas, with high temporal variability depending on the local weather conditions (typically temperature or water table depth). The emissions of CH_4 from wildfires comes from point sources and occur on

relatively short periods (Kasischke and Bruhwiler, 2002). Consequently, we do not aggregate the different types of emissions along the same spatial patterns and temporal intervals. Anthropogenic emissions are aggregated by month, while wetlands and wildfires, which have quicker time responses to meteorological changes, are grouped by periods of 10 days. In the following, we discuss the inversion results only in term of anthropogenic and wetland emissions. Indeed, as the wildfire emissions generate plumes relatively well-defined from the ambient air, the marginalized inversion exclude from the system all the emission contribution related to fires (according to the procedures described in Sect. 2.2.2).

For the computation of the observation operator \mathbf{H} (see Sect. 2.3.4), we use a regional chemistry-transport model with a domain limited in space and time. Initial and lateral boundary conditions (hereafter IC and LBC) are then also to be optimized in the system. Prior lateral mixing ratios are deduced from simulations at the global scale by the general circulation model LMDz with the assimilation of surface observations outside the domain of interest (Bousquet et al., 2011). LBC are assimilated by periods of 10 days. We arbitrarily aggregate LBC along 4 horizontal components (by side of the domain) and 2 vertical ones (1013–600 hPa and 600–300 hPa). Though we are mainly focused on Siberian lowlands, the domain of model computation has been chosen spanning over all Eurasia. This is expected to attenuate the impact of the rough global resolution in LMDz boundary conditions on the simulated variations of mixing ratios at the observation sites. Indeed, the central region is thousands of kilometres away from the sides of the domain.

To summarize, all the pieces of information in the observations are assimilated to constrain 1700 aggregated regions of flux and boundary conditions: 10×12 month regions for anthropogenic emissions, 10×36 10-day period for wildfires, 25×36 10-day period for wetlands, 9 (4 sides \times 2 horizontal levels + roof top) $\times 36$ 10-day period for the lateral boundary conditions. After the filtering of Sect. 2.2.4, the dimension of the state space is reduced from 1700 to 275.

2.3.4 The observation operator: \mathbf{H}

We explicitly define the observation operator \mathbf{H} by computing the forward atmospheric transport from the regions of aggregated emissions (defined in Sect. 2.3.3) to the observation sites. As CH_4 is a reactive species, the observation operator should include the oxidation by OH radicals. However, as the residence time of the air masses in the domain of simulation is short (a few days to a few weeks) compared to CH_4 atmospheric life time (8–10 years; Dentener et al., 2003), ignoring OH sink only generates small differences in the simulated mixing ratios. Additionally, OH sink is mostly responsible for large scale gradients while the regional inversion focuses on the synoptic scale. Thus, the regional inversion system attributes OH sink to global boundary conditions, and not to regional fluxes.

Thus, for each aggregated region, we calculate the so-called response functions using the transport module of the Eulerian meso-scale non-hydrostatic chemistry transport model CHIMERE (Vautard et al., 2001; Menut et al., 2013). This model was developed in a framework of air quality simulations (Schmidt et al., 2001a; Pison et al., 2007), but is also used for greenhouse gas studies (Broquet et al., 2011; Berchet et al., 2013b). We use a quasi-regular horizontal grid zoomed near the observation sites after Sect. 2.2 considerations. The domain of interest is of limited area and spans over the mainland of the Eurasian continent

(see Fig. VIII.4). As we are interested in meso-scale fluxes, we take a spatial resolution larger than 25 km. The average side length of the grid cells is 25 km close to the West-Siberian stations and 150 km away of the centre of the domain. The 3D-domain embraces roughly all the troposphere, from the surface to 300 hPa (~ 9000 m), with 29 vertical layers geometrically spaced. The model time step varies dynamically from 4 to 6 min depending on the maximum wind speed in the domain. The model is an off-line model which needs meteorological fields as forcing. The forcing fields are deduced from interpolated meteorological fields from the European Centre for Medium-range Weather Forecast (ECMWF) with a horizontal resolution of $0.5^\circ \times 0.5^\circ$ every 3 h (Uppala et al., 2005).

2.3.5 Independent observations for evaluation

Any inversion has to be confronted by independent data in order to evaluate its results. Few in-situ measurements of CH_4 mixing ratios are available in Siberia. We choose to assimilate all surface observations described in Sect. 2.3.1 for the optimization of CH_4 fluxes. NIES and LSCE carry out aircraft measurements in the region (Paris et al., 2010a; Umezawa et al., 2012; Berchet et al., 2013a), but these measurements are still difficult to compare to meso-scale models. In addition, their spatial and temporal coverage is poor for the year 2010 and they are not numerous enough to get significant validation insights.

For year 2010, the only remaining observations with sufficient spatial coverage and temporal availability are the total columns retrieved by the Greenhouse Gas Observing SATellite (GOSAT). In Sect. 2.6.2, we evaluate the results of the inversion against GOSAT data. The Japanese satellite GOSAT was launched by the Japan Aerospace Exploration Agency (JAXA), NIES and the Japanese Ministry of the Environment (MOE) in January 2009. It has a polar sun-synchronous orbit at 667 km and provides a full coverage of the Earth every 3 days with a swath of 750 km and a ground pixel resolution of 10.5 km at nadir. The TANSO-FTS instrument observes the solar radiation reflected at the surface and the top of the atmosphere in the short wave infrared (SWIR) domain that allows deducing total columns of methane (XCH_4) in cloud-free and sunlight conditions. The typical size of GOSAT pixels is ~ 10 km. We use version 3.2 of the TANSO-FTS bias-corrected XCH_4 proxy retrievals performed at the University of Leicester (Parker et al., 2011). The XCH_4 retrieval algorithm uses an iterative retrieval scheme based on Bayesian optimal estimation and associated to averaging kernels and a priori profiles. The retrieval accuracy is estimated to be about 0.6% (i.e. ~ 10 ppb). The retrieval algorithm needs CO_2 mixing ratios as a proxy for the light path. We use the 4-D CO_2 analysis from the surface air-sample inversion by Chevallier et al. (2010) (MACC v10.2). We obtain ~ 25000 GOSAT observations in 2010 over the domain of interest.

In order to compare the observations of the total columns to the model, we use the averaging kernels to compute prior and posterior model equivalents. The regional CTM CHIMERE has a top layer at 300 hPa in our set-up. The stratospheric contribution to the total columns is deduced from the global model LMDz used for the initial and lateral boundary conditions (described in Sect. 2.3.3). The average observed XCH_4 is ~ 1775 ppb over the domain throughout the year. The prior average in LMDz XCH_4 is ~ 1820 ppb. This bias is attributable to the excessive injection of tropospheric air into the stratosphere in our version of LMDz. In Sect. 2.6.2, the bias on XCH_4 of 45 ppb has been corrected to allow the observation-model comparison.

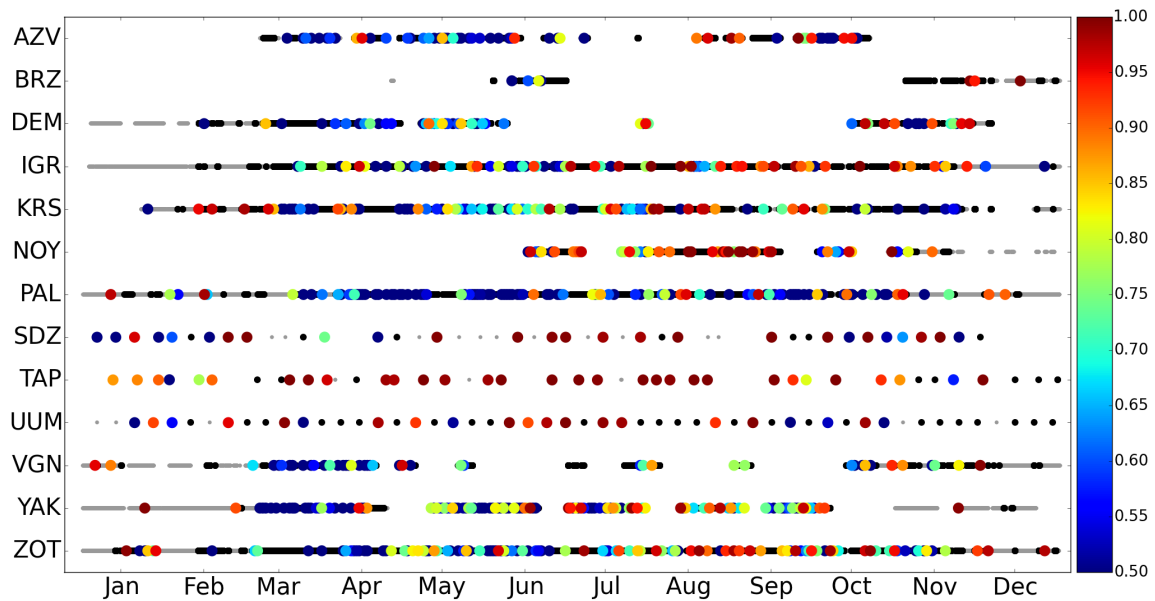


Figure VIII.3: Observation availability and contribution to the inversion for each site. Grey dots stand for available data flagged out because of a planetary boundary layer height smaller than 500 m. Black dots denote the data points that are filtered during the computation of the algorithm described in Sect. 2.2 and 2.3.1. Colour circles represent the pieces of information carried out by each selected piece of data (see Sect. 2.2.3).

2.4 Diagnostics of the marginalized inversion

The marginalized inversion described in Sect. 2.2 provides us with tolerance intervals of posterior fluxes, posterior correlations of errors and influence indicators. As the marginalized inversion filters out some data and regions, we present and analyse here the overall performance of our inversion, the effects of the data selection in the inversion and the implied limitations.

2.4.1 Observation weight in the inversion

2.4.1.1 Temporal monitoring constraints As explained in Sect. 2.2, the method developed by Berchet et al. (2014a) filters out numerous observations and emission regions. Some observations available in the domain in 2010 are set aside before the inversion because of known flaws in CTMs. But the marginalized inversion also flags out additional observations when they are measured within plumes difficult to invert. The remaining pieces of data do not have the same weight in the inversion. Contrary to most classical inversion methods which cannot afford the computation of the explicit sensitivity matrix (see Sect. 2.2.3.1) informing in the weight of individual observations, the marginalized inversion allows us to explicitly analyse the use of the observations in the system. In Fig. VIII.3, we represent the observations filtered out along the PBL height criterion before the inversion (grey points), the ones flagged out during the inversion (black dots) and the relative weight of the remaining used observations (coloured circles).

Many observations cannot be assimilated (black dots), especially in winter, when the very

cold conditions (temperatures lower than -20°C in average) related to the Siberian High generate very stable atmospheric conditions. In these conditions, the local emissions, which cannot be well assimilated in our inversion system, significantly influence the observations. In addition to the numerous not-assimilated observations, all daily observations that are not filtered do not necessarily convey the same amount of information: all the blue circles in Fig. VIII.3 depict pieces of data with a negligible influence on the inversion. The two main explanations for this inability to assimilate all the available pieces of data is the chosen scale of interest and the integrating character of the atmosphere. First, as we are interested in meso-scale fluxes, the system has been chosen with a spatial resolution of 25 to 100 km. All the variability in CH_4 mixing ratios driven by single local plumes cannot be reproduced in the system. Second, the limitation of the atmospheric inversion comes from the fact that the atmosphere behaves as an integrator, hence attenuating some information in the atmospheric signal. Tracking back the atmospheric signal to the fluxes then has intrinsic limitations. This limited capability of the system drastically reduces the number of usable pieces of information. Out of the 127000 hourly measurements available in 2010, the pre-processing (as defined in Sect. 2.3.1) only retains 2000 daily aggregates into the inversion system. The system then excludes some observations and, at the end, only 800 data points remains, with 460 pieces of information (i.e. the trace of the sensitivity matrix; see details in Cardinali et al., 2004) carried by the atmospheric signal. Many observations give redundant information in our specific inversion framework at the meso-scale. The observations that cannot be processed by the meso-scale marginalized inversion carry information about local emissions.

2.4.1.2 Network range of constraints The temporal use of the available observations matters in the inversion, as much as the relative use of the different observation sites. Fig. VIII.4 displays the location and the average weight of each observation site (the coloured background is described and discussed further in Sect. 2.4.2). It is divided into four panels in order to separate the use of information related to anthropogenic (top panels) and wetland (bottom panels) emissions. As the marginalized inversion raw results are also processed in order to detect the regions that are mis-separated from the boundary conditions (see Sect. 2.2.3.2), data from side observation sites is noticeably less used than for central sites.

Comparing anthropogenic (top panel) and wetland (bottom panel) maps, we notice that the weights of the observations are smaller for anthropogenic hot spot emissions. As expected, the inversion experiences difficulties in constraining emission hot spots, compared to diffuse fluxes. Concerning the spatial distribution of observation weights, wetland-related constraints follow the heuristics that the closer the observation site is to the fluxes, the higher the constraints to the inversion system is. Anthropogenic-related constraints do not exhibit such a pattern. For instance, NOY, close to the main oil extraction fields, has a lower observational influence than BRZ, remote from hot spots. Looking at wetland-related influences, BRZ has a bigger influence than NOY, while, as for anthropogenic emissions, wetlands emit more intensely in the vicinity of NOY. Then, a surface observation site must be not too close, but not too far, from an emission hot spot to optimally constrain it. There is no generic criterion for this optimal distance to the observation sites as it depends on the atmospheric transport or the intensity of the hot spots.

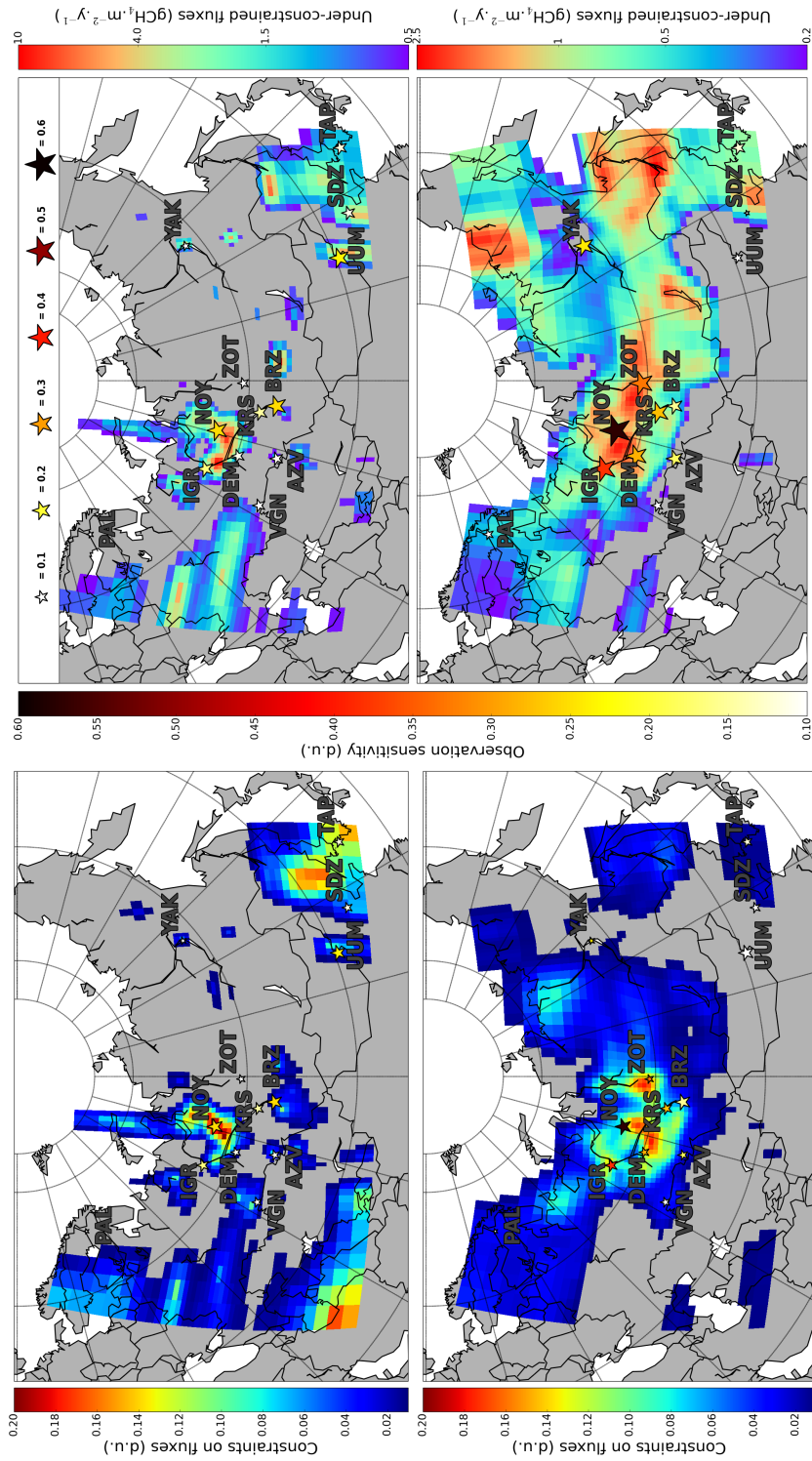


Figure VIII.4: Maps of the constraints on the fluxes (top: anthropogenic fluxes; bottom: wetland emissions) from the observations in the inversion. (left column) Constraints on the fluxes as calculated from the influence matrix (see Sect. 2.2.3.1) and sensitivity of the inversion to observation sites as defined by their average weight ω in Eq. VIII.4a of Sect. 2.2.3.1. (right column) Non-constrained fluxes as computed from the vector ϵ defined in Eq. VIII.4b of Sect. 2.2.3.1 in gCH₄.m⁻².y⁻¹.

Looking at the differences in the relative weights of the observation sites between the raw inversion results and the LBC-separated ones (not shown here), one can notice that the sites at the edge of the domain of interest are logically dedicated to constraining the LBC. Even the relative weight of the observation sites surrounding the Siberian Lowlands is significantly reduced. Additional observations away from the region of interest would be necessary to overcome the influence of the LBC.

2.4.2 Constrained regions

The spatial distribution of the observational constraints on the fluxes is calculated from the sensitivity matrix (see Sect. 2.2.3). The information in this matrix is convolved with the prior distribution of the fluxes to deduce the maps in the right column of Fig. VIII.4. The spatial distribution of the constraints on the fluxes depends on the intensity of the emissions and the distance to the observation sites. In the left column of Fig. VIII.4, the constraints are higher close to the Siberian network than away from it independent of the intensity of emissions. For example, the Western part of the Russian Federation contains most of the anthropogenic emissions of the country (roughly 20 TgCH₄/y according to EDGAR FT2010). But the constraints are lower for this region than for the Siberian Lowlands with smaller emissions (8 TgCH₄/y according to EDGAR FT2010). The post-processing excluding the regions that are mis-separated from the LBC by the inversion highlights this pattern. The observation sites within the denser part of the network seem to constrain emissions within a radius of roughly 500 km.

In the right column of Fig. VIII.4, we display the average fluxes which are not considered as constrained by the inversion (as detailed in Sect. 2.2.3.1). Despite the limited range of the observation sites and the high number of filtered out data points, wetlands in the Siberian Lowlands are significantly constrained. The remaining unconstrained fluxes are of the same order of magnitude as minor wetland emitting region, such as in the far Eastern parts of Russian shores on the Arctic ocean and the Pacific ocean. As a consequence, with the existing network constraining major wetland areas, minor wetland regions now contribute equally to the uncertainties on Siberian CH₄ budget. This points to a needed extension of the monitoring network toward these minor emitting wetland areas. For anthropogenic emissions, the constraints on oil and gas related emissions are still too low. Anthropogenic emissions in Siberian Lowlands are still to be inquired into to reduced uncertainties on Russian CH₄ emissions.

In Fig. VIII.6 described in Sect. 2.2.3.3, we also explicitly compute the portion of constrained emissions per month. On average, the major part of the emissions is not constrained by the inversion. The maximum proportion of constrained emissions is reached in summer with 50–70% of the emissions constrained. In contrast, in winter, only a small part of the emissions are constrained. The proportion is critical in January and November when only 0–9% of the emissions are constrained. In Sect. 2.4.1.1, we noticed that most of the observations are flagged out in winter because of the very low boundary layer. Consequently, emissions in winter months are not well constrained.

Overall, the configuration of the network gives valuable insights on Siberian Lowland emissions, but it is not entirely adapted to our objective of constraining the Siberian Lowland CH₄ budget, even during the period when supply issues do not prevent the acquisition

of observations and when the atmosphere is mixed enough for the CTM to accurately reproduce the transport patterns. Additional observation sites would be needed for a complete resolution of the regional fluxes.

2.4.3 Solved time and space resolution

In order to inquire into possible improvements in the regional inversion, we compute the typical temporal and spatial resolutions the inversion can solve. These scales are plotted in Fig. VIII.5.

For each pixel of the domain of interest, we consider the groups the pixel is within at the different periods of the year (months for anthropogenic pixels, 10-day periods for wetland emissions). We then average the size of the selected groups along the year. We do the same for the duration of each period. The lowest common time step for anthropogenic and wetland emissions is month. Anthropogenic are thus consistently solved at a monthly scale. As wetland emissions can be grouped with anthropogenic ones, their solved time resolution can be increased from the 10-day basis. The temporal resolution is then computed only for wetland emissions in Fig. VIII.5.

As expected, the resolved spatial resolution is better close to the network. Thus, most of Siberian Lowlands are constrained with a typical resolution below 2×10^6 km². The best resolution in our system configuration is roughly 700000 km². As we chose aggregation patterns with a mesh of about 300000 km² in Sect. 2.2.4.2, this confirms that our aggregation procedures are not too coarse.

However, the resolved spatial resolution suggests that numerous additional monitoring sites would be required to identify emission patterns in relation to hydrological and meteorological parameters as wetlands can react quickly and with high gradients to changes in the weather or in the water table depth. A high temporal resolution would be required in the inversion to link wetland emissions to these physical parameters. We see in Fig. VIII.5 that most wetland fluxes are constrained with a temporal resolution of typically two weeks, which is too long to resolve quick changes in emissions. Wetlands along the Yenissei river, far from anthropogenic emissions are solved with a better temporal resolution than wetlands in central Lowlands. Unfortunately, this encouraging resolution is compensated by very high posterior uncertainties in the Yenissei flux. Overall, wetland emissions are resolved at a temporal resolution that allows the detection of the seasonal cycle, but not sufficient for linking wetland emissions to physical parameters varying at the synoptic scale.

To summarize, the inversion approach that we developed allows a precise quantification of the use of the observations. We also can deduce where the inversion results are the most reliable from the spatial influence of the network. At the meso-scale and in the Siberian framework, it appears that i) hourly and even daily measurements are difficult to assimilate, ii) anthropogenic emission hot spots require observation sites remote from them to be inverted, iii) diffuse wetland emissions can be constrained with sites located close to them, iv) the observation sites constrain fluxes within a radius of ~ 500 km in our meso-scale inversion. This knowledge could find applications in network design and in the choice of the type of measurements.

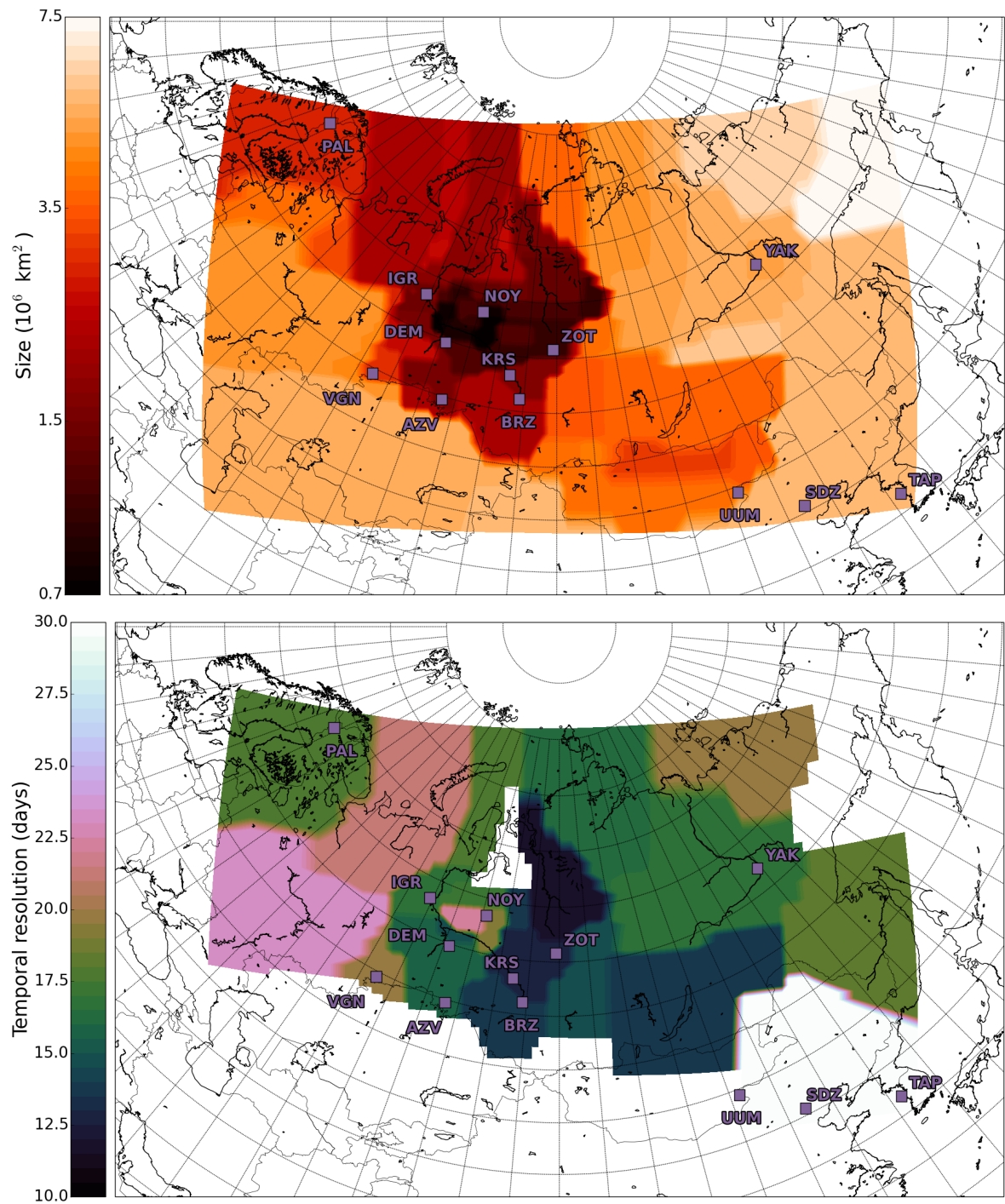


Figure VIII.5: Temporal and spatial resolutions solved by the inversion. (top) Typical sizes of aggregation groups (as defined in Sect. 2.2.3.2); for each pixel, the size of the group the pixel belongs to is averaged along the whole period of study. (bottom) Time-scales resolved for wetland emissions; the period covered by pixel groups is averaged along the year.

2.5 Results of the marginalized inversion

In Sect. 2.5.1, we describe the methane fluxes resulting from the inversion system that are explicitly constrained and not mis-separated from the lateral boundary conditions. We then extrapolate the fluxes on the constrained regions to the Siberian Lowlands and discuss the total budget of CH₄ of the region (Sect. 2.5.2).

2.5.1 Inverted fluxes

We defined successive filtering and aggregating procedures in Sect. 2.2. At the end of the inversion, out of the 275 constrained aggregated regions, only 166 groups of regions are kept: 35 anthropogenic groups, 101 wetland groups and 30 mixed groups. Constraints, ill-separated regions and posterior fluxes are summarized in Fig. VIII.6 described in Sect. 2.2.3.3.

In summer, when wetlands are active, only August and September provide dominant wetland signals whereas winter months (December to April) show dominant anthropogenic sources. The rest of the year (May to July and October to November), the mixed source dominates, indicating that the inversion meets with difficulties in separating anthropogenic from wetland emissions. In the configuration we use, as expected with a sparse network, the usable pieces of information are not sufficient to fully reach the objective to separate emission types. In general, as we filtered out hot spots too close from the observation sites, only distant emissions and diffuse ones are constrained. Thus, the inversion system tries to separate similar mixed atmospheric signals from different sources. With no reliable information on the temporal profile of the different co-located emissions or any measurements of the isotopic atmospheric composition, separating co-located emissions with atmospheric meso-scale inversion is not possible.

In almost all cases, the intensity of the prior fluxes is close to the middle of the posterior tolerance interval. This means that the inventories and models used to get the prior fluxes in the region are reasonably close to the physical fluxes. The main output of the marginalized inversion compared to model or bottom-up approaches is the explicit computation of tolerance intervals of posterior fluxes on the regions that are explicitly constrained. The posterior tolerance intervals are significantly reduced compared to the prior uncertainties computed from the maximum likelihood. In most cases, the maximum likelihood algorithm suggests prior uncertainties more than 100% of the prior emissions. This is consistent with the uncertainties in the wetland models and in the anthropogenic emission inventories in the region (e.g., Hayhoe et al., 2002; Melton et al., 2013). After the inversion, on the constrained regions, the tolerance intervals are equivalent to uncertainties below 100%. For anthropogenic emissions, the posterior uncertainties are in the range of 20–50%, while for wetland emissions, the uncertainties span from 60 to 120%. The large ranges of uncertainties attributed to wetland emissions can be explained by the high temporal and spatial variability of the fluxes. The real temporal profiles and spatial patterns of wetland emissions differ from the ones used by the inversion system. The inversion thus experiences difficulties in precisely reproducing wetland fluxes, but consistently compute high posterior uncertainties. Despite the mis-separation of the co-located emissions, the marginalized inversion thus provides reliable estimates of fluxes uncertainties at the regional scale.

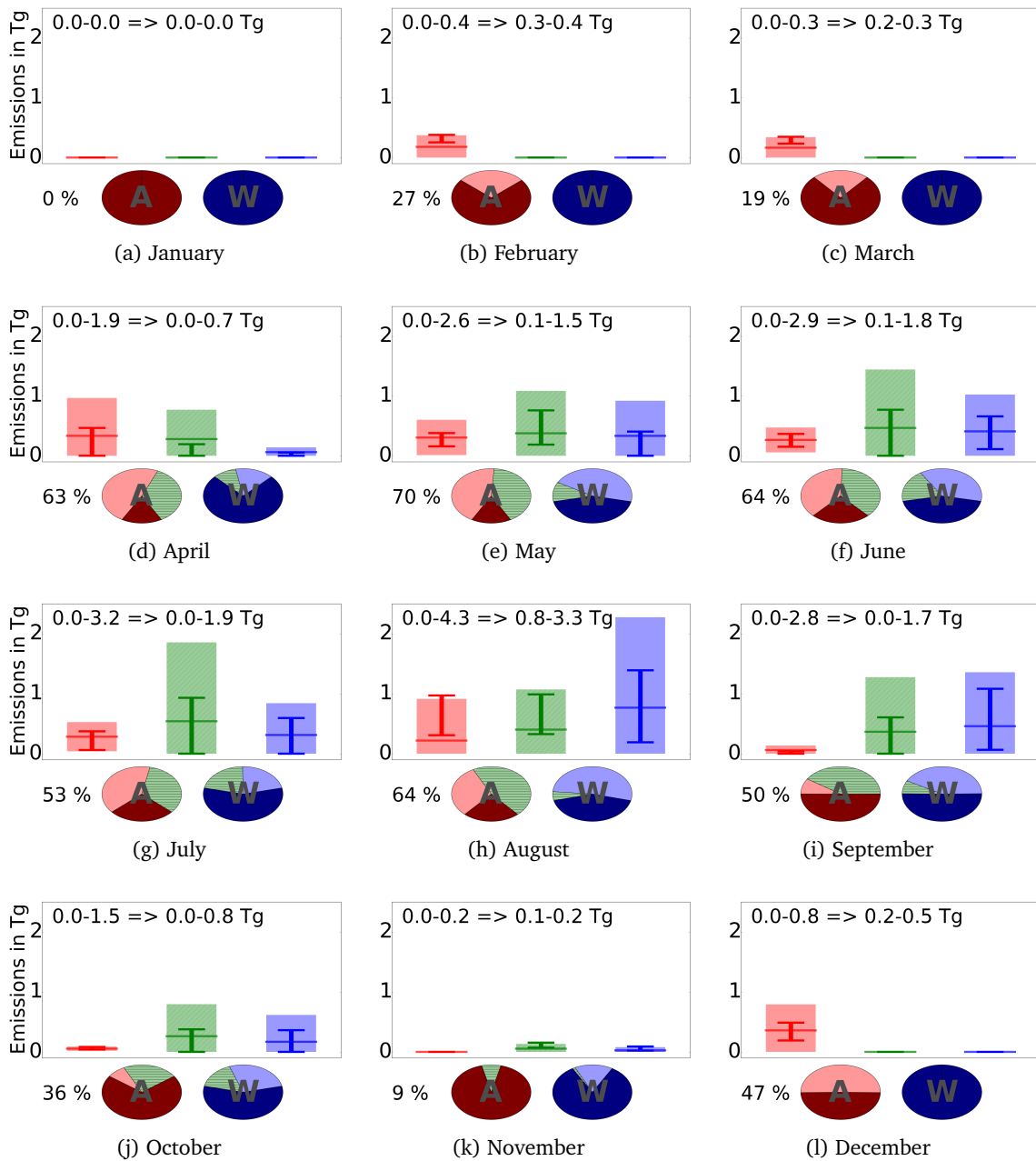


Figure VIII.6: Inversion results for the twelve months of 2010 in the Siberian Lowlands (hatched area in Fig. VIII.2). For each month, the bar plots show the prior (shade) and posterior (bar) tolerance intervals for anthropogenic (left), mixed (middle) and wetland (right) emissions. The pie charts depict the proportion of prior emissions within regions constrained by the inversion: (left, A) anthropogenic, (right, W) wetlands. Dark portions mean no constraint; lighter shade shows constrained regions; the hatched green portions are for the emissions the type of which the inversion cannot separate. The percentage indicates the fraction of the total emissions that are constrained each month. The given intervals in Tg are for the prior and posterior constrained Siberian budgets. See Sect. 2.2.3.3 for details.

2.5.2 Siberian Lowland CH₄ budget

2.5.2.1 Seasonal cycle and yearly emissions The marginalized inversion gives explicit tolerance intervals only on the constrained regions, that embody less than two thirds of the regional emissions. We then do not have more information than the prior emissions about the regions that are not upwind of the the observation sites or that are filtered out because of the plume criterion. Nevertheless, we are inquiring into the regional budget of methane. Extrapolations are necessary to infer regional emissions from the partial vision given by Fig. VIII.6.

Here, for each month of the year 2010, we decide to extrapolate the corrections applied in Sect. 2.5.1 to all the fluxes only when a significant part of the prior emissions are constrained (hereafter 20%). As the prior emissions are close to the middle of the posterior tolerance intervals, this extrapolation does not radically change the inverted budget, but it allows the computation of regional uncertainties. The time series of the extrapolated anthropogenic, wetland and total emissions in the Siberian Lowlands (hatched area in Fig. VIII.2) are displayed in Fig. VIII.7. For each type of emissions, the posterior tolerance intervals are computed only for months with more than 20% of constrained emissions for the associated type. The total annual budgets in the legend of Fig. VIII.7 are computed from the annual corrections on the constrained region for anthropogenic, wetland and total emissions.

Wetland emissions are negligible in winter and intense in summer. In the prior emission, the seasonal cycle of the regional emissions is smooth, the emitting season beginning in March. The marginalized inversion suggests a shift to April-May for the start of the CH₄ emissions from wetlands. Anthropogenic emissions are supposed to be constant in the prior. Posterior emissions in February and March are 20–30% higher than the average posterior emissions. In a region of dense oil and gas extraction, the emissions can be partly correlated to the production, but also to household heating consumption. As the demand of gas is higher in winter, the production is 15–20% bigger during the winter quarter (as suggested by GAZPROM annual reports, the main Russian gas company; www.gazprom.com). This could explain the higher posterior estimates in winter. However, local biogenic processes emitting CH₄ have also been observed in winter, which perturb the measured surface mixing ratios (Arshinov et al., 2009; Winderlich, 2012). Limited bacterial activity below the snow cover can produce CH₄, which perspires to the atmosphere. As the atmosphere is very stable in winter, this phenomenon can generate significant local increases of CH₄ mixing ratios close to the observation sites. As the model does not account for such emissions, the inversion would attribute them to the anthropogenic emissions. This should be furtherly inquired into with isotopic measurements. Nevertheless, most problematic observations with possible local biogenic influence within very stable planetary boundary layers have been flagged out from the inversion due to the mis-representation of the vertical mixing in the model. Therefore, the winter increase of CH₄ emissions in the inversion can be attributed to a real increase in anthropogenic emissions due to the cumulated increased production and consumption of gas for heating.

Overall, on a yearly basis, posterior anthropogenic and wetland emissions are roughly equal as in the prior. The tolerance interval is 1–13 TgCH₄/y for wetland emissions, and 6–16 TgCH₄/y for anthropogenic emissions. The ranges of uncertainties are reduced by 40% for wetland emissions and 57% for anthropogenic emissions from prior to posterior fluxes. Siberian oil and gas extracting activity and population centres are then responsible of 1.5–

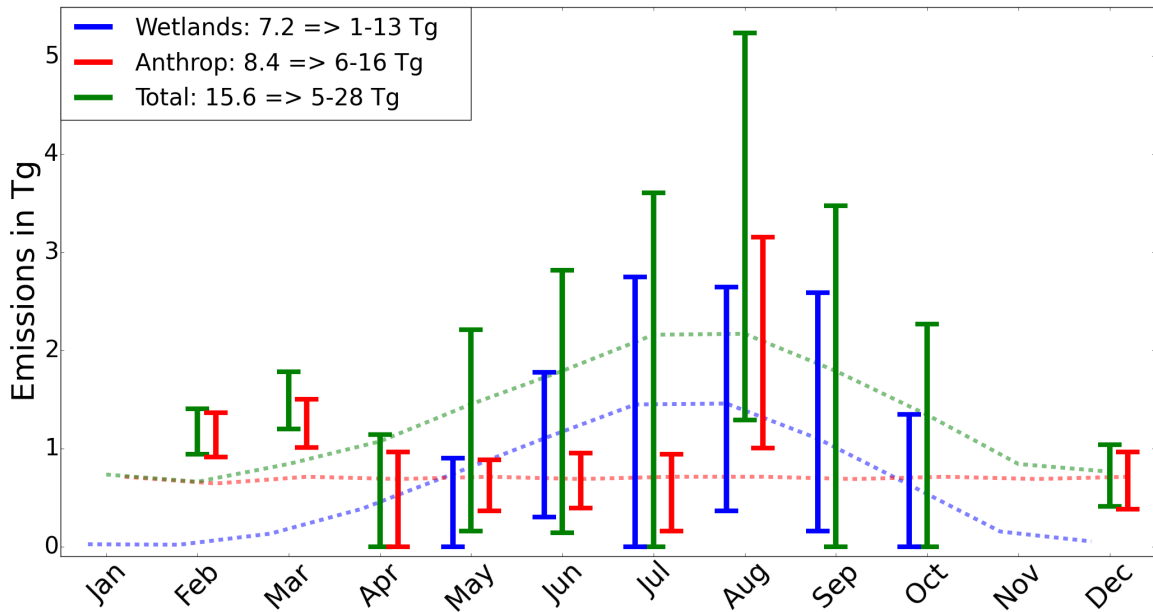


Figure VIII.7: Time series of the posterior emissions on the Siberian Lowlands. The emissions are reported in TgCH_4 for each month. The correction factors computed on the constrained regions (in Sect. 2.5) are extrapolated to the Siberian Lowlands (see Fig. VIII.2). The extrapolation is carried out only when more than 20% of the prior emissions are constrained. The error bars depict the tolerance interval of posterior emissions. The dotted lines denote the prior fluxes. The yearly prior emissions and posterior tolerance intervals in TgCH_4 are noted in the legend.

4.2% of the global anthropogenic emissions. As a large portion of the oil and gas extracted in Siberia is exported to Europe, the inclusion of the emissions from the Siberian extraction process would have a significant impact on the total European emissions of $\sim 27 \text{TgCH}_4$ (as computed from EDGAR v4.2 for all Europe apart from Russia, Ukraine and Belarus). The uncertainties of the wetland emissions are still too high to provide valuable insights for the modelling of these emissions, such as the start of the emitting season, or correlation to the precipitation rates or temperatures.

2.5.2.2 Wildfire influence A peak of anthropogenic emissions with large uncertainties occurs in August. The large tolerance interval is still compatible with the prior scenario. Though, analysing into details posterior fluxes in August aggregation groups, we can propose two plausible explanation to this increase un-captured by prior data sets.

First, an increase of 0.7TgCH_4 in August anthropogenic emissions is suggested in oil and gas extraction regions. This could be explained by the activities of prospecting companies. They make use of the decrease in the household demand in summer to carry out maintenance and welling operations on the oil and gas welling sites. These operations are related to punctual leaks and purging releases of gas.

Second, August is also a month with numerous forest fires. In particular, very large wild-fires occurred in Western Russia in August 2010, upwind our observation network. As said

in Sect. 2.3.3, wildfires are systematically excluded from the marginalized inversion because they generate intense plumes difficult to take into account in our model. But, as West Russian wildfires took place far away of the observing sites, emitted plumes could have been well mixed before reaching these sites. Thus, the inversion system would have failed in eliminating observations very influenced by wildfires. The increase in the mixing ratios could then have been wrongly attributed by the inversion to an increase in anthropogenic emissions. As a confirmation, an aggregation group in August embraces anthropogenic emissions from Western Russia, grouped with anthropogenic and wetland emissions in the Siberian Lowlands. The inversion suggests that the emissions are increased by +150% for this group.

2.6 Evaluation of the inversion

2.6.1 Performance on filtered out data

For any inversion system, the inversion results are to be evaluated with independent datasets. A usual way to evaluate inversion results is to carry out a leave-one-out experiment. It usually consists in 1) flagging out all the observations from a site, 2) computing the inversion with the reduced data set, and 3) comparing the prior and posterior simulated mixing ratios to the observed ones at the left-out site. In principle, the leave-one-out inversion should improve the simulated mixing ratios at the left-out sites. That is to say, the posterior mixing ratios should be closer to and more consistent with the observations at the left-out site than the prior ones, even if not assimilated in the inversion. In our case, the observation sites are far from each other and not numerous in regard to the size of the domain. Moreover, as developed in Sect. 2.4.2, the marginalized inversion explicitly informs about the constrained regions. Flagging out one site modifies the inversion results mainly in the surrounding regions where the uncertainty reduction becomes negligible compared to the complete dataset. Therefore, the optimized fluxes in our leave-one-out experiments (not shown) remain within the range of uncertainty of another one for constrained regions. This confirms that the method we use consistently accounts for the uncertainties, but is not sufficient to quantitatively evaluate the optimized fluxes.

We can also use the data points which have been filtered out by our system in order to evaluate the inversion results. As the number of filtered out observations is high, sampling biases may be expected from the filtering procedures. In addition, as only a few data points are assimilated, unrealistic fluxes could have been inferred by the inversion to fit the assimilated data leading to a flawed reproduction of the remaining observations. As one can see in Tab. VIII.2, the marginalized inversion significantly improves the simulated mixing ratios at the sites where data is used as expected. As a proof on realistic flux prescription despite the filter on the observations, for all the remaining data, the model results are also well improved for unused data.

This confirms that our method does not create sampling biases despite the high number of filtered data points. It also confirms that the increments on the fluxes are realistic from the point of view of our network.

2.6.2 GOSAT evaluation

Genuinely independent observations are required for a better evaluation of the inversion results. Long-term monitoring surface sites are scarce in Siberia, and airborne (e.g., Paris

Table VIII.2: Correlations of observed and simulated mixing ratios, prior and posterior to the inversion. Correlation coefficient r is presented for only the filtered data points used in the inversion, but also for the data points filtered out from the inversion. The number of daily available observations is also reported alongside with the number of data used in the inversion.

Site ID	Correlations			Number of observations	
	Prior	Posterior Used	Filtered	Daily raw	Used
AZV	0.70	0.89	0.89	131	54
BRZ	0.79	0.77	0.71	66	7
DEM	0.69	0.84	0.84	184	49
IGR	0.58	0.78	0.77	349	71
KRS	0.62	0.79	0.89	307	87
NOY	0.56	0.82	0.78	122	44
PAL	0.53	0.76	0.75	362	102
SDZ	0.58	0.99	0.99	32	26
TAP	0.70	0.99	0.99	41	31
UUM	0.09	0.65	0.69	51	21
VGN	0.64	0.82	0.81	204	44
YAK	0.17	0.91	0.90	272	115
ZOT	0.64	0.90	0.92	363	107

et al., 2010a; Berchet et al., 2013a) or train (e.g., Tarasova et al., 2006) measurement campaigns only provide snapshots of the Siberian atmospheric composition. Therefore, we try to evaluate the marginalized inversion results with satellite data. We choose GOSAT total column biased-corrected retrievals (see Sect. 2.3.5) and compare them to their simulated equivalents. In Fig. VIII.8, we compare the prior and posterior simulated total columns to the observed columns. Since the inversion only constrains the emissions that are upwind the assimilated surface observation sites, we focus on the GOSAT columns that are gathered downwind of the regions constrained by the surface sites. To do so, we select only the CH₄ columns, of which the model equivalent is computed with a contribution from the constrained regions larger than 2 ppb. This criterion keeps 3000 of the 25000 available GOSAT measurements in the domain of interest during the year 2010. More than 96% of the selected prior and posterior simulated columns are less than 10 ppb different from each other. This shift is not significant compared with the ~ 10 ppb of uncertainties in the GOSAT retrievals.

GOSAT zonal coverage is rather uniform. GOSAT columns thus carry information on the spatial gradients of the atmospheric methane composition. As analysing the individual total columns does not provide sufficient information about the regional emissions, we therefore now focus on the spatial gradients of the total columns. In the observed CH₄ columns, we notice large-scale negative Eastward (from Europe to West Siberia) and Northward (from China to North-East Siberia) gradients within measured total columns. These gradients of about 5 ppb per 1000 km are related to the large-scale mixing of the polluted air masses from Europe and China into clean background air masses. The contribution of the local emissions to the observed total columns is thus computed here as the local increase from the regional background in order to bypass the very large scale gradients. In Fig. VIII.9, we represent

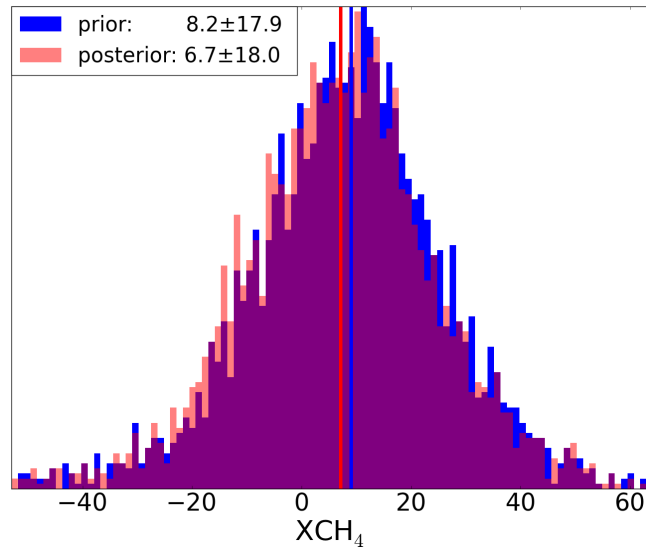


Figure VIII.8: Distribution of the differences between the observed and simulated prior and posterior GOSAT total columns. The differences are computed only for the GOSAT observations downwind the emissions constrained by the surface sites, i.e. with a simulated contribution from these regions to the total column bigger than 2 ppb.

the simulated and observed contributions of the Siberian local emissions to the GOSAT total columns. As the wetlands emit small amounts of CH_4 ($< 10^{-3} \text{ kg.m}^{-2}.\text{month}^{-1}$) spread over very large areas ($\sim 2 \times 10^6 \text{ km}^2$), their contribution to CH_4 columns is very smooth. In the model, the contribution of wetlands to the total columns does not exceed 3 ppb over the Siberian Lowlands. Therefore, the gradients in the CH_4 columns due to regional wetland emissions cannot be separated from the large scale gradients of 5 ppb per 1000 km related to the emissions outside the region of interest.

The anthropogenic hot spots are observable in both measured and simulated total columns. Above the hot spots in the Siberian Lowlands, the observed total columns shift from the observed background by up to 15 ppb. In the model, the local contribution of anthropogenic hot spots to the total columns is significant only close to the sources, with local increases in the total column up to 6 ppb. With the spatial resolution that we chose for atmospheric transport (with grid cells of more than 25 km to be compared with GOSAT pixel size of $\sim 10 \text{ km}$), the emissions from the local hot spots mix quickly into the background. This fast dilution can explain at least part of the factor of difference of 2–3 between observed and simulated CH_4 columns.

2.6.3 Toward using satellite measurements in regional frameworks

Satellite data, as a tool for filling the lack of observations in Siberia to evaluate the inversion results, do not seem to be suited for our regional configuration. As satellite observations offer a very good potential in term of spatial coverage, in-situ observations difficult to maintain and extend in the Siberian framework should be complemented by new satellite data sets.

The inversion system could be developed in order to use GOSAT data as a proxy of the

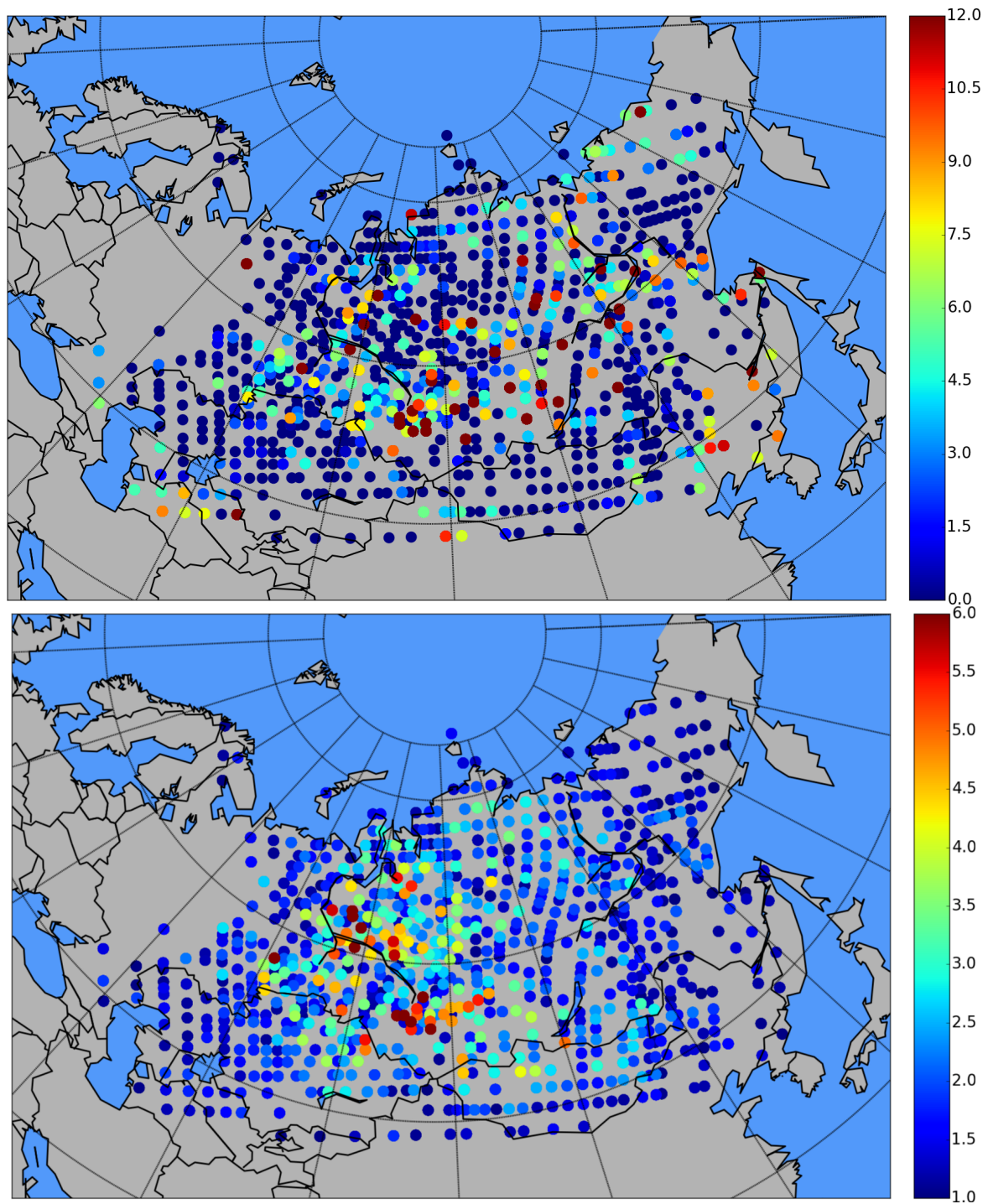


Figure VIII.9: Observed (top) and simulated (bottom) contributions of the local Siberian sources to GOSAT total columns in ppb. The observed contribution of local sources are selected as excesses from the regional measured background (defined as the average observed total column in a zone of roughly 2000 km in the zonal direction and 1000 km in the meridional one). The simulated contributions are computed from CHIMERE forward simulations of the emissions of the Siberian Lowlands. The colour scale is not the same between the two panels in order to facilitate the readability of the figure.

large scale gradients given here by a global model. With such additional observations, the LBC, representative of the emissions outside the domain of interest, would be better constrained in the inversion. As a consequence, fewer regions would be expected to be mis-separated from the LBC and the assimilated surface sites at the edge of the network would then provide more information about the emissions within the domain of interest. In addition, high resolution transport simulations close to the hot spots would better represent the dilution of the plumes. The model-observation comparison for quantifying anthropogenic sources would then be more suitable. However, the plumes generated by the hot spots are not necessarily simulated at the right location and date. Indeed, the location of emission hot spots is not always well known a priori and temporal or spatial transport mismatches can occur. Direct comparisons between the observed and simulated CH₄ columns would face the same issue as for surface sites when assimilating hot spots plumes. Recent developments (e.g., Krings et al., 2013) point to the possibility of using 1D or 2D high resolution snapshots of the hot spot plumes to infer information about very local emissions. Integrated comparison of the observed and simulated plumes in the CH₄ columns could then be implemented in a meso-scale inversion system.

With such techniques, future satellite missions with active remote sensing (e.g., the joint French-German cooperation Methane Remote Sensing LIDAR Mission, MERLIN) providing high resolution accurate 1D or 2D products could be used in regional inversions; the spatial resolution of the products to be used in such an inversion system should of the same magnitude as the meso-scale transport model, i.e. at least < 50 km.

2.7 Conclusions

We assimilated the data collected in 2010 at 8 surface observation sites measuring atmospheric CH₄ mixing ratios in the West-Siberian plain into a regional atmospheric inversion. It was the first time all these observations were used in a single study. As regional inversions suffer from mis-specified uncertainties, we implemented an enhanced Bayesian method developed by Berchet et al. (2014a) in order to get reliable results at the regional scale with an objectified quantification of the uncertainties in the system. This new method allows us to consistently evaluate the local spatial distribution of the sensitivity of the emission areas to the inversion and the usefulness of each available observation. The inversion seems to be able to primarily constrain the emissions close to the observation sites (within a radius of roughly 500 km). The inversion system assimilate daily observation aggregates to constrain the emissions. Amongst all the observation aggregates, and despite the efforts to provide precise and quasi-continuous measurements, our meso-scale inversion system properly uses only one piece of information every few days. This is mainly caused by atmospheric limitations related to transport and mixing; the fewer the assimilated data, the higher the uncertainties keep after the inversion. Even in the regions close to the stations, the posterior uncertainties (objectively quantified) thus remain larger than 20% of the prior fluxes for anthropogenic emissions and 50% for wetlands, although an important error reduction is achieved. Nevertheless, objectified uncertainties allow a robust evaluation of the wide range of proposed wetland and anthropogenic emissions in Siberia. On average, the posterior tolerance interval (defined so that 68.27% of the Monte Carlo marginalized ensemble is within the interval) on the West-Siberian plain methane budget is 5–28 TgCH₄ for the year 2010.

The year 2010 is the first year when most of the used observation sites are functional. Re-

producing our set-up to subsequent years would provide a more robust estimation of the regional fluxes and possibly valuable information about the year-to-year variability of Siberian methane fluxes. Finally, satellite platforms provide an extensive spatial coverage of observational constraints. Implementing such rather uniform observation coverage in a regional framework with few surface sites is tempting. However, with the inversion framework used here, satellite data would be useful only for constraining large scale gradients, hence the lateral boundary conditions. Further work on inversion systems is required so that satellite observations can be used to quantify local emissions in a regional framework like this one.

Acknowledgements

We thank all the PIs and supporting staff from the sites we used for maintaining the instruments and providing quasi-continuous and precise measurements. We thank Anmyeondo's (South Korea) PI for contribution to early developments of the inversion system. We thank the reviewers for their fruitful remarks and suggestions to significantly improve our first manuscript. This research was supported by the Commissariat à l'Énergie Atomique et aux Énergies Renouvelables, by the French ANR-CLIMSLIP and the YAK-AEROSIB project. The maintenance of the network is also supported by state contracts of the Ministry of Education and Science of Russia No. 14.604.21.0100, (RFMTFIBBB210290) and No. 14.613.21.0013 (RFMEFI61314X0013); and the Russian Foundation for Basic Research (grant No. 14-05-00590).

3 Des contraintes objectivées sur le budget sibérien

3.1 Des observations insuffisantes

La méthode que nous avons développée filtre les observations qu'elle ne peut assimiler. De ce fait, seule une minorité des observations est utilisée pour contraindre les flux de Sibérie. Il en résulte que les flux régionaux ne sont contraints qu'à 50–70% (en masse) par notre inversion.

Pour une région telle que la plaine de Sibérie occidentale, avec des émissions anthropiques ponctuelles et des émissions de zones inondées diffuses significatives, un réseau de huit stations est donc insuffisant. Il serait nécessaire de rajouter des stations situées optimalement pour contraindre le budget de méthane de la plaine de Sibérie dans son intégralité.

Par ailleurs, le réseau de mesures actuellement en fonction en Sibérie n'implique pas une amélioration spectaculaire de notre connaissance du budget de méthane en Sibérie. Les incertitudes quantifiées objectivement sur les régions contraintes restent élevées : 20–50% pour les émissions anthropiques et 60–120% pour les zones inondées. De plus, le système d'inversion ne peut séparer les signaux d'émissions voisines. Ainsi, il ne sait distinguer la distribution spatiale à des échelles inférieures à quelques centaines de km. Enfin, des émissions consécutives ne sont pas systématiquement disjointes par l'inversion, malgré la ventilation totale du domaine de simulations à aire limitée en quelques semaines, qui garantit une indépendance de l'état du système à deux périodes séparées du temps de ventilation. Il en découle que la variabilité temporelle détectable par notre inversion est supérieure à 2–4 semaines.

De la sorte, la résolution spatiale et temporelle de notre inversion permet de quantifier les

flux de méthane par grands ensembles de végétation et de repérer le cycle saisonnier moyen. Néanmoins, cette résolution temporelle et spatiale de l'inversion est insuffisante pour repérer le début et la fin de la saison d'émission des zones inondées en Sibérie et de séparer les émissions par processus. Un grand nombre de stations supplémentaires est nécessaire pour atteindre notre objectif d'estimation de la variabilité temporelle et de la distribution spatiale des flux de méthane en Sibérie à partir du signal atmosphérique.

3.2 Incertitude réduite sur le budget annuel de méthane en Sibérie

Nous avons donc poussé notre configuration d'inversion à ses limites. Néanmoins, si l'on estime que les flux en Sibérie se comportent d'une façon en moyenne similaire partout dans la plaine de Sibérie, nous pouvons extrapoler les résultats de l'inversion depuis les zones contraintes vers les zones invisibles pour l'inversion. Cette extrapolation repose sur une hypothèse forte, mais peut être utilisée pour estimer le budget sibérien dans son intégralité à l'échelle de l'année.

De la sorte, l'inversion suggère un intervalle de confiance pour le budget de la plaine de Sibérie occidentale de 5–28 TgCH₄.a⁻¹. Pour rappel, dans le Chapitre II, l'intervalle de confiance pour la Sibérie dans son ensemble était de 20–75 TgCH₄.a⁻¹.

La plaine de Sibérie accueillant la majeure partie des sources sibériennes de méthane, le nouvel intervalle de confiance atteste d'une nette réduction (de plus de 50%) des incertitudes sur son budget régional de méthane. Ce résultat peut être utilisé pour mieux étalonner les inventaires d'émissions anthropiques et les modèles de processus dans la région.

Quatrième partie

Conclusions et perspectives



Conclusions

La Sibérie, zone clé dans la quantification du budget mondial de méthane

Le cadre général du travail présenté dans ce rapport était l'amélioration de notre compréhension du budget mondial du méthane atmosphérique et la réduction des incertitudes dans la quantification de celui-ci. En effet, la contribution relative au budget mondial de chaque processus d'émission de méthane est encore l'objet d'incertitudes trop grandes pour garantir des projections d'émissions précises dans le cadre d'exercices de simulations climatiques à long terme. En particulier, les émissions de méthane par les zones inondées naturelles sont très sensibles – avec de surcroît une sensibilité mal chiffrée – aux conditions météorologiques et hydrologiques et donc à un changement climatique. En outre, les accords de réduction des émissions anthropiques de gaz à effet de serre nécessite des évaluations scientifiques indépendantes des déclarations officielles d'émissions de chaque pays.

Afin de diminuer les incertitudes sur les émissions mondiales de méthane, nous nous sommes intéressés à une zone clé pour le budget mondial : la Sibérie. Cette région contribue de manière notable au budget de méthane, et plus particulièrement aux incertitudes sur celui-ci. Malgré l'importance de cette région, les études s'y consacrant sont rares et ne permettent pas une bonne documentation du budget régional. De plus, la plupart des tentatives de quantification du budget régional se sont appuyées sur des approches dites « montantes » qui nécessitent d'être étalonnées.

Pour ces raisons, nous avons choisi de quantifier les émissions de méthane en Sibérie par inversion atmosphérique à la méso-échelle.

Analyse de données aéroportées pour caractériser les structures sibériennes du transport atmosphérique

L'inversion atmosphérique repose sur l'analyse statistique de la variabilité de concentrations de méthane observées en différents points. Nous avons commencé par identifier les différentes zones et types d'émissions que nous serions capables d'étudier à partir des données disponibles en Sibérie.

Pour cela, nous avons analysé les observations effectuées lors de la campagne de mesures aéroportées du projet YAK-AEROSIB d'avril 2010. Ces observations ont révélé des masses d'air dont la composition différait de manière notable par rapport à la composition atmosphérique moyenne de fond. Nous avons déterminé l'historique du transport de ces masses d'air à l'aide d'un modèle de transport lagrangien, le modèle de dispersion FLEXPART. De ces structures de transport, nous avons été à même de déduire différentes sources locales (feux de biomasse, pollution anthropiques) et distantes (pollution importée depuis la Chine et l'Europe) .

L'analyse d'observations aéroportée, doublée d'un travail de simulation du transport, nous a permis de confirmer que l'approche atmosphérique était pertinente en Sibérie. Néanmoins, nous avons aussi établi que l'analyse statistique du signal atmosphérique devrait impérativement prendre en compte l'influence des sources distantes de méthane, en plus des sources régionales que l'on cherche à déterminer.

Développement d'un nouveau contexte théorique pour les inversions bayésiennes à la méso-échelle : la marginalisation des erreurs

Les inversions de méthane classiques à la méso-échelle sont basées sur l'approche bayésienne qui permet d'inclure dans la résolution du problème les incertitudes inhérentes à celui-ci. Néanmoins, ces inversions bayésiennes sont très sensibles aux statistiques d'incertitudes spécifiées dans le système.

Les inversions de méthane à la méso-échelle sont peu courantes, en particulier en Sibérie où notre travail est pionnier à cette échelle. De ce fait, notre connaissance de ces statistiques d'incertitudes à spécifier dans l'inversion est très limitée. Notre « jugement expert » lacunaire aurait limité la pertinence des flux optimisés par inversion. Par conséquent, nous avons développé un nouveau contexte théorique aux inversions bayésiennes afin de quantifier objectivement les incertitudes liées à la mauvaise connaissance des statistiques d'erreurs.

Cette nouvelle approche bayésienne d'inversion s'appuie sur une marginalisation des statistiques d'erreurs. Toutes les statistiques d'erreurs que l'on peut spécifier dans le système sont prises en compte et leur influence sur les flux après inversion est pondérée par leur probabilité d'occurrence. De la sorte, les incertitudes sont prises en compte de manière exhaustive.

Pour des raisons pratiques et pour éviter d'avoir à nouveau recours à du « jugement expert », nous estimons la probabilité d'occurrence des statistiques à l'aide de critères statistiques objectifs. Plusieurs critères objectifs existent dans la littérature de l'assimilation de données pour les géosciences. En conséquence, nous avons testé sur un cas réaliste différents critères et en avons choisi un, le maximum de vraisemblance, pour la caractérisation des statistiques d'erreurs. Le maximum de vraisemblance apporte en effet des contraintes importantes sur ces statistiques sans pour autant saturer nos ressources de calculs numériques, à l'inverse de critères tels que celui du χ^2 trop simple ou l'analyse des propriétés algébriques des matrices de covariances d'erreurs trop complexe.

En outre, la marginalisation repose sur un échantillonnage de Monte-Carlo de la densité de probabilité des statistiques d'erreurs. La taille de l'ensemble nécessaire pour l'échantillonnage et le temps de calcul demandé par chaque échantillon dépendent de la dimension de notre problème. Pour réduire la taille de l'échantillonnage, nous avons choisi d'approximer la densité de probabilité des statistiques d'erreurs par une densité de matrices de covariances

diagonales dont les termes diagonaux peuvent être décrit par des variables aléatoires réelles indépendantes des autres termes diagonaux et suivant une loi de type χ^2 . Le temps de calcul d'inversion pour chaque échantillon est limité en projetant l'espace d'état sur un espace réduit à l'aide d'agrégation des flux a priori. Le motif de l'agrégation est choisi de manière heuristique en suivant des critères rigoureux développés par la communauté.

Une fois développée dans son ensemble, notre nouvelle approche par marginalisation a été testée et validée dans un cas académique réaliste. Une « réalité » virtuelle pour les flux et les observations a été générée. Puis, des expériences de simulation de systèmes d'observations ont été effectuées avec divers modèles de transport et flux a priori utilisés par le système d'inversion. Le système d'inversion par marginalisation s'est révélé robuste pour reproduire les erreurs dans le transport, dans la distribution des flux a priori, dans la reproduction des observations par le modèle, etc.

Réduction notable des incertitudes sur le budget régional de méthane par inversion marginalisée

La robustesse de la méthode développée nous a conforté dans notre idée d'utiliser objectivement et efficacement les mesures de concentrations de méthane collectées depuis les années 2000 en Sibérie. Avant l'année 2010, la collecte des données a été très parcellaire du fait de l'installation progressive des sites d'observations, ainsi que de nombreuses contraintes logistiques et des défaillances instrumentales. Après l'année 2010, les délais administratifs pour le transfert de données depuis la Fédération de Russie vers les instituts partenaires limitent l'accès à la base de données sibériennes. Nous avons donc appliqué notre nouveau système d'inversion par marginalisation à la Sibérie sur l'année 2010 durant laquelle le nombre de sites d'observations disponibles est le plus important.

Sur cette année, nous avons pu quantifier un budget régional de méthane dans un intervalle de confiance compris entre 5 et 28 TgCH₄.a⁻¹. Cet intervalle de confiance représente une réduction de 50% des incertitudes par rapport aux incertitudes dans les approches « montantes » consacrées à la Sibérie. Notre inversion est de plus en mesure de détecter des structures d'émissions par zones de quelques centaines de km et leur variabilité à une résolution de 2–4 semaines. De la sorte, la variabilité saisonnière des émissions des zones inondées peut être bornée par le signal atmosphérique et les émissions par grandes familles de processus peuvent être disjointes.

Nous avons donc implémenter avec succès notre approche par marginalisation nouvellement développée et avons été en mesure de quantifier efficacement et précisément le budget régional de méthane en Sibérie.



Perspectives

Mise en place systématique de l'inversion marginalisée en Sibérie

Notre nouveau système d'inversion marginalisée appliqué à la Sibérie sur un cycle saisonnier complet a débouché sur des résultats encourageants. Cependant, le budget sibérien de méthane n'est encore que partiellement contraint par les observations tels que nous les analysons.

Notre système d'inversion nous informe en effet explicitement sur les zones contraintes par les observations et sur les zones mal séparées dans le signal atmosphérique. De la sorte, nous avons vu que seuls 50–70% (en masse) des émissions de la plaine de Sibérie occidentale sont contraintes par les observations telles que nous les assimilons. De plus, les résolutions spatiales et temporelles garanties par notre système d'inversion sont encore insuffisantes pour séparer précisément les processus d'émissions (émissions par écosystèmes, fuites de gaz au niveau de l'extraction ou de la distribution des hydrocarbures, etc.) et pour analyser leurs variabilités temporelles à haute fréquence. En outre, la fenêtre d'inversion choisie du fait de l'accès difficile aux données collectées dans la région empêche une caractérisation de la variabilité des émissions à l'échelle inter-annuelle.

Des sites d'observations supplémentaires sont donc requis pour palier ces lacunes dans notre configuration d'inversion. À court terme, l'ajout de nouvelles stations de mesure des concentrations de méthane n'est cependant pas à l'ordre du jour à cause des fortes contraintes logistiques en Russie, mais aussi à cause des longues procédures administratives dans ce pays.

En attendant la disponibilité hypothétique de données supplémentaires en Sibérie, un travail approfondi doit donc être effectué sur les données existantes. Il est par exemple envisageable d'appliquer notre système d'inversion sur une fenêtre commençant en 2010 et s'achevant avec les dernières données accessibles dans la région (2012–2013). Une fois ces données collectées depuis les différents instituts responsables, nous serions à même d'estimer la capacité de notre système d'inversion d'analyser la variabilité inter-annuelle des émissions de méthane en Sibérie. Ce point est déterminant pour caractériser la sensibilité des émissions naturelles régionales aux conditions climatiques et ainsi estimer l'évolution future de ces

émissions. L'allongement de la fenêtre d'inversion en Sibérie est aussi nécessaire pour confirmer le rapport entre la production d'hydrocarbures dans la région et les fuites de méthane dans l'atmosphère, et ainsi reporter de manière précise ces fuites sur les bilans d'émissions de méthane des pays utilisateurs d'hydrocarbures sibériens (principalement la Fédération de Russie et l'Union Européenne).

Par ailleurs, dans notre système d'inversion, une partie des informations contenues dans les concentrations de méthane mesurées n'est pas utilisée : il s'agit de la variabilité à haute fréquence caractéristique des émissions locales aux sites de mesure. D'autres systèmes d'inversion que le notre, qui fonctionneraient à très haute résolution (et qu'il reste à développer), seraient susceptibles d'analyser cette variabilité à haute fréquence afin d'évaluer la réponse d'écosystèmes spécifiques aux conditions météorologiques et hydrologiques et donc d'améliorer les modèles de processus. Néanmoins, des modèles de transport à résolution très fine seraient nécessaires dans ces systèmes d'inversion et les incertitudes pourraient se révéler difficiles à caractériser. Un travail d'étude systématique de la représentation du transport et des émissions de méthane à fine échelle est nécessaire pour rendre opérationnel un tel système d'inversion. Par exemple, Matthes et al. (2014) amorce une approche en ce sens.

Une dernière approche découlant de notre travail et pouvant améliorer notre connaissance du cycle régional sibérien de méthane serait le couplage de notre système d'inversion atmosphérique à un modèle de processus. À l'heure actuelle, les émissions des modèles de processus sont considérées comme une entrée figée du système ajustée par l'inversion. Considérer les paramètres du modèle de processus comme l'état du système plutôt que les émissions seules pourrait nous permettre d'ajuster dans le modèle de processus la distribution des zones inondées et les facteurs d'émissions de méthane en fonction des précipitations, de la température, des écosystèmes, etc. Des études telles que Kuppel et al. (2012) utilisent déjà ce type d'approches avec des tours à flux.

Utilisation des observations satellitaires et aéroportées dans l'inversion à la méso-échelle

Jusqu'à présent, nous nous sommes restreints à l'implémentation d'observations de concentrations de méthane à proximité de la surface terrestre. En effet, nous avons éliminé de notre système d'inversion les données satellitaires (pour le peu de contraintes qu'elles apportent à des émissions régionales) et les données aéroportées (pour l'absence de mesures la majeure partie de l'année). Cependant, ces données sont disponibles et contiennent de l'information sur l'état de l'atmosphère sibérienne. De part la rareté des observations en Sibérie, la recherche de moyens d'utiliser efficacement et objectivement les données satellitaires et aéroportées est essentielle.

Les missions satellitaires

La faible précision des instruments d'observation des concentrations de méthane embarqués sur les satellites actuellement en orbite et l'intégration verticale du signal atmosphérique dans la mesure satellitaire empêche aujourd'hui de détecter précisément le signal de sources régionales. De plus, des biais importants et mal quantifiés apparaissent dans les observations satellitaires aux hautes latitudes (Cressot et al., 2014), ce qui dégrade encore l'information régionale apportée par les satellites. Néanmoins, la couverture large des observations satel-

litaires pourrait être utilisée pour déduire un indicateur des gradients à grande échelle des concentrations de méthane au-dessus de la Sibérie. Ces gradients seraient facilement implémentables dans une inversion marginalisée et ajouteraient des informations sur les conditions aux limites du domaine et sur les émissions totales régionales.

Par ailleurs, à moyen terme, de nouvelles missions satellitaires (par exemple, le projet franco-allemand MERLIN) proposeront de larges jeux de données à haute résolution spatiale. Cependant, le système d'inversion que nous avons développé ne peut assimiler un grand nombre de données. Des données satellitaires précises pourraient être utilisées afin de caractériser les panaches issus des sources régionales. La description des panaches observés et simulés avec un nombre réduit de paramètres (p. ex. Krings et al., 2013) pourrait être utilisées pour compléter les observations de surface.

Les campagnes de mesures aéroportées

À l'inverse des données satellitaires, les données aéroportées sont précises, avec une résolution verticale fine, mais disponibles sur des fenêtres de temps très réduites. Ces caractéristiques les rendent difficiles à implémenter dans notre système d'inversion. En effet, la structure verticale des concentrations de méthane est plus sensible à la distribution des flux de surface et à la simulation du transport que la variabilité des concentrations de méthane mesurées sur un site de surface. Par conséquent, la séparation des erreurs dans la simulation du transport et dans les flux a priori par le système d'inversion est difficile pour de telles observations. De plus, la disponibilité réduite de ces observations limite leur intérêt pour la contrainte des flux régionaux. Pour ces raisons, les données aéroportées sont en général utilisées afin d'évaluer les résultats des inversions plutôt que d'être assimilées au même titre que les observations de surface.

Néanmoins, des campagnes aéroportées telles que celles du projet YAK-AEROSIB mettent à notre disposition une documentation précise de la distribution en trois dimensions des concentrations de méthane. Des instruments de mesure de concentration de méthane sont en effet opérationnels depuis la campagne YAK-AEROSIB de juillet 2012. Nous avons vu que l'un des facteurs limitant de l'inversion atmosphérique est la qualité de la reproduction du transport par le modèle atmosphérique utilisé. Une analyse systématique des écarts entre les concentrations simulées et observées en trois dimensions lors des campagnes intensives pourrait aider notre système à caractériser au mieux les erreurs faites lors du transport de méthane.

Exploration de nouvelles hypothèses générales de marginalisation

Outre l'ajout d'informations dans notre système d'inversion, la méthode de marginalisation comme nous l'avons développée nécessite des améliorations théoriques et pratiques.

En effet, tous nos calculs reposent sur des hypothèses fortes quant aux formes des statistiques d'erreurs. Nous n'avons par exemple imposé aucune contrainte de positivité sur les variables du problème (flux et concentrations). Ce choix était motivé par l'utilisation de l'hypothèse gaussienne autorisant des valeurs négatives non-physiques pour les flux et les concentrations. Il a récemment été montré que le cadre classique d'inversion peut être étendu par des contraintes de positivité, sans alourdir les temps de calcul de manière critique (p. ex.

Koohkan et al., 2013). Ceci, appliqué à notre méthode de marginalisation, accroîtrait significativement la robustesse de nos résultats.

De plus, nous avons supposé que les matrices d'erreurs sont diagonales. Mais surtout, nous avons admis que les variables aléatoires associées à chaque terme des diagonales des matrices de covariance d'erreurs suivent une loi de χ^2 indépendante des autres termes des diagonales. La forme des lois décrivant ces variables aléatoires et l'interdépendance des termes diagonaux des matrices doivent être étudiées systématiquement afin d'affiner les hypothèses sous-jacentes à notre approche théorique. Ce travail exploratoire peut s'avérer lourd en calculs. Il devra donc dans un premier temps être effectué dans un cadre d'inversion académique simplifié.

Par ailleurs, nous avons développé un système d'inversion objectif, mais qui s'appuie sur des choix heuristiques pour les motifs d'agrégation des flux a priori. Ces choix subjectifs diminuent le caractère général de la méthode. Des travaux ont déjà été menés pour choisir optimalement les motifs d'agrégation dans une inversion, mais pour un coût numérique élevé. Une solution intermédiaire doit être explorée afin d'agréger de manière objective les flux a priori, sans dépasser nos ressources numériques. Stohl et al. (2009) et Thompson and Stohl (2014) décrivent une manière d'utiliser le modèle de dispersion FLEXPART afin de déduire une représentation simplifiée du vecteur dans l'inversion.

Transfert de la marginalisation à d'autres contextes d'inversion atmosphérique

L'application de notre méthode à la Sibérie peut être améliorée par l'apport de nouvelles observations et la méthode elle-même nécessite des ajustements pour généraliser les hypothèses sur lesquelles elle s'appuie. Néanmoins, l'inversion atmosphérique par marginalisation est une méthode générale qui peut être appliquée à d'autres espèces que le méthane et à d'autres échelles.

Le transfert à d'autres espèces et d'autres échelles permettra de valider la méthode ou de trouver des pistes d'amélioration. En effet, des artefacts ont pu rester indétectables à notre échelle et pourraient apparaître à des échelles plus fines ou plus larges. De plus, les corrélations des erreurs sur les flux gouvernant le motif d'agrégation varient d'une espèce à l'autre. L'utilisation de l'inversion marginalisée sur diverses espèces peut nous mener à des principes généraux d'agrégation des flux a priori.

À l'échelle mondiale, une inversion marginalisée peut être facilement implémentée avec les systèmes d'inversion analytiques disponibles dans la communauté (p. ex. Bousquet et al., 2011). Les erreurs calculées dans l'inversion et le transport de traceurs pourraient être mises en perspectives avec des études d'intercomparaison de modèles mondiaux comme Locatelli et al. (2013) ou Peylin et al. (2013).

À l'échelle continentale, l'intercomparaison de systèmes d'inversion atmosphérique du projet InGOS propose un cadre d'inversion européen avec des observations plus denses qu'en Sibérie et des flux mieux connus a priori. Le LSCE prend part à cette intercomparaison avec le système d'inversion variationnelle PYVAR-LMDz et l'inversion marginalisée. Nous avons choisi d'implémenter notre inversion marginalisée avec un motif d'agrégation fin du fait de la densité du réseau européen. Ce motif nous permettra d'estimer la résolution vue par notre système d'inversion en présence d'un réseau de mesures assez dense et d'explorer une mé-

thode pour un choix objectif du motif d'agrégation.

À plus fine échelle, le projet ESPiGRAD (Estimation des Sources et Puits de Gaz à effet de serre – CO₂, CH₄ et N₂O – à l'échelle Régionale par Assimilation de Données atmosphériques in-situ) porté par Isabelle Pison et Marielle Saunois (LSCE) vise à inverser les flux de CO₂, CH₄ et N₂O à haute résolution sur le territoire français. Pour cela, l'approche de l'inversion atmosphérique a été choisie avec la disponibilité de données ICOS en France et à proximité des frontières de la métropole. Ce projet est l'occasion de tester l'inversion marginalisée à fine résolution sur d'autres espèces que le méthane. La sensibilité de l'inversion au motif d'agrégation à fine échelle peut être explorée dans un tel contexte.



Bibliographie

- Environment Protection Agency : Inventory of U.S. greenhouse gas emissions and sinks : 1990-2009, Tech. Rep. 430-R-11-005, U.S. EPA, Washington D.C., 2011.
- Ahmadov, R., Gerbig, C., Kretschmer, R., Koerner, S., Neininger, B., Dolman, A. J., and Sarrat, C. : Mesoscale covariance of transport and CO₂ fluxes : Evidence from observations and simulations using the WRF-VPRM coupled atmosphere-biosphere model, *J. Geophys. Res.*, 112, D22 107, doi :10.1029/2007JD008552, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/2007JD008552/abstract>, 2007.
- Ahmadov, R., Gerbig, C., Kretschmer, R., Körner, S., Rödenbeck, C., Bousquet, P., and Ramonet, M. : Comparing high resolution WRF-VPRM simulations and two global CO₂ transport models with coastal tower measurements of CO₂, *Biogeosciences*, 6, 807–817, doi :10.5194/bg-6-807-2009, URL <http://www.biogeosciences.net/6/807/2009/>, 2009.
- Allen, D. T., Torres, V. M., Thomas, J., Sullivan, D. W., Harrison, M., Hendler, A., Herndon, S. C., Kolb, C. E., Fraser, M. P., Hill, A. D., Lamb, B. K., Miskimins, J., Sawyer, R. F., and Seinfeld, J. H. : Measurements of methane emissions at natural gas production sites in the United States, *PNAS*, p. 201304880, doi :10.1073/pnas.1304880110, URL <http://www.pnas.org/content/early/2013/09/10/1304880110>, 2013.
- Alvarado, M. J., Logan, J. A., Mao, J., Apel, E., Riemer, D., Blake, D., Cohen, R. C., Min, K.-E., Perring, A. E., Browne, E. C., Wooldridge, P. J., Diskin, G. S., Sachse, G. W., Fuelberg, H., Sessions, W. R., Harrigan, D. L., Huey, G., Liao, J., Case-Hanks, A., Jimenez, J. L., Cubison, M. J., Vay, S. A., Weinheimer, A. J., Knapp, D. J., Montzka, D. D., Flocke, F. M., Pollack, I. B., Wennberg, P. O., Kurten, A., Crounse, J., Clair, J. M. S., Wisthaler, A., Mikoviny, T., Yantosca, R. M., Carouge, C. C., and Le Sager, P. : Nitrogen oxides and PAN in plumes from boreal fires during ARCTAS-B and their impact on ozone : an integrated analysis of aircraft and satellite observations, *Atmos. Chem. Phys.*, 10, 9739–9760, doi :10.5194/acp-10-9739-2010, URL <http://www.atmos-chem-phys.net/10/9739/2010/>, 2010.
- Amante, C. and Eakins, B. W. : ETOPO1 1 arc-minute global relief model : procedures, data sources and analysis, US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, National Geophysical Data Center, Marine Geology and Geophysics Division, 2009.
- Ancellet, G., Beekmann, M., and Papayannis, A. : Impact of a cutoff low development on downward transport of ozone in the troposphere, *J. Geophys. Res.*, 99, 3451–3451, URL <http://www.physics.ntua.gr/~papayannis/5-Ancellet%20et%20al%201994-93JD02551.pdf>, 1994.
- Antokhin, P. N., Arshinov, M. Y., Belan, B. D., Davydov, D. K., Zhidovkin, E. V., Ivlev, G. A., Kozlov, A. V., Kozlov, V. S., Panchenko, M. V., Penner, I. E., et al. : Optik-\E AN-30 Aircraft Laboratory for Studies of the Atmospheric Composition, *J. Atmos. Ocean. Technol.*, URL <http://journals.ametsoc.org/doi/pdf/10.1175/2011JTECHA1427.1>, 2011.

- Arshinov, M. Y., Belan, B. D., Davydov, D. K., Inouye, G., Krasnov, O. A., Maksyutov, S., Machida, T., Fofonov, A. V., and Shimoyama, K. : Spatial and temporal variability of CO₂ and CH₄ concentrations in the surface atmospheric layer over West Siberia, *Atmos Ocean Opt*, 22, 84–93, doi :10.1134/S1024856009010126, URL <http://link.springer.com/article/10.1134/S1024856009010126>, 2009.
- Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., and Zhu, Z. : TransCom 3 inversion intercomparison : Impact of transport model errors on the interannual variability of regional CO₂ fluxes, 1988–2003, *Global Biogeochem. Cycles*, 20, GB1002, doi :10.1029/2004GB002439, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/2004GB002439/abstract>, 2006.
- Beekmann, M., Ancellet, G., Blonsky, S., De Muer, D., Ebel, A., Elbern, H., Hendricks, J., Kowol, J., Mancier, C., Sladkovic, R., et al. : Regional and global tropopause fold occurrence and related ozone flux across the tropopause, *J. Atmos. Chem.*, 28, 29–44, URL <http://www.springerlink.com/index/H0JH43714774M247.pdf>, 1997.
- Belan, B. D., Zadde, G. O., Ivlev, G. A., Krasnov, O. A., Pirogov, V. A., Simonenkov, D. V., Tolmachev, G. N., and Fofonov, A. V. : Complex assessment of the conditions of the air basin over Norilsk industrial region. Part 5. Impurities in the atmospheric boundary layer. The correspondence of air composition to hygienic norms. Recommendations., *Opt. Atmos. Okeana*, 20, 119–129, 2007.
- Belikov, I. B., Brenninkmeijer, C. a. M., Elansky, N. F., and Raľko, A. A. : Methane, carbon monoxide, and carbon dioxide concentrations measured in the atmospheric surface layer over continental Russia in the TROICA experiments, *Izv. Atmos. Ocean. Phys.*, 42, 46–59, doi :10.1134/S000143380601004X, URL <http://link.springer.com/article/10.1134/S000143380601004X>, 2006.
- Berchet, A., Paris, J.-D., Ancellet, G., Law, K. S., Stohl, A., Nédélec, P., Arshinov, M. Y., Belan, B. D., and Ciais, P. : Tropospheric ozone over Siberia in spring 2010 : remote influences and stratospheric intrusion, *Tellus B*, 65, doi :10.3402/tellusb.v65i0.19688, URL <http://www.tellusb.net/index.php/tellusb/article/view/19688/0>, 2013a.
- Berchet, A., Pison, I., Chevallier, F., Bousquet, P., Conil, S., Geever, M., Laurila, T., Lavrič, J., Lopez, M., Moncrieff, J., Necki, J., Ramonet, M., Schmidt, M., Steinbacher, M., and Tarniewicz, J. : Towards better error statistics for atmospheric inversions of methane surface fluxes, *Atmos. Chem. Phys.*, 13, 7115–7132, doi :10.5194/acp-13-7115-2013, URL <http://www.atmos-chem-phys.net/13/7115/2013/>, 2013b.
- Berchet, A., Pison, I., Chevallier, F., Bousquet, P., Bonne, J.-L., and Paris, J.-D. : Objectified quantification of uncertainties in Bayesian atmospheric inversions, *Geosci. Model Dev. Discuss.*, 7, 4777–4827, doi :10.5194/gmdd-7-4777-2014, URL <http://www.geosci-model-dev-discuss.net/7/4777/2014/>, 2014a.
- Berchet, A., Pison, I., Chevallier, F., Paris, J.-D., Bousquet, P., Bonne, J.-L., Arshinov, M. Y., Belan, B. D., Cressot, C., Davydov, D. K., Dlugokencky, E. J., Fofonov, A. V., Galanin, A., Lavrič, J., Machida, T., Parker, R., Sasakawa, M., Spahni, R., Stocker, B. D., and Winderlich, J. : Natural and anthropogenic methane fluxes in Eurasia : a meso-scale quantification by generalized atmospheric inversion, *Biogeosciences Discuss.*, 11, 14587–14637, doi :10.5194/bg-d-11-14587-2014, URL <http://www.biogeosciences-discuss.net/11/14587/2014/>, 2014b.
- Bergamaschi, P., Brenninkmeijer, C. a. M., Hahn, M., Röckmann, T., Scharffe, D. H., Crutzen, P. J., Elansky, N. F., Belikov, I. B., Trivett, N. B. A., and Worthy, D. E. J. : Isotope analysis based source identification for atmospheric CH₄ and CO sampled across Russia using the Trans-Siberian railroad, *Journal of Geophysical Research : Atmospheres*, 103, 8227–8235, doi :10.1029/97JD03738, URL <http://onlinelibrary.wiley.com/doi/10.1029/97JD03738/abstract>, 1998.
- Bergamaschi, P., Krol, M., Dentener, F., Vermeulen, A., Meinhardt, F., Graul, R., Ramonet, M., Peters, W., and Dlugokencky, E. J. : Inverse modelling of national and European CH₄ emissions using the atmospheric zoom model TM5, *Atmos. Chem. Phys.*, 5, 2431–2460, doi :10.5194/acp-5-2431-2005, URL <http://www.atmos-chem-phys.net/5/2431/2005/>, 2005.

- Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Villani, M. G., Houweling, S., Dentener, F., Dlugokencky, E. J., Miller, J. B., Gatti, L. V., Engel, A., and Levin, I. : Inverse modeling of global and regional CH₄ emissions using SCIAMACHY satellite retrievals, *Journal of Geophysical Research : Atmospheres*, 114, doi :10.1029/2009JD012287, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/2009JD012287/abstract>, 2009.
- Bergamaschi, P., Krol, M., Meirink, J. F., Dentener, F., Segers, A., van Aardenne, J., Monni, S., Vermeulen, A. T., Schmidt, M., Ramonet, M., Yver, C., Meinhardt, F., Nisbet, E. G., Fisher, R. E., O'Doherty, S., and Dlugokencky, E. J. : Inverse modeling of European CH₄ emissions 2001-2006, *J. Geophys. Res.*, 115, doi :10.1029/2010JD014180, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/2010JD014180/abstract>, 2010.
- Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R. A., Dlugokencky, E., Wofsy, S. C., Kort, E. A., Sweeney, C., Schuck, T., Brenninkmeijer, C., Chen, H., Beck, V., and Gerbig, C. : Atmospheric CH₄ in the first decade of the 21st century : Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements, *J. Geophys. Res. Atmos.*, 118, 7350–7369, doi :10.1002/jgrd.50480, URL <http://onlinelibrary.wiley.com/doi/10.1002/jgrd.50480/abstract>, 2013.
- Bocquet, M. : Toward Optimal Choices of Control Space Representation for Geophysical Data Assimilation, *Monthly Weather Review*, 137, 2331–2348, doi :10.1175/2009MWR2789.1, URL <zotero://attachment/324/>, 2009.
- Bocquet, M. : Ensemble Kalman filtering without the intrinsic need for inflation, *Nonlin. Processes Geophys.*, 18, 735–750, doi :10.5194/npg-18-735-2011, URL <http://www.nonlin-processes-geophys.net/18/735/2011/>, 2011.
- Bocquet, M. : Parameter-field estimation for atmospheric dispersion : application to the Chernobyl accident using 4D-Var, *Q.J.R. Meteorol. Soc.*, 138, 664–681, doi :10.1002/qj.961, URL <http://onlinelibrary.wiley.com/doi/10.1002/qj.961/abstract>, 2012.
- Bocquet, M., Wu, L., and Chevallier, F. : Bayesian design of control space for optimal assimilation of observations. Part I : Consistent multiscale formalism, *Q.J.R. Meteorol. Soc.*, 137, 1340–1356, doi :10.1002/qj.837, URL <http://onlinelibrary.wiley.com/doi/10.1002/qj.837/abstract>, 2011.
- Boersma, K. F., Jacob, D. J., Eskes, H. J., Pinder, R. W., Wang, J., and van der A, R. J. : Intercomparison of SCIAMACHY and OMI tropospheric NO₂ columns : Observing the diurnal evolution of chemistry and emissions from space, *J. Geophys. Res.*, 113, D16S26, doi :10.1029/2007JD008816, URL <http://onlinelibrary.wiley.com/doi/10.1029/2007JD008816/abstract>, 2008.
- Bohn, T. J., Lettenmaier, D. P., Sathulur, K., Bowling, L. C., Podest, E., McDonald, K. C., and Friborg, T. : Methane emissions from western Siberian wetlands : heterogeneity and sensitivity to climate change, *Environ. Res. Lett.*, 2, 045015, doi :10.1088/1748-9326/2/4/045015, URL <http://iopscience.iop.org/1748-9326/2/4/045015>, 2007.
- Bousquet, P., Ciais, P., Peylin, P., Ramonet, M., and Monfray, P. : Inverse modeling of annual atmospheric CO₂ sources and sinks : 1. Method and control inversion, *J. Geophys. Res.*, 104, 26 161–26 178, doi :10.1029/1999JD900342, URL <http://onlinelibrary.wiley.com/doi/10.1029/1999JD900342/abstract>, 1999a.
- Bousquet, P., Peylin, P., Ciais, P., Ramonet, M., and Monfray, P. : Inverse modeling of annual atmospheric CO₂ sources and sinks : 2. Sensitivity study, *J. Geophys. Res.*, 104, 26 179–26 193, doi :10.1029/1999JD900341, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/1999JD900341/abstract>, 1999b.
- Bousquet, P., Ciais, P., Miller, J. B., Dlugokencky, E. J., Hauglustaine, D. A., Prigent, C., Werf, G. R. V. d., Peylin, P., Brunke, E.-G., Carouge, C., Langenfelds, R. L., Lathière, J., Papa, F., Ramonet, M., Schmidt, M., Steele, L. P., Tyler, S. C., and White, J. : Contribution of anthropogenic and natural sources to atmospheric methane variability, *Nature*, 443, 439–443, doi :10.1038/nature05132, URL <http://www.nature.com/nature/journal/443/n7110/full/nature05132.html>, 2006.

- Bousquet, P., Ringeval, B., Pison, I., Dlugokencky, E. J., Brunke, E. G., Carouge, C., Chevallier, F., Fortems-Cheiney, A., Frankenberg, C., Hauglustaine, D. A., et al. : Source attribution of the changes in atmospheric methane for 2006-2008, *Atmos. Chem. Phys.*, 11, 3689–3700, URL <http://www.atmos-chem-phys.org/11/3689/2011/acp-11-3689-2011.pdf>, 2011.
- Brandt, J., Christensen, J. H., and Frohn, L. M. : Modelling transport and deposition of caesium and iodine from the Chernobyl accident using the DREAM model, *Atmos. Chem. Phys.*, 2, 397–417, doi :10.5194/acp-2-397-2002, URL <http://www.atmos-chem-phys.net/2/397/2002/>, 2002.
- Brandt, J., Silver, J. D., Christensen, J. H., Andersen, M. S., Bønløkke, J. H., Sigsgaard, T., Geels, C., Gross, A., Hansen, A. B., Hansen, K. M., Hedegaard, G. B., Kaas, E., and Frohn, L. M. : Assessment of past, present and future health-cost externalities of air pollution in Europe and the contribution from international ship traffic using the EVA model system, *Atmos. Chem. Phys.*, 13, 7747–7764, doi :10.5194/acp-13-7747-2013, URL <http://www.atmos-chem-phys.net/13/7747/2013/>, 2013.
- Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T., and Ciais, P. : A European summertime CO₂ biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, *Journal of Geophysical Research : Atmospheres*, 116, n/a–n/a, doi :10.1029/2011JD016202, URL <http://onlinelibrary.wiley.com/doi/10.1029/2011JD016202/abstract>, 2011.
- Browning, K. A. : The dry intrusion perspective of extra-tropical cyclone development, *Meteorol. Appl.*, 4, 317–324, URL <http://onlinelibrary.wiley.com/doi/10.1017/S1350482797000613/abstract>, 1997.
- Burg, J., Luenberger, D., and Wenger, D. : Estimation of structured covariance matrices, *Proceedings of the IEEE*, 70, 963–974, doi :10.1109/PROC.1982.12427, 1982.
- Cardinali, C., Pezzulli, S., and Andersson, E. : Influence-matrix diagnostic of a data assimilation system, *Q. J. R. Meteorol. Soc.*, 130, 2767–2786, doi :10.1256/qj.03.205, URL <http://onlinelibrary.wiley.com/doi/10.1256/qj.03.205/abstract>, 2004.
- Chapnik, B., Desroziers, G., Rabier, F., and Talagrand, O. : Properties and first application of an error-statistics tuning method in variational assimilation, *Q. J. R. Meteorol. Soc.*, 130, 2253–2275, doi :10.1256/qj.03.26, URL <http://onlinelibrary.wiley.com/doi/10.1256/qj.03.26/abstract>, 2004.
- Chen, Y.-H. and Prinn, R. G. : Estimation of atmospheric methane emissions between 1996 and 2001 using a three-dimensional global chemical transport model, *Journal of Geophysical Research : Atmospheres*, 111, n/a–n/a, doi :10.1029/2005JD006058, URL <http://onlinelibrary.wiley.com/doi/10.1029/2005JD006058/abstract>, 2006.
- Cherepanova, Y., Artemieva, I. M., Thybo, H., and Chemia, Z. : Crustal structure of the Siberian craton and the West Siberian basin : An appraisal of existing seismic data, *Tectonophysics*, 609, 154–183, doi :10.1016/j.tecto.2013.05.004, URL <http://www.sciencedirect.com/science/article/pii/S0040195113002989>, 2013.
- Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F. M., Chédin, A., and Ciais, P. : Inferring CO₂ sources and sinks from satellite observations : Method and application to TOVS data, *J. Geophys. Res.*, 110, D24 309, doi :10.1029/2005JD006390, URL <http://www.agu.org/pubs/crossref/2005/2005JD006390.shtml>, 2005.
- Chevallier, F., Viovy, N., Reichstein, M., and Ciais, P. : On the assignment of prior errors in Bayesian inversions of CO₂ surface fluxes, *Geophys. Res. Lett.*, 33, L13 802, doi :10.1029/2006GL026496, URL <http://www.agu.org/pubs/crossref/2006/2006GL026496.shtml>, 2006.
- Chevallier, F., Bréon, F.-M., and Rayner, P. J. : Contribution of the Orbiting Carbon Observatory to the estimation of CO₂ sources and sinks : Theoretical study in a variational data assimilation framework, *J. Geophys. Res.*, 112, D09 307, doi :10.1029/2006JD007375, URL <http://onlinelibrary.wiley.com/doi/10.1029/2006JD007375/abstract>, 2007.
- Chevallier, F., Feng, L., Bösch, H., Palmer, P. I., and Rayner, P. J. : On the impact of transport model errors for the estimation of CO₂ surface fluxes from GOSAT observations, *Geophysical Research Letters*, 37, doi :10.1029/2010GL044652, URL <http://onlinelibrary.wiley.com/doi/10.1029/2010GL044652/abstract>, 2010.

- Ciais, P., Sabine, C., Govindasamy, B., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., and Heimann, M. : Chapter 6 : Carbon and Other Biogeochemical Cycles, IPCC (2013) Climate Change, pp. 2–3, 2013.
- Cicerone, R. J. and Oremland, R. S. : Biogeochemical aspects of atmospheric methane, *Global Biogeochem. Cycles*, 2, 299–327, doi :10.1029/GB002i004p00299, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/GB002i004p00299/abstract>, 1988.
- Cohen, J., Saito, K., and Entekhabi, D. : The role of the Siberian high in northern hemisphere climate variability, *Geophysical Research Letters*, 28, 299–302, doi :10.1029/2000GL011927, URL <http://onlinelibrary.wiley.com/doi/10.1029/2000GL011927/abstract>, 2001.
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A., and Wehner, M. : Long-term Climate Change : Projections, Commitments and Irreversibility., in : *Climate Change 2013 : The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., Cambridge University Press, 2013.
- Collins, W. J., Derwent, R. G., Johnson, C. E., and Stevenson, D. S. : The impact of human activities on the photochemical production and destruction of tropospheric ozone, *Quarterly Journal of the Royal Meteorological Society*, 126, 1925–1951, doi :10.1002/qj.49712656618, URL <http://onlinelibrary.wiley.com/doi/10.1002/qj.49712656618/abstract>, 2000.
- Cooper, O., Derwent, D., Collins, B., Doherty, R., Stevenson, D., Stohl, A., and Hess, P. : Chapter 1 : Conceptual Overview of Hemispheric or Intercontinental Transport of Ozone and Particulate Matter, in : *Hemispheric Transport of Air Pollution*, no. 17 in *Air Pollution Studies*, p. 304, F. Dentener, T. Keating, and H. Akimoto, task force on hemispheric transport of air pollution edn., 2010.
- Cooper, O. R., Moody, J. L., Parrish, D. D., Trainer, M., Ryerson, T. B., Holloway, J. S., Hübler, G., Fehsenfeld, F. C., and Evans, M. J. : Trace gas composition of midlatitude cyclones over the western North Atlantic Ocean : A conceptual model, *J. Geophys. Res.*, 107, ACH 1–1, doi :10.1029/2001JD000901, URL <http://onlinelibrary.wiley.com/doi/10.1029/2001JD000901/abstract>, 2002.
- Courtier, P., Thépaut, J.-N., and Hollingsworth, A. : A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteorol. Soc.*, 120, 1367–1387, doi :10.1002/qj.49712051912, URL <http://onlinelibrary.wiley.com/doi/10.1002/qj.49712051912/abstract>, 1994.
- Cressot, C., Chevallier, F., Bousquet, P., Crevoisier, C., Dlugokencky, E. J., Fortems-Cheiney, A., Frankenberg, C., Parker, R., Pison, I., Scheepmaker, R. A., Montzka, S. A., Krummel, P. B., Steele, L. P., and Langenfelds, R. L. : On the consistency between global and regional methane emissions inferred from SCIAMACHY, TANSO-FTS, IASI and surface measurements, *Atmos. Chem. Phys.*, 14, 577–592, doi :10.5194/acp-14-577-2014, URL <http://www.atmos-chem-phys.net/14/577/2014/>, 2014.
- Cristofanelli, P., Bonasoni, P., Collins, W., Feichter, J., Forster, C., James, P., Kentarchos, A., Kubik, P. W., Land, C., Meloan, J., et al. : Stratosphere-to-troposphere transport : A model and method evaluation, *J. Geophys. Res.*, 108, 8525, URL <http://www.agu.org/journals/jd/jd0305/2002JD002600/3.shtml>, 2003.
- Crosson, E. R. : A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor, *Appl. Phys. B*, 92, 403–408, doi :10.1007/s00340-008-3135-y, URL <http://link.springer.com/article/10.1007/s00340-008-3135-y>, 2008.
- Curry, C. L. : Modeling the soil consumption of atmospheric methane at the global scale, *Global Biogeochem. Cycles*, 21, GB4012, doi :10.1029/2006GB002818, URL <http://onlinelibrary.wiley.com/doi/10.1029/2006GB002818/abstract>, 2007.
- Dedikov, J., Akopova (VNIIGaz), G., Gladkaja (VNIIGaz), N., Piotrovskij (Tyumentransgaz), A., Markellov (Volgotransgaz), V., Salichov (Yamburggazdabuicha), S., Kaesler, H., Ramm, A., Müller von Blumencron, A., and Lelieveld, J. : Estimating methane releases from natural gas production and transmission in Russia, *Atmospheric Environment*, 33, 3291–3299, doi :10.1016/S1352-2310(98)00388-4, URL <http://www.sciencedirect.com/science/article/pii/S1352231098003884>, 1999.

- Dee, D. P. : On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Weather Rev.*, 123, 1128–1145, URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.53.7382&rep=rep1&type=pdf>, 1995.
- Dee, D. P. : Bias and data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 131, 3323–3343, doi :10.1256/qj.05.137, URL <http://onlinelibrary.wiley.com/doi/10.1256/qj.05.137/abstract>, 2005.
- Denman, K. L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., Hauglustaine, D., Heinze, C., Holland, E., and Jacob, D. : Couplings between changes in the climate system and biogeochemistry, *Climate change*, 2007, 541–584, 2007.
- Denning, A. S., Holzer, M., Gurney, K. R., Heimann, M., Law, R. M., Rayner, P. J., Fung, I. Y., Fan, S.-M., Taguchi, S., Friedlingstein, P., Balkanski, Y., Taylor, J., Maiss, M., and Levin, I. : Three-dimensional transport and concentration of SF₆, *Tellus B*, 51, 266–297, doi :10.1034/j.1600-0889.1999.00012.x, URL <http://onlinelibrary.wiley.com/doi/10.1034/j.1600-0889.1999.00012.x/abstract>, 1999.
- Dentener, F., Peters, W., Krol, M., Van Weele, M., Bergamaschi, P., and Lelieveld, J. : Interannual variability and trend of CH₄ lifetime as a measure for OH changes in the 1979-1993 time period, *J. Geophys. Res.*, 108, 4442, doi :10.1029/2002JD002916, URL <http://www.agu.org/pubs/crossref/2003/2002JD002916.shtml>, 2003.
- Desroziers, G. and Ivanov, S. : Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation, *Q. J. R. Meteorol. Soc.*, 127, 1433–1452, doi :10.1002/qj.49712757417, URL <http://onlinelibrary.wiley.com/doi/10.1002/qj.49712757417/abstract>, 2001.
- Desroziers, G., Berre, L., Chapnik, B., and Poli, P. : Diagnosis of observation, background and analysis-error statistics in observation space, *Q. J. R. Meteorol. Soc.*, 131, 3385–3396, doi :10.1256/qj.05.108, URL <http://onlinelibrary.wiley.com/doi/10.1256/qj.05.108/abstract>, 2005.
- Dlugokencky, E. J., Steele, L. P., Lang, P. M., and Masarie, K. A. : The growth rate and distribution of atmospheric methane, *Journal of Geophysical Research : Atmospheres*, 99, 17 021–17 043, doi :10.1029/94JD01245, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/94JD01245/abstract>, 1994.
- Dlugokencky, E. J., Myers, R. C., Lang, P. M., Masarie, K. A., Crotwell, A. M., Thoning, K. W., Hall, B. D., Elkins, J. W., and Steele, L. P. : Conversion of NOAA atmospheric dry air CH₄ mole fractions to a gravimetrically prepared standard scale, *J. Geophys. Res.*, 110, D18 306, doi :10.1029/2005JD006035, URL <http://europa.agu.org/?uri=/journals/jd/jd0518/2005JD006035/2005JD006035.xml&view=article>, 2005.
- Dlugokencky, E. J., Bruhwiler, L., White, J. W. C., Emmons, L. K., Novelli, P. C., Montzka, S. A., Masarie, K. A., Lang, P. M., Crotwell, A. M., Miller, J. B., and Gatti, L. V. : Observational constraints on recent increases in the atmospheric CH₄ burden, *Geophys. Res. Lett.*, 36, L18 803, doi :10.1029/2009GL039780, URL <http://onlinelibrary.wiley.com/doi/10.1029/2009GL039780/abstract>, 2009.
- Elansky, N. : Russian studies of atmospheric ozone in 2007-2011, *Izv. Atmos. Ocean. Phys.*, 48, 281–298, doi :10.1134/S0001433812030024, URL <http://www.springerlink.com/content/g46784q27ln477h0/abstract/>, 2012.
- Engelen, R. J., Denning, A. S., and Gurney, K. R. : On error estimation in atmospheric CO₂ inversions, *Journal of Geophysical Research : Atmospheres*, 107, ACL 10–1–ACL 10–13, doi :10.1029/2002JD002195, URL <http://onlinelibrary.wiley.com/doi/10.1029/2002JD002195/abstract>, 2002.
- Engvall-Stjernberg, A.-C., Skorokhod, A., Paris, J.-D., Elansky, N., Nédélec, P., and Stohl, A. : Low concentrations of near-surface ozone in Siberia, *Tellus B*, 64, URL <http://www.tellusb.net/index.php/tellusb/article/view/11607>, 2012.
- Enting, I. G. : Inverse problems in atmospheric constituent transport, Cambridge University Press, URL http://books.google.fr/books?hl=fr&lr=&id=b61nA7D1XXUC&oi=fnd&pg=PP1&dq=enting+inverse+problem+atmospheric+constituent+transport+&ots=vc8hwiQ_YK&sig=u5npCEt14WdBNLrRDn9V2EfD1SQ, 2002.

- Enting, I. G., Trudinger, C. M., Francey, R. J., and Granek, H. : Synthesis inversion of atmospheric CO₂ using the GISS tracer transport model, Tech. Rep. 29, Division of Atmospheric Research Technical Paper, CSIRO, Australia, URL <http://www.csa.com/partners/viewrecord.php?requester=gs&collection=TRD&recid=N9414095AH>, 1993.
- Enting, I. G., Trudinger, C. M., and Francey, R. J. : A synthesis inversion of the concentration and $\delta^{13}\text{C}$ of atmospheric CO₂, *Tellus B*, 47, 35–52, doi :10.1034/j.1600-0889.47.issue1.5.x, URL <http://onlinelibrary.wiley.com/doi/10.1034/j.1600-0889.47.issue1.5.x/abstract>, 1995.
- Enting, I. G., Rayner, P. J., and Ciais, P. : Carbon Cycle Uncertainty in REgional Carbon Cycle Assessment and Processes (RECCAP), *Biogeosciences*, 9, 2889–2904, doi :10.5194/bg-9-2889-2012, URL <http://biogeosciences.net/9/2889/2012/bg-9-2889-2012.html>, 2012.
- Esler, J. G., Haynes, P. H., Law, K. S., Barjat, H., Dewey, K., Kent, J., Schmitgen, S., and Brough, N. : Transport and mixing between airmasses in cold frontal regions during Dynamics and Chemistry of Frontal Zones (DCFZ), *J. Geophys. Res.*, 108, 4142, doi :10.1029/2001JD001494, URL <http://onlinelibrary.wiley.com/doi/10.1029/2001JD001494/abstract>, 2003.
- Etheridge, D. M., Steele, L. P., Francey, R. J., and Langenfelds, R. L. : Atmospheric methane between 1000 A.D. and present : Evidence of anthropogenic emissions and climatic variability, *J. Geophys. Res.*, 103, 15 979–15 993, doi :10.1029/98JD00923, URL <http://onlinelibrary.wiley.com/doi/10.1029/98JD00923/abstract>, 1998.
- Etiopie, G., Lassey, K. R., Klusman, R. W., and Boschi, E. : Reappraisal of the fossil methane budget and related emission from geologic sources, *Geophys. Res. Lett.*, 35, L09 307, doi :10.1029/2008GL033623, URL <http://onlinelibrary.wiley.com/doi/10.1029/2008GL033623/abstract>, 2008.
- Evensen, G. : The Ensemble Kalman Filter : theoretical formulation and practical implementation, *Ocean Dynamics*, 53, 343–367, doi :10.1007/s10236-003-0036-9, URL <http://link.springer.com/article/10.1007/s10236-003-0036-9>, 2003.
- Feng, Z. and Kobayashi, K. : Assessing the impacts of current and future concentrations of surface ozone on crop yield with meta-analysis, *Atmospheric Environment*, 43, 1510–1519, doi :10.1016/j.atmosenv.2008.11.033, URL <http://www.sciencedirect.com/science/article/pii/S1352231008010972>, 2009.
- Field, C. B., Barros, V. R., Dokken, D. J., Mach, K. J., Mastrandrea, M. D., Bilir, T. E., Chatterjee, M., Ebi, K. L., Estrada, Y. O., Genova, R. C., Girma, B., Kissel, E. S., Levy, A. N., MacCracken, S., Mastrandrea, P. R., and White, L. : Summary for policymakers, *Climate Change 2014 : Impacts, Adaptation, and Vulnerability. Part A : Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 2014.
- Fisher, R. E., Sriskantharajah, S., Lowry, D., Lanoisellé, M., Fowler, C. M. R., James, R. H., Hermansen, O., Lund Myhre, C., Stohl, A., Greinert, J., Nisbet-Jones, P. B. R., Mienert, J., and Nisbet, E. G. : Arctic methane sources : Isotopic evidence for atmospheric inputs, *Geophys. Res. Lett.*, 38, L21 803, doi :10.1029/2011GL049319, URL <http://onlinelibrary.wiley.com/doi/10.1029/2011GL049319/abstract>, 2011.
- Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., and Myhre, G. : Changes in atmospheric constituents and in radiative forcing, *Climate change*, 20, URL <http://cfpub.epa.gov/ncea/hero/index.cfm?action=search.view=92936>, 2007.
- Frey, K. E. and Smith, L. C. : How well do we know northern land cover ? Comparison of four global vegetation and wetland products with a new ground-truth database for West Siberia, *Global Biogeochemical Cycles*, 21, doi : 10.1029/2006GB002706, URL <http://onlinelibrary.wiley.com/doi/10.1029/2006GB002706/abstract>, 2007.
- Ganesan, A. L., Rigby, M., Zammit-Mangion, A., Manning, A. J., Prinn, R. G., Fraser, P. J., Harth, C. M., Kim, K.-R., Krummel, P. B., Li, S., Mühle, J., O'Doherty, S. J., Park, S., Salameh, P. K., Steele, L. P., and Weiss, R. F. : Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods, *Atmos. Chem. Phys.*, 14, 3855–3864, doi :10.5194/acp-14-3855-2014, URL <http://www.atmos-chem-phys.net/14/3855/2014/>, 2014.

- Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P., Vermeulen, A. T., Dargaville, R., Aalto, T., Brandt, J., Christensen, J. H., Frohn, L. M., Haszpra, L., Karstens, U., Rödenbeck, C., Ramonet, M., Carboni, G., and Santaguida, R. : Comparing atmospheric transport models for future regional inversions over Europe - Part 1 : mapping the atmospheric CO₂ signals, *Atmos. Chem. Phys.*, 7, 3461–3479, doi :10.5194/acp-7-3461-2007, URL <http://www.atmos-chem-phys.net/7/3461/2007/>, 2007.
- Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A. : Toward constraining regional-scale fluxes of CO₂ with atmospheric observations over a continent : 1. Observed spatial variability from airborne platforms, *J. Geophys. Res.*, 108, 4756, doi :10.1029/2002JD003018, URL <http://www.agu.org/pubs/crossref/2003/2002JD003018.shtml>, 2003a.
- Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A. : Toward constraining regional-scale fluxes of CO₂ with atmospheric observations over a continent : 2. Analysis of COBRA data using a receptor-oriented framework, *J. Geophys. Res.*, 108, 4757, URL <http://www.agu.org/pubs/crossref/2003/2003JD003770.shtml>, 2003b.
- Giglio, L., Descloitres, J., Justice, C. O., and Kaufman, Y. J. : An enhanced contextual fire detection algorithm for MODIS, *Rem. Sens. Env.*, 87, 273–282, URL <http://www.sciencedirect.com/science/article/pii/S0034425703001846>, 2003.
- Giglio, L., Loboda, T., Roy, D. P., Quayle, B., and Justice, C. O. : An active-fire based burned area mapping algorithm for the MODIS sensor, *Remote Sensing of Environment*, 113, 408–420, doi :10.1016/j.rse.2008.10.006, URL <http://www.sciencedirect.com/science/article/pii/S0034425708003180>, 2009.
- Giglio, L., Randerson, J. T., van der Werf, G. R., Kasibhatla, P. S., Collatz, G. J., Morton, D. C., and DeFries, R. S. : Assessing variability and long-term trends in burned area by merging multiple satellite fire products., *Biogeosciences*, 7, URL <http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=17264170&AN=50295878&h=k98btkRpIk1X98YiNfsPBsazNt%2B0Z7S82fi2%2FC2eR4bykr0JyrGFgSZfqnHjI7y0q6AZNPd0HFjVNXtIoHqzBw%3D%3D&crl=c>, 2010.
- Giglio, L., Randerson, J. T., and van der Werf, G. R. : Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), *Journal of Geophysical Research : Biogeosciences*, 118, 317–328, doi :10.1002/jgrg.20042, URL <http://onlinelibrary.wiley.com/doi/10.1002/jgrg.20042/abstract>, 2013.
- Gilman, J. B., Burkhardt, J. F., Lerner, B. M., Williams, E. J., Kuster, W. C., Goldan, P. D., Murphy, P. C., Warneke, C., Fowler, C., Montzka, S. A., et al. : Ozone variability and halogen oxidation within the Arctic and sub-Arctic springtime boundary layer, *Atmos. Chem. Phys.*, 10, 10–223, URL <http://atmos-chem-phys-discuss.net/10/15885/2010/acpd-10-15885-2010.pdf>, 2010.
- Giustiniani, M., Tinivella, U., Jakobsson, M., and Rebesco, M. : Arctic Ocean Gas Hydrate Stability in a Changing Climate, *Journal of Geological Research*, 2013, e783 969, doi :10.1155/2013/783969, URL <http://www.hindawi.com/journals/jgr/2013/783969/abs/>, 2013.
- Glagolev, M., Kleptsova, I., Filippov, I., Maksyutov, S., and Machida, T. : Regional methane emission from West Siberia mire landscapes, *Environ. Res. Lett.*, 6, 045 214, doi :10.1088/1748-9326/6/4/045214, URL <http://iopscience.iop.org/1748-9326/6/4/045214>, 2011.
- Godwin, C. M., McNamara, P. J., and Markfort, C. D. : Evening methane emission pulses from a boreal wetland correspond to convective mixing in hollows, *Journal of Geophysical Research : Biogeosciences*, 118, 994–1005, doi :10.1002/jgrg.20082, URL <http://onlinelibrary.wiley.com/doi/10.1002/jgrg.20082/abstract>, 2013.
- Grace, J. D. : Russian oil supply : performance and prospects, Oxford University Press, USA, URL <http://books.google.fr/books?hl=fr&lr=&id=Wfre0JqKAjwC&oi=fnd&pg=PR9&ots=7xxUcjrFYS&sig=EtG8o4-GPMed2wB3gHhzRRbrWKU>, 2005.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.-W. : Towards robust

- regional estimates of CO₂ sources and sinks using atmospheric transport models, *Nature*, 415, 626–630, doi : 10.1038/415626a, URL <http://www.nature.com/nature/journal/v415/n6872/abs/415626a.html>, 2002.
- Göckede, M., Turner, D. P., Michalak, A. M., Vickers, D., and Law, B. E. : Sensitivity of a subregional scale atmospheric inverse CO₂ modeling framework to boundary conditions, *J. Geophys. Res.*, 115, D24 112, doi : 10.1029/2010JD014443, URL <http://www.agu.org/pubs/crossref/2010/2010JD014443.shtml>, 2010.
- Hanson, R. S. and Hanson, T. E. : Methanotrophic bacteria., *Microbiol. Rev.*, 60, 439–471, URL <http://mmbr.asm.org/content/60/2/439>, 1996.
- Hargreaves, K. J. and Fowler, D. : Quantifying the effects of water table and soil temperature on the emission of methane from peat wetland at the field scale, *Atmospheric Environment*, 32, 3275–3282, doi :10.1016/S1352-2310(98)00082-X, URL <http://www.sciencedirect.com/science/article/pii/S135223109800082X>, 1998.
- Harris, I., Jones, P., Osborn, T., and Lister, D. : Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, *Int. J. Climatol.*, 34, 623–642, doi :10.1002/joc.3711, URL <http://onlinelibrary.wiley.com/doi/10.1002/joc.3711/abstract>, 2014.
- Harvey, S., Gowrishankar, V., and Singer, T. : Leaking Profits. The U.S. Oil and Gas Industry Can Reduce Pollution, Conserve Resources, and Make Money by Preventing Methane Waste, Tech. rep., Natural Resources Defense Council, 2012.
- Hauglustaine, D. A., Lathière, J., Szopa, S., and Folberth, G. A. : Future tropospheric ozone simulated with a climate-chemistry-biosphere model, *Geophysical Research Letters*, 32, n/a–n/a, doi :10.1029/2005GL024031, URL <http://onlinelibrary.wiley.com/doi/10.1029/2005GL024031/abstract>, 2005.
- Hayhoe, K., Kheshgi, H. S., Jain, A. K., and Wuebbles, D. J. : Substitution of Natural Gas for Coal : Climatic Effects of Utility Sector Emissions, *Climatic Change*, 54, 107–139, doi :10.1023/A:1015737505552, URL <http://link.springer.com/article/10.1023/A:1015737505552>, 2002.
- Hein, R., Crutzen, P. J., and Heimann, M. : An inverse modeling approach to investigate the global atmospheric methane cycle, *Global Biogeochem. Cycles*, 11, 43–76, doi :10.1029/96GB03043, URL <http://www.agu.org/pubs/crossref/1997/96GB03043.shtml>, 1997.
- Holmes, C. D., Prather, M. J., Søvde, O. A., and Myhre, G. : Future methane, hydroxyl, and their uncertainties : key climate and emission parameters for future predictions, *Atmos. Chem. Phys.*, 13, 285–302, doi :10.5194/acp-13-285-2013, URL <http://www.atmos-chem-phys.net/13/285/2013/>, 2013.
- Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L. : Stratosphere-troposphere exchange, *Rev. Geophys.*, 33, 403–439, doi :10.1029/95RG02097, URL <http://www.agu.org/pubs/crossref/1995/95RG02097.shtml>, 1995.
- Hourdin, F. and Armengaud, A. : The Use of Finite-Volume Methods for Atmospheric Advection of Trace Species. Part I : Test of Various Formulations in a General Circulation Model, *Mon. Wea. Rev.*, 127, 822–837, doi :10.1175/1520-0493(1999)127<0822:TUOFVM>2.0.CO;2, URL [http://journals.ametsoc.org/doi/abs/10.1175/1520-0493\(1999\)127%3C0822%3ATUOFVM%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1999)127%3C0822%3ATUOFVM%3E2.0.CO%3B2), 1999.
- Hourdin, F. and Talagrand, O. : Eulerian backtracking of atmospheric tracers. I : Adjoint derivation and parametrization of subgrid-scale transport, *Q.J.R. Meteorol. Soc.*, 132, 567–583, doi :10.1256/qj.03.198.A, URL <http://onlinelibrary.wiley.com/doi/10.1256/qj.03.198.A/abstract>, 2006.
- Hourdin, F., Talagrand, O., and Idelkadi, A. : Eulerian backtracking of atmospheric tracers. II : Numerical aspects, *Q.J.R. Meteorol. Soc.*, 132, 585–603, doi :10.1256/qj.03.198.B, URL <http://onlinelibrary.wiley.com/doi/10.1256/qj.03.198.B/abstract>, 2006.
- Houtekamer, P. L. and Mitchell, H. L. : Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Wea. Rev.*, 126, 796–811, doi :10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, URL [http://journals.ametsoc.org/doi/abs/10.1175/1520-0493\(1998\)126%3C0796:DAUAEK%3E2.0.CO;2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1998)126%3C0796:DAUAEK%3E2.0.CO;2), 1998.

- Houweling, S., Kaminski, T., Dentener, F., Lelieveld, J., and Heimann, M. : Inverse modeling of methane sources and sinks using the adjoint of a global transport model, *J. Geophys. Res.*, 104, 26 137–26 160, doi :10.1029/1999JD900428, URL <http://www.agu.org/pubs/crossref/1999/1999JD900428.shtml>, 1999.
- Houweling, S., Röckmann, T., Aben, I., Keppler, F., Krol, M. C., Meirink, J. F., Dlugokencky, E. J., and Frankenberg, C. : Atmospheric constraints on global emissions of methane from plants, *Geophys. Res. Lett.*, 33, L15 821–1, doi :10.1029/2006GL026162, URL <http://igitur-archive.library.uu.nl/phys/2010-0511-200218/UUindex.html>, 2006.
- Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C. : Unified notation for data assimilation : operational, sequential and variational, *J. Meteorol. Soc. Japan*, pp. 181–189, URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.47.2607&rep=rep1&type=pdf>, 1997.
- Ishijima, K., Patra, P. K., Takigawa, M., Machida, T., Matsueda, H., Sawa, Y., Steele, L. P., Krummel, P. B., Langenfelds, R. L., Aoki, S., and Nakazawa, T. : Stratospheric influence on the seasonal cycle of nitrous oxide in the troposphere as deduced from aircraft observations and model simulations, *J. Geophys. Res.*, 115, D20 308, doi : 10.1029/2009JD013322, URL <http://onlinelibrary.wiley.com/doi/10.1029/2009JD013322/abstract>, 2010.
- Jaffe, D. A. and Wigder, N. L. : Ozone production from wildfires : A critical review, *Atmospheric Environment*, 51, 1–10, doi :10.1016/j.atmosenv.2011.11.063, URL <http://www.sciencedirect.com/science/article/pii/S1352231011012507>, 2012.
- James, P., Stohl, A., Forster, C., Eckhardt, S., Seibert, P., and Frank, A. : A 15-year climatology of stratosphere-troposphere exchange with a Lagrangian particle dispersion model : 2. Mean climate and seasonal variability, *J. Geophys. Res.*, 108, 38–55, URL <http://www.agu.org/journals/jd/jd0305/2002JD002639/2.shtml>, 2003.
- Joabsson, A., Christensen, T. R., and Wallén, B. : Vascular plant controls on methane emissions from northern peatforming wetlands, *Trends in Ecology & Evolution*, 14, 385–388, doi :10.1016/S0169-5347(99)01649-3, URL <http://www.sciencedirect.com/science/article/pii/S0169534799016493>, 1999.
- Kaminski, T., Heimann, M., and Giering, R. : A coarse grid three-dimensional global inverse model of the atmospheric transport : 2. Inversion of the transport of CO₂ in the 1980s, *J. Geophys. Res.*, 104, 18 555–18 581, doi : 10.1029/1999JD900146, URL <http://onlinelibrary.wiley.com/doi/10.1029/1999JD900146/abstract>, 1999.
- Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G. : On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.*, 106, 4703–4715, doi :10.1029/2000JD900581, URL <http://onlinelibrary.wiley.com/doi/10.1029/2000JD900581/abstract>, 2001.
- Kasischke, E. S. and Bruhwiler, L. P. : Emissions of carbon dioxide, carbon monoxide, and methane from boreal forest fires in 1998, *Journal of Geophysical Research : Atmospheres* (1984–2012), 107, FFR–2, URL <http://www.agu.org/journals/jd/jd0223/2001JD000461/>, 2002.
- Keppler, F., Hamilton, J. T. G., McRoberts, W. C., Vigano, I., Braß, M., and Röckmann, T. : Methoxyl groups of plant pectin as a precursor of atmospheric methane : evidence from deuterium labelling studies, *New Phytologist*, 178, 808–814, doi :10.1111/j.1469-8137.2008.02411.x, URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2008.02411.x/abstract>, 2008.
- Kim, H.-S., Maksyutov, S., Glagolev, M. V., Machida, T., Patra, P. K., Sudo, K., and Inoue, G. : Evaluation of methane emissions from West Siberian wetlands based on inverse modeling, *Environ. Res. Lett.*, 6, 035 201, doi :10.1088/1748-9326/6/3/035201, URL <http://iopscience.iop.org/1748-9326/6/3/035201>, 2011.
- Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., Cameron-Smith, P., Castaldi, S., Chevallier, F., Feng, L., Fraser, A., Heimann, M., Hodson, E. L., Houweling, S., Josse, B., Fraser, P. J., Krummel, P. B., Lamarque, J.-F., Langenfelds, R. L., Le Quééré, C., Naik, V., O'Doherty, S., Palmer, P. I., Pison, I., Plummer, D., Poulter, B., Prinn, R. G., Rigby, M., Ringeval, B., Santini, M., Schmidt, M., Shindell, D. T., Simpson, I. J., Spahni, R., Steele, L. P., Strode, S. A., Sudo, K., Szopa, S., van der Werf, G. R., Voulgarakis, A., van Weele, M., Weiss, R. F., Williams, J. E., and Zeng, G. : Three decades of global methane sources and sinks, *Nature Geosci*, 6, 813–823, doi :10.1038/ngeo1955, URL <http://www.nature.com/ngeo/journal/v6/n10/full/ngeo1955.html>, 2013.

- Koohkan, M. R. and Bocquet, M. : Accounting for representativeness errors in the inversion of atmospheric constituent emissions : application to the retrieval of regional carbon monoxide fluxes, *Tellus B*, 64, doi : 10.3402/tellusb.v64i0.19047, URL <http://www.tellusb.net/index.php/tellusb/article/view/19047/0>, 2012.
- Koohkan, M. R., Bocquet, M., Roustan, Y., Kim, Y., and Seigneur, C. : Estimation of volatile organic compound emissions for Europe using data assimilation, *Atmos. Chem. Phys.*, 13, 5887–5905, doi :10.5194/acp-13-5887-2013, URL <http://www.atmos-chem-phys.net/13/5887/2013/>, 2013.
- Korontzi, S., McCarty, J., Loboda, T., Kumar, S., and Justice, C. : Global distribution of agricultural fires in croplands from 3 years of Moderate Resolution Imaging Spectroradiometer (MODIS) data, *Global Biogeochem. Cycles*, 20, GB2021, URL <http://www.agu.org/pubs/crossref/2006.../2005GB002529.shtml>, 2006.
- Kremenetski, K., Velichko, A., Borisova, O., MacDonald, G., Smith, L., Frey, K., and Orlova, L. : Peatlands of the Western Siberian lowlands : current knowledge on zonation, carbon content and Late Quaternary history, *Quaternary Science Reviews*, 22, 703–723, doi :10.1016/S0277-3791(02)00196-8, URL <http://www.sciencedirect.com/science/article/pii/S0277379102001968>, 2003.
- Krings, T., Gerilowski, K., Buchwitz, M., Hartmann, J., Sachs, T., Erzinger, J., Burrows, J. P., and Bovensmann, H. : Quantification of methane emission rates from coal mine ventilation shafts using airborne remote sensing data, *Atmos. Meas. Tech.*, 6, 151–166, doi :10.5194/amt-6-151-2013, URL <http://www.atmos-meas-tech.net/6/151/2013/>, 2013.
- Kukavskaya, E. A., Soja, A. J., Petkov, A. P., Ponomarev, E. I., Ivanova, G. A., and Conard, S. G. : Fire emissions estimates in Siberia : evaluation of uncertainties in area burned, land cover, and fuel consumption, *Can. J. For. Res.*, 43, 493–506, doi :10.1139/cjfr-2012-0367, URL <http://www.nrcresearchpress.com/doi/abs/10.1139/cjfr-2012-0367>, 2012.
- Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, A. D. : Constraining a global ecosystem model with multi-site eddy-covariance data, *Biogeosciences*, 9, 3757–3776, doi :10.5194/bg-9-3757-2012, URL <http://www.biogeosciences.net/9/3757/2012/>, 2012.
- Kvenvolden, K. A. : Gas hydrates—geological perspective and global change, *Rev. Geophys.*, 31, 173–187, doi : 10.1029/93RG00268, URL <http://onlinelibrary.wiley.com/doi/10.1029/93RG00268/abstract>, 1993.
- Lamarque, J.-F., Kyle, G. P., Meinshausen, M., Riahi, K., Smith, S. J., Vuuren, D. P. v., Conley, A. J., and Vitt, F. : Global and regional evolution of short-lived radiatively-active gases and aerosols in the Representative Concentration Pathways, *Climatic Change*, 109, 191–212, doi :10.1007/s10584-011-0155-0, URL <http://link.springer.com/article/10.1007/s10584-011-0155-0>, 2011.
- Lauvaux, T. and Davis, K. J. : Planetary boundary layer errors in mesoscale inversions of column-integrated CO₂ measurements, *Journal of Geophysical Research : Atmospheres*, 119, 490–508, doi :10.1002/2013JD020175, URL <http://onlinelibrary.wiley.com/doi/10.1002/2013JD020175/abstract>, 2014.
- Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K. J., Ciais, P., Denning, A. S., and Rayner, P. J. : Mesoscale inversion : first results from the CERES campaign with synthetic data, *Atmos. Chem. Phys.*, 8, 3459–3471, doi :10.5194/acp-8-3459-2008, URL <http://www.atmos-chem-phys.net/8/3459/2008/>, 2008.
- Lauvaux, T., Pannekoucke, O., Sarrat, C., Chevallier, F., Ciais, P., Noilhan, J., Rayner, P. J., et al. : Structure of the transport uncertainty in mesoscale inversions of CO₂ sources and sinks using ensemble model simulations, *Biogeosciences*, 6, 1089–1102, URL <http://content.imamu.edu.sa/Scholars/it/net/bg-6-1089-2009.pdf>, 2009.
- Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., Sweeney, C., Diaz, L. I., Martins, D., Shepson, P. B., and Davis, K. J. : Constraining the CO₂ budget of the corn belt : exploring uncertainties from the assumptions in a mesoscale inverse system, *Atmos. Chem. Phys.*, 12, 337–354, doi :10.5194/acp-12-337-2012, URL <http://www.atmos-chem-phys.net/12/337/2012/>, 2012.

- Lauvernet, C., Brankart, J.-M., Castruccio, F., Broquet, G., Brasseur, P., and Verron, J. : A truncated Gaussian filter for data assimilation with inequality constraints : Application to the hydrostatic stability condition in ocean models, *Ocean Modelling*, 27, 1–17, doi :10.1016/j.oceomod.2008.10.007, URL <http://www.sciencedirect.com/science/article/pii/S1463500308001583>, 2009.
- Law, R. M., Chen, Y.-H., Gurney, K. R., and 3 Modellers, T. : TransCom 3 CO₂ inversion intercomparison : 2. Sensitivity of annual mean results to data choices, *Tellus B*, 55, 580–595, doi :10.1034/j.1600-0889.2003.00053.x, URL <http://onlinelibrary.wiley.com/doi/10.1034/j.1600-0889.2003.00053.x/abstract>, 2003.
- Lechtenböhmer, S., Dienst, C., Fishedick, M., Hanke, T., Langrock, T., Assonov, S. S., and Brenninkmeijer, C. : Greenhouse gas emissions from the Russian natural gas export pipeline system : results and extrapolation of measurements and surveys in Russia, p. 40, URL <http://epub.wupperinst.org/frontdoor/index/index/docId/2136>, 2005.
- Lehner, B. and Döll, P. : Development and validation of a global database of lakes, reservoirs and wetlands, *Journal of Hydrology*, 296, 1–22, doi :10.1016/j.jhydrol.2004.03.028, URL <http://www.sciencedirect.com/science/article/pii/S0022169404001404>, 2004.
- Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A. : A near-field tool for simulating the upstream influence of atmospheric observations : The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *Journal of Geophysical Research : Atmospheres*, 108, doi :10.1029/2002JD003161, URL <http://onlinelibrary.wiley.com/doi/10.1029/2002JD003161/abstract>, 2003.
- Lobert, J. M., Scharffe, D. H., Hao, W., Kuhlbusch, T. A., Seuwen, R., Warneck, P., and Crutzen, P. J. : Experimental evaluation of biomass burning emissions : Nitrogen and carbon containing compounds, URL http://inis.iaea.org/Search/search.aspx?orig_q=RN:23066772, 1991.
- Locatelli, R., Bousquet, P., Chevallier, F., Fortems-Cheney, A., Szopa, S., Saunois, M., Agusti-Panareda, A., Bergmann, D., Bian, H., Cameron-Smith, P., Chipperfield, M. P., Gloor, E., Houweling, S., Kawa, S. R., Krol, M., Patra, P. K., Prinn, R. G., Rigby, M., Saito, R., and Wilson, C. : Impact of transport model errors on the global and regional methane emissions estimated by inverse modelling, *Atmospheric Chemistry and Physics Discussions*, 13, 10961–11 021, doi :10.5194/acpd-13-10961-2013, URL <http://www.atmos-chem-phys-discuss.net/13/10961/2013/acpd-13-10961-2013.html>, 2013.
- Locatelli, R., Bousquet, P., Hourdin, F., Saunois, M., Cozic, A., Couvreux, F., Grandpeix, J.-Y., Lefebvre, M.-P., Rio, C., Bergamaschi, P., Chambers, S. D., Karstens, U., Kazan, V., van der Laan, S., Meijer, H. A. J., Moncrieff, J., Ramonet, M., Scheeren, H. A., Schlosser, C., Schmidt, M., Vermeulen, A., and Williams, A. G. : Atmospheric transport and chemistry of trace gases in LMDz5B : evaluation and implications for inverse modelling, *Geosci. Model Dev. Discuss.*, 7, 4993–5048, doi :10.5194/gmdd-7-4993-2014, URL <http://www.geosci-model-dev-discuss.net/7/4993/2014/>, 2014.
- Lopez, M. : Estimation des \`emissions de gaz \`a effet de serre \`a diff\`erentes \`echelles en France \`a l\`aide d\`observations de haute pr\`ecision., Ph.D. thesis, Univ. Paris-Sud, 2012.
- Lopez, M., Schmidt, M., Yver, C., Messenger, C., Worthy, D., Kazan, V., Ramonet, M., Bousquet, P., and Ciais, P. : Seasonal variation of N₂O emissions in France inferred from atmospheric N₂O and ²²²Rn measurements, *J. Geophys. Res.*, 117, D14 103, doi :10.1029/2012JD017703, URL <http://onlinelibrary.wiley.com/doi/10.1029/2012JD017703/abstract>, 2012.
- Lujala, P., Rod, J. K., and Thieme, N. : Fighting over Oil : Introducing a New Dataset, *Conflict Management and Peace Science*, 24, 239–256, doi :10.1080/07388940701468526, URL <http://www.tandfonline.com/doi/abs/10.1080/07388940701468526>, 2007.
- Macdonald, J. A., Fowler, D., Hargreaves, K. J., Skiba, U., Leith, I. D., and Murray, M. B. : Methane emission rates from a northern wetland ; response to temperature, water table and transport, *Atmospheric Environment*, 32, 3219–3227, doi :10.1016/S1352-2310(97)00464-0, URL <http://www.sciencedirect.com/science/article/pii/S1352231097004640>, 1998.

- Matthes, J. H., Sturtevant, C., Verfaillie, J., Knox, S., and Baldocchi, D. : Parsing the variability in CH₄ flux at a spatially heterogeneous wetland : Integrating multiple eddy covariance towers with high-resolution flux footprint analysis, *J. Geophys. Res. Biogeosci.*, p. 2014JG002642, doi :10.1002/2014JG002642, URL <http://onlinelibrary.wiley.com/doi/10.1002/2014JG002642/abstract>, 2014.
- Matthews, E. and Fung, I. : Methane emission from natural wetlands : Global distribution, area, and environmental characteristics of sources, *Global Biogeochem. Cycles*, 1, 61–86, doi :10.1029/GB001i001p00061, URL <http://onlinelibrary.wiley.com.biblioplanets.gate.inist.fr/doi/10.1029/GB001i001p00061/abstract>, 1987.
- Matthews, E., Fung, I., and Lerner, J. : Methane emission from rice cultivation : Geographic and seasonal distribution of cultivated areas and emissions, *Global Biogeochem. Cycles*, 5, 3–24, doi :10.1029/90GB02311, URL <http://onlinelibrary.wiley.com.biblioplanets.gate.inist.fr/doi/10.1029/90GB02311/abstract>, 1991.
- McGuire, A. D., Anderson, L. G., Christensen, T. R., Dallimore, S., Guo, L., Hayes, D. J., Heimann, M., Lorenson, T. D., Macdonald, R. W., and Roulet, N. : Sensitivity of the carbon cycle in the Arctic to climate change, *Ecological Monographs*, 79, 523–555, doi :10.1890/08-2025.1, URL <http://www.esajournals.org/doi/abs/10.1890/08-2025.1>, 2009.
- McRae, D. J., Conard, S. G., Ivanova, G. A., Sukhinin, A. I., Baker, S. P., Samsonov, Y. N., Blake, T. W., Ivanov, V. A., Ivanov, A. V., Churkina, T. V., Hao, W. M., Koutzenogij, K. P., and Kovaleva, N. : Variability of Fire Behavior, Fire Effects, and Emissions in Scotch Pine Forests of Central Siberia, *Mitig Adapt Strat Glob Change*, 11, 45–74, doi :10.1007/s11027-006-1008-4, URL <http://link.springer.com/article/10.1007/s11027-006-1008-4>, 2006.
- Meloan, J., Siegmund, P., van Velthoven, P., Kelder, H., Sprenger, M., Wernli, H., Kentarchos, A., Roelofs, G., Feichter, J., Land, C., Forster, C., James, P., Stohl, A., Collins, W., and Cristofanelli, P. : Stratosphere-troposphere exchange : A model and method intercomparison, *Journal of Geophysical Research : Atmospheres*, 108, n/a–n/a, doi :10.1029/2002JD002274, URL <http://onlinelibrary.wiley.com/doi/10.1029/2002JD002274/abstract>, 2003.
- Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., Eliseev, A. V., Denisov, S. N., Hopcroft, P. O., Lettenmaier, D. P., Riley, W. J., Singarayer, J. S., Subin, Z. M., Tian, H., Zürcher, S., Brovkin, V., van Bodegom, P. M., Kleinen, T., Yu, Z. C., and Kaplan, J. O. : Present state of global wetland extent and wetland methane modelling : conclusions from a model intercomparison project (WETCHIMP), *Biogeosciences*, 10, 753–788, doi :10.5194/bg-10-753-2013, URL <http://www.biogeosciences.net/10/753/2013/>, 2013.
- Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., and Vivanco, M. G. : CHIMERE 2013 : a model for regional atmospheric composition modelling, *Geosci. Model Dev.*, 6, 981–1028, doi :10.5194/gmd-6-981-2013, URL <http://www.geosci-model-dev.net/6/981/2013/>, 2013.
- Metelkin, D. V., Vernikovskiy, V. A., and Kazansky, A. Y. : Siberia-from Rodinia to Eurasia, *Tectonics*.-Sn : InTech, pp. 103–136, URL <http://cdn.intechopen.com/pdfs/14074.pdf>, 2011.
- Methven, J., Arnold, S. R., Stohl, A., Evans, M. J., Avery, M., Law, K., Lewis, A. C., Monks, P. S., Parrish, D. D., Reeves, C. E., Schlager, H., Atlas, E., Blake, D. R., Coe, H., Crosier, J., Flocke, F. M., Holloway, J. S., Hopkins, J. R., McQuaid, J., Purvis, R., Rappenglück, B., Singh, H. B., Watson, N. M., Whalley, L. K., and Williams, P. I. : Establishing Lagrangian connections between observations within air masses crossing the Atlantic during the International Consortium for Atmospheric Research on Transport and Transformation experiment, *J. Geophys. Res.*, 111, D23S62, doi :10.1029/2006JD007540, URL <http://onlinelibrary.wiley.com/doi/10.1029/2006JD007540/abstract>, 2006.
- Michalak, A. M. and Kitanidis, P. K. : A method for the interpolation of nonnegative functions with an application to contaminant load estimation, *Stoch Environ Res Ris Assess*, 19, 8–23, doi :10.1007/s00477-004-0189-1, URL <http://link.springer.com/article/10.1007/s00477-004-0189-1>, 2005.

- Michalak, A. M., Hirsch, A., Bruhwiler, L., Gurney, K. R., Peters, W., and Tans, P. P. : Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, *Journal of Geophysical Research : Atmospheres*, 110, n/a–n/a, doi :10.1029/2005JD005970, URL <http://onlinelibrary.wiley.com/doi/10.1029/2005JD005970/abstract>, 2005.
- Mobbs, S. D., Bauguitte, S. J.-B., Wellpott, A., and O'Shea, S. : Using airborne measurements and modelling to determine the leak rate of the Elgin platform in 2012, in : EGU General Assembly Conference Abstracts, vol. 15, p. 12821, URL <http://adsabs.harvard.edu/abs/2013EGUGA...1512821M>, 2013.
- Montzka, S. A., Krol, M., Dlugokencky, E., Hall, B., Jöckel, P., and Lelieveld, J. : Small Interannual Variability of Global Atmospheric Hydroxyl, *Science*, 331, 67–69, doi :10.1126/science.1197640, URL <http://www.sciencemag.org/content/331/6013/67>, 2011.
- Myhre, G., Highwood, E. J., Shine, K. P., and Stordal, F. : New estimates of radiative forcing due to well mixed greenhouse gases, *Geophys. Res. Lett.*, 25, 2715–2718, doi :10.1029/98GL01908, URL <http://onlinelibrary.wiley.com/doi/10.1029/98GL01908/abstract>, 1998.
- Myhre, G., Shindell, D., Bréon, F.-M., Collins, W. J., Fuglestedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhuang, H. : Chapter 8 : Anthropogenic and natural radiative forcing, IPCC (2013) *Climate Change*, pp. 2–3, 2013.
- Ménard, R., Cohn, S. E., Chang, L. P., and Lyster, P. M. : Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part I : formulation, *Mon. Weather Rev.*, 128, 2654–2671, URL [http://journals.ametsoc.org/doi/abs/10.1175/1520-0493\(2000\)128%3C2654%3AAGSCT0%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(2000)128%3C2654%3AAGSCT0%3E2.0.CO%3B2), 2000.
- Nakano, T., Kuniyoshi, S., and Fukuda, M. : Temporal variation in methane emission from tundra wetlands in a permafrost area, northeastern Siberia, *Atmospheric Environment*, 34, 1205–1213, doi :10.1016/S1352-2310(99)00373-8, URL <http://www.sciencedirect.com/science/article/pii/S1352231099003738>, 2000.
- Necki, J., Schmidt, M., Rozanski, K., Zimnoch, M., Korus, A., Lasa, J., Graul, R., and Levin, I. : Six-year record of atmospheric carbon dioxide and methane at a high-altitude mountain site in Poland, *Tellus B*, 55, 94–104, doi :10.1034/j.1600-0889.2003.01446.x, URL <http://onlinelibrary.wiley.com/doi/10.1034/j.1600-0889.2003.01446.x/abstract>, 2003.
- Noyce, G. L., Varner, R. K., Bubier, J. L., and Frohling, S. : Effect of *Carex rostrata* on seasonal and interannual variability in peatland methane emissions, *Journal of Geophysical Research : Biogeosciences*, pp. n/a–n/a, doi :10.1002/2013JG002474, URL <http://onlinelibrary.wiley.com/doi/10.1002/2013JG002474/abstract>, 2013.
- Nédélec, P., Cammas, J. P., Thouret, V., Athier, G., Cousin, J. M., Legrand, C., Abonnel, C., Lecoœur, F., Cayez, G., and Marizy, C. : An improved infrared carbon monoxide analyser for routine measurements aboard commercial Airbus aircraft : technical validation and first scientific results of the MOZAIK III programme, *Atmos. Chem. Phys.*, 3, 1551–1564, URL <http://adsabs.harvard.edu/abs/2003ACP...3.1551N>, 2003.
- Olivier, J. G. J., Van Aardenne, J. A., Dentener, F. J., Pagliari, V., Ganzeveld, L. N., and Peters, J. A. H. W. : Recent trends in global greenhouse gas emissions : regional trends 1970–2000 and spatial distribution of key sources in 2000, *Environmental Sciences*, 2, 81–99, doi :10.1080/15693430500400345, URL <http://dx.doi.org/10.1080/15693430500400345>, 2005.
- Oltmans, S. J., Lefohn, A. S., Harris, J. M., Tarasick, D. W., Thompson, A. M., Wernli, H., Johnson, B. J., Novelli, P. C., Montzka, S. A., Ray, J. D., Patrick, L. C., Sweeney, C., Jefferson, A., Dann, T., Davies, J., Shapiro, M., and Holben, B. N. : Enhanced ozone over western North America from biomass burning in Eurasia during April 2008 as seen in surface and profile observations, *Atmospheric Environment*, 44, 4497–4509, doi :10.1016/j.atmosenv.2010.07.004, URL <http://www.sciencedirect.com/science/article/pii/S1352231010005649>, 2010.
- Panikov, N. S. and Dedysh, S. N. : Cold season CH₄ and CO₂ emission from boreal peat bogs (West Siberia) : Winter fluxes and thaw activation dynamics, *Global Biogeochemical Cycles*, 14, 1071–1080, doi :10.1029/1999GB900097, URL <http://onlinelibrary.wiley.com/doi/10.1029/1999GB900097/abstract>, 2000.

- Paris, J. D., Ciais, P., Nédélec, P., Ramonet, M., Belan, B. D., Arshinov, M., Golitsyn, G. S., Granberg, I., Stohl, A., Cayez, G., et al. : The YAK-AEROSIB transcontinental aircraft campaigns : new insights on the transport of CO₂, CO and O₃ across Siberia, *Tellus B*, 60, 551–568, URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0889.2008.00369.x/full>, 2008.
- Paris, J. D., Ciais, P., Nédélec, P., Stohl, A., Belan, B. D., Arshinov, M. Y., Carouge, C., Golitsyn, G. S., and Granberg, I. G. : New insights on the chemical composition of the Siberian air shed from the YAK-AEROSIB aircraft campaigns, *B. Am. Meteorol. Soc.*, 91, 625–641, URL <http://zardozi.nilu.no/~andreas/publications/168.pdf>, 2010a.
- Paris, J.-D., Stohl, A., Ciais, P., Nédélec, P., Belan, B. D., Arshinov, M. Y., and Ramonet, M. : Source-receptor relationships for airborne measurements of CO₂, CO and O₃ above Siberia : a cluster-based approach, *Atmos. Chem. Phys.*, 10, 1671–1687, URL <http://www.atmos-chem-phys.net/10/1671/2010/acp-10-1671-2010.pdf>, 2010b.
- Park, S.-B., Lavric, J. V., Kolle, O., Vesala, T., Panov, A., and Heimann, M. : Greenhouse gases measurements at Zotino Tall Tower Observatory (ZOTTO) in central Siberia, in : EGU General Assembly Conference Abstracts, vol. 16, p. 12570, URL <http://adsabs.harvard.edu/abs/2014EGUGA...1612570P>, 2014.
- Parker, R., Boesch, H., Cogan, A., Fraser, A., Feng, L., Palmer, P. I., Messerschmidt, J., Deutscher, N., Griffith, D. W. T., Notholt, J., Wennberg, P. O., and Wunch, D. : Methane observations from the Greenhouse Gases Observing SATellite : Comparison to ground-based TCCON data and model calculations, *Geophysical Research Letters*, 38, n/a–n/a, doi :10.1029/2011GL047871, URL <http://onlinelibrary.wiley.com/doi/10.1029/2011GL047871/abstract>, 2011.
- Parrish, D. D., Trainer, M., Holloway, J. S., Yee, J. E., Warshawsky, M. S., Fehsenfeld, F. C., Forbes, G. L., and Moody, J. L. : Relationships between ozone and carbon monoxide at surface sites in the North Atlantic region, *J. Geophys. Res.*, 103, 13 357–13, URL <http://europa.agu.org/?uri=/journals/jd/98JD00376.xml&view=article>, 1998.
- Peregon, A., Maksyutov, S., and Yamagata, Y. : An image-based inventory of the spatial structure of West Siberian wetlands, *Environ. Res. Lett.*, 4, 045 014, doi :10.1088/1748-9326/4/4/045014, URL <http://iopscience.iop.org/1748-9326/4/4/045014>, 2009.
- Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J.-M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C., Pépin, L., Ritz, C., Saltzman, E., and Stievenard, M. : Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, *Nature*, 399, 429–436, doi :10.1038/20859, URL <http://www.nature.com/nature/journal/v399/n6735/abs/399429a0.html>, 1999.
- Peylin, P., Baker, D., Sarmiento, J., Ciais, P., and Bousquet, P. : Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO₂ data, *J. Geophys. Res.*, 107, 4385, doi :10.1029/2001JD000857, URL <http://www.agu.org/pubs/crossref/2002/2001JD000857.shtml>, 2002.
- Peylin, P., Rayner, P. J., Bousquet, P., Carouge, C., Hourdin, F., Heinrich, P., Ciais, P., et al. : Daily CO₂ flux estimates over Europe from continuous atmospheric measurements : 1, inverse methodology, *Atmos. Chem. Phys.*, 5, 1647–1678, URL <http://hal.archives-ouvertes.fr/hal-00327839/>, 2005.
- Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X. : Global atmospheric carbon budget : results from an ensemble of atmospheric CO₂ inversions, *Biogeosciences*, 10, 6699–6720, doi :10.5194/bg-10-6699-2013, URL <http://www.biogeosciences.net/10/6699/2013/>, 2013.
- Pickett-Heaps, C. A., Jacob, D. J., Wecht, K. J., Kort, E. A., Wofsy, S. C., Diskin, G. S., Worthy, D. E. J., Kaplan, J. O., Bey, I., and Drevet, J. : Magnitude and seasonality of wetland methane emissions from the Hudson Bay Lowlands (Canada), *Atmos. Chem. Phys.*, 11, 3773–3779, doi :10.5194/acp-11-3773-2011, URL <http://www.atmos-chem-phys.net/11/3773/2011/>, 2011.
- Pison, I., Menut, L., and Bergametti, G. : Inverse modeling of surface NO_x anthropogenic emission fluxes in the Paris area during the Air Pollution Over Paris Region (ESQUIF) campaign, *J. Geophys. Res.*, 112, D24 302, doi :10.1029/2007JD008871, URL <http://www.agu.org/pubs/crossref/2007/2007JD008871.shtml>, 2007.

- Pison, I., Bousquet, P., Chevallier, F., Szopa, S., and Hauglustaine, D. : Multi-species inversion of CH₄, CO and H₂ emissions from surface measurements, *Atmos. Chem. Phys.*, 9, 5281–5297, doi :10.5194/acp-9-5281-2009, URL <http://www.atmos-chem-phys.net/9/5281/2009/>, 2009.
- Pommier, M., Law, K. S., Clerbaux, C., Turquety, S., Hurtmans, D., Hadji-Lazaro, J., Coheur, P. F., Schlager, H., Ancellet, G., Paris, J. D., et al. : IASI carbon monoxide validation over the Arctic during POLARCAT spring and summer campaigns, *Atmos. Chem. Phys.*, 10, 10 655–10 678, URL <http://www.atmos-chem-phys.org/10/10655/2010/acp-10-10655-2010.pdf>, 2010.
- Pommier, M., Clerbaux, C., Law, K. S., Ancellet, G., Bernath, P., Coheur, P.-F., Hadji-Lazaro, J., Hurtmans, D., Nédélec, P., Paris, J.-D., Ravetta, F., Ryerson, T. B., Schlager, H., and Weinheimer, A. J. : Analysis of IASI tropospheric O₃ data over the Arctic during POLARCAT campaigns in 2008, *Atmos. Chem. Phys.*, 12, 7371–7389, doi :10.5194/acp-12-7371-2012, URL <http://www.atmos-chem-phys.net/12/7371/2012/>, 2012.
- Popa, M. E., Gloor, M., Manning, A. C., Jordan, A., Schultz, U., Haensel, F., Seifert, T., and Heimann, M. : Measurements of greenhouse gases and related tracers at Bialystok tall tower station in Poland, *Atmos. Meas. Tech.*, 3, 407–427, doi :10.5194/amt-3-407-2010, URL <http://www.atmos-meas-tech.net/3/407/2010/amt-3-407-2010.html>, 2010.
- Portmann, R. W., Daniel, J. S., and Ravishankara, A. R. : Stratospheric ozone depletion due to nitrous oxide : influences of other gases, *Phil. Trans. R. Soc. B*, 367, 1256–1264, doi :10.1098/rstb.2011.0377, URL <http://rstb.royalsocietypublishing.org/content/367/1593/1256>, 2012.
- Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., and Klooster, S. A. : Terrestrial ecosystem production : A process model based on global satellite and surface data, *Global Biogeochem. Cycles*, 7, 811–841, doi :10.1029/93GB02725, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/93GB02725/abstract>, 1993.
- Prather, M. J., Zhu, X., Strahan, S. E., Steenrod, S. D., and Rodriguez, J. M. : Quantifying errors in trace species transport modeling, *PNAS*, 105, 19 617–19 621, doi :10.1073/pnas.0806541106, URL <http://www.pnas.org/content/105/50/19617>, 2008.
- Quay, P., Stutsman, J., Wilbur, D., Snover, A., Dlugokencky, E., and Brown, T. : The isotopic composition of atmospheric methane, *Global Biogeochemical Cycles*, 13, 445–461, doi :10.1029/1998GB900006, URL <http://onlinelibrary.wiley.com/doi/10.1029/1998GB900006/abstract>, 1999.
- Quélo, D., Krysta, M., Bocquet, M., Isnard, O., Minier, Y., and Sportisse, B. : Validation of the Polyphemus platform on the ETEX, Chernobyl and Algeciras cases, *Atmospheric Environment*, 41, 5300–5315, doi :10.1016/j.atmosenv.2007.02.035, URL <http://www.sciencedirect.com/science/article/pii/S1352231007001677>, 2007.
- Ramonet, M., Ciais, P., Aalto, T., Aulagnier, C., Chevallier, F., Cipriano, D., Conway, T. J., Haszpra, L., Kazan, V., Meinhardt, F., Paris, J.-D., Schmidt, M., Simmonds, P., Xueref-Réemy, I., and Necki, J. N. : A recent build-up of atmospheric CO₂ over Europe. Part 1 : observed signals and possible explanations, *Tellus B*, 62, 1–13, doi :10.1111/j.1600-0889.2009.00442.x, URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0889.2009.00442.x/abstract>, 2010.
- Ramonet, M., Ciais, P., Rivier, L., Laurila, T., Vermeulen, A., Geever, M., Jordan, A., Levin, I., Laurent, O., Delmotte, M., Wastine, B., Hazan, L., Schmidt, M., Tarniewicz, J., Vuillemin, C., Pison, I., Spain, G., and Paris, J.-D. : The ICOS Atmospheric Thematic Center (ATC), GAW Report No. 206, 16th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2011), 2011.
- Rayner, P. J., Enting, I. G., Francey, R. J., and Langenfelds, R. : Reconstructing the recent carbon cycle from atmospheric CO₂, $\delta^{13}C$ and O₂/N₂ observations, *Tell. B*, 51, 213–232, doi :10.1034/j.1600-0889.1999.t01-1-00008.x, URL <http://onlinelibrary.wiley.com/doi/10.1034/j.1600-0889.1999.t01-1-00008.x/abstract>, 1999.
- Real, E., Law, K. S., Weinzierl, B., Fiebig, M., Petzold, A., Wild, O., Methven, J., Arnold, S., Stohl, A., Huntrieser, H., et al. : Processes influencing ozone levels in Alaskan forest fire plumes during long-range transport over the North Atlantic, *J. Geophys. Res.*, 112, 10, URL <http://www.agu.org/journals/jd/jd0709/2006JD007576/figures.shtml>, 2007.

- Real, E., Pisso, I., Law, K. S., Legras, B., Bousseres, N., Schlager, H., Roiger, A., Attié, J. L., et al. : Toward a novel high-resolution modeling approach for the study of chemical evolution of pollutant plumes during long-range transport, *J. Geophys. Res.*, 115, D12 302, URL <http://ceraa.enpc.fr/HomePages/reale/documents/Real10JGR.pdf>, 2010.
- Reay, D., Smith, P., and Amstel, A. v. : Methane and Climate Change, Earthscan, 2010.
- Reeves, C. E., Penkett, S. A., Bauguitte, S., Law, K. S., Evans, M. J., Bandy, B. J., Monks, P. S., Edwards, G. D., Phillips, G., Barjat, H., et al. : Potential for photochemical ozone formation in the troposphere over the North Atlantic as derived from aircraft observations during ACSOE, *J. Geophys. Res.*, 107, 25, URL <http://www.agu.org/journals/jd/jd0223/2002JD002415/4.shtm1>, 2002.
- Reimann, S., Vollmer, M., Folini, D., Steinbacher, M., Hill, M., Buchmann, B., Zander, R., and Mahieu, E. : Observations of long-lived anthropogenic halocarbons at the high-Alpine site of Jungfraujoch (Switzerland) for assessment of trends and European sources, *Science of The Total Environment*, 391, 224–231, doi :10.1016/j.scitotenv.2007.10.022, URL <http://www.sciencedirect.com/science/article/pii/S0048969707010820>, 2008.
- Reischl, G. P., Majerowicz, A., Ankilow, A., Eremenko, S., and Mavliev, R. : Comparison of the Novosibirsk automated diffusion battery with the Vienna electro mobility spectrometer, *J. Aerosol Sci.*, 22, 223–228, URL <http://cat.inist.fr/?aModele=afficheN&cpsidt=19565607>, 1991.
- Reshetnikov, A. I., Paramonova, N. N., and Shashkov, A. A. : An evaluation of historical methane emissions from the Soviet gas industry, *Journal of Geophysical Research : Atmospheres*, 105, 3517–3529, doi :10.1029/1999JD900761, URL <http://onlinelibrary.wiley.com/doi/10.1029/1999JD900761/abstract>, 2000.
- Ringeval, B., de Noblet-Ducoudré, N., Ciais, P., Bousquet, P., Prigent, C., Papa, F., and Rossow, W. B. : An attempt to quantify the impact of changes in wetland extent on methane emissions on the seasonal and interannual time scales, *Global Biogeochemical Cycles*, 24, doi :10.1029/2008GB003354, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/2008GB003354/abstract>, 2010.
- Ringeval, B., Houweling, S., van Bodegom, P. M., Spahni, R., van Beek, R., Joos, F., and Röckmann, T. : Methane emissions from floodplains in the Amazon Basin : challenges in developing a process-based model for global applications, *Biogeosciences*, 11, 1519–1558, doi :10.5194/bg-11-1519-2014, URL <http://www.biogeosciences.net/11/1519/2014/>, 2014.
- Roelofs, G. J., Kentarchos, A. S., Trickl, T., Stohl, A., Collins, W. J., Crowther, R. A., Hauglustaine, D., Klonecki, A., Law, K. S., Lawrence, M. G., von Kuhlmann, R., and van Weele, M. : Intercomparison of tropospheric ozone models : Ozone transport in a complex tropopause folding event, *J. Geophys. Res.*, 108, 8529, doi :10.1029/2003JD003462, URL <http://onlinelibrary.wiley.com/doi/10.1029/2003JD003462/abstract>, 2003.
- Rypdal, K. and Winiwarer, W. : Uncertainties in greenhouse gas emission inventories - evaluation, comparability and implications, *Environmental Science & Policy*, 4, 107–116, doi :10.1016/S1462-9011(00)00113-1, URL <http://www.sciencedirect.com/science/article/pii/S1462901100001131>, 2001.
- Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M. : CO₂ flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, 3, 1919–1964, doi :10.5194/acp-3-1919-2003, URL <http://www.atmos-chem-phys.net/3/1919/2003/>, 2003.
- Sadourny, R. and Laval, K. : January and July performance of the LMD general circulation model, *New Perspectives in Climate Modeling*, 1, 173–197, 1984.
- Sanderson, M. G. : Biomass of termites and their emissions of methane and carbon dioxide : A global database, *Global Biogeochem. Cycles*, 10, 543–557, doi :10.1029/96GB01893, URL <http://onlinelibrary.wiley.com/doi/10.1029/96GB01893/abstract>, 1996.
- Sarrat, C., Noilhan, J., Dolman, A. J., Gerbig, C., Ahmadov, R., Tolk, L. F., Meesters, A. G. C. A., Hutjes, R. W. A., Ter Maat, H. W., Pérez-Landa, G., and Donier, S. : Atmospheric CO₂ modeling at the regional scale : an intercomparison of 5 meso-scale atmospheric models, *Biogeosciences*, 4, 1115–1126, doi :10.5194/bg-4-1115-2007, URL <http://www.biogeosciences.net/4/1115/2007/>, 2007.

- Sasakawa, M., Shimoyama, K., Machida, T., Tsuda, N., Suto, H., Arshinov, M., Davydov, D., Fofonov, A., Krasnov, O., Saeki, T., Koyama, Y., and Maksyutov, S. : Continuous measurements of methane from a tower network over Siberia, *Tellus B*, 62, 403–416, doi :10.1111/j.1600-0889.2010.00494.x, URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0889.2010.00494.x/abstract>, 2010.
- Sasakawa, M., Ito, A., Machida, T., Tsuda, N., Niwa, Y., Davydov, D., Fofonov, A., and Arshinov, M. : Annual variation of CH₄ emissions from the middle taiga in West Siberian Lowland (2005–2009) : a case of high CH₄ flux and precipitation rate in the summer of 2007, *Tellus B*, 64, doi :10.3402/tellusb.v64i0.17514, URL <http://www.tellusb.net/index.php/tellusb/article/view/17514>, 2012.
- Schmidt, H., Derognat, C., Vautard, R., and Beekmann, M. : A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe, *Atmos. Environ.*, 35, 6277–6297, URL <http://www.sciencedirect.com/science/article/pii/S1352231001004514>, 2001a.
- Schmidt, M., Glatzel-Mattheier, H., Sartorius, H., Worthy, D. E., and Levin, I. : Western European N₂O emissions : A top-down approach based on atmospheric observations, *J. Geophys. Res.*, 106, 5507–5516, doi :10.1029/2000JD900701, URL <http://onlinelibrary.wiley.com/doi/10.1029/2000JD900701/abstract>, 2001b.
- Schuur, E. A. G., Bockheim, J., Canadell, J. G., Euskirchen, E., Field, C. B., Goryachkin, S. V., Hagemann, S., Kuhry, P., Lafleur, P. M., Lee, H., et al. : Vulnerability of permafrost carbon to climate change : Implications for the global carbon cycle, *BioScience*, 58, 701–714, URL <http://www.bioone.org/doi/abs/10.1641/B580807>, 2008.
- Schwinger, J. and Elbern, H. : Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation : A posteriori validation of error statistics in observation space, *Journal of Geophysical Research : Atmospheres*, 115, n/a–n/a, doi :10.1029/2009JD013115, URL <http://onlinelibrary.wiley.com/doi/10.1029/2009JD013115/abstract>, 2010.
- Seinfeld J. H. and Pandis S. N. : *Atmospheric Chemistry and Physics : From Air Pollution to Climate Change*, 2nd Edition, John Wiley & Sons, Inc., 2006.
- Shakhova, N., Semiletov, I., Salyuk, A., Yusupov, V., Kosmach, D., and Gustafsson, O. : Extensive Methane Venting to the Atmosphere from Sediments of the East Siberian Arctic Shelf, *Science*, 327, 1246–1250, doi :10.1126/science.1182221, URL <http://www.sciencemag.org/content/327/5970/1246>, 2010.
- Shakhova, N., Semiletov, I., Leifer, I., Sergienko, V., Salyuk, A., Kosmach, D., Chernykh, D., Stubbs, C., Nicolsky, D., Tumskoy, V., and Gustafsson, O. : Ebullition and storm-induced methane release from the East Siberian Arctic Shelf, *Nature Geosci*, 7, 64–70, doi :10.1038/ngeo2007, URL <http://www.nature.com/ngeo/journal/v7/n1/abs/ngeo2007.html>, 2014.
- Shakina, N. P., Ivanova, A. R., Elansky, N. F., and Markova, T. A. : Transcontinental observations of surface ozone concentration in the TROICA experiments : 2. The effect of the stratosphere-troposphere exchange, *Izv. Atmos. Ocean. Phys.*, 37, S39–S48, 2001.
- Shindell, D. T., Faluvegi, G., Koch, D. M., Schmidt, G. A., Unger, N., and Bauer, S. E. : Improved Attribution of Climate Forcing to Emissions, *Science*, 326, 716–718, doi :10.1126/science.1174760, URL <http://www.sciencemag.org/content/326/5953/716>, 2009.
- Soja, A. J., Cofer, W. R., Shugart, H. H., Sukhinin, A. I., Stackhouse, P. W., McRae, D. J., and Conard, S. G. : Estimating fire emissions and disparities in boreal Siberia (1998–2002), *J. Geophys. Res.*, 109, D14S06, doi :10.1029/2004JD004570, URL <http://onlinelibrary.wiley.com/doi/10.1029/2004JD004570/abstract>, 2004.
- Sommar, J., Andersson, M. E., and Jacobi, H. W. : Circumpolar measurements of speciated mercury, ozone and carbon monoxide in the boundary layer of the Arctic Ocean, *Atmos. Chem. Phys.*, 10, 5031–5045, URL <http://adsabs.harvard.edu/abs/2010ACP...10.5031S>, 2010.
- Spahni, R., Wania, R., Neef, L., van Weele, M., Pison, I., Bousquet, P., Frankenberg, C., Foster, P. N., Joos, F., Prentice, I. C., and van Velthoven, P. : Constraining global methane emissions and uptake by ecosystems, *Biogeosciences*, 8, 1643–1665, doi :10.5194/bg-8-1643-2011, URL <http://www.biogeosciences.net/8/1643/2011/>, 2011.

- Spahni, R., Joos, F., Stocker, B. D., Steinacher, M., and Yu, Z. C. : Transient simulations of the carbon and nitrogen dynamics in northern peatlands : from the Last Glacial Maximum to the 21st century, *Clim. Past*, 9, 1287–1308, doi :10.5194/cp-9-1287-2013, URL <http://www.clim-past.net/9/1287/2013/>, 2013.
- Sprenger, M., Wernli, H., and Bourqui, M. : Stratosphere-troposphere exchange and its relation to potential vorticity streamers and cutoffs near the extratropical tropopause, *J. Atmos. Sci.*, 64, 1587–1602, URL <http://journals.ametsoc.org/doi/abs/10.1175/JAS3911.1>, 2007.
- Stocker, B. D., Roth, R., Joos, F., Spahni, R., Steinacher, M., Zaehle, S., Bouwman, L., Xu-Ri, and Prentice, I. C. : Multiple greenhouse-gas feedbacks from the land biosphere under future climate change scenarios, *Nature Clim. Change*, 3, 666–672, doi :10.1038/nclimate1864, URL <http://www.nature.com/nclimate/journal/v3/n7/full/nclimate1864.html>, 2013.
- Stocker, B. D., Spahni, R., and Joos, F. : DYPTOP : a cost-efficient TOPMODEL implementation to simulate sub-grid spatio-temporal dynamics of global wetlands and peatlands, *Geosci. Model Dev. Discuss.*, 7, 4875–4930, doi :10.5194/gmdd-7-4875-2014, URL <http://www.geosci-model-dev-discuss.net/7/4875/2014/>, 2014.
- Stohl, A. : A 1-year Lagrangian "climatology" of airstreams in the Northern Hemisphere troposphere and lowermost stratosphere, *J. Geophys. Res.*, 106, 7263–7279, doi :10.1029/2000JD900570, URL <http://www.agu.org/pubs/crossref/2001/2000JD900570.shtml>, 2001.
- Stohl, A., Hittenberger, M., and Wotawa, G. : Validation of the lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data, *Atmospheric Environment*, 32, 4245–4264, doi :10.1016/S1352-2310(98)00184-8, URL <http://www.sciencedirect.com/science/article/pii/S1352231098001848>, 1998.
- Stohl, A., Wernli, H., James, P., Bourqui, M., Forster, C., Liniger, M. A., Seibert, P., and Sprenger, M. : A new perspective of stratosphere-troposphere exchange, *B. Am. Meteorol. Soc.*, 84, 1565–1574, URL <http://zardoz.nilu.no/~andreas/publications/70.pdf>, 2003.
- Stohl, A., Forster, C., Frank, A., Seibert, P., Wotawa, G., et al. : Technical note : The Lagrangian particle dispersion model FLEXPART version 6.2, *Atmos. Chem. Phys.*, 5, 2461–2474, URL <http://hal-insu.archives-ouvertes.fr/hal-00301615/>, 2005.
- Stohl, A., Berg, T., Burkhardt, J. F., Forster, C., Herber, A., Hov, O., Lunder, C., Mcmillan, W. W., Oltmans, S., Shiobara, M., et al. : Arctic smoke : record high air pollution levels in the European Arctic due to agricultural fires in Eastern Europe in spring 2006, *Atmos. Chem. Phys.*, 7, 511–534, URL <http://hal-insu.archives-ouvertes.fr/hal-00296130/>, 2007.
- Stohl, A., Seibert, P., Arduini, J., Eckhardt, S., Fraser, P., Grealley, B. R., Lunder, C., Maione, M., Mühle, J., O'Doherty, S., Prinn, R. G., Reimann, S., Saito, T., Schmidbauer, N., Simmonds, P. G., Vollmer, M. K., Weiss, R. F., and Yokouchi, Y. : An analytical inversion method for determining regional and global emissions of greenhouse gases : Sensitivity studies and application to halocarbons, *Atmos. Chem. Phys.*, 9, 1597–1620, doi : 10.5194/acp-9-1597-2009, URL <http://www.atmos-chem-phys.net/9/1597/2009/>, 2009.
- Stohl, A., Prata, A. J., Eckhardt, S., Clarisse, L., Durant, A., Henne, S., Kristiansen, N. I., Minikin, A., Schumann, U., Seibert, P., Stebel, K., Thomas, H. E., Thorsteinsson, T., Tørseth, K., and Weinzierl, B. : Determination of time- and height-resolved volcanic ash emissions and their use for quantitative ash dispersion modeling : the 2010 Eyjafjallajökull eruption, *Atmos. Chem. Phys.*, 11, 4333–4351, doi :10.5194/acp-11-4333-2011, URL <http://www.atmos-chem-phys.net/11/4333/2011/>, 2011.
- Sugawara, S., Nakazawa, T., Inoue, G., Machida, T., Mukai, H., Vinnichenko, N. K., and Khattatov, V. U. : Aircraft measurements of the stable carbon isotopic ratio of atmospheric methane over Siberia, *Global Biogeochemical Cycles*, 10, 223–231, doi :10.1029/96GB00038, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/96GB00038/abstract>, 1996.
- Suto, H. and Inoue, G. : A New Portable Instrument for In Situ Measurement of Atmospheric Methane Mole Fraction by Applying an Improved Tin Dioxide-Based Gas Sensor, *Journal of Atmospheric and Oceanic Technology*, 27, 1175–1184, doi :10.1175/2010JTECHA1400.1, URL <http://journals.ametsoc.org/doi/abs/10.1175/2010JTECHA1400.1>, 2010.

- Talagrand, O. : A posteriori evaluation and verification of analysis and assimilation algorithms, in : Workshop on Diagnosis of Data Assimilation Systems, pp. 2–4, 1998.
- Tarantola, A. : Inverse Problem Theory, Elsevier, New York, 1987.
- Tarasova, O., Brenninkmeijer, C., Assonov, S., Elansky, N., Röckmann, T., and Brass, M. : Atmospheric CH₄ along the Trans-Siberian railroad (TROICA) and river Ob : Source identification using stable isotope analysis, *Atmospheric Environment*, 40, 5617–5628, doi :10.1016/j.atmosenv.2006.04.065, URL <http://www.sciencedirect.com/science/article/pii/S1352231006004614>, 2006.
- Tarasova, O. A., Houweling, S., Elansky, N., and Brenninkmeijer, C. A. M. : Application of stable isotope analysis for improved understanding of the methane budget : comparison of TROICA measurements with TM3 model simulations, *J Atmos Chem*, 63, 49–71, doi :10.1007/s10874-010-9157-y, URL <http://link.springer.com/article/10.1007/s10874-010-9157-y>, 2009.
- Tarnocai, C., Canadell, J. G., Schuur, E. a. G., Kuhry, P., Mazhitova, G., and Zimov, S. : Soil organic carbon pools in the northern circumpolar permafrost region, *Global Biogeochemical Cycles*, 23, n/a–n/a, doi :10.1029/2008GB003327, URL <http://onlinelibrary.wiley.com/doi/10.1029/2008GB003327/abstract>, 2009.
- Taylor, K. E. : Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, 106, 7183–7192, doi :10.1029/2000JD900719, URL <http://onlinelibrary.wiley.com/doi/10.1029/2000JD900719/abstract>, 2001.
- Thomas, J. L., Raut, J.-C., Law, K. S., Marelle, L., Ancellet, G., Ravetta, F., Fast, J. D., Pfister, G., Emmons, L. K., Diskin, G. S., Weinheimer, A., Roiger, A., and Schlager, H. : Pollution transport from North America to Greenland during summer 2008, *Atmos. Chem. Phys.*, 13, 3825–3848, doi :10.5194/acp-13-3825-2013, URL <http://www.atmos-chem-phys.net/13/3825/2013/>, 2013.
- Thompson, R. L. and Stohl, A. : FLEXINVERT : an atmospheric Bayesian inversion framework for determining surface fluxes of trace species using an optimized grid, *Geosci. Model Dev. Discuss.*, 7, 3751–3801, doi :10.5194/gmdd-7-3751-2014, URL <http://www.geosci-model-dev-discuss.net/7/3751/2014/>, 2014.
- Tissot, B. P. and Welte, D. H. : Petroleum formation and occurrence, URL <http://www.osti.gov/scitech/biblio/6632464>, 1984.
- Tohjima, Y., Maksyutov, S., Machida, T., and Inoue, G. : Airborne measurements of atmospheric methane over oil fields in western Siberia, *Geophys. Res. Lett.*, 23, 1621–1624, doi :10.1029/96GL01027, URL <http://onlinelibrary.wiley.com/doi/10.1029/96GL01027/abstract>, 1996.
- Tolk, L. F., Meesters, A. G. C. A., Dolman, A. J., and Peters, W. : Modelling representation errors of atmospheric CO₂ mixing ratios at a regional scale, *Atmos. Chem. Phys.*, 8, 6587–6596, doi :10.5194/acp-8-6587-2008, URL <http://www.atmos-chem-phys.net/8/6587/2008/>, 2008.
- Umezawa, T., Machida, T., Aoki, S., and Nakazawa, T. : Contributions of natural and anthropogenic sources to atmospheric methane variations over western Siberia estimated from its carbon and hydrogen isotopes, *Global Biogeochemical Cycles*, 26, doi :10.1029/2011GB004232, URL <http://onlinelibrary.wiley.com/biblioplanets.gate.inist.fr/doi/10.1029/2011GB004232/abstract>, 2012.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, I., Janssen, P. a. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J. : The ERA-40 re-analysis, *Q.J.R. Meteorol. Soc.*, 131, 2961–3012, doi :10.1256/qj.04.176, URL <http://onlinelibrary.wiley.com/doi/10.1256/qj.04.176/abstract>, 2005.
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T. : Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys.*, 10, 11707–11735, doi :10.5194/acp-10-11707-2010, URL <http://www.atmos-chem-phys.net/10/11707/2010/>, 2010.

- Van Dingenen, R., Dentener, F. J., Raes, F., Krol, M. C., Emberson, L., and Cofala, J. : The global impact of ozone on agricultural crop yields under current and future air quality legislation, *Atmospheric Environment*, 43, 604–618, doi :10.1016/j.atmosenv.2008.10.033, URL <http://www.sciencedirect.com/science/article/pii/S1352231008009424>, 2009.
- Vasileva, A. and Moiseenko, K. : Methane emissions from 2000 to 2011 wildfires in Northeast Eurasia estimated with MODIS burned area data, *Atmospheric Environment*, 71, 115–121, doi :10.1016/j.atmosenv.2013.02.001, URL <http://www.sciencedirect.com/science/article/pii/S1352231013000897>, 2013.
- Vaughan, G., Garland, W. E., Dewey, K. J., and Gerbig, C. : Aircraft Measurements of a Warm Conveyor Belt – A Case Study, *Journal of Atmospheric Chemistry*, 46, 117–129, doi :10.1023/A:1026025516092, URL <http://link.springer.com/article/10.1023/A:1026025516092>, 2003.
- Vautard, R., Beekmann, M., Roux, J., and Gombert, D. : Validation of a hybrid forecasting system for the ozone concentrations over the Paris area, *Atmospheric Environment*, 35, 2449–2461, doi :10.1016/S1352-2310(00)00466-0, URL <http://www.sciencedirect.com/science/article/pii/S1352231000004660>, 2001.
- Verma, S., Worden, J., Pierce, B., Jones, D. B. A., Al-Saadi, J., Boersma, F., Bowman, K., Eldering, A., Fisher, B., Jourdain, L., Kulawik, S., and Worden, H. : Ozone production in boreal fire smoke plumes using observations from the Tropospheric Emission Spectrometer and the Ozone Monitoring Instrument, *J. Geophys. Res.*, 114, D02303, doi :10.1029/2008JD010108, URL <http://onlinelibrary.wiley.com/doi/10.1029/2008JD010108/abstract>, 2009.
- Vermeulen, A. T., Team, T. C., and Pieterse, G. : Tall tower observations of greenhouse gases in Europe : possibilities for emission verification, 7th International CO₂ Conference, 2005.
- Vigano, I., van Weelden, H., Holzinger, R., Keppler, F., McLeod, A., and Röckmann, T. : Effect of UV radiation and temperature on the emission of methane from plant biomass and structural components, *Biogeosciences*, 5, 937–947, doi :10.5194/bg-5-937-2008, URL <http://www.biogeosciences.net/5/937/2008/>, 2008.
- Villani, M. G., Bergamaschi, P., Krol, M. C., Meirink, J. F., and Dentener, F. : Inverse modeling of European CH₄ emissions : sensitivity to the observational network, *Atmos. Chem. Phys.*, 10, 1249–1267, URL <http://igitur-archive.library.uu.nl/phys/2011-0323-200350/UUindex.html>, 2010.
- Wahba, G., Johnson, D. R., Gao, F., and Gong, J. : Adaptive tuning of numerical weather prediction models : Part I : randomized GCV and related methods in three and four dimensional data assimilation, Tech. rep., Citeseer, URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.6606&rep=rep1&type=pdf>, 1994.
- Wang, Y. and Jacob, D. J. : Anthropogenic forcing on tropospheric ozone and OH since preindustrial times, *J. Geophys. Res.*, 103, 31123–31135, doi :10.1029/1998JD100004, URL <http://onlinelibrary.wiley.com/doi/10.1029/1998JD100004/abstract>, 1998.
- Wang, Y., Munger, J. W., Xu, S., McElroy, M. B., Hao, J., Nielsen, C. P., and Ma, H. : CO₂ and its correlation with CO at a rural site near Beijing : implications for combustion efficiency in China, *Atmos. Chem. Phys.*, 10, 8881–8897, doi :10.5194/acp-10-8881-2010, URL <http://www.atmos-chem-phys.net/10/8881/2010/>, 2010.
- Wania, R., Melton, J. R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Chen, G., Eliseev, A. V., Hopcroft, P. O., Riley, W. J., Subin, Z. M., Tian, H., van Bodegom, P. M., Kleinen, T., Yu, Z. C., Singarayer, J. S., Zürcher, S., Lettenmaier, D. P., Beerling, D. J., Denisov, S. N., Prigent, C., Papa, F., and Kaplan, J. O. : Present state of global wetland extent and wetland methane modelling : methodology of a model inter-comparison project (WETCHIMP), *Geoscientific Model Development*, 6, 617–641, doi :10.5194/gmd-6-617-2013, URL <http://www.geosci-model-dev.net/6/617/2013/gmd-6-617-2013.html>, 2013.
- Wernli, H. and Sprenger, M. : Identification and ERA-15 Climatology of Potential Vorticity Streamers and Cutoffs near the Extratropical Tropopause, *J. Atmos. Sci.*, 64, 1569–1586, doi :10.1175/JAS3912.1, URL <http://journals.ametsoc.org/doi/abs/10.1175/JAS3912.1>, 2007.
- Wild, O., Pochanart, P., and Akimoto, H. : Trans-Eurasian transport of ozone and its precursors, *J. Geophys. Res.*, 109, n/a–n/a, doi :10.1029/2003JD004501, URL <http://onlinelibrary.wiley.com/doi/10.1029/2003JD004501/abstract>, 2004.

- Williams, I. N., Riley, W. J., Torn, M. S., Biraud, S. C., and Fischer, M. L. : Biases in regional carbon budgets from covariation of surface fluxes and weather in transport model inversions, *Atmospheric Chemistry and Physics Discussions*, 13, 19 051–19 083, doi :10.5194/acpd-13-19051-2013, URL <http://www.atmos-chem-phys-discuss.net/13/19051/2013/acpd-13-19051-2013.html>, 2013.
- Winderlich, J. : Setup of a CO₂ and CH₄ measurement system in Central Siberia and modeling of its results, Ph.D. thesis, Univ. of Hamburg, Germany, URL <http://ediss.sub.uni-hamburg.de/volltexte/2012/5533/>, 2012.
- Winderlich, J., Chen, H., Gerbig, C., Seifert, T., Kolle, O., Lavrič, J. V., Kaiser, C., Höfer, A., and Heimann, M. : Continuous low-maintenance CO₂/CH₄/H₂O measurements at the Zotino Tall Tower Observatory (ZOTTO) in Central Siberia, *Atmospheric Measurement Techniques*, 3, 1113–1128, doi :10.5194/amt-3-1113-2010, URL <http://www.atmos-meas-tech.net/3/1113/2010/amt-3-1113-2010.html>, 2010.
- Winderlich, J., Gerbig, C., Kolle, O., and Heimann, M. : Inferences from CO₂ and CH₄ concentration profiles at the Zotino Tall Tower Observatory (ZOTTO) on regional summertime ecosystem fluxes, *Biogeosciences*, 11, 2055–2068, doi :10.5194/bg-11-2055-2014, URL <http://www.biogeosciences.net/11/2055/2014/>, 2014.
- Winiarek, V., Bocquet, M., Saunier, O., and Mathieu, A. : Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant : Application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant, *J. Geophys. Res.*, 117, D05 122, doi :10.1029/2011JD016932, URL <http://onlinelibrary.wiley.com/doi/10.1029/2011JD016932/abstract>, 2012.
- WMO/GAW : Guidelines for the Measurement of Methane and Nitrous Oxide and their Quality Assurance (WMO/TD-No. 1478), Tech. Rep. 185, World Meteorological Organization Global Atmospheric Watch, Geneva, Switzerland, URL http://www.wmo.int/pages/prog/arep/gaw/documents/Final_GAW_206_web.pdf, 2009.
- WMO/GAW : 16th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases, and Related Measurement Techniques (GGMT-2011), Tech. Rep. 206, World Meteorological Organization Global Atmospheric Watch, Wellington, New Zealand, URL http://www.wmo.int/pages/prog/arep/gaw/documents/Final_GAW_206_web.pdf, 2011.
- Wu, L., Bocquet, M., Lauvaux, T., Chevallier, F., Rayner, P., and Davis, K. : Optimal representation of source-sink fluxes for mesoscale carbon dioxide inversion with synthetic data, *J. Geophys. Res.*, 116, D21 304, URL <http://cerae.enpc.fr/HomePages/wu/file/wu11carbon.pdf>, 2011.
- Wu, L., Bocquet, M., Chevallier, F., Lauvaux, T., and Davis, K. : Hyperparameter estimation for uncertainty quantification in mesoscale carbon dioxide inversions, *Tellus B*, 65, doi :10.3402/tellusb.v65i0.20894, URL <http://www.tellusb.net/index.php/tellusb/article/view/20894>, 2013.
- Wunch, D., Wennberg, P. O., Toon, G. C., Keppel-Aleks, G., and Yavin, Y. G. : Emissions of greenhouse gases from a North American megacity, *Geophys. Res. Lett.*, 36, L15 810, doi :10.1029/2009GL039825, URL <http://onlinelibrary.wiley.com/doi/10.1029/2009GL039825/abstract>, 2009.
- Yamada, K., Yoshida, N., Nakagawa, F., and Inoue, G. : Source evaluation of atmospheric methane over western Siberia using double stable isotopic signatures, *Organic Geochemistry*, 36, 717–726, doi :10.1016/j.orggeochem.2005.01.016, URL <http://www.sciencedirect.com/science/article/pii/S0146638005000355>, 2005.
- Zhou, L. X., Kitzis, D., and Tans, P. P. : Report of the fourth WMO round-robin reference gas intercomparison, 2002–2007, in : Report of the 14th WMO Meeting of Experts on Carbon Dioxide Concentration and Related Tracer Measurement Techniques, WMO/GAW Rep, vol. 186, pp. 40–43, 2009.
- Zimov, S. A., Schuur, E. A. G., and Chapin, F. S. : Permafrost and the Global Carbon Budget, *Science*, 312, 1612–1613, doi :10.1126/science.1128908, URL <http://www.sciencemag.org/content/312/5780/1612>, 2006.
- Zuev, V. E., Belan, B. D., Kabanov, D. M., Kovalevskii, V. K., Luk'ianov, O. I., Meleshkin, V. E., Mikushev, M. K., Panchenko, M. V., Penner, I. E., and Pokrovskii, E. V. : The Optik-E AN-30 laboratory-airplane for ecological studies, *Op. Atmos. Okeana*, 5, 1012–1021, URL <http://adsabs.harvard.edu/abs/1992OpAtO...5.1012Z>, 1992.

Annexes

Acronymes et notations courantes

TABLE A.1: Acronymes et abréviations utilisés dans le rapport

Acronyme	Signification
AGAGE	Advanced Global Atmospheric Gases Experiment (http://agage.eas.gatech.edu/)
ANR	Agence Nationale de la Recherche
BLUE	Best Linear Unbiased Estimator
CASA	Carnegie-Ames-Stanford-Approach model
CEA	Commissariat à l'Énergie Atomique et aux Énergies Renouvelables
CEPMMT	Centre Européen de Prévision Météorologique à Moyen Terme
CLIMSLIP	CLimate IMPacts of Short-Lived pollutants In the Polar region
CNRS	Centre National de la Recherche Scientifique
CRU TS	Climatic Research Unit Time-Series doi : 10.5285/D0E1585D-3417-485F-87AE-4FCECF10A992
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CTM	Chemistry Transport Model
ECMWF	European Centre for Medium-range Weather Forecasts
EDGAR	Emission Database for Global Atmospheric Research
ENVISAT	ENVironmental SATellite (ESA)
EPA	U.S. Environmental Protection Agency
ERA-INTERIM	interim ECMWF Re-Analysis
ESA	European Space Agency
ESPIGRAD	Estimation des Sources et Puits de Gaz à effet de serre à l'échelle Régionale par Assimilation de Données atmosphériques in situ
EUMETSAT	Organisation européenne pour l'exploitation des satellites météorologiques
FLEXPART	FLEXible PARTicle dispersion model
GAW	Global Atmospheric Watch
GCM	Global Circulation Model

GFED	Global Fire Emission Database
GIEC	Groupe d'experts Intergouvernemental sur l'Évolution du Climat
GLCC	Global Land Cover Characteristics database
GLWD	Global Lake and Wetland Database
GOSAT	Greenhouse gases Observing SATellite (JAXA)
IASI	Interféromètre Atmosphérique de Sondage Infrarouge
ICOS	Integrated Carbon Observing System
InGOS	Integrated non CO ₂ Greenhouse gas Observing System
IPCC	Intergovernmental Panel for Climate Change
IPSL	Institut Pierre Simon Laplace
JAXA	Japan Aerospace Exploration Agency
JR-STATIONS	Japan-Russia Siberian Tall Tower Inland Observation Network
LMDz	Modèle de circulation générale zoomé du Laboratoire de Météorologie Dynamique
LPJ	Modèle de processus Lund-Potsdam-Jena
LPX	Modèle « Land surface Processes and eXchanges »
LSCE	Laboratoire des Sciences du Climat et de l'Environnement
MCG	Modèle de Circulation Générale
MCT	Modèle de Chimie Transport
MetOp	Meteorological Operational satellites (ESA/EUMETSAT)
MODIS	Moderate Resolution Imaging Spectroradiometer
MPI	Max Planck Institute
NASA	National Aeronautics and Space Administration
NIES	National Institute for Environmental Studies
NOAA	National Ocean and Atmosphere Administration
NOAA/ESRL	NOAA Earth System Research Laboratory
OMM	Organisation Météorologique Mondiale
OSSE	Observing System Simulation Experiment
PYVAR	module d'inversion VARIationnelle basé sur PYthon
RAS	Russian Academy of Science
SCHIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY
STE	Stratospheric–Tropospheric Exchange
TANSO	Thermal And Near infrared Sensor for carbon Observation
TROICA	Trans-Siberian Observations Into the Chemistry of the Atmosphere
UCI	University of California Irvine
USGS	U.S. Geological Survey
UVSQ	Université de Versailles – Saint-Quentin-en-Yvelines
VOC	Volatile Organic Compounds
WCB	Warm Conveyor Belt
WDCGG	World Data Center for Greenhouse Gases
WMO	World Meteorological Organization
WSLPD	West Siberian Lowland Peatland Database
YAK-AEROSIB	Airborne Extensive Regional Observations in SIBeria
ZOTTO	Zotino Tall Tower Observatory

e.g. *exempli gratia* = par exemple

et al.	<i>et alii</i> = et collaborateurs
i.e.	<i>id est</i> = c'est-à-dire
p. ex.	par exemple

TABLE A.2: Notations scientifiques et mathématiques (conformes à Ide et al., 1997).

Notation	Signification
CH ₄	méthane
CO	monoxyde de carbone
CO ₂	dioxyde de carbone
O ₃	ozone
COV	Composés Organiques Volatiles
NO _x	famille des oxydes d'azotes (NO + NO ₂)
UV	ultra-violet
Pg	petagramme 1 Pg = 10 ¹⁵ g
Tg	téragramme 1 Tg = 10 ¹² g
ppb	part per billion 1 ppb = 10 ⁻⁹
ppm	part per million 1 ppm = 10 ⁻⁶
ln	logarithme népérien
tr(·)	opérateur trace
(·) ^T	opérateur transposée
·	opérateur déterminant
ℰ[·]	espérance mathématique
<i>p</i> (X Y)	densité de probabilité de la variable X sachant Y
$\mathcal{N}(\bar{x}, \sigma_x)$	loi normale de mode \bar{x} et d'écart-type σ_x
I _{<i>n</i>}	matrice identité de dimension $n \times n$
ϵ_{phys}	erreurs dans le modèle physique
ϵ_{num}	erreurs dans le calcul numérique
ϵ_{discr}	erreurs dans la discrétisation et la comparaison aux observations
ϵ_{obs}	dans les mesures
\mathcal{H}	opérateur d'observation
H	matrice jacobienne de l'opérateur d'observation
x	vecteur d'état
x ^b	vecteur d'état a priori
x ^t	vecteur d'état réel
x ^b	vecteur d'état analyse, c'est-à-dire après l'inversion
ϵ^b	erreur sur l'état à priori = $\mathbf{x}^b - \mathbf{x}^t$
B	matrice de covariance des erreurs d'état = $\mathbb{E}[\epsilon^b(\epsilon^b)^T]$
y ^o	vecteur d'observations
<i>p</i>	nombre d'observations disponibles dans le système
ϵ^o	erreur d'observations = $\mathbf{y}^o - \mathcal{H}(\mathbf{x})$

R	matrice de covariance des erreurs d'observations = $\mathbb{E}[\epsilon^o(\epsilon^o)^T]$
K	matrice de gain de Kalman = $\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$
S	$= (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)^T$
S_{R,B}	$= \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$
J(x)	fonction de coût de l'inversion bayésienne
	$= \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x})$

Liste des communications

Composition of the spring Siberian troposphere during YAK-AEROSIB 2010 :
Influence of biomass burning, stratospheric intrusion and the Eyjafjöll eruption
2010, *American Geosciences Union Fall Meeting, San Francisco (présentation)*

J.-D. Paris, A. **Berchet**, M. Yu. Arshinov, P. Nédélec, A. Stohl, G. Ancellet, K. S. Law, B. D. Belan, M. Ramonet, P. Ciais.

Despite their potential relevance to atmospheric environment issues, measurements in Siberia, in the outflow of Europe and upstream of Eastern Asia and the Arctic, are sparse. Impacts of remote and regional sources on this air shed's composition remain virtually unexplored. We document the tropospheric composition and address these issues using CO₂, CO, O₃ and novel CH₄ measurement data from the 2010 YAK-AEROSIB intensive airborne campaign over Siberia. The campaign took place in April, spring being highly relevant here because of longer lifetime of pollutants, transported especially to the Arctic. We analysed the data by comparing trace gases concentrations enhancements, FLEXPART simulations of backward transport and by comparison with other campaigns. We illustrate our findings by three case studies. Typical observed mixing ratios of trace gases were 394 ppmCO₂, 140 ppbCO, 59 ppbO₃ and 1843 ppbCH₄ across the campaign. CO₂ increased by ~ 4 ppm relative to a similar 2006 spring campaign (Paris et al., 2010a), a slightly slower pace than the hemispheric increase of 1.75 ppm/yr (as estimated over the period 2006–2008). This difference could be due to different transport pattern, with more air from the Arctic, and hence less affected by pollution, than in 2006 ; vegetation respiration is low in largely snow-covered Siberia. CO mixing ratio is also ~ 5 ppb lower than our 2006 spring values, possibly also explained by a pervasive airmass of Arctic origin in 2010. But large enhancements of CO and other trace gases have been observed. CO enhancements were connected to biomass burning through tracer correlation and FLEXPART simulations convolved with fire emissions. We find that contributions from fires in agricultural areas in Western Russia and Northern Kazakhstan influence significantly CO concentrations over Siberia (+50 ppb) strongly correlated ($r = 0.86$) to an O₃ enhancement of ~ +5 ppb (regression slope 2.5 ppb/ppb). O₃ concentrations as high as 210 ppb were observed during flight from Bratsk to Novosibirsk, simultaneously with lower

CO and CO₂ concentrations. This high concentration occurred at 6 km altitude immediately before crossing a cold front and are interpreted as a stratospheric dry intrusion in the troposphere. We use this event to discuss the diffusivity of Flexpart in the analysis of such events. Finally, a consistent strong O₃ depletion at 6 km altitude over Siberia is tentatively attributed to the Eyjafjöll eruption based on FLEXPART backward simulation.

Tropospheric ozone over Siberia in spring 2010 :
long-range transport of biomass burning and anthropogenic emissions,
stratospheric intrusion and remote boundary layer influence
2012, *European Geosciences Union General Assembly, Vienna (poster)*

A. Berchet, J.-D. Paris, G. Ancellet, K. S. Law, A. Stohl, P. Nédélec, M. Yu. Arshinov, B. D. Belan, P. Ciais.

Atmospheric pollution, including tropospheric ozone, has an adverse effect on humans and their environment. The Siberian air shed covers about 10% of Earth's land surface. Therefore, it can contribute significantly to the global tropospheric ozone budget due, in the region, to vast deposition losses on the boreal forest vegetation in the atmospheric surface layer on the one hand, and in-situ photochemical production from ozone precursors emitted by Siberian terrestrial ecosystems, and the influx of stratospheric ozone to the troposphere on the other hand. We have identified and characterized factors that influenced the tropospheric ozone budget over Siberia during spring 2010 by analyzing in-situ measurements of ozone, carbon dioxide, carbon monoxide, and methane mixing ratios collected by continuous analyzers during an intensive airborne measurement campaign of the YAK-AEROSIB Project, carried out between 15 and 18 April 2010. The observations, spanning over 3000 km and stretching from 800 to 6700 m above ground level, were analyzed using the Lagrangian model FLEXPART to simulate backward air mass transport. The analysis of trace gas variability and simulated origin of air masses origins showed that biomass burning and anthropogenic activity expectedly increased carbon monoxide and dioxide concentrations. Also, such plumes coming from east and west of West Siberian plain and from North-Eastern China were shown to increase ozone mixing ratio owing to photochemical processes taking place along the transport route. In the case of low ozone mixing ratios observed over a large area (800 × 200 km) in the upper troposphere above 5500 m the air masses transported to the region under study were likely influenced by an Arctic ozone depletion event transported to lower latitudes and advected to the upper troposphere. The stratospheric source of ozone to the troposphere was observed directly in a well-defined stratospheric intrusion. Numerical simulations of this event suggest an input of 2.56×10^7 kg of ozone associated to a regional downward flux of 9.75×10^{10} molecules cm⁻² s⁻¹.

Monitoring methane plumes with a ground network :
the example of the Elgin platform
2012, *Internation summer school in advanced data assimilation for geosciences*,
Les Houches (poster)

A. Berchet, P. Bousquet, I. Pison, F. Chevallier.

Optimized atmospheric inversion for methane flux quantification in Eurasia
2013, *European Geosciences Union General Assembly, Vienna (présentation)*

A. Berchet, I. Pison, F. Chevallier, P. Bousquet, S. Conil, M. Geever, T. Laurila, J. Valentin Lavric, M. Lopez, J. Moncrieff, J. Necki, M. Ramonet, M. Schmidt, M. Steinbacher, and J. Tarniewicz.

Distinguishing between the various natural and anthropogenic sources contributing to methane surface fluxes is a prerequisite for closing the methane budget in Eurasian Boreal environments. Regional atmospheric inversions at a high resolution provide a mean for improving the inventories of methane emissions and the process-based models for land surface exchanges. Though promising, the atmospheric approach relies on performant chemistry-transport models and a network of precise in-situ observation sites. The inversion then suffer from all the uncertainties within the model and the observations. In the Bayesian inversion framework, an advanced and objective understanding of the statistics of the errors in the inversion system (such as the transport, representation, aggregation or observation errors) is required. Acquiring such an understanding rises many theoretical and practical difficulties. In most previous inversion studies, the statistics of the errors are built relying on expert knowledge on the behaviour of the models and the inversion systems. In this work, optimal matrices of error covariances are computed using general statistical methods. The error statistics are optimally estimated with a minimal set of physical hypotheses on the patterns of errors. Complex patterns and correlations of errors are then entirely retrieved. The reliability of our method was first tested on a well-documented European domain with numerous observations and precise inventories. We assimilated the data from 13 observation sites with quasi-continuous measurements to infer the European anthropogenic fluxes during a period of two weeks in March-April 2012. Errors caused by mis-representations in the transport model or by its inability to reproduce the situations of steep gradients in the air mass composition were retrieved. An analysis of the sensitivity of the inversion to each individual observation suggested guidelines for the selection of data in regional inversions. With this simple but efficient and optimal inversion system applied to a domain covering most of Siberia and Scandinavia, we inquire into the possibility of retrieving improved knowledge on the Eurasian Boreal methane fluxes.

One year of atmospheric inversion of natural and anthropogenic methane fluxes
at the meso-scale over Eurasia in 2010

2013, *International ZOTTO Workshop, Krasnoyarsk (présentation)*

A. **Berchet**, I. Pison, Paris J.-D., F. Chevallier, Bonne J.-L., Sasakawa M., Machida T., Lavrič J., Winderlich J., Arshinov M. Yu.

Distinguishing between the various natural and anthropogenic sources contributing to methane surface fluxes is a prerequisite for closing the methane budget in Eurasian Boreal environments. Anthropogenic sources of methane in Eurasia are dominated by oil and gas extraction and transport to the North and by industrial and urban emissions to the South. These sources are active 5 during all seasons. Natural emissions due to wetlands occur primarily in snow-free ecosystems, hence roughly from May to September in boreal Eurasia. Millions of hectares burn every summer in boreal Siberia, causing strong methane emissions distributed as point sources. Independent estimations of the temporal variability of all contributions to methane fluxes can help to validate or correct the anthropogenic inventories of emissions and the process-based models of land surface exchanges. In particular, the atmospheric inversion method allows efficient estimation of the surface-atmosphere exchanges at different scales. Our inversion is based on the assimilation of in-situ quasi-continuous observations and flask samples of the atmospheric mole fraction of methane. The network of quasi-continuous measurements is located in the vicinity of West Siberian Lowlands close to intensive emissions from oil and gas extraction and wetlands. It is maintained by the National Institute for Environmental Studies (Japan), the Max-Planck Institute for Biogeochemistry (Germany) and the Institute of Atmospheric Optics (Russia). Flask measurements constrain the lateral boundary conditions of the domain of interest and are extracted from the World Data Center for Greenhouse Gases. Here, a full year of inversion of methane fluxes at meso-scales should allow separating the continuous anthropogenic fluxes and the sporadic natural emissions. Further, we deduce methane fluxes at weekly and monthly resolutions in order to improve our understanding of the seasonal cycles of all the contributions to methane fluxes. Finally, the temporal and spatial resolutions make possible the identification of the location and the beginning and termination of natural processes emitting methane.

A model-based data analysis of the atmospheric methane above Siberia during
YAK-AEROSIB airborne campaign in summer 2012

2013, *European Geosciences Union General Assembly, Vienna (poster)*

E. Arzoumanian, Paris J.-D., A. Pruvost, A. **Berchet**, I. Pison, Arshinov M. Yu, Belan B. D.

High latitude regions are large sources of CH₄ in the atmosphere, both natural from boreal wetlands and wildfires, and anthropogenic from natural gas extraction industry, especially in the Russian Arctic. Our current understanding of the extent and amplitude of the natural sources, as well as the large scale driving factors, remain highly uncertain (Kirschke et al., 2013). After a decade of pause, atmospheric methane seems to be increasing again, with a possible significant contribution from the wetlands of the northern high latitudes initiated by an unusual rise of regional temperatures in 2007 (Dlugokencky et al., 2009). This work aims at better understanding high latitude CH₄ sources and sinks using atmospheric measurements and transport model. YAK-AEROSIB atmospheric airborne campaigns have been

performed in order to provide observational data about the composition of Siberian air. In this work, we focus on the 2012 campaign which has been conducted on July 31st and August 1st. It consisted of five flights, performed in the troposphere from the boundary layer up to 8.5 km, connecting Novosibirsk to Yakutsk and back. This particular campaign was dominated by wildfires in Western and central Siberia. Therefore a chemistry-transport model (CHIMERE), combined with datasets for anthropogenic (EDGAR) emissions and models for wetlands (ORCHIDEE) and wildfire (GFED), has been used to interpret the collected data. From tagged tracers and model observation mismatch we describe results concerning CH₄ fluxes in the Siberian territory.

Greenhouse gas aerosols and pollutants over Siberia : the YAK-AEROSIB French
Russian Joint International Laboratory
2013, *European Geosciences Union General Assembly, Vienna (poster)*

Paris J.-D., Belan B. D., G. Ancellet, P. Nédélec, Arshinov M. Yu., A. Pruvost, **A. Berchet**, E. Arzoumanian, I. Pison, P. Ciais, K. S. Law.

Despite the unique scientific value of better knowing atmospheric composition over Siberia, regional observations of the tropospheric composition over this region are still lacking. Large local anthropogenic emissions, strong ecosystem gas exchange across the vast forest expanse, and processes feeding back to global climate such as wetlands CH₄ emissions, seabed hydrates destabilization and degrading permafrost make this region particularly crucial to investigate. We aim at addressing this need in the YAK-AEROSIB program by collecting high-precision in-situ measurements of the vertical distribution of CO₂, CH₄, CO, O₃, black carbon and ultrafine particles distribution in the Siberian troposphere, as well as other parameters including aerosol lidar profiles, on a pan-Siberian aircraft transect. Campaigns are performed almost annually since 2006 until now on this regular route, while special campaigns are occasionally arranged to sample the troposphere elsewhere (e.g. Russian Arctic coast). We show the background tropospheric composition obtained from these surveys, the variability and the impact of large-scale transport of anthropogenic emissions from Europe and Asia, as well as the impact of biomass burning plumes both from local wildfires (2012) and from remote sources elsewhere in Asia. Long range transport of anthropogenic emissions is shown to have a discernible impact on O₃ distribution, although its lower-tropospheric variability is largely driven by surface deposition. Regional sources and sinks drive the lower troposphere CO₂ and CH₄ concentrations. Recent efforts aim at better understanding the respective role of CH₄ emission processes (including methanogenesis in wetlands and emissions by wildfires) in driving its large scale atmospheric variability over the region. Generally, the YAK AEROSIB provide unique observations over Siberia, documenting both direct impact of regional sources and aged air masses experiencing long range transport toward the high Arctic.

Précisions techniques

Le travail décrit dans ce rapport repose sur des modèles numériques et des codes informatiques qu'il a fallu développer entièrement ou depuis des programmes pré-existants. Nous décrivons succinctement ci-dessous les aspects techniques sous-jacents à notre travail.

Le modèle CHIMERE

Le modèle de chimie-transport eulérien régional CHIMERE est constitué d'un cœur de code de 15 000 lignes environ. Autour de ce code, s'articulent des scripts de préparations des entrées du modèle (grille du domaine, émissions, conditions aux bords, champs météorologiques, etc.) représentant 85 000 lignes de codes. CHIMERE est aussi doté d'un modèle tangent linéaire et d'un adjoint pour effectuer des inversions variationnelles à la méso-échelle. Ces deux modèles associés ont une taille comparable au modèle de référence (15 000 lignes).

Le travail de simulations aux hautes latitudes a requis des modifications au sein de tous ces sous-scripts du modèle. Ces modifications ont concerné des morceaux isolés dans l'ensemble du code de CHIMERE et sont chronophages en temps de programmation et en validation.

Définition de domaines polaires CHIMERE a été développé pour des simulations aux moyennes latitudes et plutôt en Europe. De ce fait, les problèmes de trigonométrie et de géométrie sphérique induits par la ligne de changement de date (180°E – 180°O) et par le pôle ont été éludés.

En soit, CHIMERE ne connaît pas la réalité et mélange une atmosphère virtuelle dans une boîte. Il suffit donc de lui spécifier des flux aux bords de cette boîte et des champs de vents à l'intérieur.

En revanche, les scripts de préparation doivent prendre en compte la géométrie réelle, ce qui n'est pas le cas dans la version de base de CHIMERE. Pour cette raison, j'ai généralisé les calculs de préparation des entrées de CHIMERE afin que des valeurs non physiques n'apparaissent pas au niveau de la ligne de changement de date et du pôle.

Généralisation de la géométrie des domaines La majorité des scripts de préparation de CHIMERE et internes à CHIMERE s'appuyait sur une définition d'un domaine quasi rectangulaire. Or, dans certains cas, cette hypothèse est difficilement conciliable avec le problème représenté. Par exemple, aux hautes latitudes, un domaine régulier en degré aura une forme très anguleuse.

Par conséquent, j'ai implémenté de nouveaux scripts de préparation et légèrement modifié le cœur de CHIMERE afin d'accepter des domaines qui ne soient pas parfaitement rectangulaires, ou au contraire afin de forcer un domaine de haute latitude à être rectangulaire (c'est-à-dire régulier en kilomètres) Pour cela, j'ai généralisé les scripts de définition des grilles de CHIMERE, ainsi que la prise en compte de la géométrie du domaine par le cœur de CHIMERE. En particulier, CHIMERE reconstruisait une grille virtuelle à partir des centres des mailles sans connaître les côtés de chaque mailles. Cela pouvait impliquer des artefacts importants dans les calculs avec des mailles non rectangulaires. J'ai donc pris en compte les coins des mailles de la grille définie et propagé cette information au fil du code CHIMERE afin de contraindre la forme de la grille.

Par ailleurs, les calculs des émissions sur une grille kilométrique depuis une grille en degré (comme c'est le cas pour la plupart des modèles de processus et des inventaires d'émission) peuvent induire des calculs de géométrie sphérique compliqués. Isabelle Pison a développé un algorithme de Monte-Carlo pour interpoler les émissions entre deux grilles tout en conservant la masse émise à une précision donnée. J'ai adapté ce code aux grilles kilométriques et l'ai optimisé afin de garantir une précision donnée pour un minimum de temps de calcul.

Paramétrisation de couche limite et de convection profonde CHIMERE mélange l'atmosphère verticalement à l'aide d'une paramétrisation de la convection profonde et du mélange dans la couche limite atmosphérique. Ce modèle a été développé dans un contexte où les données sur les flux de convection profonde n'étaient pas disponibles et/ou calculées de manière précise dans les modèles météorologiques. De même, la hauteur de couche limite n'était pas nécessairement proposée par les services météorologiques sollicités pour interpoler les champs de vent. Pour ces raisons, CHIMERE disposait d'une paramétrisation interne de la convection profonde et de la couche limite en fonction d'autres paramètres connus (température de l'air, pression, humidité, etc.).

Aujourd'hui, le CEPMMT propose explicitement des flux de masse de convection profonde et des hauteurs de couche limite assez précise. J'ai donc désactivé les scripts de calcul interne de la couche limite et de la convection profonde. En remplacement, j'ai adapté les variables proposées par le CEPMMT afin de les rendre lisibles par CHIMERE. CHIMERE utilise donc maintenant directement les paramètres du CEPMMT pour la convection profonde et le mélange dans la couche limite.

Colonnes satellitaires CHIMERE extrait au fil de ses calculs les concentrations simulées pour lesquelles on dispose d'un équivalent observé. Ces calculs ne pouvaient se faire que pour des données in situ et non pour des colonnes satellitaires.

J'ai adapté un script développé par Audrey Fortems-Cheney afin d'extraire les concentrations intégrées sur des colonnes d'air équivalentes aux colonnes mesurées par les instruments embarqués sur des satellites.

Le système d'inversion PYVAR-CHIMERE

Développement et maintenance du système Frédéric Chevallier a développé un système d'inversion variationnel basé sur des scripts codés en Python et le modèle de circulation général LMDz. Ce système a été dénommé PYVAR-LMDz.

Isabelle Pison a repris ce système afin de l'adapter à des inversions régionales avec le modèle CHIMERE. Le système ainsi développé, PYVAR-CHIMERE, représente environ 5 000 lignes de code et effectue automatiquement tous les calculs intermédiaires d'une inversion variationnelle. PYVAR-CHIMERE est aussi à même d'effectuer des simulations directes quasi « Plug and Play » du modèle CHIMERE. J'ai participé à la conception, la maintenance et à l'optimisation de ce système.

Articulation de PYVAR-CHIMERE dans l'inversion marginalisée La méthode par marginalisation que j'ai développée s'appuie sur le calcul explicite de la jacobienne de l'opérateur d'observation \mathbf{H} . Ce calcul est effectué en effectuant pour chaque composante du vecteur d'état une simulation directe et en extrayant les concentrations simulées aux sites d'observations. Cette méthode nécessite donc des milliers de simulations directes du modèle CHIMERE.

J'ai donc encapsulé le système PYVAR-CHIMERE dans un système calculant automatiquement l'opérateur \mathbf{H} . Les simulations directes sont calculées en parallèle sur le système de calcul du LSCE. Les résultats de ces simulations sont ensuite introduites dans la matrice \mathbf{H} qui est enfin utilisée pour effectuer l'inversion marginalisée.

Estimation des émissions de méthane en mer de Laptev

Comme évoqué dans le chapitre V, j'ai utilisé la version modifiée du modèle CHIMERE pour les configurations polaires afin d'évaluer les émissions de méthane venant d'hydrates de mer de Laptev. Une étude statistique simple des corrélations et biais entre modèle et mesures autour de l'Arctique a été mise en place.

Cette étude, *Revised methane emissions from the East Siberian Arctic Shelf*, a été soumise aux *Geophysical Research Letters* et est retranscrite ci-dessous.

Les auteurs suivant ont participé à cette étude :

Antoine Berchet,¹ Philippe Bousquet,¹ Isabelle Pison,¹ Robin Locatelli,¹ Frédéric Chevalier,¹ Jean-Daniel Paris,¹ Ed Dlugokencky,² Tuomas Laurila,³ Yrjo Viisanen,³ Doug Worthy,⁴ Euan Nisbet,⁵ Rebecca Fisher,⁵ James France,⁵ and David Lowry⁵.

Abstract

Massive sedimentation from the Lena River onto the East Siberian Arctic Ocean Continental Shelf (ESAS) provides labile carbon for generation of methane by methanogens at the bottom of the Arctic ocean, which can be emitted to the atmosphere, mainly through bubbling. ESAS may then contribute a significant fraction of regional methane fluxes. Previous studies based on interpolated oceanographic campaigns estimated emissions from this area at 8–17 TgCH₄.y⁻¹. Here, we propose a method of evaluating these estimates by comparing observations and regional simulations of atmospheric methane mole fractions. The analysis of this data set confirms the high variability and heterogeneity of the methane releases from this region. Based on a statistical analysis of the observations and of the simulations, we revise

¹Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CNRS-CEA-UVSQ, Orme des Merisiers, CEA Saclay, Gif-sur-Yvette, France.

²NOAA Earth System Research Laboratory, Global Monitoring Division, Boulder, Colorado, USA

³Climate and Global Change Research, Finnish Meteorological Institute, Helsinki, Finland.

⁴Environment Canada, Toronto, Ontario, Canada.

⁵Department of Earth Sciences, Royal Holloway, University of London, Egham, United Kingdom.

methane emissions from ESAS at 1.6–2.6 TgCH₄.y⁻¹, 3 to 10 times smaller than previously estimated.

1 Introduction

Most long-range global climate projections forecast a warming in the Arctic of at least a few °C over the next decades (Collins et al., 2013). Warmer Arctic temperatures could induce the melting of continental and submarine permafrost and the destabilization of marine hydrates, causing massive methane emissions into the atmosphere, and hence, generating positive feedbacks to the regional and global warming. Monitoring methane emissions at high latitudes in the North Hemisphere is therefore of critical importance to anticipate and to interpret future changes of the climate.

The present methane budget for the Arctic is rather consistently quantified between the different modelling approaches used for its assessment. For example, on the one side, emissions of methane from the Arctic tundra estimated by flux observations and process-based models (i.e. bottom-up approaches) for the 2000s have been synthesized respectively at 20 [11 to 51] TgCH₄.y⁻¹ and 28 [18 to 37] TgCH₄.y⁻¹ (McGuire et al., 2009). On the other side, top-down atmospheric inversions, based on methane atmospheric observations, show a range for total natural Arctic methane emissions north of 60°N of 16 [12–28] TgCH₄.y⁻¹ (Kirschke et al., 2013), i.e. smaller and slightly narrower than the bottom-up range, but still statistically consistent with bottom-up estimates. In addition, anthropogenic emissions are estimated at 9 [7 to 11] TgCH₄.y⁻¹ above 60°N by top-down inversions (Kirschke et al., 2013).

Methane emissions from the Arctic Ocean are lower than land emissions, but much more uncertain, as synthesized by McGuire et al. (2009), with a range of 1–12 TgCH₄.y⁻¹. Based on oceanographic measurements performed over almost a decade, Shakhova et al. (2010) suggested that the East Siberian Arctic Shelf (ESAS), which covers 2 × 10⁶ km² or 14% of the Arctic Ocean, emits 8 TgCH₄.y⁻¹ into the Arctic atmosphere, that is 2/3 of the 1–12 TgCH₄.y⁻¹ range by McGuire et al. (2009). Shakhova et al. (2014) revised ESAS emissions upwards at 17 TgCH₄.y⁻¹, accounting for methane emissions above several oceanic hotspots due to bubbling in the water-column and methane degassing to the atmosphere during storms (Shakhova et al., 2014).

Atmospheric observations of methane mole fractions and of methane isotopes in the Arctic can improve our understanding of ESAS emissions (Fisher et al., 2011). North of 50°N, 22 atmospheric stations measure methane mole fractions, among which 6 sites provide continuous observations and 3 sites sample the isotopic composition of air on weekly basis or during intensive campaigns. Although sparse, these stations are well illuminated by ocean and land methane emissions because of fast horizontal transport of air masses around the North Pole (e.g., Bousquet et al., 2011).

In this paper, atmospheric methane observations and high-resolution simulations of atmospheric transport in the Arctic are combined to evaluate the potential of a 8 TgCH₄.y⁻¹ source from ESAS and to propose a revised estimate of ESAS methane emissions

TABLE D.1: Observation site characteristics. The site location is displayed in Fig. D.1.

Station	ID	Location			Network / Institute (Data server)	Reference correlations r	
		Lon (° E)	Lat (° N)	Alt (m a.s.l)		Winter	Summer
Alert	ALT	-62.5	82.5	200	EC (WDCCGG ¹)	0.79	0.56
Barrow	BRW	-156.6	71.3	11	NOAA/ESRL (ESRL ²)	0.76	-
Tiksi	TIK	128.9	71.6	8	FMI (ESRL ³)	0.56	-0.04
Pallas	PAL	24.12	68.0	560	FMI (WDCCGG ¹)	0.89	0.63
Zeppelin	ZEP	11.9	79.9	475	RHUL (Fisher et al., 2011)	-	-

¹ World Data Center for Greenhouse Gases (<http://ds.data.jma.go.jp/gmd/wdccc/>)

² Dlugokencky et al., 1995, 2014. ftp://aftp.cmdl.noaa.gov/data/trace_gases/ch4/

³ ftp://ftp.etl.noaa.gov/psd3/arctic/tiksi/greenhouse_gas/ghg_concentration/raw/

2 Materials and Methods

2.1 Observation sites

Out of the 22 active observation sites carrying out atmospheric measurements of methane mole fractions around the Arctic Ocean in 2012, we selected 4 sites downwind the ESAS with continuous measurements : 1 nearby site, 2 sites remotely illuminated by ESAS emissions, and 1 background site only poorly influenced by ESAS emissions (Fig. D.1). Tiksi (TIK) is located closest to the shores of the Laptev sea, a few hundreds kilometres away from the core of the emitting region detected by Shakhova et al. (2010, 2014). Barrow (BRW) and Alert (ALT) are located at the northern edge of North America, in north Alaska and north Canada respectively, about 2000–2500 km away from the ESAS, with no major emissions along the air mass paths. Pallas (PAL), in northern Scandinavia, is taken as a distant site, with limited influence by Laptev Sea emissions (typical contributions < 2 ppb ; maximum 20 ppb in few plumes ; Fig. D.1). Here, PAL is used for evaluating the capacity of our transport model CHIMERE (Sect. 2.2) to reproduce the methane mole fraction variability at high latitudes and at synoptic scales. The methane mole fractions at the observation sites are analysed with instruments maintained by Environment Canada (EC), NOAA/Earth System Research Laboratory (NOAA/ESRL), and the Finnish Meteorological Institute (FMI). They are calibrated with standards traceable to the WMO X2004 CH₄ mole fraction scale (Dlugokencky et al., 2005). The combined standard uncertainty on individual measurement remains below the ± 3.7 ppb requested by the World Meteorological Organization (WMO/GAW, 2009).

Isotopes measurements of $\delta^{13}\text{C}_{\text{CH}_4}$ at Zeppelin observatory are carried out by the Royal Holloway University of London (RHUL). 5 litre tedlar bags are collected and analysed with modified gas chromatography isotope ratio mass spectrometry (GC-IRMS) at RHUL (Fisher et al., 2011).

2.2 Polar CHIMERE transport model

Atmospheric transport is simulated with the Eulerian meso-scale non-hydrostatic chemistry transport model CHIMERE (Vautard et al., 2001; Menut et al., 2013). The model is constrained by meteorological fields interpolated at a spatial resolution of $0.5^\circ \times 0.5^\circ$ every 3 hours from re-analyses of the European Centre for Medium-range Weather Fore-

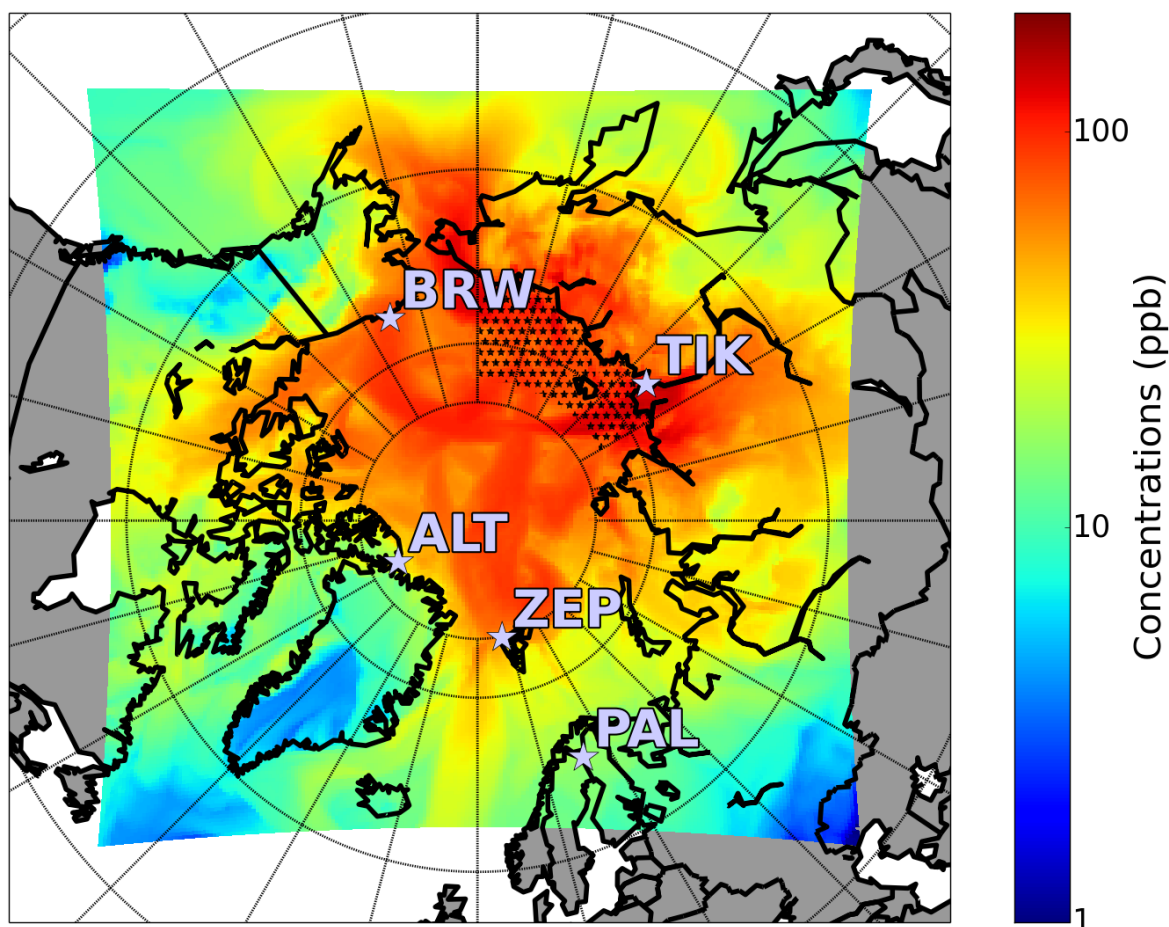


FIGURE D.1: Map of the domain with the emission zone from ESAS (black stars ; see Sect. 2.3) and the stations used in the analysis. Shaded colours show the maximum near-surface simulated influence of the ESAS methane emissions after transport for the whole year 2012.

casts (ECMWF, ERA-INTERIM; Uppala et al., 2005). The original model has been modified to simulate atmospheric transport over polar regions with a regular kilometric resolution of $35 \times 35 \text{ km}^2$ covering all latitudes from 50°N up to the North Pole (see Fig. D.1). The transport simulations represent the troposphere in the region from the surface to 300 hPa ($\sim 9000 \text{ m}$). The oxidation of CH_4 by the OH radical is not considered here since the focus is put on synoptic variations of atmospheric methane mole fractions at the surface.

2.3 Transport inputs and emission scenarios

The regional transport model CHIMERE requires boundary concentrations and surface emissions to simulate methane mole fractions in its limited domain. Boundary mole fractions are interpolated from global analyses obtained by assimilating surface mole fraction measurements (Locatelli et al., 2014). Surface emissions for the CHIMERE domain are deduced from : (1) EDGAR v4.2 FT2010 for anthropogenic emissions (<http://edgar.jrc.ec.europa.eu>), (2) LPJ global monthly climatology for wetland emissions (Spahni et al., 2011), (3) GFED v3 for fire emissions (Giglio et al., 2009; van der Werf et al., 2010), and (4) emissions from

ESAS (see below and location in Fig. D.1). These four types of emissions are run as separate passive tracers in polar CHIMERE for 2012, which allows analysing the contribution of each source separately at observation sites. The combination of the four types of emissions and of the boundary concentrations provides the modelled methane mole fractions including the ESAS contribution. The emission scenario not including ESAS emissions is hereafter called the reference scenario.

The emissions from ESAS are computed based on Shakhova et al. (2010) as they provide a detailed description of the emission zones and emission strengths per period (winter/summer). ESAS includes Laptev, east Siberian and Russian parts of Chuckchi seas. We build a methane flux map according to their Fig. 1D and to their Tab. D.1 : a zone covering 120–180°W and 68–78°N is separated into seven hotspot areas ($\sim 10\%$ of ESAS surface), the rest forming background areas (Fig. 1D of Shakhova et al., 2010). For each type (hotspot/background), a constant summer flux (mid-June to mid-September) and a constant winter flux (the rest of the year) is computed so that total hotspot emissions are 2 TgCH₄ (resp. 3.5 TgCH₄) for the summer months (resp. winter months), and total background emissions are 0.7 TgCH₄ (resp. 1.8 TgCH₄) for the summer months (resp. winter months), as given in Tab. D.1 of Shakhova et al. (2010). Annual methane emissions of ESAS are therefore 8 TgCH₄ over one year. Doing so, we underestimate the variability of methane emissions from ESAS which likely vary on shorter time scales, especially in winter in relation with sea ice breaks and ice displacements after periods of accumulation under the ice. This means that simulated concentrations are less contrasted with smaller peaks and higher background values than with a more realistic (but unknown) flux map. Finally, the simulated contribution of ESAS at observation sites are smaller than if the more recent 17 TgCH₄.yr⁻¹ budget of Shakhova et al. (2014) was used.

3 Results

In the following, we compare the simulated atmospheric mole fractions for the four contributions described above with methane observations. We also use isotopic observations to reinforce our conclusions, and present Taylor diagrams to estimate the methane emissions from the ESAS which best fit the observations.

3.1 Model–observation comparisons at four Arctic sites

At PAL, a site scarcely influenced by ESAS emissions, polar-CHIMERE computed with our reference emission scenario demonstrates a remarkable skill in representing the methane mole fractions at high latitude sites. As shown in Fig. D.2A, the variability of the daily averages observed methane mole fractions is indeed well captured by CHIMERE (annual temporal correlation of $r = 0.87$, Tab. D.1). Discrepancies between the observed signal and the simulated one at other sites can then be reasonably interpreted in terms of mis-specified regional emissions.

At ALT and BRW (Fig D.2B, C), two sites remote from ESAS but influenced by long-range transport from ESAS, winter mole fractions (i.e. the year outside June–September) are well reproduced by the reference scenario ($r = 0.79$ and $r = 0.76$, respectively). The contribution of ESAS is much too large at ALT and BRW in winter as shown by the large blue spikes of Fig. D.2 between January and April and for October–November (ALT only, no data available

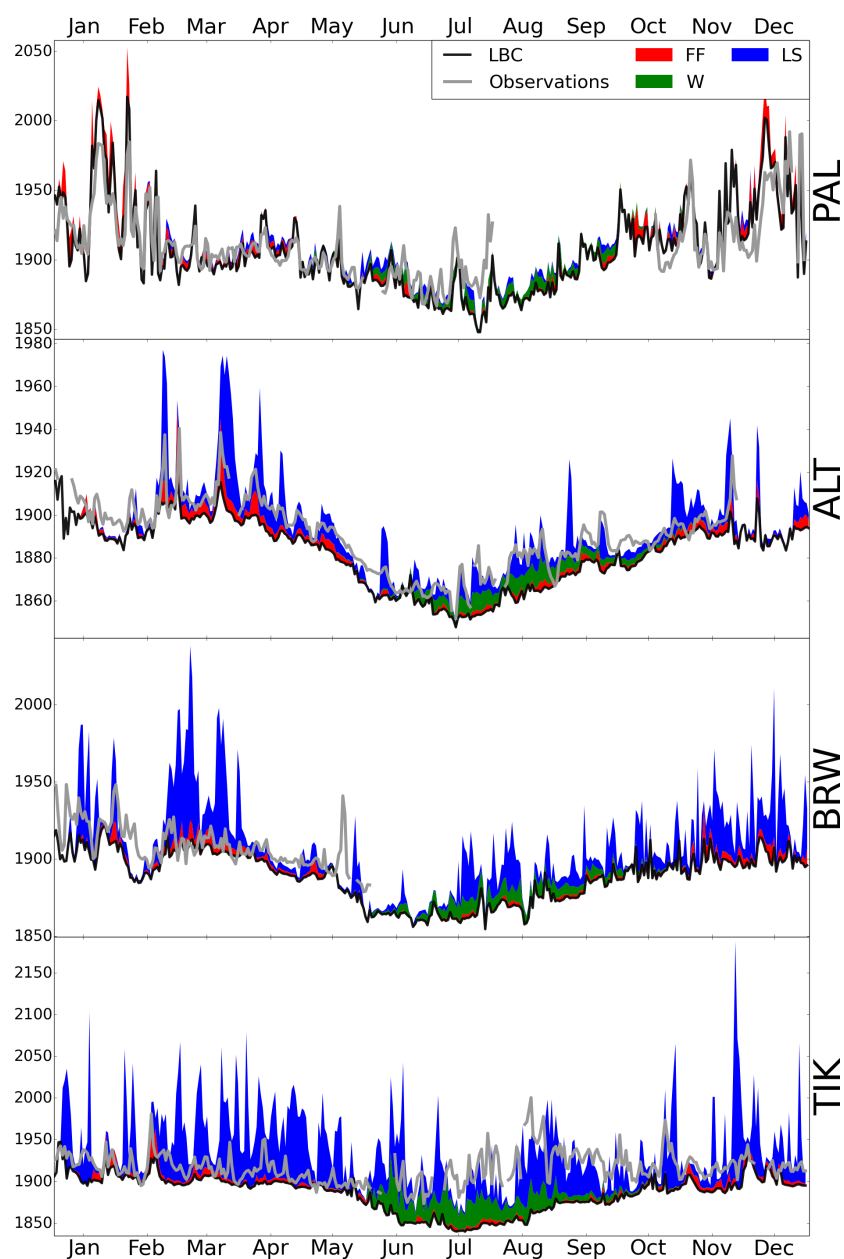


FIGURE D.2: Time series of observed and simulated methane mole fractions at four Arctic sites in 2012. The filled area depicts the daily contributions from wetlands (W), fossil fuels and other anthropogenic emissions (FF) and from the ESAS (LS; $8 \text{ TgCH}_4 \cdot \text{y}^{-1}$ scenario; Sect. 2.3). The LBC line represents the contribution of the lateral boundary conditions transported into the domain. Grey lines are observations (daily averages of continuous measurements). Fire emissions are not represented in this figure due to very low influence on the studied sites.

for BRW). As discussed in Sect. 2.3, the hot-spot distribution of ESAS emissions is not represented. A realistic distribution would have led to enhanced simulated spikes, reinforcing the inconsistency of winter ESAS fluxes.

In summer, at ALT, the fit of the reference scenario to the observations is less favourable than in winter ($r = 0.56$). Adding ESAS emissions may fill some gaps in July–August, less in June and September, though spikes are too high and phases are not always in agreement with observations. Nevertheless, even though the scenario by Shakhova et al. (2010) seems over-estimated from the point of view of BRW and ALT in winter, some summer plumes from the ESAS are well reproduced by the model (Fig. D.2). This would suggest that sudden bursts of methane may be released on short periods (typically days) with instantaneous rates corresponding to $8 \text{ TgCH}_4 \cdot \text{y}^{-1}$, but not on the long run.

Getting closer to the ESAS, TIK methane observations compared to simulations confirm that the simulated contribution of ESAS emissions from October to June is over-estimated (Fig. D.2D). Indeed, the baseline of observations is well reproduced by the reference scenario, despite some unexplained spikes in winter (Fig. D.2D), which slightly decorrelate the fit of the reference scenario to observations ($r = 0.56$, Tab. D.1). These spikes can be attributed either to small and short-term releases of CH_4 from ESAS or to other emissions not properly represented or transported to TIK. In summer, in June, the contribution from the ESAS is still too large compared to observations. However, from July to September, the observed mole fractions are higher and more variable than the reference scenario. This could suggest a strong influence from ESAS, but one can also expect a significant ill-simulated signal from wetland areas nearby. Indeed, the distribution and variability of wetland emissions in northern Yakutia, located upwind TIK during part of the summer, may be not well computed by the LPJ model. Moreover, these emissions are likely under-estimated in the global version of LPJ used here with fluxes $< 0.5 \text{ TgCH}_4 \cdot \text{y}^{-1}$ for boreal wetlands covering $\sim 400000 \text{ km}^2$. With a 2–3 month emission season, this would imply instantaneous emission rates of $< 0.5 \text{ mgC} \cdot \text{m}^{-2} \cdot \text{h}^{-1}$, 5–10 times smaller than the observed range of emissions in similar Siberian landscapes ($3\text{--}8 \text{ mgC} \cdot \text{m}^{-2} \cdot \text{h}^{-1}$; Glagolev et al., 2011)

In summary, the emission scenario from Shakhova et al. (2010) shows a large over-estimation of methane mole fractions at Arctic stations during all months but from July to September. Definite conclusion on the exact magnitude of ESAS methane releases cannot be obtained from TIK alone, due to unaccounted regional influences from natural wetlands, but also to the simplified spatial and temporal scenario used here for ESAS emissions. The distant observation sites (BRW and ALT) are more likely to provide integrated pieces of information about the methane fluxes from ESAS.

3.2 Summer isotopic observations in the Arctic

Quantitative information on Arctic emissions and the relative contribution of each process can be brought by the analysis of the isotopic composition of Arctic air ($\delta^{13}\text{C}_{\text{CH}_4}$ here), as Arctic surface emissions mixed into the atmosphere own very different isotopic signatures : typically of $-40/ -50\text{‰}$ for gas leaks, $-50/ -55\text{‰}$ for marine hydrates and $-60/ -75\text{‰}$ for wetlands (Fisher et al., 2011).

Isotopic observations of $\delta^{13}\text{C}_{\text{CH}_4}$ observations are regularly made at Zeppelin observatory (ZEP, Spitsbergen). During summer 2008, episodes with identified origin from River Ob and

Eastern Siberia pointed to dominant contribution from wetlands with a mean signature of $-65 \pm 3\%$ (Fisher et al., 2011). These observations seem not consistent with the emission scenario from Shakhova et al. (2010). In this emission scenario transported by CHIMERE, ESAS contribute to 45% of the signal at ZEP and wetland (resp. fossil fuels) to 35% (resp. 20%) in summer. This would imply a simulated $\delta^{13}\text{C}_{\text{CH}_4}$ of $-51 / -61\%$, not compatible with the observations. The Shakhova et al. (2010) scenario should be reduced by 3-5 times so that the isotopic observations are compatible to the simulations in summer.

3.3 Estimation of methane emissions from the ESAS

The previous analysis suggests substantially lower annual emissions from ESAS than the $8 \text{ TgCH}_4.\text{y}^{-1}$ proposed by Shakhova et al. (2010), especially in winter. Significant methane emissions from ESAS cannot be excluded in summer, but isotopic observations indicate a prevalence of natural wetlands in the Arctic air during summer months. A deeper statistical analysis is performed on CH_4 mole fractions by testing the model-observations agreement (variances and correlations) at three sites (ALT, TIK, PAL) for varying magnitudes of ESAS emissions from 0 to $10 \text{ TgCH}_4.\text{y}^{-1}$ (Fig. D.3). Meanwhile, other emission rates (wetlands and anthropogenic) are assumed to remain constant. BRW is not plotted because no observations are available between June and December 2012 (Fig. D.2).

The resulting trajectories for relative variances and correlations plotted in a Taylor diagram (Taylor, 2001) provide an estimate of the most likely emission strength at each station (Fig. D.3). This best guess for emissions is materialized by the point of the trajectory the closest to the full-correlation reference (standard deviation = 1, correlation = 1). As seen in Sect. 3.1, PAL is not affected by a change in ESAS emissions resulting in a very short trajectory in Fig. D.3. This confirms its status of background site regarding ESAS emissions. In winter, the trajectories in the Taylor diagram for TIK and ALT provide consistent optimal emissions with rates of 1–2 TgCH_4 for one year. Integrating this flux on the winter duration (8 months), ESAS are suggested to emit 0.6 to 1.3 $\text{TgCH}_4.\text{y}^{-1}$ every winter. In summer, due to the probable influence of local Arctic wetlands, the Taylor diagram cannot be exploited at TIK. However, using the diagram at ALT, summer emissions of 3–4 TgCH_4 can be expected from ESAS. Weighting by the duration of summer as defined above, ESAS would emit 1–1.3 $\text{TgCH}_4.\text{y}^{-1}$ in summer. Annually, our simple statistical analysis suggests methane emissions from 1.6 to 2.6 $\text{TgCH}_4.\text{y}^{-1}$ from the ESAS.

This estimate is an upper bound for ESAS emissions for the following reasons : 1) only ESAS emissions are varied to fit the observed mole fractions and not the continental wetland emissions likely under-estimated in the global LPJ version we used, and 2) the flat temporal emission profile from ESAS in our scenario underestimate the impact of ESAS region on synoptic methane variations at observation sites.

4 Conclusions

We tested the consistency of a methane emission scenario in the Arctic including a $8 \text{ TgCH}_4.\text{y}^{-1}$ source in the ESAS with atmospheric observations of CH_4 and of $\delta^{13}\text{C}_{\text{CH}_4}$ at different Arctic stations, close or remote from ESAS, and a high resolution model of atmospheric transport. The analysis of the modelled and observed time series suggests a large overestimation of ESAS emissions in winter and a possible contribution of ESAS emissions from July to Septem-

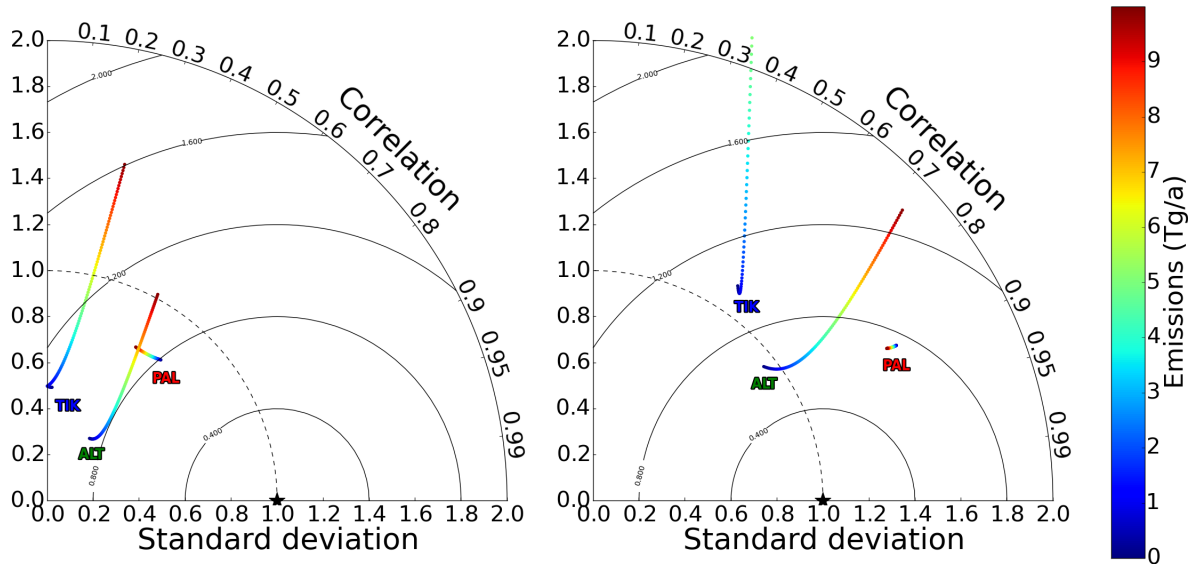


FIGURE D.3: Taylor diagrams using varying emission scenarios for ESAS from 0 to 10 $\text{TgCH}_4\cdot\text{y}^{-1}\text{r}$ (see Sect. 3.3) at three observation sites for summer (JJAS ; left panel) and winter (the rest of the year ; right panel) of year 2012. The trajectories in the Taylor diagram depict the compatibility with the observations of the different emissions scenarios. To compare the different sites, each trajectory has been normalized by the site standard deviation. For each site, the point of the emission trajectory closest to the reference point (black star) suggests the most plausible emission rate.

ber, also consistent with isotopic observations. Over 2012, a simple statistical analysis of the simulated and observed time series suggests emissions from ESAS of 1.6 to 2.6 $\text{TgCH}_4\cdot\text{y}^{-1}$. Although significant at the regional scale, these revised emissions are 3 to 5 times smaller than previous estimates from Shakhova et al. (2010) and 6 to 10 times smaller than the most recent estimates (Shakhova et al., 2014). The time series from the different sites also confirm a very likely heterogeneous temporal variability and spatial distribution, with very short and local methane releases from ESAS but also possibly from nearby natural wetlands. A multi-year analysis with more observation sites should be carried out in order to properly identify the sensitivity of the emissions to the ice cover or to other meteorological conditions and the distribution and short-scale variability of the fluxes. The development of continuous $^{13}\text{CH}_4$ observations at Arctic observation sites, now possible through cavity ring down spectrometry, would provide additional constraints for partitioning emissions between marine hydrates, gas leaks, and continental wetlands. Finally, the observatories operated around the Arctic Ocean could also provide more quantitative estimates of Arctic emissions from ESAS using direct and inverse modelling of both CH_4 and $^{13}\text{CH}_4$ observations.

Acknowledgments

We thank the PI from the observation sites we used for maintaining methane measurements at high latitudes and sharing their data. We also thank Patrick Crill (Department of Geological Sciences, Stockholm University) for his kind advices and remarks on the manuscript and F. Marabelle (LSCE) and the LSCE IT team for the computer resources. This study

was supported by the CEA and ANR-CLIMSLIP project.

