FINNISH METEOROLOGICAL INSTITUTE
CONTRIBUTIONS
No. 141

APPLICATION OF THE ENSEMBLE KALMAN FILTER IN THE ESTIMATION OF GLOBAL METHANE BALANCE

Aki Tsuruta

Doctoral Programme in Mathematics and Statistics
Department of Mathematics and Statistics
Faculty of Science
University of Helsinki

Academic dissertation presented
for the degree of Doctor of Philosophy

To be presented for public examination with the permission of
the Faculty of Science of the University of Helsinki
in the Auditorium Brainstorm at the Finnish Meteorological Institute, Helsinki,
on 21 December 2017 at 12 o'clock.

Finnish Meteorological Institute
Helsinki, 2017

Author's Address: Climate Research Unit,
Finnish Meteorological Institute,
PL503, FI-00101 Helsinki, Finland.
aki.tsuruta@fmi.fi

Supervisors: Docent Tuula Aalto, Ph.D.
Department of Physics, University of Helsinki
Climate Research Unit, Finnish Meteorological Institute

Leif Backman, Ph.D.
Climate Research Unit, Finnish Meteorological Institute

Professor Jukka Corander, Ph.D.
Department of Mathematics and Statistics,
University of Helsinki

Reviewers: Professor Heikki Haario, Ph.D.
School of Engineering Science,
Lappeenranta University of Technology

Docent Marko Scholze, Ph.D.
Department of Physical Geography and Ecosystem Science,
University of Lund

Opponent: Shamil Maksyutov, Ph.D.
Head of Biogeochemical Cycle Modeling and Analysis Section,
Center for Global Environmental Research,
National Institute for Environmental Studies

Custos: Professor Samuli Siltanen, Ph.D.
Department of Mathematics and Statistics,
University of Helsinki

ISBN 978-952-336-040-2 (paperback)
ISSN 0782-6117
Erweko Oy
Helsinki 2017

ISBN 978-952-336-041-9 (PDF)
https://ethesis.helsinki.fi/
Helsingin yliopiston verkkojulkaisut
Helsinki 2017

FINNISH METEOROLOGICAL INSTITUTE

Published by Finnish Meteorological Institute
(Erik Palménin aukio 1) , P.O. Box 503
FIN-00101 Helsinki, Finland

Series title, number and report code of publication
Finnish Meteorological Institute
Contributions 141, FMI-CONT-141
Date
December 2017

Author(s)
Aki Tsuruta

Title
Application of Ensemble Kalman Filter in Estimation of Global Methane Balance

Abstract

Ensemble Kalman filter (EnKF) is a useful Bayesian inverse modelling method to make inference of the states of interest from observations, especially in non-linear systems with a large number of states to be estimated. This thesis presents an application of EnKF in estimation of global and regional methane budgets, where methane fluxes are inferred from atmospheric methane concentration observations. The modelling system here requires a highly non-linear atmospheric transport model to convert the state space on to the observation space, and an optimization in both spatial and temporal dimensions is desired.

Methane is an important greenhouse gas, strongly influenced by anthropogenic activities, whose atmospheric concentration increased more than twice since pre-industrial times. Although its source and sink processes have been studied extensively, the mechanisms behind the increase in the 21st century atmospheric methane concentrations are still not fully understood. In this thesis, contributions of anthropogenic and natural sources to the increase in the atmospheric methane concentrations are studied by estimating the global and regional methane fluxes from anthropogenic and biospheric sources for the 21st century using an EnKF based data assimilation system (CarbonTracker Europe-CH_4; CTE-CH_4). The model was evaluated using assimilated in situ atmospheric concentration observations and various non-assimilated observations, and the model sensitivity to several setups and inputs was examined to assess the consistency of the model estimates.

The key findings of this thesis include: 1) large enough ensemble size, appropriate prior error covariance, and good observation coverage are important to obtain consistent and reliable estimates, 2) CTE-CH_4 was able to identify the locations and sources of the emissions that possibly contribute significantly to the increase in the atmospheric concentrations after 2007 (the Tropical and extra Tropical anthropogenic emissions), 3) Europe was found to have an insignificant or negative influence on the increase in the atmospheric CH_4 concentrations in the 21st century, 4) CTE-CH_4 was able to produce flux estimates that are generally consistent with various observations, but 5) the estimated fluxes are still sensitive to the number of parameters, atmospheric transport and spatial distribution of the prior fluxes.

Publishing unit
Climate Research

Classification (UDC)
681.5.015.44, 551.501, 547.211

Keywords
Bayesian inversion, ensemble Kalman filter, data assimilation, GHG flux inversion, methane

ISSN and series title
0782-6117 Finnish Meteorological Institute Contributions

ISBN
978-952-336-040-2 (paperback)
978-952-336-041-9 (pdf)

Language
English

Pages
166

Acknowledgements

This PhD work was widely supported by many scientists and other people around me, and several educational and non-educational institutes and organisations from several countries. I greatly appreciate all the support, opportunities, and knowledge provided during this study.

Thank you Prof. Ari Laaksonen (FMI: Finnish Meteorological Institute) and Tuomas Laurila (FMI) for letting me start this work at FMI. It was a great privilege to work in the Tuomas' group, which was a combination of modellers and instrumentalists. This experience has surely widened my ability to think about problems from both sides.

I am grateful to the supervisors; Prof. Jukka Corander (University of Helsinki), Doc. Tuula Aalto (FMI), and Dr. Leif Backman (FMI) for their generous support. Thanks to Jukka, who approved the subject; I really enjoyed working on this application study. Thank you Tuula and Leif for consistent daily supervisions, from valuable scientific discussions to daily matters, which have enormously helped me go through this study and accustom to the working life at FMI.

This work was greatly supported by co-authors. I would like to thank all the co-authors, but especially Prof. Wouter Peters (WUR: Wageningen University and Research), Prof. Maarten Krol (WUR), Dr. Ingrid T. van der Laan-Luijkx (WUR), Dr. Sander Houweling (Netherlands Institute for Space Research) and Dr. Edward Dlugokencky (National Oceanic and Atmospheric Administration). They provided support not only for the manuscript preparation and technical developments, but also hosted me during several short visits to their institutes. It was a valuable experience to work in so many different environments.

I appreciate the pre-examiners Prof. Heikki Haario (Lappeenranta University of Technology) and Doc. Marko Scholze (University of Lund) for their comments, which improved the thesis greatly. I would also like to thank Dr. Janne Hakkarainen (FMI), Dr. Marko Laine (FMI) and Dr. Toni Viskari (FMI) for their unofficial, but important comments for the Introductory part. Honourable Shamil Maksyutov (National Institute for Environmental Studies), thank you for accepting to be the opponent for the public examination.

The financial support for this work has been provided by FMI, the Maj and Tor Nessling foundation, the NCoE projects (DEFROST, eSTICC), the Finnish Academy project (CARB-ARC) and the EU FP7 project (InGOS).

Peer support from FMI colleagues, especially those from the Carbon Cycle Modelling Group and Greenhouse Gas Group, have encouraged me during hard times. I really enjoyed the cultural evenings with them - great food, fun play and lots of laughs. Hope we can continue this strong and enjoyable collaboration in the future.

Last, but not least, I would like to thank all my friends and families, both in Finland and abroad. They have always believed in me, and supported my way of life with lots of love. Thank you especially to my mother, father, little sister, and my dearest husband.

Helsinki, December 2017
Aki Tsuruta

Abbreviations and symbols

CTE	CarbonTracker-Europe data assimilation system
DYPTOP	Dynamical Peatland Model Based on TOPMODEL
ECMWF	European Centre for Medium-range Weather Forecasts
EnKF	Ensemble Kalman Filter
EnSRF	Ensemble square root Filter
ERA-Interim	European Reanalysis climate data
GHG	Greenhouse Gas
GOSAT	Greenhouse Gases Observing Satellite
KF	Kalman Filter
LPJG	Lund-Potsdam-Jena General Ecosystem Simulator
LPX-Bern	Land surface Processes and eXchanges process model
NOAA	National Oceanic and Atmospheric Administration
NOAA/ESRL	NOAA's Earth System Research Laboratory
TCCON	Total Carbon Column Observing Network
WDCGG	World Data Centre for Greenhouse Gases
WHyMe	Wetland Hydrology and Methane Dynamic Global Vegetation Model
d.o.f.	degree of freedom
pdf	probability density function
$\boldsymbol{x} \in \mathbb{R}^N$	a state vector
$\boldsymbol{y} \in \mathbb{R}^S$	an observation vector
$\boldsymbol{P} \in \mathbb{R}^{N \times N}$	a model error covariance matrix
$\boldsymbol{R} \in \mathbb{R}^{S \times S}$	an observation error covariance matrix
$\boldsymbol{Q}$	background error covariance matrix
$\mathcal{H}$	observation operator
$\mathcal{M}$	state dynamical model
X^{-1}	inverse of a matrix X
X^T	transpose of a matrix X
$p(x)$	pdf of a random variable x
$p(x\|y)$	conditional pdf of a random variable x given y

Contents

List of publications and author's contribution

The dissertation consists of an introduction and the following four peer-reviewed research articles. The aim of the introduction is to give an overview of the contents of the papers. The articles are referenced in the text by the Roman numerals **I-IV**.

I Tsuruta, A., Aalto, T., Backman, L., Peters, W., Krol, M., van der Laan-Luijkx, I. T., Hatakka, J., Heikkinen, P., Dlugokencky, E. J., Spahni, R. and Paramonova, N.: Evaluating atmospheric methane inversion model results for Pallas, northern Finland., Boreal Environ. Res., 20(4), 506–525, 2015.

II Tsuruta, A., Aalto, T., Backman, L., Hakkarainen, J., van der Laan-Luijkx, I. T., Krol, M. C., Spahni, R., Houweling, S., Laine, M., Dlugokencky, E., Gomez-Pelaez, A. J., van der Schoot, M., Langenfelds, R., Ellul, R., Arduini, J., Apadula, F., Gerbig, C., Feist, D. G., Kivi, R., Yoshida, Y. and Peters, W.: Global methane emission estimates for 2000–2012 from CarbonTracker Europe-CH_4 v1.0, Geosci. Model Dev., 10(3), 1261–1289, doi:10.5194/gmd-10-1261-2017, 2017.

III Bergamaschi, P., Karstens, U., Manning, A. J., Saunois, M., Tsuruta, A., Berchet, A., Vermeulen, A. T., Arnold, T., Janssens-Maenhout, G., Hammer, S., Levin, I., Schmidt, M., Ramonet, M., Lopez, M., Lavric, J., Aalto, T., Chen, H., Feist, D. G., Gerbig, C., Haszpra, L., Hermansen, O., Manca, G., Moncrieff, J., Meinhardt, F., Necki, J., Galkowski, M., O'Doherty, S., Paramonova, N., Scheeren, H. A., Steinbacher, M. and Dlugokencky, E.: Inverse modelling of European CH_4 emissions during 2006-2012 using different inverse models and reassessed atmospheric observations, Atmos. Chem. Phys., in print, 2017.

IV Tsuruta, A., Aalto, T., Backman, L., Krol., M. C., Peters, W., Lienert, S., Joos F., Miller, P. A., Zhang, W., Laurila, T., Hatakka, J., Leskinen, A., Lehtinen, K., Peltola, O., Vesala, T., Levula, J., Dlugokencky, E., Heimann, M., Kozlova, L., Aurela, M., Lohila, A., Kauhaniemi, M. and Gomez-Pelaez, A. J.: Methane budget estimates in Finland from the CarbonTracker Europe-CH_4 data assimilation system, manuscript submitted to Tellus B, 2017.

As a principal author of **Papers I**, **II** and **IV**, I have worked on the code development and implementation, processing input data, planning and carrying out computer simulations, analysis of the results and manuscript preparation together with the co-authors. In **Paper III**, I have provided CTE-CH_4 results by implementing the codes specifically for this study, planning and carrying out computer simulations, including several multi-year inversion experiments. In addition, I have contributed to analysis and discussion of the results throughout the manuscript preparation.

1 Introduction

Bayesian inverse modelling is a powerful method to make inferences about certain states or parameters based on some observed data, which has been applied to real-life problems in a wide range of scientific fields, including physics, meteorology and environmental sciences, to name a few. The method is based on probability theories, and offers a possibility to study highly ill-posed problems that involve non-linear systems with a large number of states or parameters to be estimated.

The aim of the Bayesian inverse modelling is to estimate a probability density function (pdf) of variables of interest conditional on the observations. There are two popular algorithms to derive such estimates; variational methods that are based on numerical iterations, and other methods that are based on filtering theory. One of the most used filtering method is the Kalman filter (KF), which derives a maximum likelihood estimator of the pdf of our interest, first introduced by Rudolf E. Kalman (Kalman, 1960). Although this gives the optimal estimate for linear systems, the linearity assumption limits its application to real-life problems because they often involve non-linear systems. In addition, KF is computationally demanding for high dimension systems where the number of parameters is large.

One alternative to KF was introduced by Evensen (1994), which is a Monte Carlo approximation for KF, called an ensemble Kalman filter (EnKF). In EnKF, the pdf of interest is approximated with a limited number of random samples, reducing the computational demand for high dimensional cases, and the assumption of linearity is not required. The EnKF is therefore applicable in various real-life problems which involve complicated non-linear systems, such as numerical weather prediction (Buehner et al., 2016; Rabier, 2005; Lorenc, 2003), oceanography (Park and Kaneko, 2000; Echevin et al., 2000; Keppenne and Rienecker, 2003) and atmospheric physics (Peters et al., 2005; Bruhwiler et al., 2014). In addition, those applications are often very high dimensional, on the global scale in the horizontal, vertical and temporal dimensions, where KF would not be appropriate.

Estimation of greenhouse gas (GHG) budgets, which requires temporally evolving high resolution estimates and involves non-linear climate systems, is not an exception here. Historically, GHG fluxes have been estimated using process-based models, where the fluxes are estimated based on biogeochemical and physical theories. Those process-based models and reported statistics based inventories have been useful in estimating GHG fluxes from various source and sink processes. However, those process-based models target certain types of processes, such as biogenic, anthropogenic and oceanic, but do not take all processes into account to give the "whole picture".

The Bayesian atmospheric GHG inverse models, on the other hand, aim to estimate the budgets which are the aggregated burden from both sources and sinks. Although

inverse models have difficulties in giving detailed estimates for each processes, they have the possibility to give more accurate total budgets, which are constrained by measured atmospheric GHG concentrations. The inverse modelling of GHG budget based on the variational method have been developed for various applications (Houweling et al., 1999, 2014; Bergamaschi et al., 2005, 2009; Bousquet et al., 2011). In addition, Bruhwiler et al. (2005) developed a KF based system in an application for global carbon dioxide (CO_2) budget estimates, which was further improved to an EnKF based system by Peters et al. (2005) (called CarbonTracker). The CarbonTracker system is now developed for various applications for regional and global GHG budget estimates (Peters et al., 2007, 2010; van der Laan-Luijkx et al., 2015, 2017; Zhang et al., 2014), including methane (CH_4) (Bruhwiler et al., 2014, and those presented in this thesis) and sulphur hexafluoride (SF_6) (van der Veen, 2013).

CH_4 is the second most important greenhouse gas after CO_2, strongly influenced by anthropogenic activities, such as fossil fuel use, agriculture, and landfills (Myhre et al., 2013). In addition, natural sources, such as wetlands and peatlands, contribute significantly to the global and regional CH_4 budget, with a strong temporal and spatial variability (Myhre et al., 2013). Atmospheric CH_4 concentrations have more than doubled since pre-industorial times, and continue to increase even today. In addition, the mechanisms behind the atmospheric CH_4 growth in the 21st century are still not fully understood (Heimann, 2011), where recent studies point out that the anthropogenic sources (Saunois et al., 2016a), biospheric sources (Dlugokencky et al., 2011; Schwietzke et al., 2016) and sinks processes (Ghosh et al., 2015; Montzka et al., 2011) could all be contributing.

In this thesis, an EnKF based data assimilation system, CarbonTracker Europe-CH_4 (CTE-CH_4), is further developed, and the method and applications of the system to estimate regional and global CH_4 budgets are presented. This development aims to give further insight into the global and regional CH_4 budgets for the the 21st century, which would then increase understanding about the reasons for the recent-year atmospheric CH_4 growth. The thesis specifically looked at CH_4 emissions of the largest contributors: anthropogenic and natural biospheric sources. The magnitude, spatial distribution and interannual and seasonal variability of those sources still have high uncertainty, especially on regional and country-scales. Therefore, the system was developed to give regional and grid-wise CH_4 flux estimates, where the anthropogenic and natural biospheric sources are optimised simultaneously.

In **Paper I**, the sensitivity of one site in northern Finland to regional CH_4 budget was examined, where the site was found to be essential for constraining regional biospheric emissions. In **Paper II**, the global and regional CH_4 budgets for 2000-2012 were examined in detail. The estimated emission trend suggested a possible increase in the anthropogenic emissions from northern temperate and tropical regions after 2007. The results were evaluated with various observations, including in-situ and aircraft CH_4 ob-

servations, and satellite- and ground-based dry-air total column CH_4 retrievals. The evaluation showed a good performance of the model in general, but found that a faster vertical mixing scheme in atmospheric transport model gives better agreement with the observations than a slower vertical mixing scheme. In **Paper III**, the European CH_4 budgets were examined from an ensemble of seven inversion systems assimilating newly harmonised observational sets from Europe. The estimated total European budget showed no strong trend, and the estimates were close to each other despite the differences in the modelling systems. However, it was also shown that the estimates were sensitive to atmospheric transport, and the differences in the background concentrations could possibly explain part of the discrepancies. In **Paper IV**, a country-scale budget for Finland, driven by grid-based inversion was evaluated. Although the country-scale budget was still sensitive to the priors and observations assimilated, the example showed that the model is applicable not only for estimating the global budget, but also for the regional budgets.

2 Bayesian inverse modelling

2.1 Bayes' formula

Bayes' formula is a fundamental tool in statistics and in Bayesian inverse modelling in particular, where an unknown parameter θ or state x is inferred from observed data y. The relation between those can be written as

$$y = \mathcal{H}(\theta, x) + \varepsilon, \tag{2.1}$$

where $\mathcal{H}$ is a measurement function (also called the observation operator) which transforms state space to observation space, and ε is an error term. In the case of state estimation, our interest is a probability density function (pdf) of the variable x given the observation y, $p(x|y)$. From the Bayesian probability theory,

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \tag{2.2}$$

where $p(x)$ is a pdf of the state of interest, and $p(y)$ is a pdf of the observed data, and $p(y|x)$ is a conditional pdf of the observations given the state, called a likelihood function. The conditional pdf $p(x|y)$ is often called a posterior pdf, and $p(x)$ a prior pdf. Note that the words "posterior" and "prior" will refer to the pdf of the state to be estimated, but also of "a priori" or "a posteriori" fluxes in this thesis. Note that the interest in the optimization could also be the parameter θ, with a known model state x, but this thesis will be devoted only to state optimization, where θ is omitted from the equation (2.1).

2.2 Cost function

Let $\boldsymbol{x} \in \mathbb{R}^N$ be a vector in a state space. Typically, the pdf $p(\boldsymbol{x})$ is written as

$$p(\boldsymbol{x}) \propto \exp\Big(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^p)\boldsymbol{P}^{-1}(\boldsymbol{x} - \boldsymbol{x}^p)\Big), \tag{2.3}$$

where $\boldsymbol{P} \in \mathbb{R}^{N\times N}$ is an error covariance matrix in the state space, and $\boldsymbol{x}^p$ is a vector of prior states. Similarly, let $\boldsymbol{y} \in \mathbb{R}^S$ be a vector in a observation space, and the conditional pdf $p(\boldsymbol{y}|\boldsymbol{x})$ is typically written as

$$p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\Big(-\frac{1}{2}(\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}))\boldsymbol{R}^{-1}(\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}))\Big), \tag{2.4}$$

where $\boldsymbol{R} \in \mathbb{R}^{S\times S}$ is an error covariance matrix in the observation space, $\mathcal{H}$: $\mathbb{R}^N \to \mathbb{R}^S$ converts the state space to the observation space. In many real-life applications,

the observation and the state spaces are often different. In the case of GHG flux inversion, for example, the observations are atmospheric concentrations and the states are fluxes.

From the Bayes' theorem and probability theory of joint distribution, the posterior pdf $p(\boldsymbol{x}|\boldsymbol{y})$ is proportional to the product of (2.3) and (2.4):

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \propto \exp\Big(-\frac{1}{2}\mathcal{J}(\boldsymbol{x})\Big), \tag{2.5}$$

where the cost function $\mathcal{J}(x)$ is:

$$\mathcal{J}(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{x}^p)^T\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{x}^p) + (\boldsymbol{y}-\mathcal{H}(\boldsymbol{x}))^T\boldsymbol{R}^{-1}(\boldsymbol{y}-\mathcal{H}(\boldsymbol{x})). \tag{2.6}$$

The equation assumes that all error terms are Gaussian. Note that the posterior pdf is at its maximum when the cost function is minimised.

2.3 Kalman filter

Kalman filtering (KF; Kalman, 1960) provides an optimal least-square solution of the cost function for linear operational cases. A general filtering formula consists of a pair of equations for the states and observations:

$$\begin{cases} \boldsymbol{x}_k &= \mathcal{M}(\boldsymbol{x}_{k-1}) + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim N(\boldsymbol{0}, \boldsymbol{Q}) \\ \boldsymbol{y}_k &= \mathcal{H}(\boldsymbol{x}_k) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{R}), \end{cases} \tag{2.7}$$

where $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$ are a state and observation vectors for a discrete time step k, $\mathcal{M}$: $\mathbb{R}^N \to \mathbb{R}^N$ is a dynamical model that describes evolution of the states in time, $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ are random errors of the states and observations respectively, $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ is the background error covariance matrix, and $\boldsymbol{R} \in \mathbb{R}^{S\times S}$ is a observation operator error covariance matrix.

KF is a sequential data assimilation system that consists of two phases: prediction and analysis (update). In the prediction phase, the posterior state and its covariance are moved forward in time based on posterior states from a previous time step before new observations are provided:

$$\boldsymbol{x}_k^p = \boldsymbol{M}\boldsymbol{x}_{k-1}^a \tag{2.8}$$

$$\boldsymbol{P}_k^p = \boldsymbol{M}\boldsymbol{P}_{k-1}^a\boldsymbol{M}^T + \boldsymbol{Q}, \tag{2.9}$$

where $\boldsymbol{M}$ is a linearlised dynamical model, $\boldsymbol{x}_k^p$ and $\boldsymbol{P}_k^p$ are prior (p) states and its error covariance matrix at time k, $\boldsymbol{x}_{k-1}^a$ and $\boldsymbol{P}_{k-1}^a$ are posterior (a) states and its error covariance matrix at time $k-1$.

Once a set of observations are provided, the prior state and its covariance are updated based on the prior states and covariance, provided from the prediction phase, using the Kalman filtering formulae:

$$\begin{aligned} \boldsymbol{x}_k^a &= \boldsymbol{x}_k^p + \boldsymbol{G}_k(\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}_k^p)), \\ \boldsymbol{P}_k^a &= (\boldsymbol{I} - \boldsymbol{G}_k\boldsymbol{H})\boldsymbol{P}_k^p. \end{aligned} \tag{2.10}$$

Here $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{H}$ is a linearised observation operator, and the matrix $\boldsymbol{G}_k$ is called Kalman gain matrix:

$$\boldsymbol{G}_k = \boldsymbol{P}_k^p\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}_k^p\boldsymbol{H}^T + \boldsymbol{R})^{-1}. \tag{2.11}$$

2.4 Ensemble Kalman filter

Problems of KF are a linearity assumption in the system and computational cost in the systems with large number of unknowns. In many real-life applications, including atmospheric inversions, the systems are often non-linear, and linearlisation of the dynamical model $\mathcal{M}$ and observation operator $\mathcal{H}$ may not be possible. In such systems, the predicted state and covariance, and the Kalman gain matrix cannot be calculated explicitly.

Let $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_L)$ be a set of L random samples (ensemble) of the states, drawn from a known pdf, e.g. $N(0, 1)$. In ensemble Kalman filter (EnKF; Evensen, 1994, 2003), the sample covariance $\boldsymbol{P} = \boldsymbol{X}\boldsymbol{X}^T$ is calculated from the ensemble:

$$\boldsymbol{X} = (\boldsymbol{x} - \overline{\boldsymbol{x}})/\sqrt{L-1}, \tag{2.12}$$

where, $\overline{\boldsymbol{x}}$ is a vector of the ensemble sample means. Based on the law of large numbers, the sample covariance becomes the full covariance as $L \to \infty$, and the error due to sampling decreases proportional to $1/\sqrt{L}$. The larger the size of ensemble is, therefore, the better the posterior error covariance is represented, and one should avoid using too small ensemble sizes (van Leeuwen, 1999; Houtekamer and Mitchell, 1998). It is often chosen considering the balance of the representation error and the computational cost.

EnKF is useful for systems where state dimension is large because the computational costs depends on ensemble size, rather than the number of parameters. In the global atmospheric inversion, the state dimension depends on horizontal (latitude and longitude) resolution, and even at $1^\circ \times 1^\circ$ resolution, the dimension of $\boldsymbol{P}$ becomes 64800×64800, which makes the Kalman gain calculation computationally demanding even with today's computational resources.

The prediction and analysis can then be calculated for each ensemble members independently, where the predicted states are represented based on the dynamical model:

$$\boldsymbol{x}_l^p = \mathcal{M}(\boldsymbol{x}_l^a) + \boldsymbol{\eta}_l, \tag{2.13}$$

from which the matrix $\boldsymbol{P}^p$ is formed. Due to the nature of EnKF, the variance of the posterior ensemble decreases from the prior, and therefore, the randomisation by the background error $\boldsymbol{\eta}_l$ plays an important role in avoiding the variance of the ensemble to converge to unrealistically small values.

Various dynamical model $\mathcal{M}$ can be applied, but only a simple averaging for the mean states is applied in this thesis following Peters et al. (2007) (see Section 4.1 for details). A choice of dynamical model for the mean is rather straightforward, but not for the covariance because problems such as under- or overestimation of the sample variance, and severe decreases in the degree of freedom could easily occur. In those cases, an additional method is required to inflate or deflate the sample deviation and to regain a sufficient degree of freedom in the covariance matrix. In atmospheric inversion systems, constructing an appropriate dynamical model for sample covariance matrix can be challenging, and this method is not applied in this thesis.

For the analysis phase, we need realisations of the observations $\mathcal{H}(\boldsymbol{x}_l^p)$ from the ensemble. The analysis phase then becomes:

$$\boldsymbol{x}_l^a = \boldsymbol{x}_l^p + \boldsymbol{G}(\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}_l^p) + \boldsymbol{\varepsilon}_l). \tag{2.14}$$

Here, it is important to treat the observations as random variables by adding the observation errors $\boldsymbol{\varepsilon}_l$ and generating randomised observations for each ensemble member (Evensen, 2003).

In the model presented in this thesis, the states $\boldsymbol{x}$ to be optimised are scaling factors for global CH_4 flux fields, where "prior" knowledge of CH_4 fluxes were obtained from process-based models. The dimension of the state vector at single time t is not high-dimensional for **Papers I**, **II** and **III**, as CH_4 fluxes were optimised regionally. However, for **Papers IV**, the dimension is increased substantially by optimising CH_4 fluxes on $1^\circ \times 1^\circ$ scale over Europe. The dynamical model $\mathcal{M}$ is close to the identity matrix $\boldsymbol{I}$, the observations $\boldsymbol{y}$ are the measurements of atmospheric CH_4 concentrations, and the observation operator $\mathcal{H}$ is a highly non-linear atmospheric chemistry transport model (see Section 4 for detailed description).

2.5 Ensemble square root filter

One of the problems in the traditional EnKF is so called "inbreeding" problem, where the same ensemble is used to calculate the Kalman gain and update the states. This

creates a systematic underestimation in the analysis error covariance matrix $\boldsymbol{P}^a$ and leads to filter divergence, unless the observations for each ensemble member are treated as random variables (Houtekamer and Mitchell, 1998; Burgers et al., 1998).

Whitaker and Hamill (2002) introduced an alternative filtering approach (the ensemble square root filter; EnSRF), where the observation perturbation in (2.14) is not needed, but the analysis error covariance is still properly estimated assuming that the observations are uncorrelated.

In EnSRF, the matrices $\boldsymbol{HPH}^T$ and $\boldsymbol{PH}^T$ needed for the Kalman gain calculation (2.11) are represented from the ensembles (Whitaker and Hamill, 2002):

$$\boldsymbol{HP}^p\boldsymbol{H}^T \approx \frac{1}{L-1}(\mathcal{H}(\boldsymbol{x}'_1),...,\mathcal{H}(\boldsymbol{x}'_L)) \cdot (\mathcal{H}(\boldsymbol{x}'_1),...,\mathcal{H}(\boldsymbol{x}'_L))^T \tag{2.15}$$

$$\boldsymbol{P}^p\boldsymbol{H}^T \approx \frac{1}{L-1}(\boldsymbol{x}'_1,...,\boldsymbol{x}'_L) \cdot (\mathcal{H}(\boldsymbol{x}'_1),...,\mathcal{H}(\boldsymbol{x}'_L))^T, \tag{2.16}$$

where $\boldsymbol{x}'_l = \boldsymbol{x}_l - \overline{\boldsymbol{x}}$ are sample deviations.

While EnKF uses the same gain to update the mean states and the sample deviations, EnSRF employs a scaled gain for the sample deviations. In EnSRF, the mean states are updated in the same way as the traditional EnKF, i.e. the following equation (2.14), and the sample deviations $\boldsymbol{x}'_l$ are updated as:

$$\boldsymbol{x}'^a_l = \boldsymbol{x}'^p_l + \tilde{\boldsymbol{G}}(\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}'^p_l)), \tag{2.17}$$

where the revised gain $\tilde{\boldsymbol{G}}$ is

$$\begin{aligned} \tilde{\boldsymbol{G}} &= \alpha\boldsymbol{G} \\ \alpha &= \left(1 + \sqrt{\frac{\boldsymbol{R}}{\boldsymbol{HP}^p\boldsymbol{H}^T + \boldsymbol{R}}}\right)^{-1}. \end{aligned} \tag{2.18}$$

In EnSRF, each individual observations are processed sequentially one at a time, and therefore, $\boldsymbol{R}$ and $\boldsymbol{HP}^p\boldsymbol{H}^T$ are scalars and α is simply a constant.

Although this is an efficient algorithm, the calculation of $\mathcal{H}(\boldsymbol{x}'^b_l)$ by reapplying the observation operator $\mathcal{H}$ is computationally expensive in atmospheric inversion. Therefore, the modelled realisation of the observations yet to be assimilated $\mathcal{H}(\boldsymbol{x}_k)_m$ is updated following Peters et al. (2005). From the mean states,

$$\mathcal{H}(\boldsymbol{x}^a_k)_m = \mathcal{H}(\boldsymbol{x}^p_k)_m + \mathcal{H}_m\boldsymbol{G}(\boldsymbol{y}_k - \mathcal{H}(\boldsymbol{x}^p_k), \tag{2.19}$$

and from the deviations,

$$\mathcal{H}(\boldsymbol{x}'^a_k)_m = \mathcal{H}(\boldsymbol{x}'^p_k)_m + \mathcal{H}_m\tilde{\boldsymbol{G}}(\mathcal{H}(\boldsymbol{x}'^p_k). \tag{2.20}$$

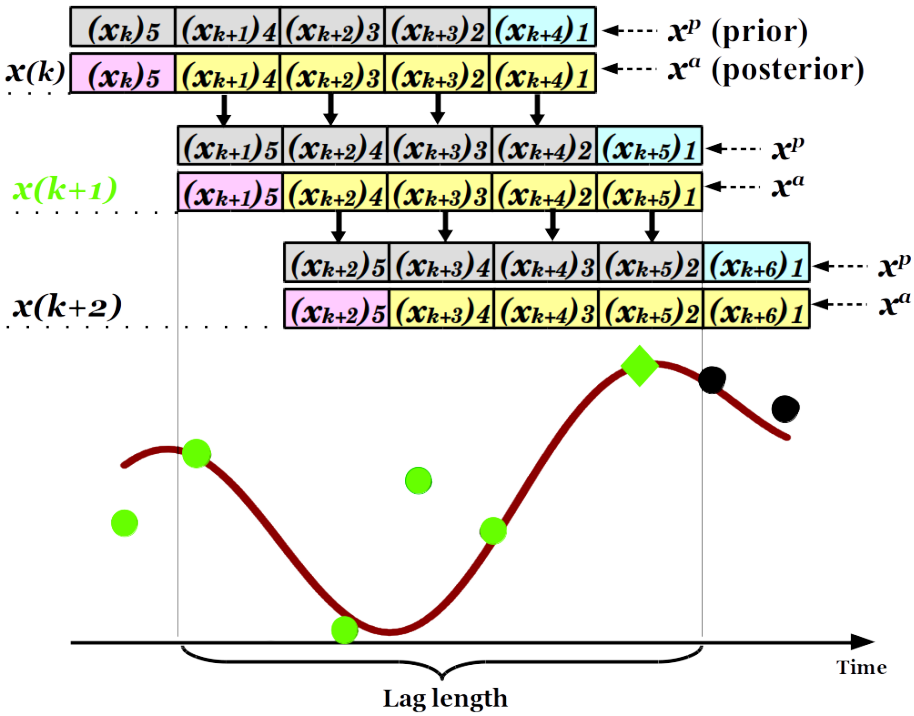


Figure 1: Representation of fixed-lag filtering with a lag length of 5. The state vectors $\boldsymbol{x}(k)$, $\boldsymbol{x}(k+1)$ and $\boldsymbol{x}(k+2)$ are joint state vectors at time k, $k+1$ and $k+2$, containing five state vectors illustrated in boxes. State vectors in grey and light blue boxes are priors $\boldsymbol{x}^p$ which are used to calculate posterior states $\boldsymbol{x}^a$. The red posterior states are the final results, while the yellow posterior states are inferred to next time step as priors illustrated in grey. The time series shows states to be optimised (red line), and observation sets (dots). At time $(k+1)$, the new observations at $k+lag$ (green diamond) are assimilated, i.e. the observation sets up to the final lag (green circles) have been assimilated.

2.6 Fixed-lag filtering

In the traditional KF, the states $\boldsymbol{x}_k$ only depend on observations up to time k, i.e. the posterior pdf to be estimated is $p(\boldsymbol{x}_k|\boldsymbol{y}_1, ..., \boldsymbol{y}_k)$. However, it is known that the smoothed estimates which take future observations into account are more accurate for intermediate times (van Leeuwen and Evensen, 1996; Evensen and van Leeuwen, 2000; Evensen, 2003). In this thesis, a fixed-lag filtering is applied following Peters et al. (2005), where the state vector that contains future time steps is estimated.

Let $\boldsymbol{y}_{1:K} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_K)$ be all available observations and $\boldsymbol{x}_{1:K} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_K)$ be the states to be optimised at discrete times $k \in [1, K]$. The prediction and analysis are done based on equations (2.8) and (2.10), but the state vector to be optimised contains

future states:

$$\boldsymbol{x}(k) = \boldsymbol{x}_{k:k+\tau}, \tag{2.21}$$

where τ is called assimilation window or lag length and defines a length of time window up to which the observations have influence over, and the observations used to update are

$$\boldsymbol{y}(k) = \begin{cases} \boldsymbol{y}_{k:k+\tau} & \text{for } k = 0 \\ \boldsymbol{y}_{k+\tau} & \text{for } k > 0 \end{cases} \tag{2.22}$$

Then, the state vector $\boldsymbol{x}(k+1)$ contains state vectors from time $k+1, ..., k+1+\tau$, where $x_{k+1+\tau}$ is a new state, and others $(x_{k+1}, ..., x_{k+\tau})$ are those inferred from the previous time step. The analysis equation for $k > 0$ can be written as:

$$\begin{bmatrix} (\boldsymbol{x}_k^a)_\tau \\ \vdots \\ (\boldsymbol{x}_{k+\tau}^a)_1 \end{bmatrix} = \begin{bmatrix} (\boldsymbol{x}_k^p)_\tau \\ \vdots \\ (\boldsymbol{x}_{k+\tau}^p)_1 \end{bmatrix} + \boldsymbol{G}_{k+\tau}(\boldsymbol{y}_{k+\tau} - \mathcal{H}(\boldsymbol{x}_{k+\tau}^p)). \tag{2.23}$$

The "intermediate" states $(\boldsymbol{x}_{k+1}^a)_{\tau-1}, ..., (\boldsymbol{x}_{k+\tau}^a)_1$ at time $k+1, ..., k+\tau$ are those updated $\tau-1, ..., 1$ times (Fig. 1), respectively, but not the final result. For each state, updating is done τ times, and therefore only $(\boldsymbol{x}_k^a)_\tau$ is the final result at time (k). The prior state $(\boldsymbol{x}_{k+\tau}^p)_1$ is the first initial state, which is used to calculate the posterior state $(\boldsymbol{x}_{k+\tau}^a)_1$. The intermediate prior states are those inferred from previous time steps (Fig. 1), i.e.

$$\begin{bmatrix} (\boldsymbol{x}_k^p)_\tau \\ \vdots \\ (\boldsymbol{x}_{k+\tau-1}^p)_2 \end{bmatrix} = \begin{bmatrix} (\boldsymbol{x}_k^a)_{\tau-1} \\ \vdots \\ (\boldsymbol{x}_{k+\tau-1}^a)_1 \end{bmatrix}. \tag{2.24}$$

In another words, the posterior states $(\boldsymbol{x}_k^a)_{\tau-1}, ..., (\boldsymbol{x}_{k+\tau-1}^a)_1$ in $\boldsymbol{x}(k)$ is considered as "prior" in $\boldsymbol{x}(k+1)$. In this thesis, this approach is considered as filtering rather than smoothing because the backward calculus, required for proper smoothing, is not applied.

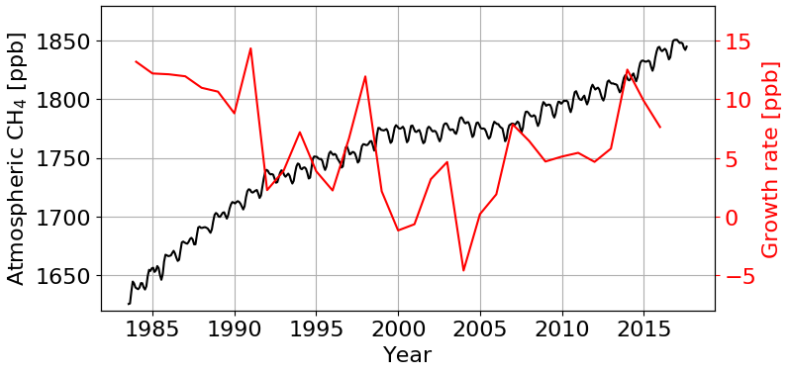


Figure 2: Global average surface atmospheric CH_4 concentration and its growth rate. Data source: https://www.esrl.noaa.gov/gmd/.

3 CH_4 balance and atmospheric concentrations in the 21st century

3.1 Atmospheric CH_4

Methane (CH_4) is an important greenhouse gas (GHG), which is directly influenced by anthropogenic emissions and its atmospheric concentrations have increased substantially in recent centuries. The global mean atmospheric CH_4 increased from about 700 ppb in pre-industrial times to 1843 ppb in 2016, and continues to increase even today. The effective radiative forcing of CH_4 since pre-industrial times to 2016 is +0.507 ± 0.05 W m^{-2} (update of Hofmann et al. (2006), https://www.esrl.noaa.gov/gmd/aggi/aggi.html). The growth rate (GR) of atmospheric CH_4 varies interannually due to interannual variability in the CH_4 fluxes and atmospheric sinks, but the highest GR measured before the 21st century was 14.33 ppb yr^{-1} in 1991, while the lowest was 2.25 ppb yr^{-1} in 1996 (NOAA: globally averaged marine surface annual mean, https://www.esrl.noaa.gov/gmd/ccgg/trends_ch4/) (Fig. 2).

One of the issues of interest about the 21st century atmospheric CH_4 is its GR. Following years of a steady period during 1999-2006, when the atmospheric CH_4 stayed around 1771–1774 ppb, the atmospheric CH_4 started to increase again in 2007 with a GR of about 5 ppb yr^{-1} or even higher (Dlugokencky et al., 2011) (Fig. 2), suggesting that a significant change has occurred in the global CH_4 budget. The average GR during 2007-2016 was 7.08 ppb yr^{-1}, which is even higher than that of the 1990s (6.34 ppb yr^{-1}). The mechanisms behind this growth are still not sufficiently explained, and various potential reasons have been discussed (Heimann, 2011). Saunois et al. (2016a) examined the GR from an ensemble of various process-based and inverse models, which

showed a significant increase in the anthropogenic sources. Dlugokencky et al. (2011) found that the increase in the atmospheric CH_4 growth corresponds to a decrease in $^{13}\delta$C-CH_4 isotopic signals especially in the Tropics, suggesting a potential increase in biogenic sources. Assuming fossil fuel based anthropogenic CH_4 emissions to be smaller than those estimated by earlier studies (Schwietzke et al., 2016), the increase could potentially be caused by agricultural CH_4 emissions (Saunois et al., 2016b). In addition, a change in OH concentrations could also cause changes in atmospheric CH_4 growth (Ghosh et al., 2015; Dalsøren et al., 2016), and the OH concentrations were estimated to have decreased after 2005 (Montzka et al., 2011).

CH_4 is a reactive chemical compound, which is removed by the hydroxyl (OH) radical, chlorine (Cl) and electronically excited atomic oxygen ($O(^1D)$) in the atmosphere.

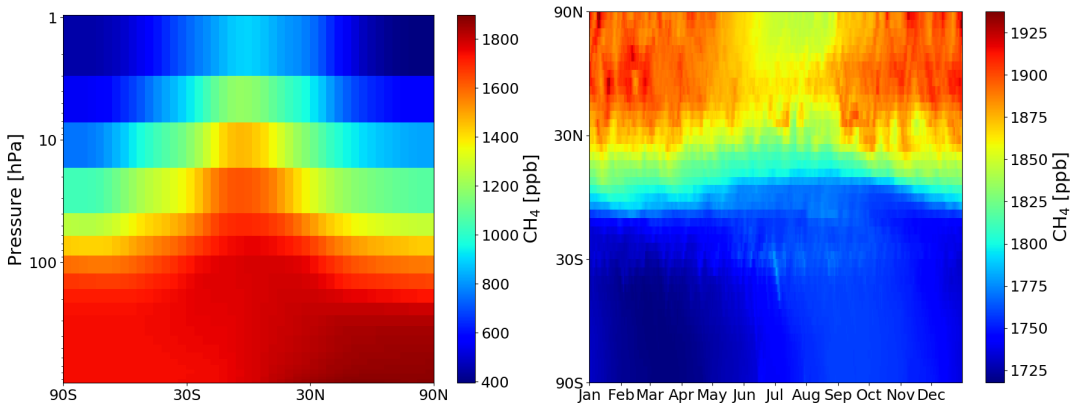
$$\begin{aligned} CH_4 + OH &\longrightarrow CH_3 + H_2O \\ CH_4 + Cl &\longrightarrow CH_3 + HCl \\ CH_4 + O(^1D) &\longrightarrow CH_3 + OH \end{aligned} \tag{3.1}$$

The CH_4 removal due to chemical reaction with tropospheric OH is the largest CH_4 sink, which is about 90% of the total sink. Other chemical reactions are small sinks, but important in stratospheric chemistry. Due to these removals, emitted CH_4 is estimated to stay in the troposphere only for about 9 years. The OH concentrations vary seasonally, as OH is produced through photodissociation of ozone.

$$\begin{aligned} O_3 + \text{sunlight} &\longrightarrow O_2 + O(^1D) \\ O(^1D) + H_2O &\longrightarrow 2\,OH \end{aligned} \tag{3.2}$$

Therefore, the removal by OH is the highest during summers, in the upper troposphere and the Tropics, where solar radiation is intense and water vapour concentrations are high. On the other hand, the OH concentrations are low during winters, and lower in the northern and southern high latitudes then in the Tropics.

The spatial distribution of atmospheric CH_4 depends on several factors: emissions, sinks and transport. Atmospheric CH_4 is high in the troposphere where the surface emissions of CH_4 are transported to the atmosphere and mixed well, but much lower in the stratosphere (Fig. 3(a)). The vertical distribution of the CH_4 concentrations show that CH_4 decreases above the tropopause with a much faster rate than below it (Fig. 3(a)). In addition, the atmospheric CH_4 is higher in the Northern Hemisphere (NH) than in the Southern Hemisphere (SH) (Fig. 3(b)). This is mainly due to emission distribution; most of the CH_4 sources are located in the NH and the Tropics, and much less in the SH. The seasonal cycle of the atmospheric CH_4 shows high concentrations



(a) (b)

Figure 3: Atmospheric CH_4 concentration distributions. (a) Vertical profile of zonal mean atmospheric CH_4, and (b) Lower troposphere (lowest 5 levels, approximately >850 hPa) zonal mean atmospheric CH_4 over a year.

during winter and low concentrations in the summer, which is mainly driven by the atmospheric sink (Fig. 3(b)).

3.2 Global and regional CH_4 budgets

Average annual global CH_4 emission for 2000–2012 is assumed to be about 526–582 Tg CH_4 yr^{-1} based on an ensemble of several inverse models (Saunois et al., 2016a; Kirschke et al., 2013). The largest source of CH_4 is anthropogenic emissions, such as fugitive emissions from solid fuels, leaks from gas extraction and distribution, agriculture, landfills and waste water management, which in total account for more than half of the global total emissions (Saunois et al., 2016a; Kirschke et al., 2013; Ciais et al., 2013). The anthropogenic emissions have an increasing trend that is closely related to economical and population growth. Although the seasonal cycle of global anthropogenic emissions is assumed to be small, the emissions from agriculture, especially from rice cultivation, have a strong seasonal cycle depending on the rice growing seasons. The emissions from oil and gas could also have seasonal cycles e.g. in northern countries, where emissions from heating are possibly high during winter. The spatial pattern or regional budgets therefore differ between continents and countries (Fig. 5), and also depends on the policies applied and available energy sources.

The second largest source is natural biospheric emissions from wetlands and peatlands, which accounts for about 30% of global total emissions (Saunois et al., 2016a; Kirschke et al., 2013). The natural biospheric GHG emissions are very sensitive to climate con-

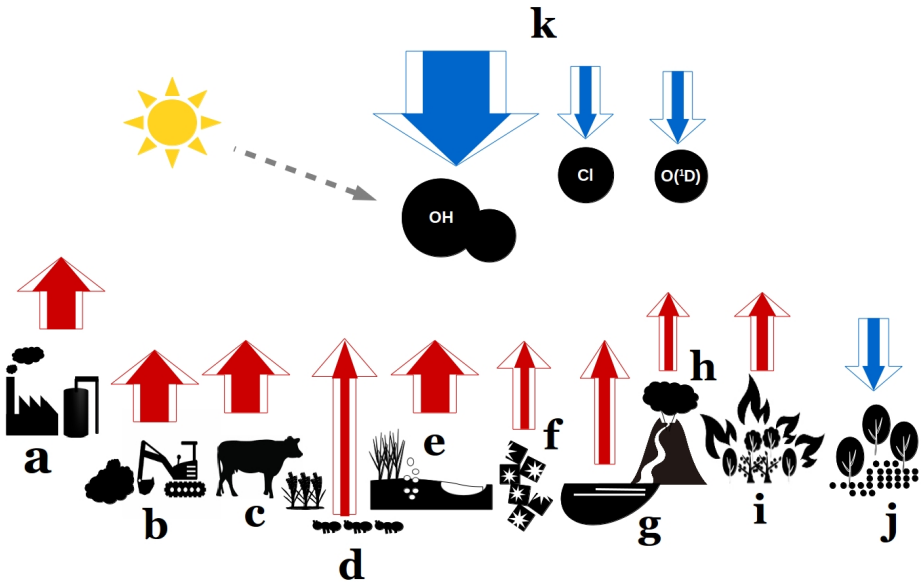


Figure 4: Sources (red) and sinks (blue) of CH_4. The width of the arrows approximately illustrates the relative magnitude of the fluxes (filled) and their uncertainty (non-filled). a) solid fuels, oil and gas extraction and distribution b) landfills and waste management c) agriculture d) termites e) wetlands, peatlands and fresh water f) permafrost and CH_4 hydrates g) open ocean h) natural geological i) biomass burning j) forests and dry mineral soils k) atmospheric chemistry.

ditions, and therefore, it is important to understand the feedbacks and interactions of those ecosystems with atmosphere (Heimann and Reichstein, 2008). In the natural biogenic ecosystems, CH_4 is produced as a result of micro organic (methanogen) respiration. In anaerobic conditions, methanogens use oxygen to grow, and produce CH_4 as a result (methanogenesis). This process also applies to biogenic anthropogenic sources, such as those from rice paddies, livestock herbivores (cows, sheep, deer, etc) and waste treatment, where methanogens are active in paddy sediments, digestive systems and waste matter. In addition, special methanogens live in termites. The CH_4 emissions from termites are not significantly large in the global budget (about 5-10%), but are important for the regional budget (Jamali et al., 2013; Khalil et al., 1990; Jamali et al., 2011). In wetlands and peatlands, CH_4 is produced in the soil sediments, and emitted mainly by diffusion through the surface water layer, ebullition from the soil layer, and transport through the aerenchyma in plant stems. The natural biospheric CH_4 emissions from wetlands and peatlands are highly sensitive to water table depth, precipitation and soil temperature which affect the activity level of methanogens. In general, the higher the soil temperature is, the more CH_4 is emitted. This creates a clear seasonal cycle

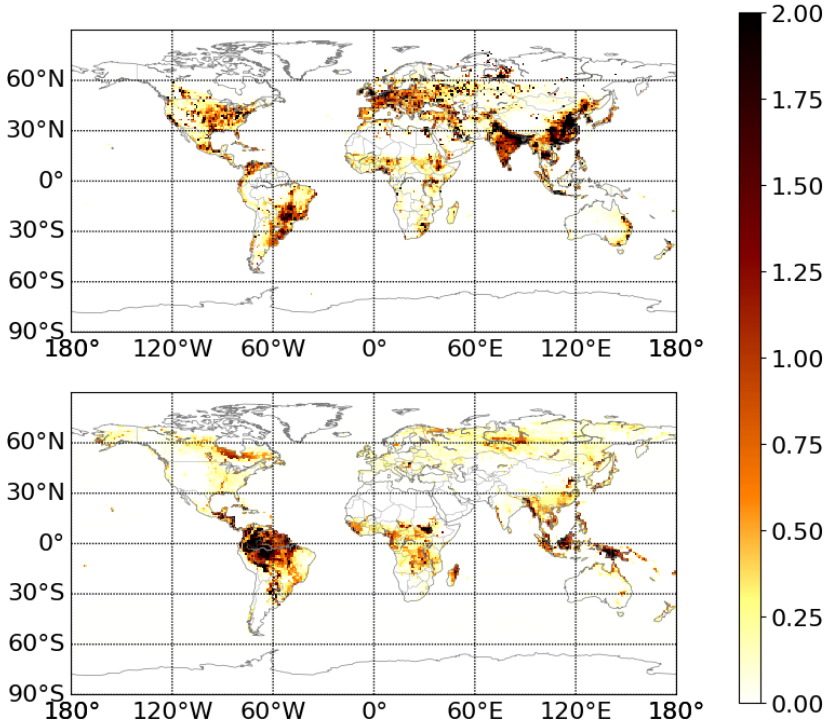


Figure 5: Global CH_4 emission distributions of prior anthropogenic (top), and prior natural (wetlands and peatlands, fire, termites and ocean) (bottom) sources from CTE-CH_4 in the unit of [$10^{-8} mol\ m^{-2}\ s^{-1}$].

in the natural biospheric emission that is high during summer and low in winter. In addition, the hydrology such as the water table depth and soil moisture also affect the seasonal cycle in methane emissions. However, not only the meteorological conditions, but also the soil properties such as amount of nutrients and carbon, and vegetation types also influence the situation on, and therefore, the actual seasonal cycle and its amplitude have a high interannual and spatial variability. A large area of wetlands is located in the Tropics and there are peatlands in the northern Boreal regions (Fig. 5). Peatlands are assumed to store about one fifth of global terrestrial carbon (Ciais et al., 2013), which could possibly expand with global warming (Walter et al., 2006; McGuire et al., 2012; Johansson et al., 2006).

Other than wetlands and peatlands, similar micro organic processes also occur in lake bottoms. Although the magnitude of contribution from inland water emissions is still uncertain, they could have a significant contribution globally and especially in the northern high latitudes (Thonat et al., 2017; Walter et al., 2006), where about 40% of the total area of inland waters is located. In addition, the upland mineral soils and forests

contribute to natural biospheric CH_4 budget, where in general wet mineral soils are often a net source of CH_4, and dry soils and forests are a net sink (Yavitt et al., 1990; Guckland et al., 2009; Dutaur and Verchot, 2007; Lohila et al., 2016). In dry soils, the methanogenesis is often exceeded by oxidation in methanotrophy (microbial consumption), where a significant percentage of the CH_4 produced in soils and sediments could be consumed (Conrad, 1996).

Other natural sources, such as natural biomass burning and geological sources are minor sources of the global CH_4 budget, but are significant in regional and seasonal budgets. The natural biomass burning occurs in ecosystems such as forests, savannas, grasslands and peatland during dry seasons. The emissions are therefore high during summer around the Tropics, where large areas of forests and savannas are located, but contributions are also made by temperate and boreal forests. The CH_4 emissions from forest fires have significant interannual variability, and the annual CH_4 emission varies from 11 to 20 Tg CH_4 yr^{-1} (during 2000-2014; Giglio et al., 2013). Natural geological sources, such as volcanic eruptions, can occasionally be significant sources of CH_4. The emission source is often local, but the emitted CH_4 spreads regionally and globally in the atmosphere though specific transport patterns, that affect the atmospheric CH_4 levels in the short (hours) and long (several years) term. For example, the Mount Pinatubo eruption in 1991 resulted in an increase in CH_4 GR shortly after the event, but the long-term effect was a decrease in the GR due to depletion of the stratospheric ozone (Bândă et al., 2013).

Unlike CO_2, the ocean is a minor contributor to the CH_4 budget because CH_4 emitted from the sea sediments is mostly oxidised before it reaches the atmosphere. However, the Arctic ocean could become a larger source of CH_4, as significant amounts of CH_4 hydrates are located in the sea sediments (Kvenvolden, 1988). CH_4 emissions from the Arctic sea are sensitive to climate change because the increase in the Arctic sea temperature could directly affect destabilisation of the CH_4 hydrates (Biastoch et al., 2011) and the extent of sea-ice. As the Arctic ocean is shallower than other open oceans, the oxidative water depth is shallow, and the CH_4 release from the hydrates due to warming could directly affect the CH_4 emissions to the atmosphere. The extent of sea-ice is decreasing due to global warming, which could also release CH_4 that would otherwise be trapped in the ice (Kort et al., 2012).

3.3 Modelling of CH_4 budget

Modelling is necessary in order to understand global or regional CH_4 budgets, as spatial coverage of direct flux measurements is limited. The global and regional budgets can be estimated by various models that can mainly be put into two categories: inventories or process-based models that derive flux estimates based on process-related theories of each

source, and data-driven atmospheric inverse models that infer emission estimates from atmospheric concentration observations. Although this thesis focuses on the inverse models, the estimates from inventories and process-based models are often used as prior fluxes in the inverse models, and therefore, it is important to understand the basic mechanisms in both models.

For the estimation of anthropogenic emissions, information such as regional statistics from counties, states and municipalities, together with population and other known spatial distributions can be used. The country statistics such as fossil fuels, gas and oil production and use, amount of landfill and waste water, and livestock population are often used as emission scaling factors to convert each metric to amounts of GHG emissions. The country GHG emission statistics are then reported to the United Nations Framework Convention on Climate Change (UNFCCC), for example. The country statistics can be distributed with some suitable information on spatial distribution to derive grid-based estimates. For example, the landfill and waste water distribution can be derived using population distribution and information about the location of treatment plants. The livestock statistics can be distributed with the animal density map or agriculture distribution map. Although information from developed countries is often reliable, the information from developing countries is often missing or has high uncertainty.

The process-based models are useful in estimating natural emissions. Information on meteorological and climate conditions, soil and vegetation types and conditions, and their distributions are used to estimate CH_4 emissions by modelling biogeochemical processes and transport in the soil and water layers. It is important to note that the regional estimates from process-based models could differ significantly between models (Bohn et al., 2015). One important factor that affects the CH_4 emission estimates is the extent of wetlands and peatlands. Those models use either prescribed or dynamically estimated distribution of those extents, which are used to scale the emission in each grid cell.

A simple way of estimating ocean emissions is by calculating the product of gas transfer velocity, gas solubility and pressure differences in the sea and atmosphere. Although the spatial distribution of ocean emissions is not accurately known, its contribution to total global budget is small that it will not bring significant additional uncertainty in global estimates.

However, the process-based models are not designed to have a closed global budget. Because their interest is only in a part of the emission sources or sinks, the other sources and sinks that contribute to the global budget are not considered. Therefore, this can create a large range of the emission estimates (Saunois et al., 2016a), which could be unrealistic when associated with other source information, i.e. the estimated atmospheric CH_4 concentrations from those emissions may not agree with the observed

levels.

Here, the atmospheric inverse models become useful. Although they cannot estimate processes separately as in the process-based models, the inverse models can take several sources and sinks into account at once. This results in global budget that is better closed, where the range of estimates between the models are smaller than those estimated from process-based models (Saunois et al., 2016a; Kirschke et al., 2013). Atmospheric inverse models often use estimates from inventories and process-based models and constrain total budget using atmospheric concentration measurements. The information such as spatial distribution, seasonal cycle, interannual variability, trend, and the magnitude of the emission from the inventories and process-based model could be provided to inverse models. The mathematics behind the inverse models is based on Bayesian probability theory, as explained in Section 2.1–2.2. A traditional way of solving the cost function is called the variational method (e.g. 4DVAR; Houweling et al., 1999, 2014; Bergamaschi et al., 2005, 2009; Bousquet et al., 2011), where the derivative of the cost function is calculated based on the numerical minimisation method. Another method is statistical filtering (Peters et al., 2005; Bruhwiler et al., 2005, 2014; Chen and Prinn, 2006), which is applied in this thesis (Section 2.3–2.4), where the maximum estimator of the cost function is estimated based on probability theories. Both methods have been found to derive similar flux estimates, although temporal correlation was better resolved in 4DVAR and computational efficiency was better in the EnFK-based model (Babenhauserheide et al., 2015).

In the inverse models, the choices of optimization resolution (temporal and horizontal), prior state covariance, observation error covariance, ensemble size in the ensemble filtering, all affect the results (Peters et al., 2005; Babenhauserheide et al., 2015). Furthermore, the inputs such as prior fluxes and the observations affect the results (Houweling et al., 2014). Although inversions should ideally give the same emission estimates regardless of the prior fluxes, the results are still affected by the choice especially in regions where observation constrains are not enough (Bergamaschi et al., 2005). Although high optimization resolution often helps to better close the budget and to resolve spatial distributions (Bergamaschi et al., 2015), the prior flux estimates are especially important in regions where observation constraints are weak. In addition, their spatial distributions directly affect the results in regional-based optimization.

Since it is important to produce realistic atmospheric CH_4 at observed times and places, atmospheric inverse models are often associated with atmospheric transport models (ATMs), which are used as the observation operator $\mathcal{H}$. The ATM calculates atmospheric states and gas concentrations based on physical theories, constrained by meteorological inputs. The atmospheric sink is often taken into account through ATMs in the inverse models by either using prescribed removal rate, or calculating the removal rate dynamically. Note that despite its importance, the spatial distribution and interannual variation of the atmospheric sink have high uncertainties. There exits a variety

of ATMs, and the inversion estimates can vary up to 150% in the regional level by using different ATMs (Locatelli et al., 2013). The differences are associated with model parameterisation, horizontal, vertical and temporal resolution, and meteorological constraints.

4 CarbonTracker Europe-CH_4

CarbonTracker Europe-CH_4 (CTE-CH_4) is the main inverse modelling tool used in **Paper I**, **II** and **IV**, and it contributed to the model comparison study in **Paper III**. The model is based on ensemble Kalman fixed-lag filtering, developed to make inference over global and regional CH_4 budgets by assimilating atmospheric CH_4 concentration observations.

4.1 State optimization in CTE-CH_4

In CTE-CH_4, the state to be optimised x is a scaling factor for first guess (prior) flux estimates F^p:

$$f(x, F^p) = x \times F^p, \tag{4.1}$$

and the cost function (2.6) becomes:

$$J(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}')^T \boldsymbol{P}^{-1}(\boldsymbol{x} - \boldsymbol{x}') + (\boldsymbol{y} - \mathcal{H}(f(\boldsymbol{x}, \boldsymbol{F^p})))^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathcal{H}(f(\boldsymbol{x}, \boldsymbol{F^p})))), \tag{4.2}$$

where $\boldsymbol{y}$ is a set of atmospheric CH_4 observations, the observation operator $\mathcal{H}$ is an ATM which transforms the flux estimates to the observation space. CTE-CH_4 is also a time evolving model, where CH_4 fluxes are optimised sequentially at weekly resolution, with a lag of 5 weeks as the default (see Section 4.1.1 for the sensitivity to lag length). Note that although scaling factors $\boldsymbol{x}$ are the states to be optimized, we discuss the actual fluxes, i.e. $f(\boldsymbol{x}, \boldsymbol{F^p})$, in terms of CH_4 emissions.

As discussed in Section 3.3, large uncertainty in CH_4 flux estimates is associated with to two major sources: anthropogenic and natural. Those source distributions and temporal variability are challenging to understand accurately due to limited coverage of observations and information. CTE-CH_4 is therefore designed to optimise the two sources simultaneously. Since spatial distribution of the two sources can be found from inventories and process-based models, the two sources were optimised region-wise based on modified TransCom regions (mTC) and a terrestrial land-ecosystem type (LET) map (e.g. Fig. 3 and 4 in **Paper I**). The original TransCom regions consist of terrestrial and ocean regions used in the TransCom project (http://transcom.project.asu.edu/). The LET region in CTE-CH_4 is first defined based on soil types used in process-based models. The LET soil type definitions include e.g. peatland, land with mineral soil and inundated land (wetland). The LET also contains "anthropogenic" land, such as cities and rice fields. In the case of **Paper II**, the number of mTCs (N_{mTC}) is 20 and the number of LET (N_{LET}) is 5, and the total number of optimization regions in theory is $N_{\mathrm{all}} = N_{\mathrm{mTC}} \times N_{\mathrm{LET}} = 100$. However, the actual number of optimization regions was $N_{\mathrm{all}} = 62$ because not all mTCs contained all LETs. This region-wise optimization

generally works, but as shown in **Paper III**, the spatial distribution very much depends on the prior estimates, and may not be able to resolve the emission distribution well. Therefore, in **Paper IV**, Europe was further divided into grid-based optimization regions. **Paper IV** showed that the grid-based inversion works well for regions where the observation network is dense. The state vector in CTE-CH_4 then becomes $\boldsymbol{x} = (\boldsymbol{x}_{\text{anth}}, \boldsymbol{x}_{\text{bio}})$, where

$$\begin{aligned} F^a_{\text{total}}(k,r) = x^a_{\text{anth}}(k,r) \times F^p_{\text{anth}}(k,r) + x^a_{\text{bio}}(k,r) \times F^p_{\text{bio}}(k,r) \\ + F^p_{\text{fire}}(k,r) + F^p_{\text{termites}}(k,r) + F^p_{\text{ocean}}(k,r). \end{aligned} \tag{4.3}$$

Here, $x(k,r)$ denotes a scaling factor at time (week) k and region (or grid) r for anthropogenic (anth) and biospheric (bio) emissions, F^a is optimised regional (or grid) total emissions, and F^p is the prior emission estimated from inventories and process models.

CTE-CH_4 applies a simple dynamical model for the mean states following Peters et al. (2007) by taking averages of previously updated states:

$$\overline{\boldsymbol{x}}^p_k = (\overline{\boldsymbol{x}}^a_{k-1} + \overline{\boldsymbol{x}}^a_{k-2} + \boldsymbol{I})/3.0, \quad \text{for } t \leq 2. \tag{4.4}$$

The identity matrix $\boldsymbol{I}$ is added to regularise the prior around 1, such that if no information is obtained from the observations, the scaling factor equals one and thus the flux remains as the prior. Note that this dynamical model is not applied to the ensemble deviations, and new deviations are randomly drawn at each time step (2.13).

The ensemble of the prior deviations $\boldsymbol{x}'^p_k$ is drawn from a normal distribution, $N(\boldsymbol{0}, \boldsymbol{Q})$. The matrix $\boldsymbol{Q}$ is a background error covariance matrix that defines prior state variance and spatial correlation between the optimization regions. Note that in EnKF, the prior error covariance $\boldsymbol{P}^p$ is built from the ensemble of the samples $\boldsymbol{x}^p_k = \overline{\boldsymbol{x}}^p_k + \boldsymbol{x}'^p_k$.

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{\text{anth}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_{\text{bio}} \end{pmatrix}, \tag{4.5}$$

where $\boldsymbol{Q}_{\text{anth}}$ and $\boldsymbol{Q}_{\text{bio}}$ are the covariance matrices for anthropogenic and biospheric emissions, respectively, and those emissions are assumed to be uncorrelated in space. The diagonals of $\boldsymbol{Q}$ represent variance, and the off-diagonals representing spatial correlation between optimization regions are defined based on the great circle distance d_{r^1,r^2} between the two regions (r^1, r^2):

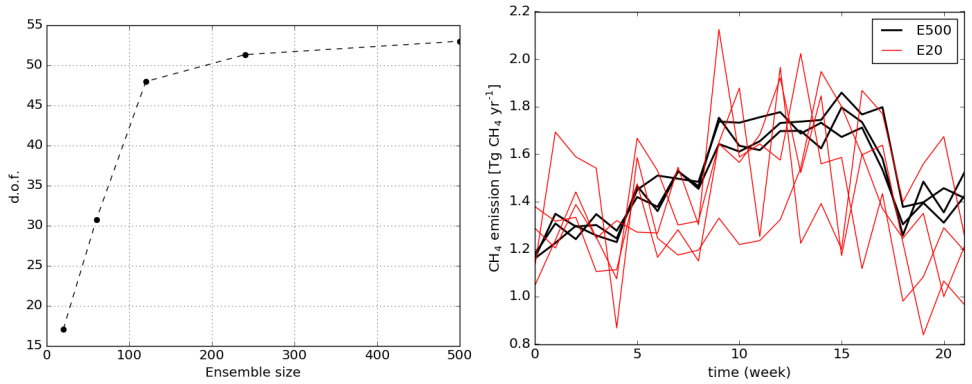
$$q_{r^1,r^2} = q_{r^1} \times \exp(-d_{r^1,r^2}/\Lambda), \tag{4.6}$$

where q_{r^1} is a diagonal element (prior uncertainty) for region r^1, and Λ is a pre-defined spatial correlation length. The correlation length Λ could be chosen based on the horizontal resolution of the atmospheric transport model, and the distances between

observations. In CTE-CH_4, where the coarsest horizontal resolution in TM5 ATM is $6°\times4°$, the correlation length is chosen to be 500 km over land, and 900 km over ocean (distance between two centres of TM5 grids at the Equator is approximately 670 km). However, as discussed in **Paper IV**, a shorter correlation length would be more appropriate for grid-based inversions with a dense observation network because too long correlation length makes it difficult to resolve local fluxes from the multiple observations.

The definition of the background error covariance matrix $\boldsymbol{Q}$ is somewhat arbitrary and subjective, and those used in the papers presented in this thesis differ slightly in each study. In **Paper II**, the effect of the covariance matrix is briefly studied by comparing results using a diagonal matrix and another matrix including off-diagonals. As shown in **Paper II** the structure of the prior covariance does not affect the mean estimates much, but the uncertainty estimates are significantly reduced when an informative covariance matrix with off-diagonals is used. This is also shown in **Paper IV**, where $\boldsymbol{Q}_{\text{anth}}$ is assumed to be a diagonal matrix and $\boldsymbol{Q}_{\text{bio}}$ contains off-diagonal elements. There, the resulting uncertainty reduction $(1-(\sigma^{\text{posterior}})^2/(\sigma^{\text{prior}})^2$, i.e. a relative variance reduction from prior to posterior) is much higher in biospheric emission estimates.

An important development in this study was the ability to differentiate scaling factors for anthropogenic and natural sources through inverse modelling. The atmospheric CH_4 observations capture joint CH_4 signals from all sources and sinks. Although trajectories can differentiate major source signals captured in each observations, it is not possible to distinguish between all signals separately in the observations. In CTE-CH_4, the scaling factors for anthropogenic and natural sources are separated by optimization regions. One approach is to optimise either anthropogenic or biospheric fluxes per optimization region (A1). Here, anthropogenic emissions are optimised if the LET is e.g. either anthropogenic, rice, water, or ocean, and biospheric fluxes are optimised elsewhere. Another approach is to optimise both sources per region (A2), employed in part of **Papers II** and **IV**. In this approach, two different maps were used for anthropogenic and biospheric emissions. Here, it is important not to use exactly the same map for both sources, in order not to introduce additional ill-posedness in the system. In **Paper II**, inversions using both approaches were compared, which showed an improvement in agreement with the observations using approach A2. However, it was also shown that the flux estimates may be influenced too strongly by the observations, resulting in unrealistic interannual variability in some regional flux estimates. Therefore, in **Paper IV**, a combination of the two approaches were used: A2 is applied over Europe, which is our focus region that has a dense observation network with high quality measurements, and A1 elsewhere globally.



(a) (b)

Figure 6: Performance of CTE-CH_4. (a) degrees of freedom in covariance matrix using different ensemble sizes. (b) flux estimates with ensemble size of 20 (E20) and 500 (E500) using different random samples.

4.1.1 Ensemble size and lag length

The ensemble size is an important factor in EnKF when it comes to stabilising the results. As studied in **Paper II**, using small ensemble size that only gives few degrees of freedom (d.o.f.) in the optimization may misrepresent the posterior distribution, and the optimised results become sensitive to the random samples. In CTE-CH_4, the d.o.f., calculated from the singular values v_i of matrix $\boldsymbol{X}^a$, $\left((\sum_{i=1}^{K} v_i)^2 / \sum_{i=1}^{K} v_i^2\right)$, successfully increased towards an equilibrium, and around 250 or larger would be a sufficient size of ensemble (Fig. 6). Note that an informative covariance including off-diagonals is used, and therefore, the d.o.f. is smaller than the state vector length (number of optimization regions). The weekly emissions also show that the estimates are more stable with larger ensemble sizes (Fig. 6). **Paper II** showed that using large enough ensemble size is especially important in regions with sparse observation networks. Based on these analysis, CTE-CH_4 uses 500 ensemble members as the default.

The lag length can be chosen based on physical parameters, such as the trajectories of atmospheric CH_4 between the observational sites. From the previous studies, e.g. Bergamaschi et al. (2005) and Bruhwiler et al. (2014), where the same transport model is used, the lag length of 5 weeks is found to be appropriate in most of the cases. However, as shown in Babenhauserheide et al. (2015), longer a lag length would be more appropriate for the Tropics, where the observation network is sparse, and southern high latitudes, where local sources are low and need longer transport time for the source signals to reach the observation locations. In **Paper II**, the effect of lag length was studied by using 5 and 12 weeks, but due to the short study period (half a year), the

effect was not clear.

4.2 Atmospheric transport model

In many atmospheric inverse models, an atmospheric chemistry transport model (ACTM) is used as an observation operator in the cost function (2.6), which links the flux estimates to atmospheric concentration observations. In CTE-CH_4, the TM5 global ACTM is employed (Krol et al., 2005). TM5 is an off-line Euclidean model, where the atmospheric GHG concentration is calculated for each grid box, conserving the atmospheric mass. The off-line mode uses meteorological constrains from pre-calculated 3-hourly ECMWF ERA-Interim meteorological fields.

TM5 has the ability to zoom over a part of the globe, and CTE-CH_4 uses a $1°\times1°$ zoom over Europe (21°W–45°E, 24°N–74°N), framed by an $3°\times2°$ inter-mediate region (36°W–54°E, 2°N–82°N), and $6°\times4°$ globally. The spatial resolution of the ACTM is independent of model states, i.e. it is mathematically possible to estimate fluxes in smaller spatial resolutions than that of the ACTM. However, the spatial resolution is important in resolving atmospheric states well, and therefore, the flux optimization resolution is often chosen based on that of ACTM. In **Paper IV**, $1°\times1°$ grid-based optimization is applied over Europe, which is the smallest optimization resolution in CTE-CH_4. For the vertical resolution, 25 layers were used, where there are 19-20 layers in the troposphere and 5-6 in the stratosphere. Although stratospheric CH_4 has less importance than that of the troposphere in resolving atmospheric CH_4 at surface observational sites, overestimation of the stratospheric CH_4 in the model was found when the vertical profiles were compared against aircraft observations, especially in the estimates using the G2000 convection scheme (**Paper II**, Fig. 4). However, increasing the vertical resolution can improve the vertical profile in the stratosphere.

The parameterisation of convection is important in resolving inter-hemispheric exchange. The two convection schemes in TM5 are examined in **Paper II** by applying Tiedtke (1989) (hereafter T1989) and Gregory et al. (2000) (hereafter G2000) convection schemes. The two versions differ mainly in vertical mixing in the troposphere, where G2000 produces a faster mixing compared to T1989. Therefore, atmospheric CH_4 at the surface using G2000 is smaller in the Northern Hemisphere, and inter hemispheric exchange times are faster than using the T1989 scheme. **Paper II** showed that the dry air total column-averaged CH_4 mole fractions (XCH_4) better matched the satellite and ground based retrievals when the G2000 scheme with the faster vertical mixing was used. The XCH_4 seasonal cycle, especially around the Tropics, showed two peaks, one of which was only captured in the inversions using the G2000 scheme.

The atmospheric sink of CH_4 due to oxidation by OH, Cl and $O(^1D)$ is taken into account in TM5. The reaction rate coefficient κ of each compound j is calculated based

on the Arrhenius equation:

$$\kappa_j = A \times \exp(-\beta_j / R\mathcal{T}), \tag{4.7}$$

where A is a pre-exponential factor, ζ_j is the Activation energy [J mol^{-1}] of the compound j, R is the universal gas constant [J K^{-1} mol^{-1}], and $\mathcal{T}$ is the absolute air temperature [K]. The reaction rate between two compounds j and j' is

$$\lambda_{j,j'} = \kappa[C_j][C_{j'}], \tag{4.8}$$

where $[C_j]$ is atmospheric concentration of a compound j. The total removal rate λ_{total} for CH_4 then becomes

$$\begin{aligned} \lambda_{\text{total}} &= \lambda_{\text{OH,CH}_4} + \lambda_{\text{Cl,CH}_4} + \lambda_{\text{O(}^1\text{D),CH}_4} \\ &= (\kappa_{\text{OH}}[C_{\text{OH}}] + \kappa_{\text{Cl}}[C_{\text{Cl}}] + \kappa_{\text{O(}^1\text{D)}}[C_{\text{O(}^1\text{D)}}])[C_{\text{CH}_4}]. \end{aligned} \tag{4.9}$$

In TM5, the monthly atmospheric OH concentration distribution, $[C_{\text{OH}}]$, is estimated from methyl chloroform (Huijnen et al., 2010; Brühl and Crutzen, 1993)), and precalculated reaction rates with Cl and O(^{1}D), are derived from a 2D photochemical Max-Planck-Institute (MPI) model (Bergamaschi et al., 2005). The interannual variability of atmospheric sink is not taken into account, or optimised by CTE-CH_4. However, as discussed in **Paper II**, changes in OH concentration could be one reason for the flattening of atmospheric CH4 growth during 1999-2006 and the subsequent growth thereafter. Therefore, the increasing rate in CTE-CH_4 emission estimates could be overestimated.

4.3 Prior fluxes

In this context, prior refer to first guess estimates of CH_4 fluxes F^p, rather than prior states (scaling factors) to be optimised. In CTE-CH_4, fluxes from five sources are taken into account: anthropogenic, biospheric (e.g. wetlands and peatlands), fire, termites and ocean (see also Section 4.1). For the annual prior anthropogenic emissions, F^p_{anth}, the estimates from EDGAR v4.2 FT2010 inventory (edgar.jrc.ec.europa.eu/overview.php?v=42FT2010) were used. The inventory accounts for emissions from energy manufacturing and transportation, residential emissions, road transportation, non-road transportation such as ships and aircraft, fugitive from solid, oil and gas production and distribution, agriculture (enteric fermentation, manure management, and agricultural soils), landfills (solid waste disposal and waste water). Among those, the emission from fugitive, oil and gas, and agriculture are the major sources, accounting for about 70% of total anthropogenic emissions. The inventory uses national statistics, mainly from UNFCC, which are distributed spatially according to e.g. population distribution and livestock distribution.

For the monthly prior natural biospheric emissions, F^p_{bio}, the flux estimates from several process-based models were used to account for sources from wetlands, peatlands and mineral soils, and the small sink to dry soils. In **Paper I**, the estimates from LPJ-GUESS-WHyMe (LPJG-WHyMe) (Smith et al., 2001; Wania et al., 2009; McGuire et al., 2012), were used, where a possible overestimation in the estimate over Eurasian boreal region was found. In **Papers II** and **III**, the estimates from LPX-Bern v1.0 (Spahni et al., 2013) were used. **Paper II** showed that the global total was well in line with inversion estimates, but suggested redistribution of the emissions by reducing estimates over the northern high latitudes, and increasing over the Tropics. In addition, **Paper III** showed a possible overestimation in the seasonal amplitude of the LPJG-WHyMe and LPX-Bern v1.0 estimates, where inversion estimates using CTE-CH_4 and other models showed smaller amplitudes. In **Paper IV**, LPX-Bern DYPTOP (Stocker et al., 2014) estimates were used in addition to LPJG-WHyMe and LPX-Bern v1.0 estimates. The comparison showed that the differences in wetland and peatland extent and parameterisation of dependencies to climate drivers result in significant differences in their regional flux estimates. Not only the magnitude, but also interannual variability and seasonal cycle differed significantly between the process-based models (**Paper IV**, Fig. 4).

For the monthly prior fluxes for large-scale fire (F^p_{fire}), including biomass burning, the estimates from Global Fire Emissions Database (GFED; Randerson et al., 2012; van der Werf et al., 2010; Giglio et al., 2013) were used. The data accounts for natural fire emissions from savanna, grassland, shrubland, boreal forest, temperate forest and peatland. In addition, anthropogenic biomass burning emissions from deforestation, degradation and agricultural waste are taken into account. The fire emissions were estimated using source-specific burned area distribution maps and modelled source-specific processes. For annual termites emissions (F^p_{termite}), estimates from Ito and Inatomi (2012) were used, where the termite emissions were estimated based on biome-specific termite biomass density and emission factors from Fraser et al. (1986), with the interannual variability introduced using a historical land cover map (Hurtt et al., 2006). Monthly ocean prior fluxes (F^p_{ocean}) were estimated based on ECMWF ERA-Interim meteorology, zonally averaged saturation ratios from Bates et al. (1996), and the zonal monthly mean dry-air CH_4 mixing ratios from GLOBALVIEW-CH_4 (www.esrl.noaa.gov/gmd/ccgg/globalview/ch4/ch4_intro.html) (**Paper I**).

For all the prior fluxes, the spatial distribution is unified to $1^\circ \times 1^\circ$ resolution to match TM5 resolution, and the last available yearly estimates were used for extensive years. For example, the EDGAR inventory has data up to 2010, and the 2010 data was used for 2011-2014 in **Paper IV**.

4.4 Atmospheric concentration observations

The global atmospheric CH_4 concentration observations made by NOAA/ESRL and other global networks and institutions were collected mainly from the World Data Centre for Greenhouse Gas (WDCGG; http://ds.data.jma.go.jp/gmd/wdcgg/). The observation set contains measurements from discrete and continuous air samples. The discrete air samples are collected approximately once a week, and the reported continuous measurements are averaged to hourly resolution. The observations were filtered according to each contributor's flags before assimilation. Observations having high measurement errors or those strongly influenced by local sources were filtered out. In addition, hourly values were further averaged to daily using values from 12:00–16:00 local time (LT). Exceptions were high altitude sites, where day time observations are often influenced by convection from e.g. a foot of a mountain. For those sites, well-mixed air is better captured during night, and therefore, daily means were calculated from 00:00–04:00 LT values. The daily averages were used because the atmospheric transport model uncertainty for daily averages are lower than for the hourly values.

The observation uncertainty was assumed independent in space and time, and the variance (diagonals of matrix $\boldsymbol{R}$) was estimated based on site location. Note that the observation uncertainty is an aggregated uncertainty of measurement and transport model errors. The transport model errors are often larger than the measurement errors, and it is estimated for each site by running the ACTM in the forward mode. The differences between model and observed values (detrended residuals) could be used to determine the transport model error, separated from emission uncertainty. The observation error for CTE-CH_4 is mainly based on the previous study by Bruhwiler et al. (2014) who used a similar inversion system and ACTM. The observational sites were categorised into 5-6 types, where observation uncertainty varied from 4.5 ppb to 75 ppb. This observation uncertainties are aggregated values of measurement errors and the transport model errors.

Several sets of observations were examined using CTE-CH_4. In **Paper I**, NOAA/ESRL global observations were assimilated, and the significance of the Pallas observations was examined by comparing estimates with and without the Pallas observations. The study showed the significance of the Pallas observations, which resulted in smaller regional mean biospheric flux estimates while assimilating the Pallas observations, and the estimated posterior atmospheric CH_4 better matched the observed values especially in summer. In **Paper II**, the effect of two Asian sites were examined by changing observation uncertainty of the sites from 15 ppb to 1000 ppb (i.e. almost no influence in the optimization). In the paper, those sites were found to be especially effective in reducing anthropogenic emissions over Asia, which led to unrealistic interannual variability of the regional anthropogenic emissions during 2000-2012. It was also shown that the influence of those observations was stronger when both anthropogenic and biospheric

emissions were optimised per region (i.e. A2 approach; see also Section 4.1). Those results highlighted the importance of pre-filtering before assimilation, and assigning appropriate observation error. In **Paper III**, four different sets of quality-controlled and harmonised observations from the Integrated non-CO_2 Greenhouse gas Observing System (InGOS) project were used to evaluate the range in the flux estimates over Europe. The results showed that the range of the flux estimates due to differences in the assimilated observation sets is small when the atmospheric mixing is well resolved. In **Paper IV**, similar sets of global observations as in **Paper II** were used, but the number of sites for Finland and surrounding regions was extended significantly. In addition, the effect of two urban sites was examined by comparing estimates with and without those sites. The results showed a significant reduction in anthropogenic emission estimates over southern Finland by assimilating the urban observations.

4.5 Uncertainty estimates

One of the biggest issues in atmospheric inversion is the estimation of uncertainties. Although there are some ways to estimate such uncertainties, the prior uncertainties would not be estimated correctly and fully. Therefore, the model users must be aware of the limitations related to the construction and estimation of those uncertainties.

4.5.1 Prior flux uncertainty

In CTE-CH_4, the prior uncertainty (diagonals of prior covariance $\boldsymbol{P}^b$) is defined in respect to the percentage of the prior fluxes. One problem in using such uncertainty is that the prior uncertainty depends on magnitude of the prior flux estimate. Although prior uncertainty should be large enough to allow scaling factors to change, the prior uncertainty tends to be underestimated when prior fluxes are small. This problem was addressed in **Paper IV**, where the prior uncertainty for the biospheric fluxes over southern Finland, defined as 80% of the prior fluxes, was possibly underestimated. Due to the too small uncertainty, the inversion could not increase the biospheric fluxes, and anthropogenic fluxes were increased instead to compensate the budget. On the other hand, when prior uncertainty was large enough, the inversion estimates became close to each other regardless of the prior fluxes.

4.5.2 Observation operator uncertainty

The uncertainty of the observation operator $\mathcal{H}$ is an ATM uncertainty in resolving atmospheric CH_4. This consists of errors from model parameterisation, meteorology and resolution errors. Locatelli et al. (2013) showed that the standard deviation in

the estimated concentrations over a year could differ by about 100 ppb (≈6%), and the spread of the inferred fluxes could reach up to 150% of the prior fluxes. Locatelli et al. (2013) also pointed out that the driving meteorology is one of key factors, which results in similar estimates when similar meteorological fields are used. In addition, a finer model resolution in all dimensions tends to result in better estimates. The model parameterisation for vertical convection is discussed in **Paper II**, where two TM5 convection schemes were compared. The estimated concentrations showed a better agreement with the observations when the faster vertical mixing scheme was applied, where RMSE with the observations improved by more than 10 ppb in the prior (**Paper II**, Fig. 2). In addition, the observation operator errors are likely to be spatially and temporally correlated. However, such spatial correlation is difficult to assess, as resolved atmospheric CH_4 also depends on flux distribution.

5 Application of CTE-CH_4

5.1 Global CH_4 budget

The estimation and evaluation of the global CH_4 budget using CTE-CH_4 are discussed mainly in **Paper II**, and briefly in **Paper IV**. In those studies, weekly CH_4 fluxes were estimated with 500 ensemble members, drawn from a normal distribution, with a smoothing lag of 5 weeks. The annual mean global total CH_4 emission for 2000-2006 was estimated to be 507–509 Tg CH_4 yr^{-1}, with an uncertainty range of 45–62 Tg CH_4 yr^{-1}, which increased to 526-528±44-61 Tg CH_4 yr^{-1} for 2007-2012 (**Paper II**). A similar increase was also found in **Paper IV**, where global total emission increased from 510±73 to 540±84 Tg CH_4 yr^{-1} during 2004-2014. Although the model parameterisation, prior fluxes and observation sets assimilated differ in those studies, the global total is very well preserved, and in line with previous studies, e.g. Bousquet et al. (2006) and Fraser et al. (2013), indicating the adequacy of the CTE-CH_4 system for use in inversion studies. **Paper II** showed that the increase in the anthropogenic emission was the possible cause, which increased 15–27 Tg CH_4 yr^{-1} between 2000–2006 and 2007–2012. However, the trend in biospheric emissions was not clear as the estimates were sensitive to model parameterisation and observation sets to be assimilated.

The uncertainty was successfully reduced from the prior to posterior in both studies, as expected. The uncertainty reduction for global total emission estimates was about 50% (**Paper II**). The uncertainty reduction is often larger when spatial correlation is included in the prior uncertainty, the number of observations assimilated is large and the observation network is dense. However, it is hard to conclude whether the uncertainty ranges are realistic, and the estimates with smaller uncertainties are not necessary more realistic. The uncertainty estimates may be over- or underestimated since they also depend on the uncertainty estimates on the prior and observation covariance matrices, which are somewhat arbitrary. The range of estimates from an ensemble of models may be more appropriate. Estimates of global total budget from the different inverse models agree with each other within 10%, regardless of the inversion models (Saunois et al., 2016a).

5.2 Regional CH_4 budget

It is important to understand the regional CH_4 balance because the distribution of anthropogenic and natural emissions differ in each region due to economy, politics, climate, soil properties and vegetation properties, to mention a few. Those processes are often strongly connected, and they all influence the atmospheric CH_4 burden. Therefore, global inverse models have the advantage that regional budgets can be estimated within a well-constrained global budget.

Regional CH_4 balances are discussed globally in detail in **Paper II**, with a focus on Europe in **Papers I** and **III**, and for Finland in **Paper IV**. In **Paper II**, the regional budgets were found to be sensitive to the number of parameters optimised, atmospheric transport and observations. Using the faster vertical mixing in TM5, the NH emission estimates lowered by about 10 Tg CH_4 yr^{-1} and the estimates for the SH increased by a similar amount compared to the slower vertical mixing scheme. The model parameterisation, whether to optimise both or either of anthropogenic and biospheric fluxes per optimization region (approach A1 and S2), had an influence on the regional source division between anthropogenic and biospheric emissions, and their trends. For example, the 2000–2006 biospheric emissions for the Asian temperate region were significantly lower when the A2 approach, with larger number of parameters, was used compared to the estimates using the A1 approach (see Section 4.1 and Table 6 in **Paper II**). In addition, the biospheric emission increase in the Asian temperate region was much stronger using the A2 approach. This was due to too strong influence from two Asian sites (discussed also in Section 4.4).

Europe was one of the main interests in this study. The estimated total emissions over Europe (including parts of western Russia) ranged between 47–60±10–13 Tg CH_4 yr^{-1} (**Papers I** and **II**), and 20–30 Tg CH_4 yr^{-1} for EU-28 countries (10th and 90th percentiles from seven inversions; **Paper III**). In **Papers II** and **III**, the estimates showed no significant trend in total emission for Europe, suggesting that Europe is possibly not contributing to the increase in the atmospheric CH_4 concentrations. In addition, both papers addressed the uncertainty in the biospheric emission estimates, especially for northern Europe. The studies suggested a possible overestimation in the prior fluxes from the LPJG-WHyMe and LPX-Bern v1.0 process-based models, in magnitude (**Papers I** and **II**) and seasonal amplitude (**Paper III**). This was further examined in depth in **Paper IV**, which showed the inversion estimates to be in between those two process-based models and the LPX-Bern DYPTOP, supporting the findings from the previous studies.

5.3 Model evaluation

The atmospheric inverse model was developed to solve a real-life problem without a known solution. Therefore, it is not possible to validate our results, but they can only be evaluated through comparison with sets of observations and other models.

5.3.1 In-situ atmospheric observations

In the evaluation with in-situ atmospheric observations, there are two sets of observations we could use. One is of assimilated (model dependent) observations, which is used

to derive the posterior fluxes, and another is of non-assimilated (model independent) observations, which is not used in the inversion. The model dependent observations are useful for assessing general model performance, such as the quality of inversion and transport. The posterior atmospheric CH_4 often does not match the observations perfectly due to the limited number of parameters and a complicated non-linear system in the observation operator, and the d.o.f. in the data assimilation system is not large enough to match individual observations perfectly. The model independent in-situ observations are useful in assessing a potential model bias for locations and times that are not directly constrained by the closest observations.

In general, CTE-CH_4 is able to reproduce observations at assimilated sites well, but the bias in the latitudinal gradient and the seasonal cycle is still seen (Fig. 2 in **Paper II**). This could be due to transport model error, but also indicates that the inversion did not resolve the spatial and temporal discrepancies well enough. In addition, the model has difficulties reproducing CH_4 with high local influence, mainly due to spatial resolution error in the transport model. In **Paper I**, the model independent continuous observations from Pallas were used for evaluation. However, the discrete observations from the same site were assimilated, i.e. the observations were not considered to be fully independent. Nevertheless, the continuous observations showed that the atmospheric CH_4 values in summer were often lower than those observed without assimilating the Pallas discrete observations, indicating a possible underestimation in the resulting biospheric fluxes in northern Europe. In **Paper III**, both model dependent and independent in-situ observations were used for evaluation, where posterior atmospheric CH_4 showed a moderate to high correlation (0.5 to 0.79) with the observations. In **Paper IV**, model dependent in-situ observations were used for evaluation, not only to assess the model performance, but also the adequacy of the observations to be used in the inversion. The evaluation showed relatively high correlation (>0.75) at all Fennoscandian sites. However, since some bias remained at all sites, it was suspected that a shorter spatial correlation length in the prior covariance would be more appropriate.

5.3.2 Aircraft measurements

Altitude information provided by aircraft profiles is useful in understanding atmospheric CH_4 in general, but also specifically those for vertical transport. Regular profiling aircraft data operated within the European CarboEurope project at Orléans (France), Bialystok (Poland), Hegyhatsal (Hungary) and Griffin (UK) during 2006–2012, and an aircraft campaign performed within the Infrastructure for Measurement of the European Carbon Cycle (IMECC) project (Geibel et al., 2012) were used for model independent evaluation in **Papers II** and **III**.

The estimated CH_4 profile showed an improvement from the prior, and the agreement

with the observations were good at locations where surface CH_4 is well resolved, at Griffin, for example (**Paper II**). The comparison with background CH_4 derived from TM5 showed that the low vertical gradient at Griffin indicates the regional emission signals from Europe have limited influence on atmospheric CH_4 at Griffin (**Paper III**). This suggests that the emissions and transport are possibly well resolved to match background atmospheric CH_4 and emission magnitude. However, the bias in the vertical profile was larger at the sites where influence from the local sources is high. At those sites, the agreement with surface observations was often also poor, suggesting the emission estimates are not well constrained (**Paper II**, **III**). In addition, the IMECC profile showed that the estimated CH_4 burden in general matched the observations well, but a negative bias was found in the upper troposphere (**Paper II**). This was better captured when the faster vertical mixing scheme was used, but the fast vertical mixing scheme introduced a greater positive bias in the stratosphere (**Paper II**, Fig. 4).

5.3.3 Satellite and ground based retrievals

Satellite-based remote sensing retrievals have a much larger spatial coverage than in-situ or aircraft observations, providing useful information on the spatial distributions of atmospheric CH_4 on global scale. The satellite retrievals are not a direct measure of CH_4 concentrations, but those inferred from incoming light spectra of atmospheric columns. Therefore, the information about the CH_4 burden for a specific altitude derived from satellite retrievals has high uncertainty, and the dry air total column-averaged CH_4 mole fraction (XCH_4) is more adequate for the evaluation. In this thesis, the XCH_4 retrievals from the Total Carbon Column Observing Network (TCCON) (**Papers I** and **II**), and the TANSO-FTS instrument on board the Greenhouse gases Observing Satellite (GOSAT) spacecraft (Kuze et al., 2009) were used (**Paper II**). The TCCON consists of ground-based stationally sites, measuring the same quantity as the satellite-based instruments, and their retrievals are used to validate the satellite-based retrievals. Although GOSAT provides larger spatial information than any in-situ measurements, it is argued that the retrieved XCH_4 has some bias on latitudinal gradient (Yoshida et al., 2013), and therefore, it is important to use both sets in model evaluation to avoid possible misinterpretation.

The posterior XCH_4 (derived from TM5 with posterior fluxes) showed a small negative bias, but good agreements on trend and seasonal cycle with TCCON XCH_4 in the NH (**Paper II**). However, the posterior XCH_4 has a significant negative bias in the SH compared to both TCCON and GOSAT XCH_4 (**Paper II**). The possible underestimation of tropical emissions is discussed, this was about 10–20 Tg CH_4 yr^{-1} lower than inversion estimates by e.g. Houweling et al., 2014. The XCH_4 estimates were not significantly sensitive to small differences in surface flux estimates (**Papers I** and **II**), but strongly affected by the convection schemes in TM5. The negative bias in the SH

and the agreement with retrievals in the tropical seasonal cycle significantly improved when the faster vertical mixing scheme was used (**Paper II**), which indicated a possible underestimation of emissions for the Tropics and northern-latitude temperate regions derived with the T1989 scheme.

5.3.4 In-situ flux observations

The flux estimates can be evaluated with in-situ flux observations, such as those from chamber and eddy covariance (EC) measurements. The comparison with site-level measurements can be challenging, as model resolution is often much larger than the measurements spatial representativity. This is because CH_4 fluxes are highly heterogeneous in space (Frolking and Crill, 1994; Moore et al., 1994), which means they can differ significantly within a few metres (Moore et al., 1998), and therefore, the site-level flux measurements may not be representative of averaged or aggregated modelled fluxes over a model grid cell. Nevertheless, the in-situ fluxes measured at several locations, and EC measurements which could be representative of a few hundred kilometres, have been extensively studied over the northern latitude Boreal regions, and some general information can be used for evaluation.

In **Paper II**, the 2006–2007 interannual variability of the biospheric flux estimates for north American and European boreal regions was compared to previously published in-situ flux observations from Moore et al. (2011); Drewer et al. (2010) and Jackowicz-Korczyński et al. (2010). The analysis showed that the interannual variability in the posterior fluxes agreed with the measurements better than the prior, supporting adequacy of inversion results that are driven by the atmospheric concentration observations. In **Paper IV**, the regional biospheric fluxes for northern and southern Finland were compared to EC observations from two peatlands in Finland: Lompolojänkkä and Siikaneva. The interannual variability in the Lompolojänkkä EC measurements showed that the flux estimates in the LPJG-WHyMe and LPX-Bern v1.0 are possibly overestimated, and the fluxes estimated by the LPX-Bern DYPTOP and CTE-CH_4 are more reasonable. In addition, the Siikaneva EC measurements suggested that the seasonal cycle in the posterior anthropogenic fluxes is possibly unrealistic, and the increase in the summer fluxes is probably of biospheric origin.

5.4 Limitation of CTE-CH_4

5.4.1 Separation of emission sources

One limitation of CTE-CH_4 is that the separation of emission sources are still uncertain in regions where both anthropogenic and biospheric emissions are equally large, or the

emission ratio is incorrectly described in the prior fluxes. Since CTE-CH_4 is based on mathematical optimization, which do not optimise fluxes based on physical parameters, the weight of the sources can only be decided by the pre-defined background error covariance matrix $\boldsymbol{Q}$. When prior uncertainty is equally large, the source division is driven simply to find the mathematically optimal cost function. **Paper II** showed that it was difficult to revert the emission ratio in the prior with the inversion. The anthropogenic emission estimates in Africa were larger than the biospheric emission estimates, and increased during 2000-2012 in **Paper II**. On the other hand, Bruhwiler et al. (2014) showed an increase in biospheric emission estimates that were larger than the anthropogenic emission estimates in their study. Although both studies used similar optimization methods, the differences in the prior fluxes led to opposite source division in the posterior and the differences in the trends. **Paper IV** addressed a possibly unrealistic increase in posterior anthropogenic emissions during summer in southern Finland, although the signal in the observed atmospheric CH_4 is possibly of biospheric origin. This was mainly due to underestimation in the prior uncertainty for biospheric fluxes, which were smaller than those of the anthropogenic emissions.

In addition, the system could create physically incorrect fluxes, such as negative anthropogenic emissions. The prior state is assumed to be normally distributed, and the prior ensemble is drawn from the normal distribution with a mean of 1 and variance e.g. of 0.8. This could give negative prior state samples with probability of $\sim$0.13, which is insignificant. Other distributions, such as a truncated normal distribution or transfromation to the log-normal distribution (Kemp et al., 2014), could be used to make prior state samples strictly positive. However, due to normality assumption in states and observations, it is not possible to avoid having negative scaling factors in the current setup.

5.4.2 Dynamical model for prediction

Another limitation is the lack of a dynamical model $\mathcal{M}$ in predicting the state covariance matrix (2.8). Although flux estimates from a week to another are related through physical and chemical processes, which are included through prior flux estimates and the transport model, their scaling factors (except for the mean states) are assumed to be uncorrelated in time, and therefore, the posterior flux estimates still have large fluctuations between weeks. Increasing the ensemble size helps (Fig. 6), but some of the instantaneous changes in fluxes could be unrealistic and possibly resulted from overfitting the observations. Therefore, only annual and monthly estimates have been reported in the papers presented here, although fluxes were optimised weekly.

6 Summary and concluding remarks

This thesis presented the application of an ensemble Kalman filter data assimilation system, CTE-CH_4, to a real-life problem of global and regional CH_4 flux estimations. The analysis showed that consistent and reliable estimates of global and regional CH_4 budgets were obtained with a large enough ensemble size, appropriate prior error covariance, and good observations, which agreed well with a variety of model dependent and independent observations. CTE-CH_4 identified the significant contribution from the tropical and extra tropical anthropogenic emissions to the increase in the global CH_4 budget after 2007, which could possibly be a cause of the increase in atmospheric CH_4 GR after 2007. In addition, CTE-CH_4 was able to give further understanding to the regional CH_4 budgets; Europe as a whole and Finland were found to have an insignificant or negative influence on the atmospheric CH_4 growth in the 21st century.

However, the evaluation also addressed several problems in the system. The somewhat subjective prior covariance matrix could limit the inversion to change fluxes from the prior, and the inversion results still depended on the choice of prior fluxes and transport model. In addition, two major limitations of the CTE-CH_4 model were addressed: (1) uncertainty in anthropogenic and biospheric source division, and (2) incomplete temporal evolution due to the lack of a dynamical model for covariance matrix.

A better source separation could be achieved by using isotopic information. The isotopic signal of $^{13}\delta$C ($^{13}C/^{12}C$ ratio in CH_4) differs by source processes, and could be used to separate biogenic, fossil-origin and geological sources. Global measurements of the isotopic signals for about two decades are already available for the inversion through NOAA and the University of Colorado. Such development will give further understanding of the recent atmospheric CH_4 growth. Furthermore, the next generations of satellites such as GOSAT-2 would be able to provide distribution of $^{13}\delta$C signal on global scale, which could be used for evaluation or assimilation.

Improving the dynamical model to predict covariance matrices, and thus passing information on to sample deviations is challenging. A similar averaging as used for mean estimates could be applied, but it would require deflation/inflation as variance in the ensemble tends to shrink too quickly, and a way to regain sufficient d.o.f. in the states could be required with some additional tuning. However, smoother estimates for the deviations could be obtained by using e.g. temporally correlated random samples (the hidden Markov model) or using the adaptive Kalman filter. When a better temporal correlation is introduced, it would be possible to provide the estimates with shorter temporal scale, such as weekly. This would then give further opportunity to study short-temporal scale changes in CH_4 fluxes, such as during spring thaw and autumn freezing.

In addition, the estimates driven by CTE-CH_4 could be applied to regional inverse

models and evaluation of inventories and process-based models, as partly shown in this thesis. The regional inverse models, based on the regional transport model, often has advantage that the fluxes can be resolved in fine resolution on a specific domain. However, these often need "background" atmospheric concentration fields, from which backward trajectories outside of the domain are calculated. Since CTE-CH_4 is a global model, constrained by global atmospheric observations, such information can be provided for the regional inverse models to obtain flux estimates consistent with CTE-CH_4, but optimised in finer resolution. For the evaluation of inventories and process-based models, a similar study can be done for other inventories, models, and regions than those presented here, which could possibly lead to better estimates globally.

References

Babenhauserheide, A., Basu, S., Houweling, S., Peters, W., and Butz, A.: Comparing the CarbonTracker and TM5-4DVar data assimilation systems for CO2 surface flux inversions, Atmos. Chem. Phys., 15, 9747–9763, doi:10.5194/acp-15-9747-2015, 2015.

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, 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, J. Geophys. Res., 114, D22 301, doi:10.1029/2009JD012287, 2009.

Bergamaschi, P., Corazza, M., Karstens, U., Athanassiadou, M., Thompson, R. L., Pison, I., Manning, A. J., Bousquet, P., Segers, A., Vermeulen, A. T., Janssens-Maenhout, G., Schmidt, M., Ramonet, M., Meinhardt, F., Aalto, T., Haszpra, L., Moncrieff, J., Popa, M. E., Lowry, D., Steinbacher, M., Jordan, A., O'Doherty, S., Piacentino, S., and Dlugokencky, E.: Top-down estimates of European CH4 and N2O emissions based on four different inverse models, Atmos. Chem. Phys., 15, 715–736, doi:10.5194/acp-15-715-2015, 2015.

Biastoch, A., Treude, T., Rüpke, L. H., Riebesell, U., Roth, C., Burwicz, E. B., Park, W., Latif, M., Böning, C. W., Madec, G., and Wallmann, K.: Rising Arctic Ocean temperatures cause gas hydrate destabilization and ocean acidification, Geophys. Res. Lett., 38, L08 602, doi:10.1029/2011GL047222, 2011.

Bohn, T. J., Melton, J. R., Ito, A., Kleinen, T., Spahni, R., Stocker, B. D., Zhang, B., Zhu, X., Schroeder, R., Glagolev, M. V., Maksyutov, S., Brovkin, V., Chen, G., Denisov, S. N., Eliseev, A. V., Gallego-Sala, A., McDonald, K. C., Rawlins, M., Riley, W. J., Subin, Z. M., Tian, H., Zhuang, Q., and Kaplan, J. O.: WETCHIMP-WSL: intercomparison of wetland methane emissions models over West Siberia, Biogeosciences, 12, 3321–3349, doi:10.5194/bg-12-3321-2015, 2015.

Bousquet, P., Ciais, P., Miller, J. B., Dlugokencky, E. J., Hauglustaine, D. A., Prigent, C., Van der Werf, G. R., 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, 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., Krummel, P. B., Langenfelds, R. L., Ramonet, M., Schmidt, M., Steele, L. P., Szopa, S., Yver, C., Viovy, N., and Ciais, P.: Source attribution of the changes in atmospheric methane for 2006–2008, Atmos. Chem. Phys., 11, 3689–3700, doi:10.5194/acp-11-3689-2011, 2011.

Bruhwiler, L., Dlugokencky, E., Masarie, K., Ishizawa, M., Andrews, A., Miller, J., Sweeney, C., Tans, P., and Worthy, D.: CarbonTracker-CH4: an assimilation system for estimating emissions of atmospheric methane, Atmos. Chem. Phys., 14, 8269–8293, doi:10.5194/acp-14-8269-2014, 2014.

Bruhwiler, L. M. P., Michalak, A. M., Peters, W., Baker, D. F., and Tans, P.: An improved Kalman Smoother for atmospheric inversions, Atmos. Chem. Phys., 5, 2691–2702, doi:10.5194/acp-5-2691-2005, 2005.

Brühl, C. and Crutzen, P. J.: MPIC Two-dimensional model, NASA Ref. Publ., 1292, 103–104, 1993.

Buehner, M., McTaggart-Cowan, R., and Heilliette, S.: An Ensemble Kalman Filter for Numerical Weather Prediction Based on Variational Data Assimilation: VarEnKF, Mon. Wea. Rev., 145, 617–635, doi:10.1175/MWR-D-16-0106.1, 2016.

Burgers, G., Jan van Leeuwen, P., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Monthly weather review, 126, 1719–1724, 1998.

Bândă, N., Krol, M., van Weele, M., van Noije, T., and Röckmann, T.: Analysis of global methane changes after the 1991 Pinatubo volcanic eruption, Atmos. Chem. Phys., 13, 2267–2281, doi:10.5194/acp-13-2267-2013, 2013.

Chen, Y.-H. and Prinn, R. G.: Estimation of atmospheric methane emissions between 1996 and 2001 using a three-dimensional global chemical transport model, J. Geophys. Res., 111, D10 307, doi:10.1029/2005JD006058, 2006.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R. B., Piao, S., and Thornton, P.: Carbon and Other Biogeochemical Cycles, 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. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Tech. rep., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, URL https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_Chapter06_FINAL.pdf, 2013.

Conrad, R.: Soil microorganisms as controllers of atmospheric trace gases (H2, CO, CH4, OCS, N2O, and NO)., Microbiological reviews, 60, 609–640, 1996.

Dalsøren, S. B., Myhre, C. L., Myhre, G., Gomez-Pelaez, A. J., Søvde, O. A., Isaksen, I. S. A., Weiss, R. F., and Harth, C. M.: Atmospheric methane evolution the last 40 years, Atmos. Chem. Phys., 16, 3099–3126, doi:10.5194/acp-16-3099-2016, 2016.

Dlugokencky, E. J., Nisbet, E. G., Fisher, R., and Lowry, D.: Global atmospheric methane: budget, changes and dangers, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 369, 2058–2072, doi:10.1098/rsta.2010.0341, 2011.

Drewer, J., Lohila, A., Aurela, M., Laurila, T., Minkkinen, K., Penttilä, T., Dinsmore, K. J., McKenzie, R. M., Helfter, C., Flechard, C., Sutton, M. A., and Skiba, U. M.: Comparison of greenhouse gas fluxes and nitrogen budgets from an ombotrophic bog in Scotland and a minerotrophic sedge fen in Finland, European Journal of Soil Science, 61, 640–650, doi:10.1111/j.1365-2389.2010.01267.x, 2010.

Dutaur, L. and Verchot, L. V.: A global inventory of the soil CH4 sink, Global Biogeochem. Cycles, 21, GB4013, doi:10.1029/2006GB002734, 2007.

Echevin, V., De Mey, P., and Evensen, G.: Horizontal and Vertical Structure of the Representer Functions for Sea Surface Measurements in a Coastal Circulation Model, J. Phys. Oceanogr., 30, 2627–2635, doi:10.1175/1520-0485(2000)030⟨2627:HAVSOT⟩2.0.CO;2, 2000.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10 143–10 162, doi:10.1029/94JC00572, 1994.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, doi:10.1007/s10236-003-0036-9, 2003.

Evensen, G. and van Leeuwen, P. J.: An Ensemble Kalman Smoother for Nonlinear Dynamics, Mon. Wea. Rev., 128, 1852–1867, doi:10.1175/1520-0493(2000)128⟨1852:AEKSFN⟩2.0.CO;2, 2000.

Fraser, A., Palmer, P. I., Feng, L., Boesch, H., Cogan, A., Parker, R., Dlugokencky, E. J., Fraser, P. J., Krummel, P. B., Langenfelds, R. L., O'Doherty, S., Prinn, R. G., Steele, L. P., van der Schoot, M., and Weiss, R. F.: Estimating regional methane surface fluxes: the relative importance of surface and GOSAT mole fraction measurements, Atmos. Chem. Phys., 13, 5697–5713, doi:10.5194/acp-13-5697-2013, 2013.

Fraser, P. J., Rasmussen, R. A., Creffield, J. W., French, J. R., and Khalil, M. a. K.: Termites and global methane—another assessment, J Atmos Chem, 4, 295–310, doi:10.1007/BF00053806, 1986.

Frolking, S. and Crill, P.: Climate controls on temporal variability of methane flux from a poor fen in southeastern New Hampshire: Measurement and modeling, Global Biogeochem. Cycles, 8, 385–397, doi:10.1029/94GB01839, 1994.

Geibel, M. C., Messerschmidt, J., Gerbig, C., Blumenstock, T., Chen, H., Hase, F., Kolle, O., Lavrič, J. V., Notholt, J., Palm, M., Rettinger, M., Schmidt, M., Sussmann, R., Warneke, T., and Feist, D. G.: Calibration of column-averaged CH4 over European TCCON FTS sites with airborne in-situ measurements, Atmos. Chem. Phys., 12, 8763–8775, doi:10.5194/acp-12-8763-2012, 2012.

Ghosh, A., Patra, P. K., Ishijima, K., Umezawa, T., Ito, A., Etheridge, D. M., Sugawara, S., Kawamura, K., Miller, J. B., Dlugokencky, E. J., Krummel, P. B., Fraser, P. J., Steele, L. P., Langenfelds, R. L., Trudinger, C. M., White, J. W. C., Vaughn, B., Saeki, T., Aoki, S., and Nakazawa, T.: Variations in global methane sources and sinks during 1910–2010, Atmos. Chem. Phys., 15, 2595–2612, doi:10.5194/acp-15-2595-2015, 2015.

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), J. Geophys. Res. Biogeosci., 118, 317–328, doi:10.1002/jgrg.20042, 2013.

Gregory, D., Morcrette, J.-J., Jakob, C., Beljaars, A. C. M., and Stockdale, T.: Revision of convection, radiation and cloud schemes in the ECMWF integrated forecasting system, Q.J.R. Meteorol. Soc., 126, 1685–1710, doi:10.1002/qj.49712656607, 2000.

Guckland, A., Flessa, H., and Prenzel, J.: Controls of temporal and spatial variability of methane uptake in soils of a temperate deciduous forest with different abundance of European beech (Fagus sylvatica L.), Soil Biology and Biochemistry, 41, 1659–1667, doi:10.1016/j.soilbio.2009.05.006, 2009.

Heimann, M.: Atmospheric science: Enigma of the recent methane budget, Nature, 476, 157–158, doi:10.1038/476157a, 2011.

Heimann, M. and Reichstein, M.: Terrestrial ecosystem carbon dynamics and climate feedbacks, Nature, 451, 289–292, doi:10.1038/nature06591, 2008.

Hofmann, D. J., Butler, J. H., Dlugokencky, E. J., Elkins, J. W., Masarie, K., Montzka, S. A., and Tans, P.: The role of carbon dioxide in climate forcing from 1979 to 2004: introduction of the Annual Greenhouse Gas Index, Tellus B: Chemical and Physical Meteorology, 58, 614–619, doi:10.1111/j.1600-0889.2006.00201.x, 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, 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, 1999.

Houweling, S., Krol, M., Bergamaschi, P., Frankenberg, C., Dlugokencky, E. J., Morino, I., Notholt, J., Sherlock, V., Wunch, D., Beck, V., Gerbig, C., Chen, H., Kort, E. A., Röckmann, T., and Aben, I.:

A multi-year methane inversion using SCIAMACHY, accounting for systematic errors using TCCON measurements, Atmos. Chem. Phys., 14, 3991–4012, doi:10.5194/acp-14-3991-2014, 2014.

Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., de Laat, J., Boersma, F., Bergamaschi, P., van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., and Pätz, H.-W.: The global chemistry transport model TM5: description and evaluation of the tropospheric chemistry version 3.0, Geosci. Model Dev., 3, 445–473, doi:10.5194/gmd-3-445-2010, 2010.

Hurtt, G. C., Frolking, S., Fearon, M. G., Moore, B., Shevliakova, E., Malyshev, S., Pacala, S. W., and Houghton, R. A.: The underpinnings of land-use history: three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands, Global Change Biology, 12, 1208–1229, doi:10.1111/j.1365-2486.2006.01150.x, 2006.

Ito, A. and Inatomi, M.: Use of a process-based model for assessing the methane budgets of global terrestrial ecosystems and evaluation of uncertainty, Biogeosciences, 9, 759–773, doi:10.5194/bg-9-759-2012, 2012.

Jackowicz-Korczyński, M., Christensen, T. R., Bäckstrand, K., Crill, P., Friborg, T., Mastepanov, M., and Ström, L.: Annual cycle of methane emission from a subarctic peatland, J. Geophys. Res., 115, G02 009, doi:10.1029/2008JG000913, 2010.

Jamali, H., Livesley, S. J., Grover, S. P., Dawes, T. Z., Hutley, L. B., Cook, G. D., and Arndt, S. K.: The Importance of Termites to the CH4 Balance of a Tropical Savanna Woodland of Northern Australia, Ecosystems, 14, 698–709, doi:10.1007/s10021-011-9439-5, 2011.

Jamali, H., Livesley, S. J., Hutley, L. B., Fest, B., and Arndt, S. K.: The relationships between termite mound CH4/CO2 emissions and internal concentration ratios are species specific, Biogeosciences, 10, 2229–2240, doi:10.5194/bg-10-2229-2013, 2013.

Johansson, T., Malmer, N., Crill, P. M., Friborg, T., Åkerman, J. H., Mastepanov, M., and Christensen, T. R.: Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing, Global Change Biology, 12, 2352–2369, doi:10.1111/j.1365-2486.2006.01267.x, 2006.

Kalman, R. E.: A new approach to linear filtering and prediction problems, Journal of basic Engineering, 82, 35–45, 1960.

Kemp, S., Scholze, M., Ziehn, T., and Kaminski, T.: Limiting the parameter space in the Carbon Cycle Data Assimilation System (CCDAS), Geosci. Model Dev., 7, 1609–1619, doi:10.5194/gmd-7-1609-2014, 2014.

Keppenne, C. L. and Rienecker, M. M.: Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter, Journal of Marine Systems, 40, 363–380, doi:10.1016/S0924-7963(03)00025-3, 2003.

Khalil, M. a. K., Rasmussen, R. A., French, J. R. J., and Holt, J. A.: The influence of termites on atmospheric trace gases: CH4, CO2, CHCl3, N2O, CO, H2, and light hydrocarbons, J. Geophys. Res., 95, 3619–3634, doi:10.1029/JD095iD04p03619, 1990.

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 Quéré, 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, 2013.

Kort, E. A., Wofsy, S. C., Daube, B. C., Diao, M., Elkins, J. W., Gao, R. S., Hintsa, E. J., Hurst, D. F., Jimenez, R., Moore, F. L., Spackman, J. R., and Zondlo, M. A.: Atmospheric observations of Arctic Ocean methane emissions up to 82° north, Nature Geosci, 5, 318–321, doi:10.1038/ngeo1452, 2012.

Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, Atmos. Chem. Phys., 5, 417–432, doi:10.5194/acp-5-417-2005, 2005.

Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Applied Optics, 48, 6716, doi:10.1364/AO.48.006716, 2009.

Kvenvolden, K. A.: Methane hydrate—a major reservoir of carbon in the shallow geosphere?, Chemical geology, 71, 41–51, 1988.

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, Atmos. Chem. Phys., 13, 9917–9937, doi:10.5194/acp-13-9917-2013, 2013.

Lohila, A., Aalto, T., Aurela, M., Hatakka, J., Tuovinen, J.-P., Kilkki, J., Penttilä, T., Vuorenmaa, J., Hänninen, P., Sutinen, R., Viisanen, Y., and Laurila, T.: Large contribution of boreal upland forest soils to a catchment-scale CH4 balance in a wet year, Geophys. Res. Lett., 43, 2016GL067 718, doi:10.1002/2016GL067718, 2016.

Lorenc, A. C.: The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var, Q.J.R. Meteorol. Soc., 129, 3183–3203, doi:10.1256/qj.02.132, 2003.

McGuire, A. D., Christensen, T. R., Hayes, D., Heroult, A., Euskirchen, E., Kimball, J. S., Koven, C., Lafleur, P., Miller, P. A., Oechel, W., Peylin, P., Williams, M., and Yi, Y.: An assessment of the carbon balance of Arctic tundra: comparisons among observations, process models, and atmospheric inversions, Biogeosciences, 9, 3185–3204, doi:10.5194/bg-9-3185-2012, 2012.

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, 2011.

Moore, T. R., Heyes, A., and Roulet, N. T.: Methane emissions from wetlands, southern Hudson Bay lowland, J. Geophys. Res., 99, 1455–1467, doi:10.1029/93JD02457, 1994.

Moore, T. R., Roulet, N. T., and Waddington, J. M.: Uncertainty in Predicting the Effect of Climatic Change on the Carbon Cycling of Canadian Peatlands, Climatic Change, 40, 229–245, doi:10.1023/A:1005408719297, 1998.

Moore, T. R., Young, A., Bubier, J. L., Humphreys, E. R., Lafleur, P. M., and Roulet, N. T.: A Multi-Year Record of Methane Flux at the Mer Bleue Bog, Southern Canada, Ecosystems, 14, 646–657, doi:10.1007/s10021-011-9435-9, 2011.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and natural radiative forcing. 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. Ed.by Stocker, T.F.,

Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Tech. rep., Cambridge University Press., Cambridge, URL http://www.climatechange2013.org/images/report/WG1AR5_Chapter08_FINAL.pdf, 2013.

Park, J.-H. and Kaneko, A.: Assimilation of coastal acoustic tomography data into a barotropic ocean model, Geophys. Res. Lett., 27, 3373–3376, doi:10.1029/2000GL011600, 2000.

Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO2 surface fluxes from atmospheric trace gas observations, J. Geophys. Res., 110, D24 304, doi:10.1029/2005JD006157, 2005.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M., Pétron, G., Hirsch, A. I., and others: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proceedings of the National Academy of Sciences, 104, 18 925–18 930, 2007.

Peters, W., Krol, M. C., Van Der WERF, G. R., Houweling, S., Jones, C. D., Hughes, J., Schaefer, K., Masarie, K. A., Jacobson, A. R., Miller, J. B., Cho, C. H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., Di Sarra, A. G., Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H. a. J., Van Der Laan, S., Neubert, R. E. M., Jordan, A., Rodó, X., Morguí, J.-A., Vermeulen, A. T., Popa, E., Rozanski, K., Zimnoch, M., Manning, A. C., Leuenberger, M., Uglietti, C., Dolman, A. J., Ciais, P., Heimann, M., and Tans, P. P.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, Global Change Biology, 16, 1317–1337, doi:10.1111/j.1365-2486.2009.02078.x, 2010.

Rabier, F.: Overview of global data assimilation developments in numerical weather-prediction centres, Q.J.R. Meteorol. Soc., 131, 3215–3233, doi:10.1256/qj.05.129, 2005.

Randerson, J. T., Chen, Y., van der Werf, G. R., Rogers, B. M., and Morton, D. C.: Global burned area and biomass burning emissions from small fires, J. Geophys. Res., 117, G04 012, doi:10.1029/2012JG002128, 2012.

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S., Janssens-Maenhout, G., Tubiello, F. N., Castaldi, S., Jackson, R. B., Alexe, M., Arora, V. K., Beerling, D. J., Bergamaschi, P., Blake, D. R., Brailsford, G., Brovkin, V., Bruhwiler, L., Crevoisier, C., Crill, P., Covey, K., Curry, C., Frankenberg, C., Gedney, N., Höglund-Isaksson, L., Ishizawa, M., Ito, A., Joos, F., Kim, H.-S., Kleinen, T., Krummel, P., Lamarque, J.-F., Langenfelds, R., Locatelli, R., Machida, T., Maksyutov, S., McDonald, K. C., Marshall, J., Melton, J. R., Morino, I., Naik, V., O'Doherty, S., Parmentier, F.-J. W., Patra, P. K., Peng, C., Peng, S., Peters, G. P., Pison, I., Prigent, C., Prinn, R., Ramonet, M., Riley, W. J., Saito, M., Santini, M., Schroeder, R., Simpson, I. J., Spahni, R., Steele, P., Takizawa, A., Thornton, B. F., Tian, H., Tohjima, Y., Viovy, N., Voulgarakis, A., Weele, M. v., Werf, G. R. v. d., Weiss, R., Wiedinmyer, C., Wilton, D. J., Wiltshire, A., Worthy, D., Wunch, D., Xu, X., Yoshida, Y., Zhang, B., Zhang, Z., and Zhu, Q.: The global methane budget 2000–2012, Earth System Science Data, 8, 697–751, doi:10.5194/essd-8-697-2016, 2016a.

Saunois, M., Jackson, R. B., Bousquet, P., Poulter, B., and Canadell, J. G.: The growing role of methane in anthropogenic climate change, Environmental Research Letters, 11, 120 207, doi:10.1088/1748-9326/11/12/120207, 2016b.

Schwietzke, S., Sherwood, O. A., Bruhwiler, L. M. P., Miller, J. B., Etiope, G., Dlugokencky, E. J., Michel, S. E., Arling, V. A., Vaughn, B. H., White, J. W. C., and Tans, P. P.: Upward revision of global fossil fuel methane emissions based on isotope database, Nature, 538, 88–91, doi:10.1038/nature19797, 2016.

Smith, B., Prentice, I. C., and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, Global Ecology and Biogeography, 10, 621–637, doi:10.1046/j.1466-822X.2001.t01-1-00256.x, 2001.

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, 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., 7, 3089–3110, doi:10.5194/gmd-7-3089-2014, 2014.

Thonat, T., Saunois, M., Bousquet, P., Pison, I., Tan, Z., Zhuang, Q., Crill, P., Thornton, B., Bastviken, D., Dlugokencky, E. J., Zimov, N., Laurila, T., Hatakka, J., Hermansen, O., and Worthy, D. E. J.: Detectability of Arctic methane sources at six sites performing continuous atmospheric measurements, Atmos. Chem. Phys. Discuss., 2017, 1–35, doi:10.5194/acp-2017-169, 2017.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Mon. Wea. Rev., 117, 1779–1800, doi:10.1175/1520-0493(1989)117⟨1779:ACMFSF⟩2.0.CO;2, 1989.

van der Laan-Luijkx, I. T., van der Velde, I. R., Krol, M. C., Gatti, L. V., Domingues, L. G., Correia, C. S. C., Miller, J. B., Gloor, M., van Leeuwen, T. T., Kaiser, J. W., Wiedinmyer, C., Basu, S., Clerbaux, C., and Peters, W.: Response of the Amazon carbon balance to the 2010 drought derived with CarbonTracker South America, Global Biogeochem. Cycles, 29, 2014GB005 082, doi: 10.1002/2014GB005082, 2015.

van der Laan-Luijkx, I. T., van der Velde, I. R., van der Veen, E., Tsuruta, A., Stanislawska, K., Babenhauserheide, A., Zhang, H. F., Liu, Y., He, W., Chen, H., Masarie, K. A., Krol, M. C., and Peters, W.: The CarbonTracker Data Assimilation Shell (CTDAS) v1.0: implementation and global carbon balance 2001–2015, Geosci. Model Dev., 10, 2785–2800, doi:10.5194/gmd-10-2785-2017, 2017.

van der Veen, E.: Optimizing transport properties in TM5 using SF6, Master's thesis, Wageningen University, University of Twente, Enschede, the Netherlands, URL http://essay.utwente.nl/65459/, 2013.

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, 11 707–11 735, doi:10.5194/acp-10-11707-2010, 2010.

van Leeuwen, P. J.: Comment on "Data Assimilation Using an Ensemble Kalman Filter Technique", Mon. Wea. Rev., 127, 1374–1377, doi:10.1175/1520-0493(1999)127⟨1374:CODAUA⟩2.0.CO;2, 1999.

van Leeuwen, P. J. and Evensen, G.: Data assimilation and inverse methods in terms of a probabilistic formulation, Mon. Wea. Rev., 124, 2892–2913, 1996.

Walter, K. M., Zimov, S. A., Chanton, J. P., Verbyla, D., and Chapin, F. S.: Methane bubbling from Siberian thaw lakes as a positive feedback to climate warming, Nature, 443, 71–75, doi:10.1038/nature05040, 2006.

Wania, R., Ross, I., and Prentice, I. C.: Integrating peatlands and permafrost into a dynamic global vegetation model: 1. Evaluation and sensitivity of physical land surface processes, Global Biogeochemical Cycles, 23, doi:10.1029/2008GB003412, 2009.

Whitaker, J. S. and Hamill, T. M.: Ensemble Data Assimilation without Perturbed Observations, Mon. Wea. Rev., 130, 1913–1924, doi:10.1175/1520-0493(2002)130⟨1913:EDAWPO⟩2.0.CO;2, 2002.

Yavitt, J. B., Downey, D. M., Lang, G. E., and Sexstone, A. J.: Methane Consumption in Two Temperate Forest Soils, Biogeochemistry, 9, 39–52, 1990.

Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S., Bril, A., Saeki, T., Schutgens, N., Toon, G. C., Wunch, D., Roehl, C. M., Wennberg, P. O., Griffith, D. W. T., Deutscher, N. M., Warneke, T., Notholt, J., Robinson, J., Sherlock, V., Connor, B., Rettinger, M., Sussmann, R., Ahonen, P., Heikkinen, P., Kyrö, E., Mendonca, J., Strong, K., Hase, F., Dohe, S., and Yokota, T.: Improvement of the retrieval algorithm for GOSAT SWIR XCO2 and XCH4 and their validation using TCCON data, Atmos. Meas. Tech., 6, 1533–1547, doi:10.5194/amt-6-1533-2013, 2013.

Zhang, S., Yi, X., Zheng, X., Chen, Z., Dan, B., and Zhang, X.: Global carbon assimilation system using a local ensemble Kalman filter with multiple ecosystem models, J. Geophys. Res. Biogeosci., 119, 2014JG002 792, doi:10.1002/2014JG002792, 2014.