

FINNISH METEOROLOGICAL INSTITUTE
CONTRIBUTIONS 172

DIMENSION REDUCTION APPROACHES FOR ATMOSPHERIC
REMOTE SENSING OF GREENHOUSE GASES

Otto Lamminpää

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in the hall "Metsätalo, Sali 4", Unioninkatu 40, Helsinki, on June 20th 2020 at 12 o'clock noon.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF HELSINKI
HELSINKI, FINLAND

SUPERVISOR

Prof. Samuli Siltanen, University of Helsinki, Finland
Prof. Johanna Tamminen, Finnish Meteorological Institute
Prof. Marko Laine, Finnish Meteorological Institute

PRE-EXAMINERS

Prof. Erkki Somersalo, Case Western Reserve University, USA
Prof. John Bardsley, University of Montana, USA

OPPONENT

Prof. Patrick Eriksson, Chalmers University of Technology, Sweden

CUSTOS

Prof. Eero Saksman, University of Helsinki, Finland

CONTACT INFORMATION

Department of Mathematics and Statistics
P.O. Box 64 (Gustav Hällströmin katu 2)
FI-00014 University of Helsinki
Finland
URL: <http://mathstat.helsinki.fi/>
Telephone: +358 29 419 11

Finnish Meteorological Institute
P.O. Box 503 (Erik Palménin aukio 1)
00101 Helsinki
Finland
URL: <https://ilmatieteenlaitos.fi/>
Telephone: +358 29 539 1000

Copyright © 2020 Otto Lamminpää
ISSN 0782-6117
ISBN 978-952-336-121-8 (paperback)
ISBN 978-952-336-122-5 (PDF)
Helsinki 2020
Edita Prima Oy



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
Contributions, 172, FMI-CONT-172
Date 10.6.2020

Author	Title
Otto Lamminpää	Dimension reduction approaches for atmospheric remote sensing of greenhouse gases

Abstract

Carbon dioxide (CO₂) and methane (CH₄) are two most significant anthropogenic greenhouse gases contributing to climate change and global warming. Indirect remote sensing measurements of atmospheric concentrations of these gases are crucial for monitoring manmade emissions and understanding their effects and related atmospheric processes. The reliability of these studies depends largely on robust uncertainty quantification of the measurements, which provides rigorous error estimates and confidence intervals for all results.

The main goal of this work is to develop and implement rigorous, robust and computationally efficient means of uncertainty quantification for atmospheric remote sensing of greenhouse gases. We consider both CO₂ measurements by NASA's Orbiting Carbon Observatory 2 (OCO-2) and CH₄ measurements by Sodankylä Arctic Space Center's Fourier Transform Spectrometer (FTS), the latter being studied from the perspectives of both individual measurements, and the entire time series from 2009-2018. Our approach leverages recent mathematical results on dimension reduction to produce novel algorithms that are a step towards thorough and efficient operational uncertainty quantification in the field of atmospheric remote sensing.

Mathematically, the process of inferring gas concentrations from indirect measurements is an ill-posed inverse problem, meaning that a well-defined solution doesn't exist without proper regularization. Bayesian approach utilizes probability theory to provide a regularized solution to the inverse problem as a posterior probability distribution. The posterior distribution is conventionally approximated using a Gaussian distribution, and results are reported as the mean of the distribution as a point estimate, and the corresponding variance as a measure of uncertainty. In reality, due to non-linear physics models used in the computations, the posterior is not well approximated by a Gaussian distribution, and ignoring its actual shape can lead to unpredictable errors and inaccuracies in the retrieval. Markov chain Monte Carlo (MCMC) methods offer a robust way to explore the actual properties of posterior distributions, but they tend to be computationally infeasible as the dimension of the state vector increases. In this work, the low intrinsic information content of remote sensing measurements is exploited to implement the Likelihood-Informed Subspace (LIS) dimension reduction method, which increases the computational efficiency of MCMC. Novel algorithms using LIS are implemented to abovementioned atmospheric CH₄ profile and column-averaged CO₂ concentration inverse problems.

Publishing unit	Classification (UDC)
Earth Observation Research	519.676, 551.510.41

Keywords

Greenhouse gases, Markov chain Monte Carlo, inverse problems, remote sensing, computational statistics

ISSN and series title	ISBN
0782-6117	978-952-336-121-8 (paperback)
Finnish Meteorological Institute Contributions	978-952-336-122-5 (pdf)

DOI	Language	Pages
https://doi.org/10.35614/isbn.9789523361225	English	153



Julkaisija

Ilmatieteen laitos
(Erik Palménin aukio 1)
PL 503, 00101 Helsinki

Julkaisun sarja, numero ja raporttikoodi

Contributions, 172, FMI-CONT-172
Päiväys 10.6.2020

Tekijä

Nimeke

Otto Lamminpää

Dimensioreduktiomenetelmien sovelluksia kasvihuonekaasujen kaukokartoituksessa

Tiivistelmä

Hiilidioksidi (CO₂) ja metaani (CH₄) ovat merkittävimmät ihmisperäiset kasvihuonekaasut, joilla molemmilla on huomattava vaikutus ilmastomuutokseen ja ilmakehän lämpenemiseen. Näiden kaasujen pitoisuuksien epäsuorat kaukokartoitusmittaukset ovat oleellinen osa ihmisperäisten päästöjen kehityksen seurannassa. Näitä mittauksia tarvitaan myös arvioitaessa kasvihuonekaasujen vaikutusta ilmakehän prosesseihin. Edellämäin tutkimuksen luotettavuus perustuu suurilta osin mittausten epävarmuuden arvioinnin paikkansapitävyyteen, minkä takaamiseksi käytetään korkeatasoista epävarmuusanalyysiä.

Tämän väitöskirjatyön tavoitteena on kehittää ja ottaa käyttöön luotettavia ja laskennallisesti tehokkaita epävarmuusanalyysimenetelmiä sovellettuna kasvihuonekaasujen kaukokartoitukseen. Kehitetyt menetelmät perustuvat matemaattisesti käänteisongelmien teoriaan ja todennäköisyysteorian sovelluksiin. Käytämme erityisesti informaatioteoreettisia työkaluja pienentääksemme käänteisongelman ulottuvuutta. Tämä johtaa laskennalliseen ongelmaan, joka on huomattavasti nopeampi ratkaista.

Työn sovelluskohteita ovat Nasan Orbiting Carbon Observatory 2 -satelliitin hiilidioksidipitoisuusmittaukset sekä Sodankylän Arktisessa Avaruuskeskuksessa sijaitsevan spektrometrin mittaamat metaanipitoisuudet. Jälkimmäisessä keskitymme sekä yksittäisiin mittauksiin että koko aikasarjan tutkimiseen ajalta 2009–2018. Kehitetyt menetelmät toimivat erittäin hyvin käsitellyissä sovelluksissa luoden pohjaa uusille operatiivisille epävarmuusanalyysialgoritmeille.

Julkaisijayksikkö

Luokitus (UDK)

Kaukokartoitustutkimus

519.676, 551.510.41

Asiasanat

Kasvihuonekaasut, Markovin ketju Monte Carlo, inversio-ongelmat, kaukokartoitus, laskennallinen tilastotiede

ISSN ja avainnimeke

ISBN

0782-6117

978-952-336-121-8 (paperback)

Finnish Meteorological Institute Contributions

978-952-336-122-5 (pdf)

DOI

Kieli

Sivumäärä

<https://doi.org/10.35614/isbn.9789523361225>

Englanti

153

SELECTED PUBLICATIONS

- I **O. Lamminpää**, M. Laine, S. Tukiainen and J. Tamminen. *Likelihood Informed Dimension Reduction for Remote Sensing of Atmospheric Constituent Profiles*. 2017 MATRIX Annals, 2:65-78, 2019. doi: 10.1007/978-3-030-04161-8.

O.L. implemented the LIS method to existing Swirlab software, derived the formulas connecting the previous work of Rodgers et al. and Cui et al., performed all the computations, analyzed all the results, and prepared the manuscript and figures for publication.

- II T. Karppinen, **O. Lamminpää**, S. Tukiainen, P. Heikkinen, R. Kivi, J. Hatakka, H. Lindqvist, M. Laine and J. Tamminen. *Vertical distribution of Arctic methane in 2009–2018 using ground-based remote sensing*. Remote Sens. 2020, 12, 917. doi: 10.3390/rs12060917.

O.L. took part in the initial planning of the article, implemented the dimension reduction approach, translated the computational model from MATLAB to Python, implemented the DLM code, computed the TCCON and AirCore comparisons, implemented averaging kernel corrections for ACE data and contributed to the preparation of the manuscript by writing sections and plotting figures.

- III **O. Lamminpää**, J. Hobbs, J. Brynjarsdóttir, M. Laine, A. Braverman, H. Lindqvist and J. Tamminen. *Accelerated MCMC for Satellite-Based Measurements of Atmospheric CO₂*. Remote Sens. 2019, 11(17):2061, 2019. doi: 10.3390/rs11172061.

O.L. did a major part of the planning of the article, obtained access to JPL's computational resources, implemented and tested all the new methods used in the paper, did all the analyses and computations, and prepared the manuscript and figures for publication.

PARTIAL LIST OF SYMBOLS

Symbol	Meaning
$[a, b]$	$\{x \in \mathbb{R} : a \leq x \leq b\}$
\mathbb{R}	$(-\infty, \infty)$
\approx	Is approximately equal to
\propto	Is proportional to
\sim	Is distributed according to
$a \rightarrow b$	a approaches b
$\mathcal{N}(\mu, \Sigma)$	Multivariate normal distribution with mean vector μ and covariance Σ
$ K $	Determinant of matrix K
Cov	Covariance
\mathbb{E}	Expected value
\mathcal{L}	Cholesky factorization
\mathbb{I}_n	$n \times n$ unit matrix

PARTIAL LIST OF ABBREVIATIONS

Abbreviation Stands for

AM	Adaptive Metropolis algorithm
AOD	Aerosol optical depth
ATBD	Algorithm Theoretical Basis Document
CH ₄	Methane
CO ₂	Carbon Dioxide
CS	Complement Subspace
DLM	Dynamic Linear Model
ECMWF	European Center for Medium-Range Weather Forecasts
FTS	Fourier Transform Spectrometer
IPCC	Intergovernmental Panel on Climate Change
LIS	Likelihood-Informed Subspace
MAP	Maximum a posteriori
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings algorithm
OCO-2	Orbiting Carbon Observatory 2 (NASA satellite)
OE	Optimal Estimation
X_{CO_2}	Column averaged dry air mole fraction of CO ₂
SVD	Singular Value Decomposition
TCCON	Total Carbon Column Observing Network
i.i.d.	Independent and identically distributed

ACKNOWLEDGMENTS

First of all, I'd like to thank my supervisors: Prof. Samuli Siltanen, whose courses were also my original inspiration for pursuing applied sciences, Prof. Johanna Tamminen, who has provided me with excellent insights and ideas throughout this process, and Prof. Marko Laine, who has patiently assisted and mentored me with any questions I had. Also, a special thanks to Marko for all the jazz gigs.

The publications in this work would not have been possible without my co-authors. A special thanks goes to Dr. Simo Tukiainen for all the insights on Swirlab and to Dr. Jonathan Hobbs for all the mentoring on JPL systems and OCO-2 codes. Further thank you for Dr. Amy Braverman for "headhunting" me during IWGGMS in Helsinki and introducing me to the JPL family. I would also like to thank Tomi Karppinen, Pauli Heikkinen, Rigel Kivi, Juha Hatakka, Huilin Chen and Jenny Bryjarsdóttir for all the collaboration on the papers.

During my doctoral studies, I have had the pleasure to work at Finnish Meteorological Institute's Greenhouse gases and satellite methods group, lead by Hannakaisa Lindqvist who has been a phenomenal boss and a friend during this time. My gratitude also goes to my colleagues from the research group: Ella Kivimäki, Antti Mikkonen, Anu Kauppi, Janne Hakkarainen, and Sindu Parampil. Other colleagues I would like to thank include Prof. Heikki Haario especially for the invitations to Luosto, Sebastian Springer and Ramona Maraia for all the good company around the world, Jouni Susiluoto for all the helpful insights you have shared with me, and all the folks of Geodynamo for the music we got to play together.

None of this work would have been possible without the support of my amazing family and friends. I would like to thank my parents, Jukka and Anne, and my siblings, Jussi and Riikka, for always being there for me no matter what. A special thanks goes to my friends Niklas Itänen, Mika Väänänen and Joonas Viuhko who made me explain what I'm actually doing and who shared many good times and wonderful ideas with me during this project. My warmest gratitude also goes to Jhanel Stewart: you held me together during the most difficult and frustrating moments of my work, and gave me the will and motivation to go forward each and every day.

x

Finally, I would like to dedicate this work to the memory of my late grandfather Leo Gunnar Lamminpää. You believed in my graduation already when I had barely started, and I wish you could have been with us to see this day come true at last.

Turku, Finland, June 2020
Otto Lamminpää

CONTENTS

1	Introduction	1
2	Inverse problems	5
2.1	Classical inverse problems	5
2.1.1	Atmospheric remote sensing	6
2.2	Bayesian formulation	9
2.2.1	Application to atmospheric remote sensing	11
2.3	State space modeling	14
2.3.1	Dynamic linear model	15
2.4	Optimization for MAP estimates	16
2.4.1	Levenberg-Marquardt algorithm	16
2.5	Markov chain Monte Carlo	17
2.5.1	Metropolis-Hastings algorithm	18
2.5.2	Adaptive Metropolis algorithm	19
3	Dimension Reduction	21
3.1	Information content	22
3.2	Prior reduction	25
3.3	Likelihood-informed subspace	26
4	Uncertainty quantification in remote sensing of greenhouse gases	31
4.1	Ground based CH ₄ retrieval	31
4.1.1	Sodankylä Fourier Transform Spectrometer	32
4.1.2	Implementation	33
4.1.3	Overview of results	37
4.2	Satellite retrieval of CO ₂	43
4.2.1	Orbiting Carbon Observatory 2	43
4.2.2	Implementation	45

4.2.3 Overview of results	48
5 Conclusions	55
Appendices	59
A Preliminary mathematical concepts	61
A.1 Matrices	61
A.2 Random variables	62
References	65

1 INTRODUCTION

Wide effects of global warming and climate change can be largely attributed to anthropogenic release of greenhouse gases in the atmosphere. These gases have the effect of reflecting solar light back towards the surface of the planet, leading in more heat being trapped. This phenomenon has a vast body of scientific evidence in its support, including e.g. IPCC (2013), Melillo et al. (2014), Council (2008). The most prevalent greenhouse gases are carbon dioxide (CO_2) and methane (CH_4), both of which have their concentrations increasingly risen from pre-industrial times, from about 280 ppm to 410 ppm for CO_2 , and for CH_4 , from about 700 ppb to almost 1800 ppb.

Carbon dioxide is the most significant long-lived anthropogenic greenhouse gas in Earth's atmosphere. Quantifying point sources, such as power plants, as well as state-of-art modelling of sources and sinks of CO_2 require a global coverage of accurate concentration measurements. NASA's Orbiting Carbon Observatory 2 (OCO-2) satellite mission assists in this goal by measuring the attenuation of sunlight in Earth's atmosphere. From the measured *absorption spectra*, CO_2 concentrations can be inferred. For credible analysis and further research, rigorous *uncertainty quantification* is needed to correctly assess the reliability of the measurements.

Atmospheric methane is the second most important greenhouse gas in the atmosphere after CO_2 . While methane is not as abundant, it traps heat over 20 times more per unit mass than carbon dioxide. While major sources of methane are on the surface, the main loss mechanism for is the oxidation by the hydroxyl radical, mainly happening in the upper *troposphere* (altitudes from 0 to approx. 13 km on average) and in the *stratosphere* (altitudes from 13 km to approx. 50 km). Due to this phenomenon, in order to understand the carbon cycle of methane, it is important to study the changes in the profile shape of CH_4 . This is especially true in the Arctic regions like northern Finland, where a *polar vortex* forms each winter, limiting the flow of air to the region and thus causing CH_4 concentrations to decrease more rapidly during this time. This vertical profile can be extracted from the measurements

carried out by the Sodankylä Total Carbon Column Observation Network (TCCON) Fourier transform spectrometer (FTS). It is a ground based device that, similar to OCO-2, measures the attenuation of solar light in the atmosphere and requires solid uncertainty quantification for the measurements to be reliable.

The main goal of this work is to develop and implement rigorous, robust and computationally efficient means of uncertainty quantification for atmospheric remote sensing of greenhouse gases. We consider both CO₂ measurements by OCO-2 and CH₄ measurements by Sodankylä FTS mentioned above. Our approach leverages recent mathematical results on dimension reduction to produce novel algorithms that are a step towards thorough operational uncertainty quantification in this field.

The process of inferring gas concentrations from measured spectra is known as *retrieval*. The retrieval is an *ill-posed inverse problem*, meaning that a well-defined solution doesn't exist without proper regularization. Bayesian inversion casts the retrieval to the language of probability theory and provides a regularized solution to the inverse problem as posterior probability distribution. The posterior distribution is conventionally approximated using a Gaussian distribution, and results are reported as the mean of the distribution as a point estimate, and the corresponding variance as a measure of uncertainty. In reality, due to non-linear computational physics models used in the retrievals, the posterior is not well approximated by a Gaussian distribution, and ignoring its actual shape can lead to unpredictable errors and inaccuracies in the retrieval. Markov chain Monte Carlo (MCMC) methods offer a robust way to explore the actual properties of posterior distributions, but they tend to be computationally challenging as the dimension of the state vector increases.

In this work, we exploit the low intrinsic information content of remote sensing measurements and implement the Likelihood-Informed Subspace (LIS) dimension reduction method to increase the computational efficiency of MCMC retrievals. Novel algorithms using LIS are implemented to abovementioned atmospheric methane profile and column-averaged carbon dioxide concentration retrievals. The work in **Paper I** implements adaptive LIS MCMC to CH₄ profile retrieval from real world FTS data. This method additionally provides the freedom of choosing any desired prior covariance matrix. The performance of this novel algorithm is finally compared against another dimension reduction method. In **Paper II**, the dimension reduction based profile retrieval is implemented for all measurements of Sodankylä FTS from 2009 to 2018, resulting in a robust algorithm that produces a time series vertical profiles that is shown to agree well with three separate instruments measuring CH₄ concentrations. This time series is further analyzed with a dynamic linear model, which lets us obtain information about the non-linear trend of CH₄ as a function of altitude. Lastly, in **Paper III** LIS dimension reduction is implemented to a synthetic OCO-

2 CO₂ retrieval, which increases the computational efficiency of adaptive MCMC. Resulting posterior sample is then used to explore the non-Gaussian shape of the posterior, which is shown to improve the accuracy of the retrieval.

The rest of this introductory part is organized as follows: Chapter 2 introduces the concepts of inverse problems and the Bayesian formulation, also going through the optimization and MCMC methods used in this work. Chapter 3 deals with information content of a measurement and lays out how the utilized dimension reduction strategies are formulated. Chapter 4 lays out the implementation and results of our work with the Sodankylä FTS and OCO-2 retrievals, together with the DLM time series analysis method and our dimension reduction retrievals application to a long time series. The basic mathematical concepts needed to understand the theory and results of this work are briefly introduced in the Appendix.

2 INVERSE PROBLEMS

To infer the quantities of interest from the measured atmospheric absorption of radiation, we need to solve an *inverse problem*. In this Chapter, we first introduce the general definition of inverse problems, followed by a formulation of the *Bayesian framework* for obtaining solutions as probability distributions. We then describe the atmospheric remote sensing inverse problem in more detail and identify the challenging aspects which are the focus of this work. Lastly, we introduce the optimization and Markov chain Monte Carlo algorithms used in this work to solve Bayesian inverse problems.

2.1 CLASSICAL INVERSE PROBLEMS

In a general setting, the result of some indirect measurement, given as the *data vector* y , is used to obtain information about a *state vector* x . These vectors belong to the *data space* $y \in \mathbb{Y} \subseteq \mathbb{R}^m$ and the *state space* $x \in \mathbb{X} \subseteq \mathbb{R}^n$, where generally $m \neq n$. The results presented in this section are easily generalized into more general Hilbert spaces (Mueller and Siltanen, 2012), but in this overview we limit ourselves to the real valued case.

The measurement is described with the *indirect inverse model* (Mueller and Siltanen, 2012)

$$y = \mathcal{F}(x) + \varepsilon, \quad (2.1)$$

where $y \in \mathbb{Y} \subseteq \mathbb{R}^m$, $x \in \mathbb{X} \subseteq \mathbb{R}^n$ and $\mathcal{F} : \mathbb{X} \rightarrow \mathbb{Y}$ is the *forward map* which describes the physics of the measurement. In this work, we assume the measurement noise $\varepsilon \in \mathbb{R}^m$ to be additive.

In practice, in order to solve the inverse problem, we need to approximate the forward map with

$$y = F(x, \theta) + \varepsilon. \quad (2.2)$$

where $F(x, \theta) : \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}^m$ is the discretized version of the ideal forward map \mathcal{F} called the *forward model*, where some model parameters $\theta \in \mathbb{R}^l$ are fixed when building a computational approximation. In this work, we assume the forward model to be differentiable so that the first order derivative $\frac{\partial}{\partial x} F(x)$ exists.

Difficulties in solving inverse problem arise from *ill-posedness*, which is best understood as the opposite of *well-posedness*. According to Jacques Hadamard (1865-1963), a *well posed problem* (see e.g. Mueller and Siltanen (2012)) fulfils the following criteria:

H1: The problem has a solution (existence).

H2: The solution is unique (uniqueness).

H3: The solution depends continuously on the data (stability).

An *ill posed problem* can now be thought of as a problem that violates one or more of Hadamard's conditions. This means that the exact answer might not exist, or that there are several or even infinitely many states that can produce the same data. Furthermore, the problem is ill-posed if arbitrarily small errors in the measured data can cause large errors in the result, known as *unstable* problems. In practice, all the measurements are affected by some sort of noise or measurement error, and hence this is a challenge in solving inverse problems.

In practice, inverse problems tend to be ill posed. A technique called *regularization* is thus required to deal with ill-posedness. Classical regularization methods include the truncated SVD and Tikhonov regularization (e.g. Mueller and Siltanen (2012), Bardsley (2018)), but in this work we use the statistical approach and a priori constraints (Kaipio and Somersalo, 2005) for regularization. Statistical approach to inverse methods is also the standard method used in the atmospheric remote sensing literature (e.g. Rodgers (2000)).

2.1.1 ATMOSPHERIC REMOTE SENSING

We are interested in solving the inverse problem of inferring different atmospheric properties, such as temperature or trace gas concentrations, from measured attenuation of electromagnetic radiation in the atmosphere, recorded as an *absorption spectrum*. The measurements we consider in this work are done by ground based devices looking directly at the sun, and by satellites measuring solar radiation scattered back to space by Earth's surface. The latter is illustrated in Figure 2.1, with the measurement geometry of ozone column measuring *GOME* satellite. The process of solving the resulting inverse problem is referred to as a *retrieval*. The part

of the forward model describing the absorption and propagation of radiation in the atmosphere is referred to as atmospheric *radiative transfer*. In addition, the forward model includes the *instrument model* which describes effects taking place inside the measurement device. In computational implementation of radiative transfer models, choices like number of absorbing gases, means of modelling of scattering and polarization of light, effects of atmospheric aerosols, pressure and temperature, and properties of Earth's surface are made, leading to models of different accuracies.

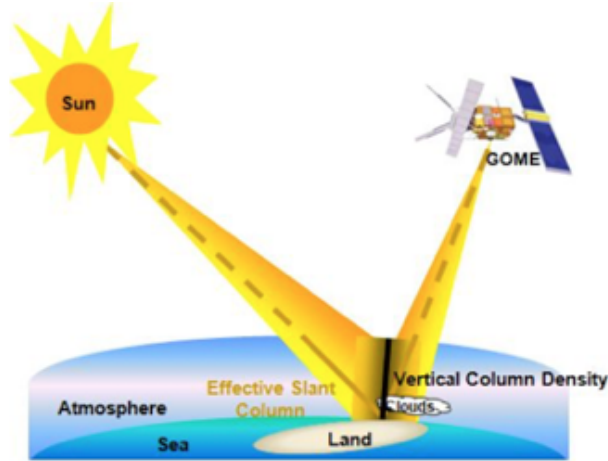


Figure 2.1: Measurement geometry of the GOME satellite, which detects solar light reflected by Earth's surface. In the retrieval process the satellite measured ozone slant column (brown path) is converted to viewing geometry independent vertical column of ozone (black path). Illustrative figure taken from Dameris and Loyola (2011).

As an illustration, let us consider the forward model for the atmospheric remote sensing problem in **Paper I** and **Paper II** (from the Swirlab software by Tukiainen et al. (2016)). The data y (Figure 2.2) is obtained as an absorption spectrum with m separate wavelengths. The solar light passing through the atmosphere and hitting the detector is modelled using the *Beer-Lambert law* (see e.g. IUPAC (1997)), which gives, for wavelengths $\lambda_j, j \in [1, \dots, m]$, the intensity of detected light as

$$I(\lambda_j) = I_0(\lambda_j) \exp \left(- \sum_{k=1}^K \int_0^{\infty} C_k(\lambda_j, z) \rho_k(z) dz \right), \quad (2.3)$$

where I_0 is the intensity of solar light, the atmosphere has K absorbing trace gases,

$C_k(\lambda_j, z)$ is the absorption coefficient of gas k , which depends on altitude z and on the wavelength λ_j , and $\rho_k(z)$ is the density of gas k at altitude z .

In a retrieval, the gas densities $\rho_k(z)$ are inferred from the measured spectra $I(\lambda_j)$. In the computational model, the solar light intensity $I_0(\lambda_j)$ is in our case assumed to be known. The integral in the Beer-Lambert law is discretized into a sum over atmospheric layers in which all properties are assumed to be constant. This discretization also determines the size of the state vector. The absorption coefficients $C_k(\lambda_j, z)$ have to be computed based on an a-priori atmosphere as they depend on the pressure and temperature of the atmosphere. A challenge in atmospheric remote sensing is that even forward models based on this simple radiative transfer model are non-linear, which necessitates either linear approximations or more complex and sophisticated methods to solve the resulting inverse problem.

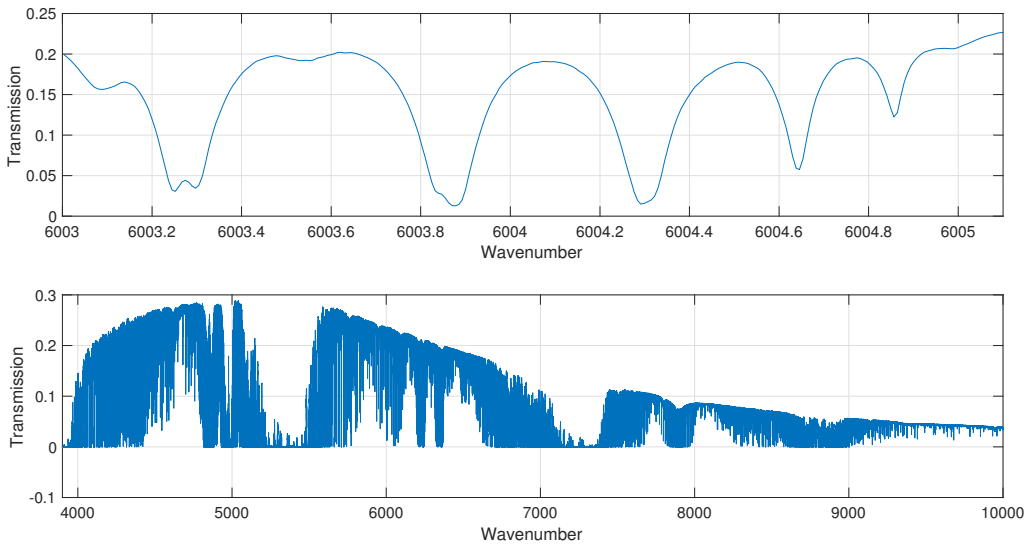


Figure 2.2: Absorption spectrum measured by the Sodankylä FTS in arbitrary units. Lower panel shows the entire spectrum from 3900cm^{-1} to 10000cm^{-1} . The upper panel shows the spectral window of 6003cm^{-1} to $6005,5\text{cm}^{-1}$ used in the CH_4 profile retrieval, since this part of the spectrum doesn't have significant absorption from other chemical compounds than CH_4 and water vapor.

2.2 BAYESIAN FORMULATION

In this work, we utilize the statistical approach to inverse problems (e.g. Kaipio and Somersalo (2005)), in which all variables are treated as random variables. This permits using the *Bayes formula* to obtain a solution to the inverse problem as *posterior* probability distribution for state x . This method combines information from the data as *likelihood* distribution, together with regularization from a *a priori* probability distribution. We will next provide a short description of the Bayesian formulation of inverse problems.

Let $Y \in \mathbb{R}^m$ be the data and $X \in \mathbb{R}^n$ be the unknown state. Let $\pi(x, y)$ be the joint probability distribution random variables X and Y . Then *the posterior density function* of $X = x$ conditional on the data $Y = y$ is given by the *Bayes formula* (e.g. Gelman et al. (1995))

$$\pi(x|y) = \frac{\pi(y|x)\pi_0(x)}{\pi(y)}, \quad (2.4)$$

where $\pi(y) = \int_{\mathbb{R}^n} \pi(x, y) dx > 0$ by assumption. The marginal density function

$$\pi_0(x) = \int_{\mathbb{R}^m} \pi(x, y) dy \quad (2.5)$$

is called the *a priori density function* (also *prior* for short). It describes the information about the unknown X before the measurement $Y = y$ is made. Physical constraints such as positivity and feasible intervals can be included in the prior to regularize an ill-posed inverse problem.

For fixed $X = x$ and $\pi_0(x) \neq 0$, we get the *likelihood function* of data $Y = y$ given x as

$$\pi(y|x) = \frac{\pi(x, y)}{\int_{\mathbb{R}^m} \pi(x, y) dy}. \quad (2.6)$$

The dependency between these variables is described with the inverse model $y = F(x, \theta) + \varepsilon$. It is common practice to write out the model parameters θ explicitly, but in the following discussions we assume them to be fixed and write the forward model as $F(x)$. The likelihood distribution is obtained using the statistics of the measurement error, since by formula (2.2) we get $y = F(x) + \varepsilon \Leftrightarrow \varepsilon = F(x) - y$.

In this work, we assume the prior density function to be Gaussian with a known mean and covariance, which is denoted as $\mathcal{N}(x_0, \Sigma_0)$. Also, the additive noise is assumed to have be zero-mean Gaussian density function with known covariance matrix, $\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$. Combining the Gaussian prior and likelihood using the Bayes

formula (2.4) we get the solution to the inverse problem as the posterior distribution

$$\pi(x|y) \propto \exp\left(-\frac{1}{2}(y - F(x))^T \boldsymbol{\Sigma}_\varepsilon^{-1}(y - F(x)) - \frac{1}{2}(x - x_0)^T \boldsymbol{\Sigma}_0^{-1}(x - x_0)\right), \quad (2.7)$$

where the proportionality \propto comes from the fact that the marginal distribution $\pi(y)$ in (2.4) can be thought of as a normalizing constant not depending on the state x and is therefore omitted.

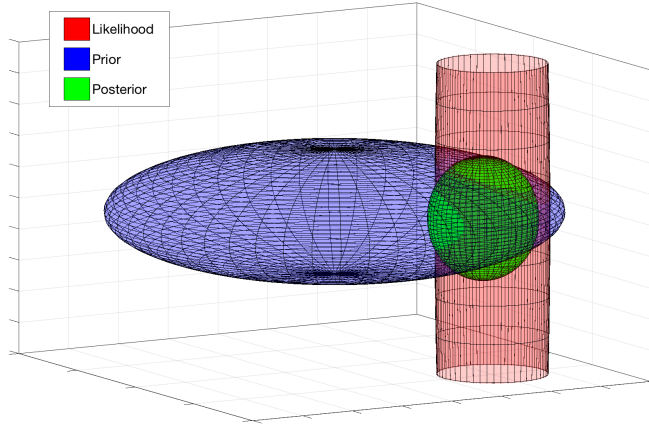


Figure 2.3: Illustration of a 3-dimensional Gaussian posterior distribution (green) as a product of a Gaussian likelihood (red) that doesn't contain information in the z -direction, making the problem ill-posed, and the Gaussian a priori distribution (blue) used to regularize the posterior.

Assume that the likelihood and prior are Gaussian. When the forward model F is linear, the solution to the inverse problem (2.2) is the Gaussian distribution $\mathcal{N}(\hat{x}, \boldsymbol{\Sigma}_{post})$ (Kaipio and Somersalo, 2005). The posterior mean and covariance are obtained analytically as

$$\hat{x} = \left(F^T \boldsymbol{\Sigma}_\varepsilon^{-1} F + \boldsymbol{\Sigma}_0^{-1}\right)^{-1} \left(F^T \boldsymbol{\Sigma}_\varepsilon^{-1} y + \boldsymbol{\Sigma}_0^{-1} x_0\right) \quad (2.8)$$

$$\boldsymbol{\Sigma}_{post} = \left(F^T \boldsymbol{\Sigma}_\varepsilon^{-1} F + \boldsymbol{\Sigma}_0^{-1}\right)^{-1} \quad (2.9)$$

When the forward model F is non-linear, the resulting posterior is not Gaussian and the formulas (2.8) don't hold. This is usually the case in atmospheric remote sensing. Several *statistics* are still obtainable if the expression (2.7) is available.

A widely used point estimator for a solution to the Bayesian inverse problem is the *maximum a posteriori estimator* (MAP), denoted \hat{x} . For posterior distribution $\pi(x|y)$ of the state $X \in \mathbb{R}^n$ it holds that

$$\hat{x} = \arg \max_{x \in \mathbb{R}^n} \pi(x|y), \quad (2.10)$$

if the maximum exists. This maximum is not necessarily unique. In this work, we use optimization methods to compute the most probable value of the posterior

Various numerical summaries of the posterior are desirable. Two of the most common such numbers are *conditional mean* $\bar{x} \in \mathbb{R}^n$ as a point estimate, and *co-variance* $\text{Cov}(X) \in \mathbb{R}^{n \times n}$ as a measure of uncertainty. These are given as expected values

$$\begin{aligned} \bar{x} &= \int_{\mathbb{R}^n} x \pi(x|y) dx \\ \text{Cov}(X) &= \int_{\mathbb{R}^n} (x - \bar{x})^2 \pi(x|y) dx \end{aligned} \quad (2.11)$$

over the posterior distribution $\pi(x|y)$. Evaluating this types of integrals is often challenging or impossible, since the result is typically not available in closed form. Therefore we will use *Markov chain Monte Carlo* (MCMC) methods (e.g. Tarantola (2005), Kaipio and Somersalo (2005) or Stuart (2010) for a comprehensive review) to obtain a set of realizations from the posterior and approximate these statistic via the result (A.14) from the preliminary chapter.

2.2.1 APPLICATION TO ATMOSPHERIC REMOTE SENSING

The retrieval problem in atmospheric remote sensing is described in detail by Rodgers (2000) using Bayesian formulation, which is a widely used framework in the remote sensing literature. The solution to the inverse problem is obtained as a *Maximum A Posteriori* estimate, denoted \hat{x} , which is the most probable state of the posterior distribution. This approach described later in Section 2.4.

The retrieval is typically an ill-posed inverse problem and requires regularization. Enforcement of feasible physical constraints, modelling choices, and approximations in radiative transfer offer general guidelines in defining a proper prior distribution which is usually assumed to be Gaussian. A further assumption that the posterior distribution can be approximated by a Gaussian distribution is also made. This assumption permits a convenient linear error characterization and propagation analysis, which we will shortly describe next.

The derivative of the forward model is given as the *Jacobian matrix* $\mathbf{J}_x \in \mathbb{R}^{m \times n}$ with elements $\mathbf{J}_{x,ij} = \frac{\partial}{\partial x_j} F_i(x)$. We further define the following quantities:

- The *Gain matrix* $\mathbf{G} := \frac{\partial \hat{x}}{\partial y}$ characterizes the linear response that a small change in the measurement causes to the retrieval:

$$\mathbf{G} = \Sigma_0 \mathbf{J}_x^T (\mathbf{J}_x \Sigma_0 \mathbf{J}_x^T + \Sigma_\epsilon)^{-1} \quad (2.12)$$

- The *Averaging kernel* $\mathbf{A} := \frac{\partial \hat{x}}{\partial x}$ characterizes the linear response that a small change in the true state causes to the retrieval:

$$\mathbf{A} = \mathbf{G} \mathbf{J}_x \quad (2.13)$$

Let us again consider the retrieval using the forward model of the Swirlab software for illustration. The software includes a methane profile retrieval, which we will shortly illustrate next. Figure 2.4 shows a retrieved CH_4 profile from 1 to 70 km and related posterior uncertainties, together with the absorption spectrum and the Gain matrix of the retrieval. We can see that the retrieval is most sensitive in altitudes from about 10 km to 40 km, whereas the highest altitudes come directly from the prior. Furthermore, information on these altitudes can be seen to be obtained largely from the width of the absorption spikes and not the entire spectrum. A further illustration of the sensitivity of the retrieved state to the changes in the underlying true value is seen in the averaging kernel of the retrieval in Figure 2.5.

Assume that the true atmospheric state is known and we know the best possible values x_{true} for the state vector and θ_{true} for the model parameters. A widely used error analysis framework introduced by Rodgers (2000)) quantifies the linear propagation of uncertainty in the retrieval as the difference

$$\begin{aligned} \hat{x} - x_{\text{true}} = & (\mathbb{I} - \mathbf{A})(x_0 - x_{\text{true}}) && \text{Smoothing error} \\ & + \mathbf{G} \mathbf{J}_\theta (\theta_{\text{true}} - \theta) && \text{Parameter error} \\ & + \mathbf{G} \Delta F(x, \theta) && \text{Approximation error} \\ & + \mathbf{G} \epsilon && \text{Instrument noise,} \end{aligned} \quad (2.14)$$

where θ is our best estimate of the "true" model parameters θ_{true} , $\mathbf{J}_\theta = \frac{\partial}{\partial \theta} F(x, \theta)$ is the sensitivity of the forward model to the model parameters, and $\Delta F(x, \theta)$ contains the errors we make when approximating the true atmospheric radiative transfer (2.1) by a computational model (2.2).

As can be seen from formula (2.14), the retrieval problem is prone to numerous different sources of error. A linear approximation of the posterior covariance matrix is given as

$$\Sigma_{\text{post}}(\hat{x}) = \left(\mathbf{J}_x^T \Sigma_\epsilon^{-1} \mathbf{J}_x + \Sigma_0^{-1} \right)^{-1} \in \mathbb{R}^{n \times n} \quad (2.15)$$

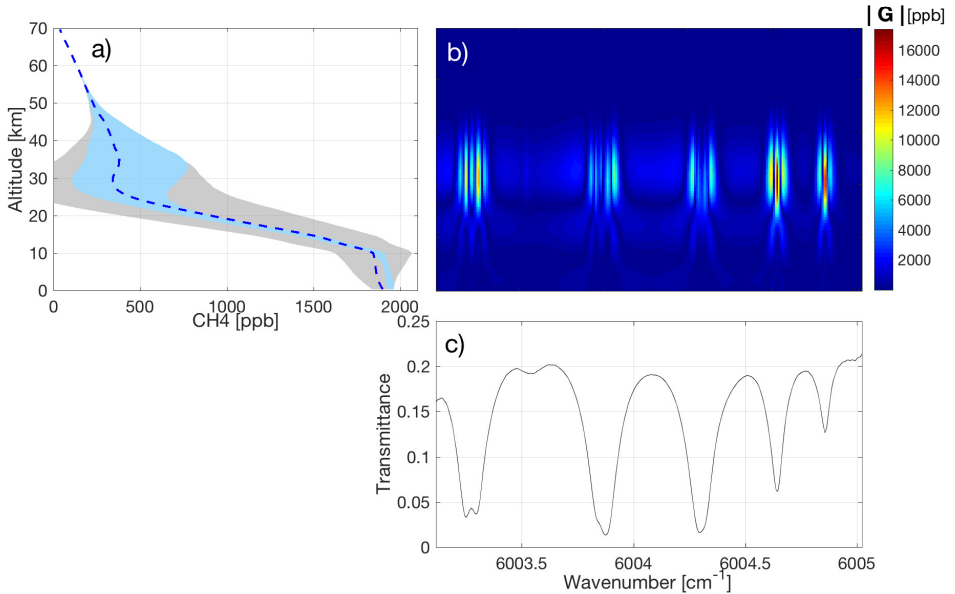


Figure 2.4: Illustration of the Gain matrix from our FTS retrieval's physics model. a) The 95% confidence limits of retrieved posterior distribution (blue) and prior (dashed blue for mean, grey for variance) of CH_4 concentration profile. b) Absolute values of the Gain matrix \mathbf{G} , showing is the sensitivity of the retrieved profile to the measured absorption spectrum, that is, $\frac{\partial \hat{x}}{\partial y}$. c) The data y , an absorption spectrum, used in the retrieval. It can be seen that the retrieval is most sensitive at altitudes from 20 km to 40 km, to changes in the data that mainly corresponds to the width and depth of absorption spikes in the spectrum.

However, this error characterization does not take into account the effect of a non-linear forward model. The main idea of this work is to use *Markov chain Monte Carlo* (MCMC) to explore the non-Gaussian shape of the posterior distribution. High dimensional MCMC suffers from slow convergence and high computational cost, so dimension reduction combined with the adaptive Metropolis algorithm is implemented to speed up the computation. Motivation for this comes from the information theoretical considerations in section 3.1, as atmospheric remote sensing problems typically have a significantly low intrinsic dimensionality and hence contain a relatively low amount of information about the features we are interested in.

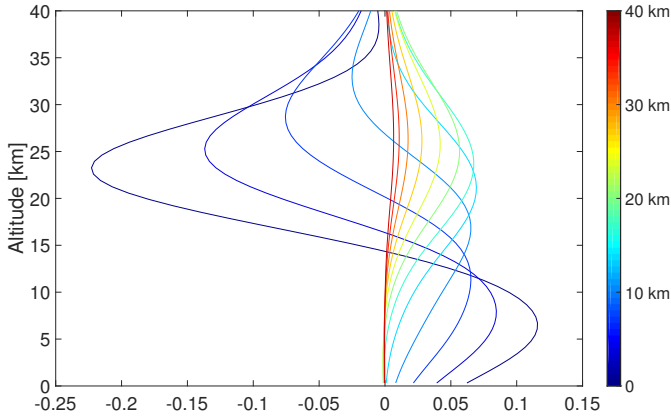


Figure 2.5: Every fifth layer of the FTS retrieval's averaging kernel matrix at $SZA = 45^\circ$ depicted in ppb units. This describes how a change at a given altitude, color coded on the right, propagates to other altitudes in the retrieval.

2.3 STATE SPACE MODELING

In the context of atmospheric remote sensing, the observed phenomena are often dependent on the dynamics of atmospheric processes, and therefore not static in relation to time. Earlier, we defined inverse problems by the equation

$$y = F(x, \theta) + \varepsilon. \quad (2.16)$$

This problem does not explicitly address any time-dependence of the state x . Furthermore, the uncertainty model parameters θ are not implicitly treated. We can, however, consider targets that evolve in time, e.g. *time series*, and formulate the inverse problem so that we obtain a solution as a function of time.

The *state space* formulation describes both how the data $y_t \in \mathbb{R}^m$ at time step t depends on the target $x_t \in \mathbb{R}^n$, as well as how the state evolved from the previous timestep x_{t-1} . We will also assign a prior distribution on parameters θ and explicitly fit them to the observations. State space models, sometimes called *hidden Markov models* or *structured time series models*, are well known and documented in time series literature (Chatfield (1989), Harvey (1991), Hamilton (1994), Migon et al. (2005)). The dynamic formulation we use Laine et al. (2014), given an assumption that the time evolution of atmospheric states can be described with linear operators, is known as the *dynamic linear model* (DLM), turning the time series analysis into a Bayesian inverse problem. This allows us to assess uncertainty in the timeseries

instead of just a point estimate solution.

We next follow the description by Laine (2019) for a brief DLM introduction. The techniques used in the software implementation include the *Kalman filter*, which solves the linear inverse problem on each timestep and uses the previous solution as a prior in next timestep, *Kalman smoother* going backwards in time again to get a smoother solution, and the *simulation smoother* to sample from the joint posterior given all the data. A more detailed description of these methods is beyond the scope of this overview

2.3.1 DYNAMIC LINEAR MODEL

In order to formulate the dynamic linear model, we need we need to assume that the evolution of the states depends on the history only through the previous time step (Kaipio and Somersalo, 2005). Let $\{X_t\}_{t=0}^{\infty}$ and $\{Y_t\}_{t=1}^{\infty}$ be stochastic processes so that $X_t \in \mathbb{R}^n$ represents a state vector and $Y_t \in \mathbb{R}^m$ is a measurement or an observation on time step t . We also assume that these processes have the following properties:

1. $\{X_t\}_{t=0}^{\infty}$ is a Markov process, s.t

$$\pi(x_{t+1}|x_0, \dots, x_t) = \pi(x_{t+1}|x_t). \quad (2.17)$$

2. $\{Y_t\}_{t=1}^{\infty}$ is a Markov process with respect to the history of $\{X_t\}_{t=0}^{\infty}$, meaning that

$$\pi(y_t|x_0, \dots, x_t) = \pi(y_t|x_t). \quad (2.18)$$

3. $\{X_t\}_{t=0}^{\infty}$ depends on the observations only through its own history, that is,

$$\pi(x_{t+1}|x_t, y_0, \dots, y_t) = \pi(x_{t+1}|x_t). \quad (2.19)$$

With the previous assumptions, we can write the dynamic linear model as a general linear state space model with an observation equation and an evolution equation as

$$\begin{aligned} y_t &= \mathbf{H}_t x_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, \mathbf{R}_t), \\ x_t &= \mathbf{M}_t x_{t-1} + E_t, & E_t &\sim \mathcal{N}(0, \mathbf{Q}_t), \end{aligned} \quad (2.20)$$

where $y_t \in \mathbb{R}^m$ contains the observations and $x_t \in \mathbb{R}^n$ the hidden and unobserved model state at time steps $t = 1, \dots, N$, the observations are gained via the *observation operator* $\mathbf{H}_t \in \mathbb{R}^{m \times n}$, the state is evolving along a *linear system operator* $\mathbf{M}_t \in \mathbb{R}^{n \times n}$. A non-linear extension exists (Cressie and Wikle (2011), Särkkä

(2013)), but in our case linear operators are sufficient. In this formulation, the observation error $\varepsilon_t \in \mathbb{R}^n$ and the model error $E_t \in \mathbb{R}^m$ are assumed to be zero mean Gaussian with positive definite, symmetric covariance matrices given by $\mathbf{R}_t \in \mathbb{R}^{n \times n}$ and $\mathbf{Q}_t \in \mathbb{R}^{m \times m}$, respectively.

The DLM formulation can be seen as a hierarchical statistical model with three levels: data y_t , process x_t and model parameters θ (Cressie and Wikle, 2011). By using the Bayes formula, we can write the state and parameter posterior distributions as a product of conditional distributions

$$\pi(x_t, \theta | y_t) \propto \pi(y_t | x_t, \theta) \pi(x_t | \theta) \pi(\theta) \quad (2.21)$$

where $\pi(y_t | x_t, \theta)$ is the likelihood describing observation uncertainty, $\pi(x_t | \theta)$ is the process uncertainty given model parameters, and $\pi(\theta)$ is the prior uncertainty for the model parameters. Explicitly, the parameters θ are used to define the operators \mathbf{H}_t and \mathbf{M}_t as well as the covariance matrices \mathbf{R}_t and \mathbf{Q}_t .

2.4 OPTIMIZATION FOR MAP ESTIMATES

Finding the MAP-estimator is an optimization problem, which can be solved using iterative algorithms to minimize the cost function

$$-2 \log \pi(x|y) = \text{constant} + (y - F(x))^T \boldsymbol{\Sigma}_\varepsilon^{-1} (y - F(x)) + (x - x_0)^T \boldsymbol{\Sigma}_0^{-1} (x - x_0), \quad (2.22)$$

which results from the formula (2.7) for the posterior distribution. This procedure is referred to as *Optimal Estimation* (OE) in the remote sensing literature (Rodgers, 2000). A large variety of methods exist, including quadratic programming, gradient descent, or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (see for example Gill et al. (1981)). In atmospheric remote sensing, a common strategy for finding MAP-estimates is to use the *Levenberg-Marquardt* (Levenberg (1944), Marquardt (1963)) algorithm.

2.4.1 LEVENBERG-MARQUARDT ALGORITHM

The Levenberg-Marquardt algorithm is a balance between the gradient descent and *Gauss-Newton* iterations. Gradient descent updates the parameter values in the direction opposite to the gradient of the cost function (2.22). The update is given by

$$x_{i+1} = x_i + \alpha \left(\mathbf{J}_{x_i}^T \boldsymbol{\Sigma}_\varepsilon^{-1} (y - F(x_i)) - \boldsymbol{\Sigma}_0^{-1} (x_i - x_0) \right). \quad (2.23)$$

where α is a scaling parameter defining the step length, and \mathbf{J}_{x_i} is the Jacobian matrix of the forward model F evaluated at x_i .

The Gauss-Newton method is a method for minimizing a sum-of-squares cost function under the assumption that the cost function is approximately quadratic near the solution (Björck, 1996). If the assumption holds, the Gauss-Newton method converges faster than the gradient descent (Marquardt, 1963) for moderate sized problems. For non-quadratic cost functions with thousands of parameters the slower converging gradient descent may be the only viable option. The Gauss-Newton update is given by

$$x_{i+1} = x_i + [\boldsymbol{\Sigma}_0^{-1} + \mathbf{J}_{x_i}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{J}_{x_i}]^{-1} \left(\mathbf{J}_{x_i}^T \boldsymbol{\Sigma}_\varepsilon^{-1} (y - F(x_i)) - \boldsymbol{\Sigma}_0^{-1} (x_i - x_0) \right) \quad (2.24)$$

The Levenberg-Marquardt algorithm is a modification of the Gauss-Newton method with an additional term $\gamma \mathbf{D}$, where $\gamma \in \mathbb{R}$ is a damping parameter and $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a scaling matrix. The value of γ is initialized to be large, resulting in a gradient descent update. If an iteration results in an increased value of the cost function, i.e. $-2 \log \pi(x_{i+1}|y) > -2 \log \pi(x_i|y)$, the value of the damping parameter is increased. This shortens the step length. Otherwise the value is decreased, resulting in a Gauss-Newton update closer to the solution, which accelerates convergence (Marquardt, 1963). Choosing $\mathbf{D} = \boldsymbol{\Sigma}_0^{-1}$ (see Rodgers (2000), Crisp et al. (2019)), we can write the iteration as

$$x_{i+1} = x_i + [(1 + \gamma) \boldsymbol{\Sigma}_0^{-1} + \mathbf{J}_{x_i}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{J}_{x_i}]^{-1} (\mathbf{J}_{x_i}^T \boldsymbol{\Sigma}_\varepsilon^{-1} (y - F(x_i)) - \boldsymbol{\Sigma}_0^{-1} (x_i - x_0)). \quad (2.25)$$

The starting value of the optimization is usually set to the prior mean x_0 . For a more comprehensive derivation, the reader is referred to e.g. Bardsley (2018).

2.5 MARKOV CHAIN MONTE CARLO

The posterior distribution contains all the information about the parameter of interest, and ideally one would want to report the entire distribution. Obtaining a large sample from the posterior and graphing pairs of parameters against each other is an excellent tool for examining the shape of an unknown posterior. The needed samples can be obtained in several ways. Some widely used methods are for example importance sampling (see e.g. Ripley (1987)), forward-backward algorithm (Scott, 2002) and sequential Monte Carlo method (Liu and Chen, 1998). In this work we employ the *Markov chain Monte Carlo* (MCMC) method for its relatively easy computational implementation and a variety of existing theoretical results making it mathematically convenient (Gelman et al., 1995).

The idea of MCMC is to simulate a random walk process $X_1, X_2, \dots, X_M \in \mathbb{R}^n$ called a *Markov chain*, which converges to the posterior distribution as the *stationary distribution* of the chain. The Markov Chain needs to be *aperiodic and irreducible* for the stationary distribution to exist and be unique. We will briefly describe this property before moving on to computational implementations (following Gamerman (1997)).

Generating a Markov chain is based on the *Markov property*. This means that every member of the sample sequence is obtained with a conditional probability that depends only on the previous member of the sequence:

$$P(x_{k+1}|x_1, \dots, x_k) = P(x_{k+1}|x_k). \quad (2.26)$$

If the conditional probability above is independent of the time k , the Markov chain is called *homogeneous*, in which case the evolution is governed by the *Markov transition kernel*

$$T(x_1, x_2) = P(x_1|x_2). \quad (2.27)$$

If for any $x_1, x_2 \in \mathbb{R}^n$ there exist an integer m such that $T^m(x_1, x_2) > 0$, the chain is called *irreducible*. This means that any states are reachable from all other states within finite time. Furthermore, a chain is called *aperiodic* if no states are visited at fixed intervals.

2.5.1 METROPOLIS-HASTINGS ALGORITHM

Several possible ways of implementing MCMC exist, including the *Gibbs sampler* (e.g. Gelman et al. (1995)) that samples one parameter dimension at the time, or the *Hamiltonian Monte Carlo* (S.Brooks et al., 2011) which uses Hamiltonian dynamics to produce posterior samples. In our applications, many of these otherwise efficient methods are less feasible since we usually don't know anything about the posterior shape beforehand. Furthermore, evaluating the marginal likelihood $\pi(y)$ in the denominator of the Bayes formula (2.4) is computationally challenging due to large data vectors y . In light of this reasoning, in this work we use *Metropolis-Hastings* (MH) algorithm (Metropolis et al. (1953), Hastings (1970)) which is the simplest MCMC algorithm for our unknown posterior. Additionally, with MH we can avoid evaluating the marginal likelihood altogether. There exists a growing amount of literature in applying MCMC to atmospheric remote sensing (Tamminen and Kyrölä (2001), Tamminen (2004), Haario et al. (1999), Haario et al. (2004), Laine and Tamminen (2008), Tukiainen et al. (2016), Hobbs et al. (2017), Brynjarsdottir et al. (2018)) We will next give a short description of the MH algorithm.

Let $X \in \mathbb{R}^n$ be a random variable with probability density function $\pi(x)$ as before. Define the *proposal distribution* $Q(x, \cdot)$ and its density function $q(x_1, x_2)$ describing the probability of moving from state $x_1 \in \mathbb{R}^n$ to state $x_2 \in \mathbb{R}^n$. Assume that at a given moment t the chain is in state $x_t \in \mathbb{R}^n$. The following state x_{t+1} is obtained with following steps:

1. Pick a random candidate z from the distribution $Q(x_t, \cdot)$.
2. The candidate z is either accepted or rejected. The *acceptance probability* is given by:

$$\alpha(x_t, z) = \begin{cases} \min\left(\frac{\pi(z)q(z, x_t)}{\pi(x_t)q(x_t, z)}, 1\right), & \text{if } \pi(x_t)q(x_t, z) > 0 \\ 1, & \text{if } \pi(x_t)q(x_t, z) = 0. \end{cases} \quad (2.28)$$

3. If the candidate z is accepted, set $x_{t+1} = z$, otherwise $x_{t+1} = x_t$.

Using a symmetric proposal distribution, with $q(x, z) = q(z, x)$, the previous rule simplifies to

$$\alpha(x, z) = \min\left(\frac{\pi(z)}{\pi(x)}, 1\right). \quad (2.29)$$

In this case the algorithm is simply called the *Metropolis algorithm*. In practice, it is often convenient to use a Gaussian proposal distribution (which is symmetric), so that the new proposal is given by $x_{t+1} \sim \mathcal{N}(x_t, \mathbf{C})$, where $\mathbf{C} \in \mathbb{R}^{n \times n}$ is the *proposal covariance matrix*.

When the target distribution is the posterior $\pi(x|y)$, the normalization constant in Bayes' formula (2.4) is cancelled out when computing the division $\pi(z|y)/\pi(x|y)$, leading to a more computationally efficient algorithm.

2.5.2 ADAPTIVE METROPOLIS ALGORITHM

A key element in the successful convergence of an MCMC algorithm is the choice of proposal distribution. In implementing MH, this is usually done by trial and error, which can be difficult in the case of an unknown shaped posterior distribution. A remedy for this problem is to use the *Adaptive Metropolis algorithm* (Haario et al., 2001), where the proposal covariance matrix is tuned by empirically calculating the covariance from the already obtained chain:

$$\mathbf{C}_t = \begin{cases} \mathbf{C}_0, & t < t_0, \\ s_d \text{Cov}(x_0, \dots, x_{t-1}) + \epsilon \mathbb{I}_n, & t \geq t_0 \end{cases} \quad (2.30)$$

where t_0 is the number of time steps the chain is run before the first covariance adaptation, $\mathbf{C}_0 \in \mathbb{R}^{n \times n}$ is some appropriate starting covariance for parameter of interest $x \in \mathbb{R}^n$, $\epsilon \in \mathbb{R}^+$ is a constant we can choose to be very small, and $s_d = (2.4)^2/n$ is a scaling parameter that optimizes the sampling efficiency in close-to-Gaussian cases (from Gelman et al. (1996)). The training time $t_0 > 0$ is free of choice and reflects our belief in the initial choice of covariance \mathbf{C}_0 .

The empirical covariance computed from points $x_0, \dots, x_k \in \mathbb{R}^n$ is given by

$$\text{Cov}(x_0, \dots, x_k) = \frac{1}{k} \left(\sum_{i=0}^k x_i x_i^T - (k+1) \bar{x}_k \bar{x}_k^T \right), \quad (2.31)$$

where $\bar{x}_k = (k+1)^{-1} \sum_{i=0}^k x_i$ and the elements x_i are considered as column vectors. Using definition 2.30 with $t > t_0 + 1$, we get a recursive formula for the covariance matrix

$$\mathbf{C}_{t+1} = \frac{t-1}{t} \mathbf{C}_t + \frac{s_d}{n} \left(t \bar{x}_{t-1} \bar{x}_{t-1}^T - (t+1) \bar{x}_t \bar{x}_t^T + x_t x_t^T + \epsilon \mathbb{I}_n \right). \quad (2.32)$$

This way, the covariance can be updated recursively without having to perform the computation using the entire history of the MCMC chain.

3 DIMENSION REDUCTION

While MCMC offers a robust way of exploring the shape of a possibly non-Gaussian posterior distribution, the method is prone to the curse of dimensionality (Roberts et al., 1997). This means that the algorithm converges increasingly slower as the dimension of the state vector x increases. In this work we solve this problem by considering the inherent low *information content* of the measurements in atmospheric remote sensing (Rodgers, 2000) and exploiting it to reduce the dimension of the state vector.

An atmospheric remote sensing measurement of absorption spectra y often has all the information constrained in a subspace that is low-dimensional compared to the size of the state vector x . This can be seen for example in the earlier Figure 2.4) where the retrieved state is seen to be mostly sensitive to changes on the measured absorption peaks. The result is mainly influenced by the data on altitudes up to 50km from the total of 70 km grid. The measurements are also corrupted by a significant amount of noise originating from the measurement device.

We next explore the information content of measurements in general, based on the discussion by Rodgers (2000) and the sources therein. It is shown that there are fewer independent *informative directions* in the state than actual state vector dimensions. We then present two different dimension reduction methods guided by this idea of low information content; one based on truncating the prior covariance matrix (Marzouk and Najm (2009), Solonen et al. (2016)) and a second based on isolating the informative directions of the likelihood (Cui et al. (2014), Spantini et al. (2015)). These ideas are then applied to a ground-based and a satellite-based retrievals together with adaptive MCMC. An illustration of the effect of dimension reduction to MCMC performed on the FTS problem is given in Figure 3.1. With the same amount of samples, LIS AM is seen to converge to correct posterior significantly faster than regular AM. In this case regular MH algorithm is producing barely any accepted states. It is worth noticing that AM is approaching the correct posterior distribution, but with the given amount of samples in this case the LIS method yields

samples from the stationary distribution of the Markov Chain faster.

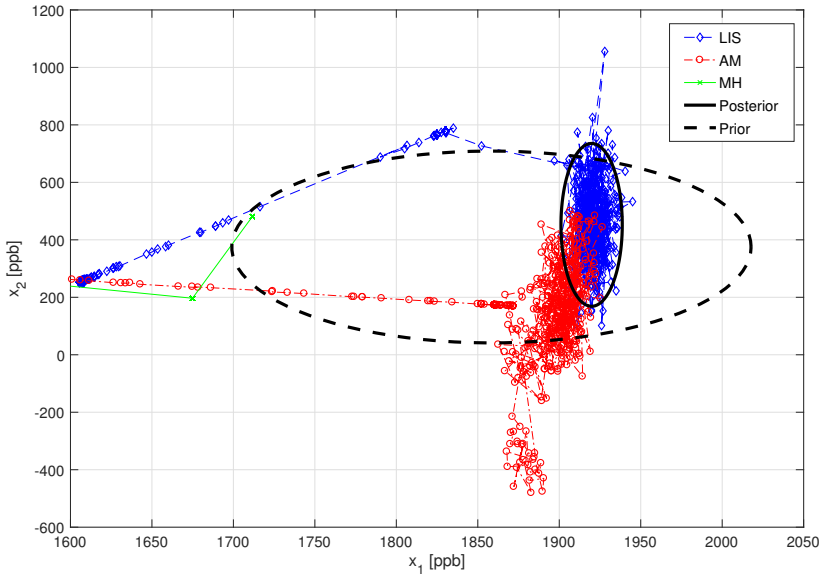


Figure 3.1: Different MCMC methods applied to the FTS retrieval. Two dimensions of the 100-dimensional CH_4 profile are given here in ppb. Shown is every 100th of the first 1 000 000 samples of the MCMC chain that was started purposefully from a bad starting value for each method. The Adaptive Metropolis algorithm with LIS-dimension reduction is shown on blue, the full-dimensional AM on red, and a full-dimensional Metropolis-Hastings MCMC on green. Dashed black line is the 95% prior probability region, and the solid black line is the approximate 95% posterior probability region obtained via MAP-estimation

3.1 INFORMATION CONTENT

In order to quantify how much information can be obtained from a measurement, we follow Rodgers (2000) and consider the *informative directions* in the posterior distribution. This quantification uses a version of the inverse problem in equation (2.2) with a linearized forward model, and Gaussian prior and likelihood, given by

$$y = \mathbf{J}_x(x - x_0) + \varepsilon, \quad (3.1)$$

where $x_0 \in \mathbb{R}^n$ is the prior mean, and $\mathbf{J}_x \in \mathbb{R}^{m \times n}$ denotes the Jacobian matrix of the forward model evaluated at $x \in \mathbb{R}^n$ with elements $\mathbf{J}_{x,ij} = \frac{\partial}{\partial x_j} F_i(x)$. We will occasionally drop the subscript for brevity. This approximation works best with posteriors that resemble a Gaussian distribution around the posterior mean.

The prior uncertainty of y is given by the *prior predictive covariance*

$$\boldsymbol{\Sigma}_y = \mathbb{E}[(\mathbf{J}_x(x - x_0) + \boldsymbol{\varepsilon})(\mathbf{J}_x(x - x_0) + \boldsymbol{\varepsilon})^T] = \mathbf{J}\mathbf{J}^T + \boldsymbol{\Sigma}_\varepsilon, \quad (3.2)$$

where $\mathbf{J} = \mathbf{J}_x(x - x_0)$. We will use this notation for the rest of this discussion for brevity. Next, we perform *diagonalization* and *pre-whitening* for y . The resulting prior predictive covariance is diagonal, and the resulting prior and noise are zero-mean Gaussian with unit covariance $\mathcal{N}(0, \mathbb{I})$. We use the Cholesky factorizations (A.1) of prior and error covariances,

$$\boldsymbol{\Sigma}_0 = \mathcal{L}_0 \mathcal{L}_0^T, \quad \boldsymbol{\Sigma}_\varepsilon = \mathcal{L}_\varepsilon \mathcal{L}_\varepsilon^T, \quad (3.3)$$

and the singular value decomposition (A.2) of the whitened Jacobian $\tilde{\mathbf{J}} \in \mathbb{R}^{m \times n}$, given by

$$\tilde{\mathbf{J}} = \mathcal{L}_\varepsilon^{-1} \mathbf{J} \mathcal{L}_0 = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T. \quad (3.4)$$

Define

$$\tilde{y} = \mathbf{U}^T \mathcal{L}_\varepsilon^{-1} y, \quad \tilde{x} = \mathbf{V}^T \mathcal{L}_0^{-1} (x - x_0) \text{ and } \tilde{\boldsymbol{\varepsilon}} = \mathbf{U}^T \mathcal{L}_\varepsilon^{-1} \boldsymbol{\varepsilon}, \quad (3.5)$$

where $\tilde{x} \sim \mathcal{N}(0, \mathbb{I}_n)$ and $\tilde{\boldsymbol{\varepsilon}} \sim \mathcal{N}(0, \mathbb{I}_m)$. Together with (3.1) and (3.4), we get

$$\begin{aligned} \tilde{y} = \mathbf{U}^T \mathcal{L}_\varepsilon^{-1} y &= \mathbf{U}^T \mathcal{L}_\varepsilon^{-1} \mathbf{J} \mathcal{L}_0 \mathcal{L}_0^{-1} (x - x_0) + \mathbf{U}^T \mathcal{L}_\varepsilon^{-1} \boldsymbol{\varepsilon} \\ &= \mathbf{U}^T (\mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T) \mathcal{L}_0^{-1} (x - x_0) + \mathbf{U}^T \mathcal{L}_\varepsilon^{-1} \boldsymbol{\varepsilon} \\ &= \boldsymbol{\Lambda} \tilde{x} + \tilde{\boldsymbol{\varepsilon}}. \end{aligned} \quad (3.6)$$

We can now derive expressions for the prior predictive covariance for whitened measurements \tilde{y} , as well as the posterior covariance of whitened state \tilde{x} :

$$\boldsymbol{\Sigma}_{\tilde{y}} = \mathbb{E}[\tilde{y} \tilde{y}^T] = \mathbb{E}[(\boldsymbol{\Lambda} \tilde{x} + \tilde{\boldsymbol{\varepsilon}})(\boldsymbol{\Lambda} \tilde{x} + \tilde{\boldsymbol{\varepsilon}})^T] = \boldsymbol{\Lambda}_m^2 + \mathbb{I}_m. \quad (3.7)$$

Referring to formula (A.3), $\boldsymbol{\Lambda}_m^2 \in \mathbb{R}^{m \times m}$ are the eigenvalues of matrix $\tilde{\mathbf{J}} \tilde{\mathbf{J}}^T$ corresponding to the eigenvectors $u_i \in \mathbb{R}^m$. The prior predictive covariance $\boldsymbol{\Sigma}_{\tilde{y}}$ (3.7) can be interpreted to show the r components of the variance in the data that can be distinguished from whitened noise, corresponding to eigenvalues $\lambda_i^2 > 1$, where $i = 1, \dots, r$.

The same treatment can be repeated for the whitened state \tilde{x} . Because the forward model is linear, we can use the formula (2.8) for $\Sigma_{post} = (F^T \Sigma_\epsilon F + \Sigma_0^{-1})^{-1}$ and write the posterior covariance as

$$\Sigma_{post}(\tilde{x}) = (\mathbf{V}^T \tilde{\mathbf{J}}^T \tilde{\mathbf{J}} \mathbf{V} + \mathbb{I}_n)^{-1} = (\Lambda_n^2 + \mathbb{I}_n)^{-1}, \quad (3.8)$$

We recall from formulas (A.3) and (A.4) that the non-zero eigenvalues of the matrix $\tilde{\mathbf{J}}^T \tilde{\mathbf{J}}$ in Λ_n^2 are the same as the non-zero eigenvalues of Λ_m^2 with $\lambda_i^2 > 1$, $i = 1, \dots, r$. The informative directions of the posterior distribution are thus the eigenvectors $v_i \in \mathbb{R}^n$, $i = 1, \dots, r$ of matrix $\tilde{\mathbf{J}}^T \tilde{\mathbf{J}}$. A two dimensional Gaussian example with one only one informative direction is presented in Figure 3.2. In this case the likelihood is effectively one dimensional and has only one informative direction x_r , so all the variability of the posterior in the non-informative direction x_\perp is resulting from the prior distribution.

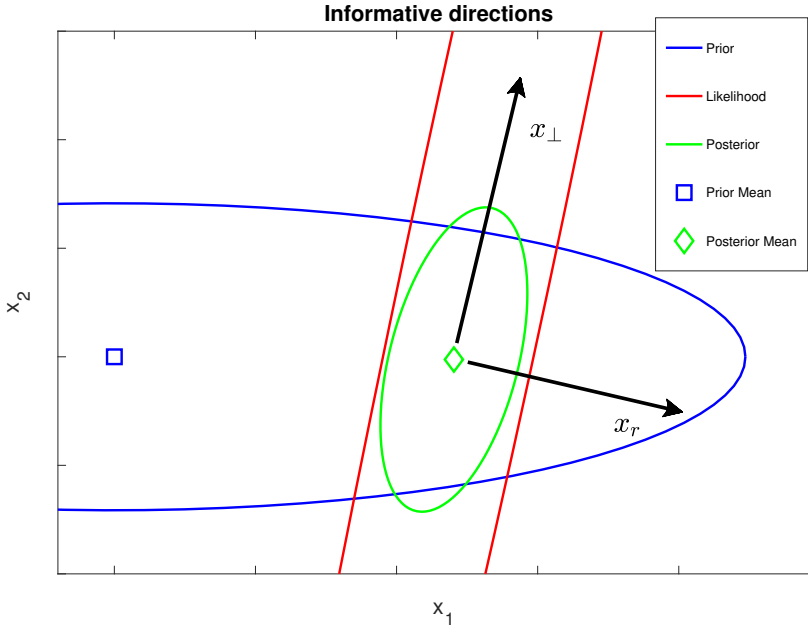


Figure 3.2: Illustration of an informative direction x_r and a non-informative direction x_\perp using a 2-dimensional Gaussian case. A Gaussian prior is shown in blue, likelihood in red and the resulting posterior in green. Here, the likelihood has only one informative component, so the remaining direction for the posterior is obtained from the prior.

3.2 PRIOR REDUCTION

Let us consider a dimension reduction method utilized by Marzouk and Najm (2009), Solonen et al. (2016) and Tukiainen et al. (2016). The idea is to construct a prior covariance matrix for the inverse problem that has a number of non-zero eigenvalues equal to the number of informative directions of the corresponding posterior distribution. We will then use the truncated *Karhunen-Loève* (K-L) decomposition (Loève, 1978) to obtain a reduced dimensional version of the problem.

The Karhunen-Loève theorem (Loève, 1978) states that for a square-integrable zero-mean random process X_t defined over a probability space $(\Omega, \mathcal{F}, \mathfrak{P})$ that is indexed over a closed and bounded interval $[a, b]$ and that has a continuous covariance function $\mathbf{C}(s, t), s, t \in [a, b]$, there exist a representation called Karhunen-Loève decomposition as follows:

$$X_t = \sum_{i=1}^{\infty} Z_i e_i(t) \quad (3.9)$$

where $Z_i = \int_a^b X_t e_i(t) dt$ are uncorrelated, zero-mean and have variance λ_i , and λ_i, e_i are the eigenvalues and functions of $\mathbf{C}(s, t)$, which means $\int_a^b \mathbf{C}(s, t) e_i(s) ds = \lambda_i e_i(t)$.

Consider the inverse problem (2.2) with $x \in \mathbb{R}^n$ and a Gaussian prior $\mathcal{N}(x_0, \Sigma_0)$, $\Sigma_0 \in \mathbb{R}^{n \times n}$. We use the eigenvalue decomposition (A.4) of the prior covariance matrix

$$\Sigma_0 = \mathbf{U} \Lambda_n^2 \mathbf{U}^T = \sum_{i=1}^n \lambda_i^2 u_i u_i^T, \quad (3.10)$$

to obtain the K-L decomposition (Solonen et al. (2016), Tukiainen et al. (2016)) of the state vector x as

$$x - x_0 = \mathbf{P} \alpha, \quad \mathbf{P} = \mathbf{U} \Lambda_n = [\lambda_1 u_1, \lambda_2 u_2, \dots, \lambda_n u_n] \in \mathbb{R}^{n \times n} \quad (3.11)$$

which satisfies the requirements of the Karhunen-Loève theorem: $x - x_0$ is zero-mean, $\alpha \in \mathbb{R}^n, \alpha \sim \mathcal{N}(0, \mathbb{I}_n)$ and each $\lambda_i u_i \alpha_i$ has a variance λ_i^2 (Solonen et al., 2016).

If the prior covariance matrix Σ_0^{-1} has numerous zero or nearly-zero eigenvalues, the K-L decomposition can be truncated to only include the first significant eigenvalues and the corresponding eigenvectors, resulting in an optimal low-dimensional approximation (Marzouk and Najm, 2009).

The state vector x now has an approximate representation by r basis vectors from the first columns of \mathbf{P} , denoted $\mathbf{P}_r \in \mathbb{R}^{n \times r}$. Using a reduced dimensional parameter $\alpha_r \in \mathbb{R}^r$ we can write it as

$$x \approx x_0 + \mathbf{P}_r \alpha_r. \quad (3.12)$$

The random vector α_r has a Gaussian prior $\alpha_r \sim \mathcal{N}(0, \mathbb{I})$, which allow us to write an approximate posterior distribution as

$$\pi(x|y) \approx \tilde{\pi}(\alpha_r|y) \propto \exp\left(-\frac{1}{2}\left((y - F(x_0 + \mathbf{P}_r\alpha_r))^T \boldsymbol{\Sigma}_\varepsilon^{-1}(y - F(x_0 + \mathbf{P}_r\alpha_r)) + \alpha_r^T \alpha_r\right)\right). \quad (3.13)$$

Now, instead running MCMC in full dimensional space, we can sample the low dimensional parameter α_r and retain the approximation of the full posterior by equation (3.12).

3.3 LIKELIHOOD-INFORMED SUBSPACE

The previous discussion on informative directions offers a direct means for reducing the dimension of the state vector. The dimension reduction method uses the *likelihood-informed subspace* based on considering only the first r right singular vectors of the whitened Jacobian (3.4). A growing amount of literature has already demonstrated the effectiveness of this approach (Spantini et al. (2015), Cui et al. (2014), Cui et al. (2016a), Cui et al. (2016b), Beskos et al. (2018), Zahm et al. (2018)), although no previous application to real remote sensing data has been implemented. This method is also similar to the widely used *active subspace* dimension reduction (Constantine et al. (2014), Constantine et al. (2016)).

Consider the low-rank approximations for the posterior of the form

$$\tilde{\pi}(x|y) \propto \pi(y|\mathbf{P}_r x) \pi_0(x), \quad (3.14)$$

where $\mathbf{P}_r \in \mathbb{R}^{r \times n}$ is rank r projection matrix. It is shown by Spantini et al. (2015) that when forward model F is linear, the optimal \mathbf{P}_r that minimizes the Hellinger distance (A.16) between a class of all rank r approximations and the full posterior can be obtained from the eigenvalue decomposition of prewhitened *Hessian* matrix of the negative log-likelihood

$$-\log\pi(y|x) = \frac{1}{2} \left((y - F(x))^T \boldsymbol{\Sigma}_\varepsilon^{-1} (y - F(x)) \right). \quad (3.15)$$

The Hessian is given by

$$\mathbf{H}(x) = -\nabla^2 \log\pi(y|x) = \mathbf{J}(x)^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{J}(x) + \boldsymbol{\Sigma}_\varepsilon^{-1} (y - F(x)) \nabla^2 F(x), \quad (3.16)$$

where $\mathbf{J}(x) \in \mathbb{R}^{m \times n}$ is the Jacobian of the forward model F . When the forward model is linear, the second term of the sum equals zero due to the second derivative, and it is justified to consider the *Gauss-Newton approximation* of the Hessian:

$$\mathbf{H}(x) = \mathbf{J}(x)^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{J}(x). \quad (3.17)$$

With same reasoning as in section 3.1, we perform prewhitening on the Gauss-Newton Hessian

$$\tilde{\mathbf{H}} := \mathcal{L}_0^T \mathbf{H} \mathcal{L}_0 = \mathcal{L}_0^T \mathbf{J}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{J} \mathcal{L}_0 = \tilde{\mathbf{J}}^T \tilde{\mathbf{J}} \quad (3.18)$$

and recall that only the components of the posterior corresponding to singular values of scaled Jacobian $\tilde{\mathbf{J}}$ that are greater than 1 can be obtained from the measurement, and the rest are given by the prior.

Informative directions of the measurement can now be identified with the first r eigenvectors of matrix $\tilde{\mathbf{H}}$ corresponding to the first r eigenvalues greater than 1. These are the first r columns of matrix \mathbf{V} given by the eigenvalue decomposition

$$\tilde{\mathbf{H}} = \mathbf{V} \boldsymbol{\Lambda}^2 \mathbf{V}^T. \quad (3.19)$$

Given $\mathbf{V}_r \in \mathbb{R}^{n \times r}$ containing the first r eigenvectors of $\tilde{\mathbf{H}}$, we can define

$$\Phi_r := \mathcal{L}_0 \mathbf{V}_r \text{ and } \Theta_r := \mathcal{L}_0^{-T} \mathbf{V}_r \quad (3.20)$$

where the superscript $^{-T}$ denotes the transpose of the inverse.

The rank r LIS projection for the low-rank posterior approximation in eq. (3.14) is given by

$$\mathbf{P}_r = \Phi_r \Theta_r^T. \quad (3.21)$$

The range $\mathbb{X}_r \subseteq \mathbb{R}^n$ of projection $\mathbf{P}_r : \mathbb{R}^n \rightarrow \mathbb{X}_r$ is a subspace of state space \mathbb{R}^n spanned by the column vectors of matrix Φ_r . We call the subspace \mathbb{X}_r the *likelihood-informed subspace (LIS)*, and its complement $\mathbb{R}^n \setminus \mathbb{X}_r$ the *complement subspace (CS)* (Spantini et al., 2015).

The matrix of singular vectors $\mathbf{V} = [\mathbf{V}_r \mathbf{V}_\perp]$, where $\mathbf{V}_\perp \in \mathbb{R}^{n \times n-r}$ contains the rest of the eigenvectors not included in \mathbf{V}_r , is an orthonormal basis for \mathbb{R}^n and we can define

$$\Phi_\perp := \mathcal{L}_0 \mathbf{V}_\perp \text{ and } \Theta_\perp := \mathcal{L}_0^{-T} \mathbf{V}_\perp \quad (3.22)$$

and the projection $\mathbb{I} - \mathbf{P}_r$ can be written as

$$\mathbb{I} - \mathbf{P}_r = \Phi_\perp \Theta_\perp^T. \quad (3.23)$$

Define the LIS-parameter $x_r \in \mathbb{R}^r$ and the CS-parameter $x_\perp \in \mathbb{R}^{n-r}$ as

$$x_r := \Theta_r^T x, \quad x_\perp := \Theta_\perp^T x. \quad (3.24)$$

The parameter x can now be decomposed as

$$\begin{aligned} x &= \mathbf{P}_r x + (\mathbb{I} - \mathbf{P}_r) x \\ &= \Phi_r x_r + \Phi_\perp x_\perp. \end{aligned} \quad (3.25)$$

Using the previous definitions, we can decompose the posterior distribution as

$$\begin{aligned}\pi(x|y) &= \pi(y|x)\pi_0(x) \\ &= \pi(y|\Phi_r x_r + \Phi_\perp x_\perp)\pi_0(\Phi_r x_r + \Phi_\perp x_\perp) \\ &= \pi(y|\Phi_r x_r)\pi(y|\Phi_\perp x_\perp)\pi_0(x_r)\pi_0(x_\perp)\end{aligned}\quad (3.26)$$

As was shown before, we know that the $\pi(y|\Phi_\perp x_\perp)$ part of the likelihood is non-informative as all the corresponding eigenvalues are less than 1. We can therefore omit this part and get the low-rank approximation of the posterior distribution as

$$\tilde{\pi}(x|y) = \pi(y|\Phi_r x_r)\pi_0(x_r)\pi_0(x_\perp). \quad (3.27)$$

Since the parameters x_r and x_\perp are independent, the inverse problem can now be solved in two parts. First, using e.g. optimization or MCMC, the desired x_r are obtained from $\pi(y|\Phi_r x_r)\pi_0(x_r)$ and projected into full space as

$$\Phi_r x_r + x_0 \in \mathbb{R}^n. \quad (3.28)$$

Second, the remaining CS part can be analytically obtained from the complement prior

$$x_\perp \sim \mathcal{N}(0, \mathbb{I}), x_\perp \in \mathbb{R}^{n-r}. \quad (3.29)$$

The sample is then projected back to full space as

$$\Phi_\perp x_\perp \in \mathbb{R}^n \quad (3.30)$$

and added to the LIS-part, yielding a solution from the optimal rank r posterior approximation $\tilde{\pi}(x|y)$.

When the forward model is non-linear, the Hessian matrix depends on the point x and is not constant on the state space. Furthermore, the Gauss-Newton approximation of the Hessian matrix is no longer accurate since the second order derivative of F on the second term of expression (3.16) is no longer zero. A non-linear extension of the LIS decomposition is provided by Cui et al. (2014): Consider the expectation of the Gauss-Newton Hessian over the posterior distribution, given as the integral

$$\hat{\mathbf{H}} = \mathcal{L}_0^T \left(\int_{\mathbb{R}^n} \mathbf{H}(x)\pi(x|y)dx \right) \mathcal{L}_0. \quad (3.31)$$

This integral is computed as sample mean

$$\hat{\mathbf{H}}_N = \mathcal{L}_0^T \left(\frac{1}{N} \sum_{k=1}^N \mathbf{H}(x^{(k)}) \right) \mathcal{L}_0, \quad (3.32)$$

where the N samples $x^{(k)}$ should ideally be drawn from the posterior. However, the goal of this construction is to avoid evaluating the full posterior. The reference distribution used in the articles in this work was taken to be the Laplace approximation $\mathcal{N}(\hat{x}, \hat{\Sigma}_{post})$, where $\hat{x} \in \mathbb{R}^n$, $\hat{\Sigma}_{post} \in \mathbb{R}^{n \times n}$ are the posterior mean and covariance obtained by solving the full-dimensional optimization problem (2.15). The eigenvalue decomposition $\hat{\mathbf{H}}_N = \hat{\mathbf{V}}\hat{\Lambda}^2\hat{\mathbf{V}}^T$ is then used to find a basis for the non-linear LIS analogously to the linear case.

4 UNCERTAINTY QUANTIFICATION IN REMOTE SENSING OF GREENHOUSE GASES

4.1 GROUND BASED CH₄ RETRIEVAL

Ground based measurements of concentration profiles of atmospheric CH₄ are important for studying the altitude-dependent dynamics of CH₄ as well for validating profile measurements by satellites. These tasks require high precision and reliable error estimates, which are achieved by rigorous uncertainty quantification via Bayesian approach. The measurements we consider in **Paper I** and **Paper II** are carried out in Arctic conditions at Finnish Meteorological Institute's Arctic Research Center in Sodankylä, Northern Finland using a *Fourier Transform Spectrometer* (FTS) (Kivi and Heikkinen, 2016).

In order to fully quantify the retrieval uncertainty while taking into account the effects of a non-linear forward model, we perform our retrievals with Markov chain Monte Carlo. We use Likelihood-informed subspace (LIS) dimension reduction to achieve a computationally fast algorithm that also offers flexibility in selection of prior distribution. Our retrieval is based on the *Swirlab* retrieval software (Tukiainen et al., 2016) developed at FMI. Prior to the work in **Paper I**, MCMC with LIS dimension reduction had been implemented by Cui et al. (2014) to the GOMOS satellite's retrieval problem (retrieval described by Tamminen and Kyrölä (2001)). This implementation was done using simulated data as a ground truth, but LIS retrievals with real data had not been performed for atmospheric remote sensing inverse problems prior to our research.

While the long-term trends of atmospheric methane have been studied (e.g. Saunio et al. (2019)), the mechanisms governing the trend remain largely unknown. Since the major sink of CH₄ is in the upper troposphere, long-term changes in the vertical distribution can give insight on the underlying phenomena. In **Paper II** we utilize the dimension reduction retrieval framework to solve for all CH₄ concentration

profiles measured from 2009 to 2018 by the Sodankylä FTS. We then implement a *Dynamic linear model* (DLM) based analysis method to investigate the non-linear, altitude-dependent background level and trend in CH₄ timeseries.

For empirical prior construction and validation of our retrievals, we use two supporting instruments providing CH₄ profiles: a occultation measurement from the *Atmospheric Chemistry Experiment* (ACE, Bernath et al. (2005)) satellite, and an in-situ balloon measurement called *AirCore* (Karion et al. (2010), Chen et al. (2016)). Both are useful together, since the satellite measurements are unable to capture the concentrations close to Earth's surface due to measurement geometry, whereas AirCore can only fly up to 30 km altitude.

4.1.1 SODANKYLÄ FOURIER TRANSFORM SPECTROMETER

Fourier Transform Spectrometer (FTS, Figure 4.1) is an interferometer for measuring light attenuation as a function of wavelength. The device divides incoming solar light into two beams; one hitting a detector directly, and another reflected from a moving mirror back to the detector, which travels a longer distance as a result. These beams are then compared, producing an interferogram, that is, intensity as a function of difference in optical path length. The interferogram is finally Fourier transformed to obtain an absorption spectrum. Up to several hundreds of measurements per day can be recorded (depending on the season and cloudiness during measurement), although sunlight is still required to do so and hence the instrument is not operated during winter.

The Sodankylä FTS is a Bruker IFS 125 HR with a A547 solar tracker and three detectors: InGaAs (12,800-4,000 cm⁻¹), Si (25,000-9000 cm⁻¹) and InSb(10,000-1,850 cm⁻¹). This instrument is capable of a spectral resolution of about 0.02 cm⁻¹. It is part of the international Total Carbon Column Observation Network (TCCON) (Wunch et al., 2011) spectrometers, which delivers the results as column-averaged dry-air mole fractions, i.e. the gas columns are divided by the dry-air column coming from O₂ measurements. These column-averages (denoted XCO₂ for carbon dioxide, XCH₄ for methane etc.) are operationally processed using the GGG14 (Wunch et al., 2015) software, which solves the retrieval inverse problem by scaling the prior atmospheric density profiles of gases of interest. Column-averaged abundance of CO₂, CH₄, N₂O, HF, CO, H₂O, and HDO are retrieved from the measured spectra.

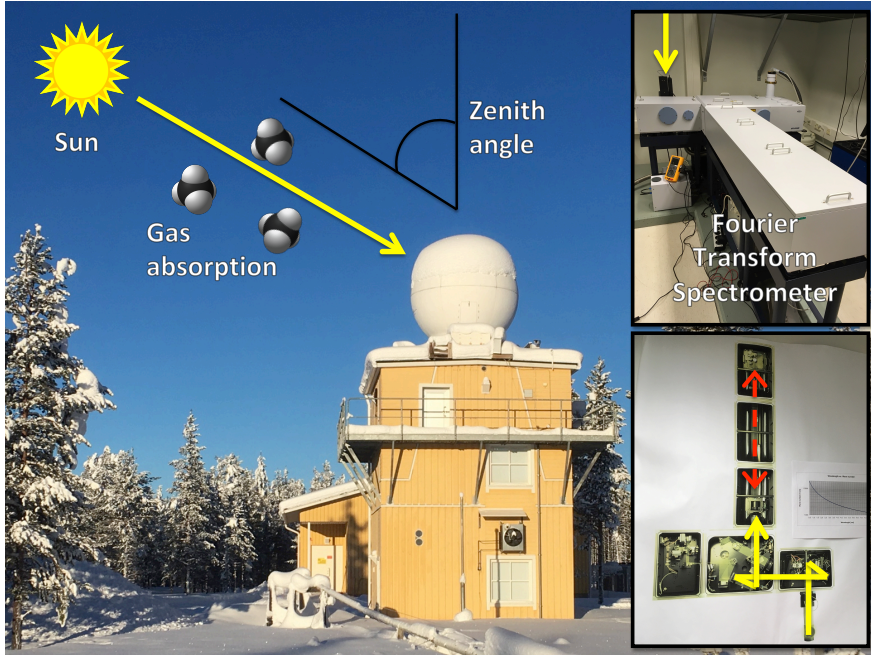


Figure 4.1: Sodankylä Fourier Transform Spectrometer measurement geometry. The solar light (yellow) enters the detector from a slit on top of the housing building. The device itself is shown in the upper right panel with the light entrance indicated by a yellow arrow. The lower left panel shows a schematic of the aperture inside the FTS, with light beam reflected along the yellow line to hit a mirror. The red dashed arrow shows the movement of the mirror, which is used to obtain different path lengths to form an interferogram. Photos by O.L.

4.1.2 IMPLEMENTATION

For comprehensive uncertainty quantification, we use the Bayes formula (2.4) to obtain the solution for the FTS retrieval problem as a posterior distribution for state $x \in \mathbb{R}^n$. The Bayesian solution is implemented as follows:

State vector: The state vector consists of CH_4 concentrations at $n = 100$ altitudes on a grid from 0 to 70km. The grid spacing is non-uniform to reflect the atmosphere getting thinner on higher altitudes.

Prior: We use a Gaussian prior $\mathcal{N}(x_0, \Sigma_0)$, with a prior mean $x_0 \in \mathbb{R}^n$ and

covariance matrix $\Sigma_0 \in \mathbb{R}^{n \times n}$ computed empirically as

$$\Sigma_0 = \frac{1}{N-1} \sum_{l=1}^N (x^{(l)} - x_0)(x^{(l)} - x_0)^T, \quad x_0 = \frac{1}{N} \sum_{l=1}^N x^{(l)}, \quad (4.1)$$

where $x^1, x^2, \dots, x^N \in \mathbb{R}^n$ are an ensemble of ACE satellite's CH₄ profiles interpolated to the 100-altitude retrieval grid. Since these profiles don't reach the ground level due to ACE's limited vertical coverage, a randomized extrapolation was performed to achieve a similar variability to the TCCON prior covariance in lower altitudes. This prior was chosen to reflect the best knowledge we have from vertical distribution of atmospheric CH₄ in contrast to previous work by Tukiainen et al. (2016), where the prior covariance was built to have a desired amount of non-zero eigenvalues by using a Gaussian process.

Additionally, it was observed that for time series studies, the empirical prior covariance matrix derived from ACE satellite measurements constrained the retrieval too much in altitudes from 1 to 10 km. Therefore, in **Paper II** we settled to using the prior reduction retrieval with the covariance matrix defined in (4.3), as it performed similarly to the LIS method when not using the empirical covariance matrix. The prior covariance Σ_0 is constructed so that it only has $r = 4$ non-zero eigenvalues, corresponding to the total number of informative directions of the posterior distribution. The diagonal elements of Σ_0 as

$$\sigma(h) = \sigma_1 \exp\left((h - h_1)^2 s_1^{-2}\right) + \left(\sigma_2 \exp(h - h_2)^2 s_2^{-2}\right), \quad (4.2)$$

where $h = 1, \dots, n$ and the parameters $\sigma_1, \sigma_2, h_1, h_2, s_1$ and s_2 are chosen appropriately to reflect the physical conditions of the retrieval (Tukiainen et al., 2016). The off-diagonal terms are further given by

$$\Sigma_0(i, j) = \sigma(i)\sigma(j) \exp\left(-\frac{1}{2} \left(\frac{d(i, j)}{l}\right)^2\right), \quad (4.3)$$

where i and j are two atmospheric layers, $d(i, j)$ is the distance of these layers in kilometers and l is an appropriately chosen correlation length.

In **Paper II**, the prior mean x_0 was estimated by combining the stratospheric part from the mean of an ensemble ACE satellite measurements and the tropospheric part from the TCCON prior profile from GGG14 algorithm. The ensemble of ACE measurements was obtained using co-location criteria of 3 degrees in latitude, 30 degrees in longitude from Sodankylä station, which is situated at 7.367 N, 26.629 E. Since ACE has been operational from 2003, we limited ourselves to using only the

data from 2009 onwards. The prior mean was calculated separately for three domains: summer (from May to August), Polar Vortex and the rest. The location inside the polar vortex was defined using ERA-Interim atmospheric model data. To assess the retrieval's sensitivity to the prior profile, we also performed the retrievals for the entire times series with a prior profiles without the additional trend, and with a constant profile shape. When compared with retrieved ACE profiles, the use of a seasonal prior mean resulted in a better agreement with ACE, while retrievals using a constant prior mean were still able to match the general characteristics in the ACE time series reasonably well.

Likelihood: We utilize likelihood derived from the assumed Gaussian distribution for the observation noise $\varepsilon \sim \mathcal{N}(0, \mathbf{\Sigma}_\varepsilon)$, $\varepsilon \in \mathbb{R}^m$, $\mathbf{\Sigma}_\varepsilon \in \mathbb{R}^{m \times m}$. This results in a likelihood proportional to $\exp\left(-\frac{1}{2}(y - F(x))^T \mathbf{\Sigma}_\varepsilon^{-1}(y - F(x))\right)$, where $y \in \mathbb{R}^m$ is the measured absorption spectrum and F is the forward model. In our case we have $m = 355$. The covariance matrix is taken to be diagonal: $\mathbf{\Sigma}_\varepsilon = c\mathbb{I}_m$, where $c \in \mathbb{R}$ is a multiplier initialized at $c = 0.0015$. This multiplier is updated by first performing an Optimal Estimation (2.4) retrieval to obtain \hat{x} , and then scaling c with the square root of the residual sum of squares $\text{RSS} = \sum_{i=1}^m (y_i - F(\hat{x}_i))^2$.

Forward model: Due to it's computational efficiency and proven functionality, we use the forward model from the Swirlab software by Tukiainen et al. (2016). This radiative transfer model consists of non-scattering atmosphere with n spherical layers from ground level up to 70 km. Absorption lines are calculated for each layer separately as Voigt profiles taking into account the temperature and pressure of the layer. Atmospheric model data from National Centers for Environmental Prediction (NCEP) are used as source for these atmospheric parameters. Using only the main beam geometry, that is, considering only the photons that are coming directly from the sun, the absorption for each wavelength λ is then calculated using the Beer-Lambert law, yielding the intensity of detected light for wavelengths $\lambda_j, j \in [1, \dots, m]$ as

$$I(\lambda_j) = I_0(\lambda_j) \exp\left(-\sum_{k=1}^K \int_0^\infty C_k(\lambda_j, z) \rho_k(z) dz\right) (a\lambda_j^2 + b\lambda_j + c) + d, \quad (4.4)$$

where I_0 is the intensity of solar light when it enters the atmosphere, the atmosphere has K absorbing trace gases, $C_k(\lambda_j, z)$ is the absorption coefficient of gas k , which depends on height z and on the wavelength λ_j , and $\rho_k(z)$ is the density of gas k at altitude z . The second degree polynomial and the constant d in (4.4) are used to describe instrument related features and the continuity properties of the spectrum. In reality, solar light is scattered on the way by atmospheric particles.

This phenomenon is relatively weak in the wavelength band we are considering in this work (see 2.2), so it is ignored for simplicity. The absorption in continuous atmosphere is modeled by discretizing the integral in equation (4.4) into a sum over atmospheric layers and assuming a constant absorption for each separate layer. As an update to description by Tukiainen et al. (2016), the absorption coefficients are calculated using HITRAN2016 line database (Gordon et al., 2017).

Dimension reduction: For enhancing the efficiency of MCMC computations, we use LIS dimension reduction to exploit the low intrinsic dimensionality of the retrieval problem. The corresponding Gauss-Newton Hessian (3.17) in this case generally has three singular values greater than unity, but we set $r = 4$ to make sure that all the information in the likelihood is used. The global basis (3.32) for non-linear LIS was computed by first solving for posterior estimates \hat{x} and Σ_{post} with Optimal Estimation (2.4) and then drawing samples from the Laplace approximation $\mathcal{N}(\hat{x}, \Sigma_{post})$ for the sample mean computation. The LIS basis is used to obtain the projection matrices $\Phi_r \in \mathbb{R}^{n \times r}$ and $\Phi_{\perp} \in \mathbb{R}^{n \times n-r}$ such that $x = \Phi_r x_r + \Phi_{\perp} x_{\perp}$ for the LIS parameter $x_r \in \mathbb{R}^r$ and the CS parameter $x_{\perp} \in \mathbb{R}^{n-r}$.

MCMC: For efficient sampling of the high-dimensional, possibly non-Gaussian posterior, we use Adaptive Metropolis algorithm (2.5.2) for the r -dimensional LIS parameter x_r . A starting value $x_1 = (0, 0, \dots, 0)^T \in \mathbb{R}^r$ is used, since the state has been centered at the prior mean by the formulation in Chapter 3. Due to pre-whitening in this formulation, we also use the r dimensional unit covariance matrix \mathbb{I}_r as the training covariance for AM. To ensure that all uncertainties are taken into account, we add the polynomial and offset parameters a, b, c, d from the Beer-Lambert equation (4.4) to the state vector with a white noise prior $\mathcal{N}(0, \mathbb{I})$ for each parameter (see Tukiainen et al. (2016) for details). As we want to compare the results of LIS dimension reduction to the full-dimensional retrieval, MCMC was run for a total of $N_M = 1000000$ timesteps in order to ensure the convergence of the full-dimensional chains. MCMC chains $x_r^{(l)}$, $l = 1, \dots, N_M$ are projected back to full-dimensional space according to

$$x^{(l)} = x_0 + \Phi_r x_r^{(l)} + \Phi_{\perp} x_{\perp}^{(l)}, \quad (4.5)$$

where $x_{\perp}^{(l)}$ is an ensemble of realizations drawn from the complement prior $\mathcal{N}(0, \mathbb{I}_{n-r})$. In order to compare the performance of MCMC methods we have additionally defined the *sample speed* of a MCMC run via the *effective sample size* N_{eff} as

$$N_{\text{eff}} = \frac{N_M}{1 + s \sum_{k=1}^{\infty} \rho_k(x)}, \quad (4.6)$$

where N_M is the length of the MCMC chain and $\rho_k(x)$ is lag- k autocorrelation for parameter x in the chain (Ripley, 1987). Then the sample speed of an MCMC chain is obtained as

$$\mathbb{V} = \frac{N_{\text{eff}}}{t_M}, \quad (4.7)$$

where t_M is the total computation time of the MCMC chain.

SZA Correction The FTS collects measurements throughout the day, and it is well documented (Wunch et al., 2011) that the measured concentrations have a solar zenith angle (SZA) dependence. We utilized a statistical correction to adjust all the measured profiles for SZA dependence to match the smallest SZA for each given day. After this, a daily average was calculated and used as the timeseries for fitting the DLM model. The fit was performed on a layer-by-layer basis, and prior assumptions on the beginning of each fit were adjusted by hand based on knowledge from adjacent atmospheric layers.

DLM: The time series model for Sodankylä FTS time series was of the form

$$y_t = \mu_t + \gamma_t + \epsilon_t \quad (4.8)$$

where y_t are the retrieved methane concentrations, μ_t is a second order random walk component that describes the slowly varying trend in the data, γ_t is a seasonal component, and the Gaussian error term ϵ_t is allowed to have AR(1) autoregressive temporal correlation. The fitting was done using the DLM approach (see Kivimäki et al. (2019) and Laine (2019) for details) which utilizes Kalman filter and smoother formulas to estimate the time series components in (4.8). The structural parameters of the model that constrain the time-varying aspects were estimated by MCMC.

4.1.3 OVERVIEW OF RESULTS

We performed the MCMC retrieval on real FTS data using LIS dimension reduction, as well the prior reduction and full dimensional MCMC retrievals implemented in the Swirlab software. We used the empirical prior for all retrievals, and the comparison of the retrieved posterior distributions is shown in Figure 4.2 together with an AirCore sounding performed at the same time as the FTS measurement. The AirCore profile is seen to be better contained within the 95% posterior confidence interval with full dimensional and LIS retrievals, in comparison to prior reduction that performs worse. We can also observe that the LIS method agrees remarkably well with the full dimensional MCMC up to the altitude of 35 km, whereas the retrieval performed with prior reduction did not produce a similar posterior. This disagreement can likely

be attributed to the usage of an empirical covariance matrix, which demonstrates the flexibility and robustness of the LIS method.

The differences of prior reduction and LIS based dimension reduction approaches to the MCMC retrieval were further explored by measuring the distance of the approximate posterior to the full dimensional one using the Hellinger distance (A.16). In short, for two probability distributions that give a non-zero probability for exactly the same events, their Hellinger distance is 0, and if the distributions don't align at all, the Hellinger between them is 1. In Figure 4.3, the upper-left panel shows the Hellinger distances of the dimension-reduced MCMC posterior distributions from prior reduction and LIS to the full-dimensional posterior. We performed the comparison with different selections of the number r of singular vectors used in the retrieval. As seen in the graph, LIS method yields a Hellinger distance close to 0 with just 4 basis vectors used, while prior reduction performs poorly in comparison. This can be partially explained by the shapes of the basis vectors shown in the upper-right and lower-right panels of Figure A.16, where it can be seen that the LIS basis vectors more closely resemble the shapes of the concentration profiles shown in Figure 4.2).

The lower-left panel of Figure 4.3 shows the corresponding sample speeds of the MCMC retrievals as a function of basis vectors used. The two methods perform equally well, but as the LIS method provides a more accurate posterior approximation it is to be preferred while using an empirical prior. However, using the prior covariance derived previously by Tukiainen et al. (2016), these methods performed equally well. For this reason, prior reduction with was chosen as a basis for our approach for time series analysis.

In **Paper II**, we implemented the dimension reduction based CH_4 concentration profile retrieval for the entire time series of FTS measurements from 2009 to 2018, which was processed without significant algorithmic failures. While time series of column-averaged CH_4 have been retrieved and studied by e.g. the TCCON network, a times eries of profiles in the Arctic region has not been available prior to this study. Our analysis discovered that while the CH_4 concentrations close to the surface are decreasing, the concentrations on upper troposphere and lower stratosphere are increasing. This observation agrees with the in-situ mast measurements as well as an independent satellite borne measurement by ACE, thus the results provide a convincing argument for further research on this phenomenon and its causes. A comment on the resolution of our retrieval is also merited: As is apparent from the averaging kernels in Figure 2.5, the retrieval is well informed on the lower levels of the atmosphere. For altitudes higher 40 km, the retrieved profile no longer contains information obtained from the measurements and is defined by the prior. This limits the usability of the developed methodology to altitudes lower than 40 km.

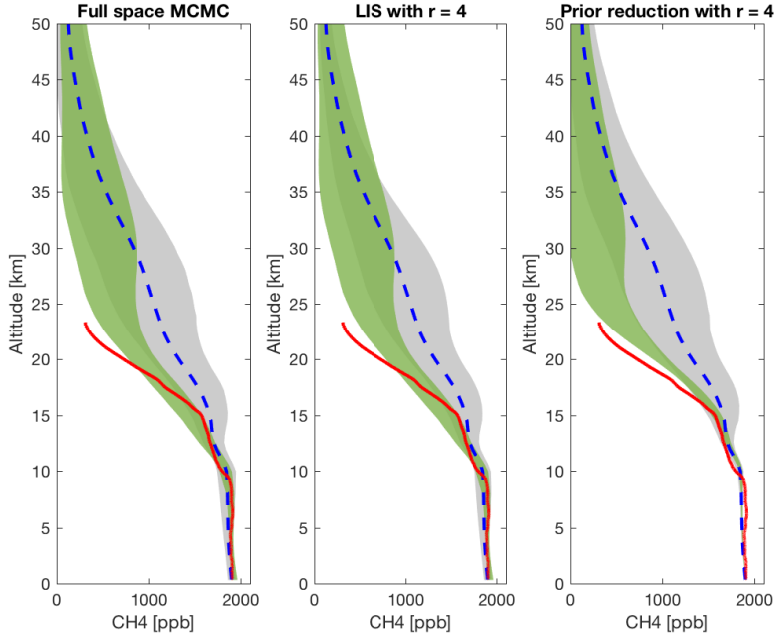


Figure 4.2: Posterior distributions (green) from the FTS CH₄ profile retrievals shown as 95% confidence intervals. From left to right, the results of MCMC applied to full dimensional, LIS, and prior reduction retrievals. Also shown are the prior distributions (95% confidence intervals (grey) and prior mean (dashed blue)) and an in-situ AirCore measurement (red) for reference.

As in **Paper I** we compared the dimension reduction retrievals to AirCore balloon measurements on 22 separate measurement times ranging from 2013 to 2017 and spanning all seasons the measurements are done. It was observed that Optimal Estimation (2.4) yielded results comparable to those from MCMC, which justified using OE to efficiently retrieve the entire time series. The comparison of MCMC solutions to the AirCore measurements is shown in Figure 4.4.

Figure 4.5 summarizes the results of the time series analysis. For the trend analysis, we draw samples from the posterior distribution of the level parameter μ_t using MCMC within the DLM software package. For each sample we computed the yearly trend at time step t as

$$\text{trend}_t = \mu_{t+s/2} - \mu_{t-s/2}, \quad (4.9)$$

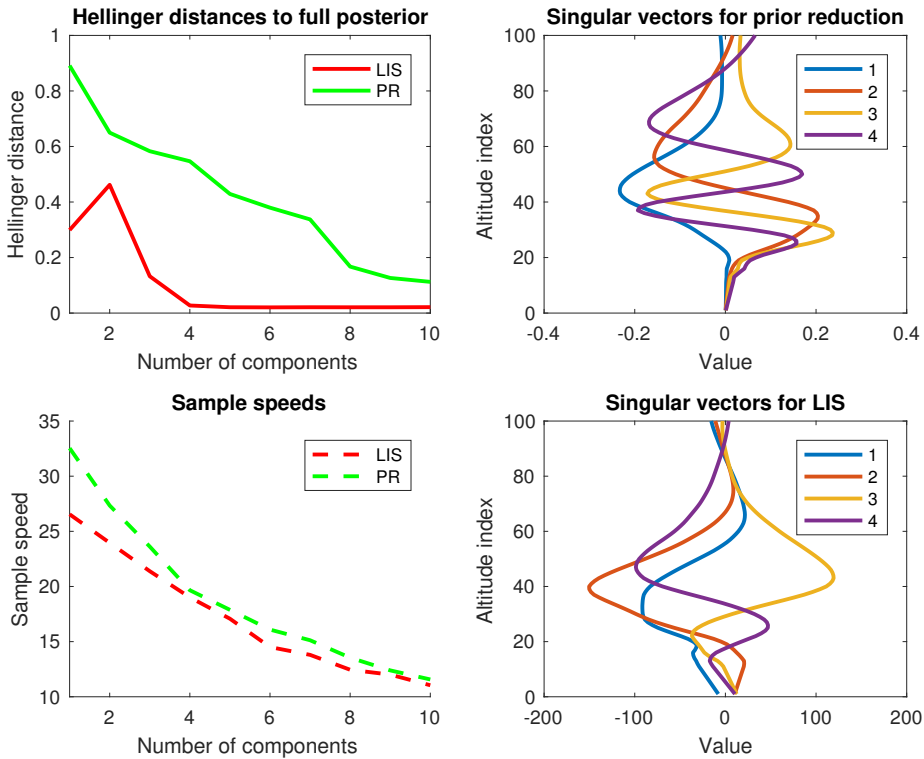


Figure 4.3: Upper left panel: Hellinger distances to the full posterior as a function of number of basis vectors used in the retrieval, with prior (green) and LIS (red) based dimension reductions. Lower left panel: sample speeds of the MCMC chains as a function of number of basis vectors used in the retrieval, with prior (green) and LIS (red) based dimension reductions. Upper right panel: 4 first basis vectors used in prior based dimension reduction. Lower right panel: 4 first basis vectors used in LIS dimension reduction.

where s is the number of days considered. To assess the trend and the uncertainty, we ran MCMC for 1000 steps and calculated the average trend and the standard deviation of the trend for every time step.

Comparison of the lowest altitude of the retrieved CH_4 profile against an in situ mast measurement from 50 m shows that while the spread and temporal coverage of our retrieval is inferior, the agreement in the background level and seasonal cycle is remarkable. DLM analysis of both instruments pointed out a decreasing trend from 2016 onwards. While artifacts are common in the beginning and end of time series

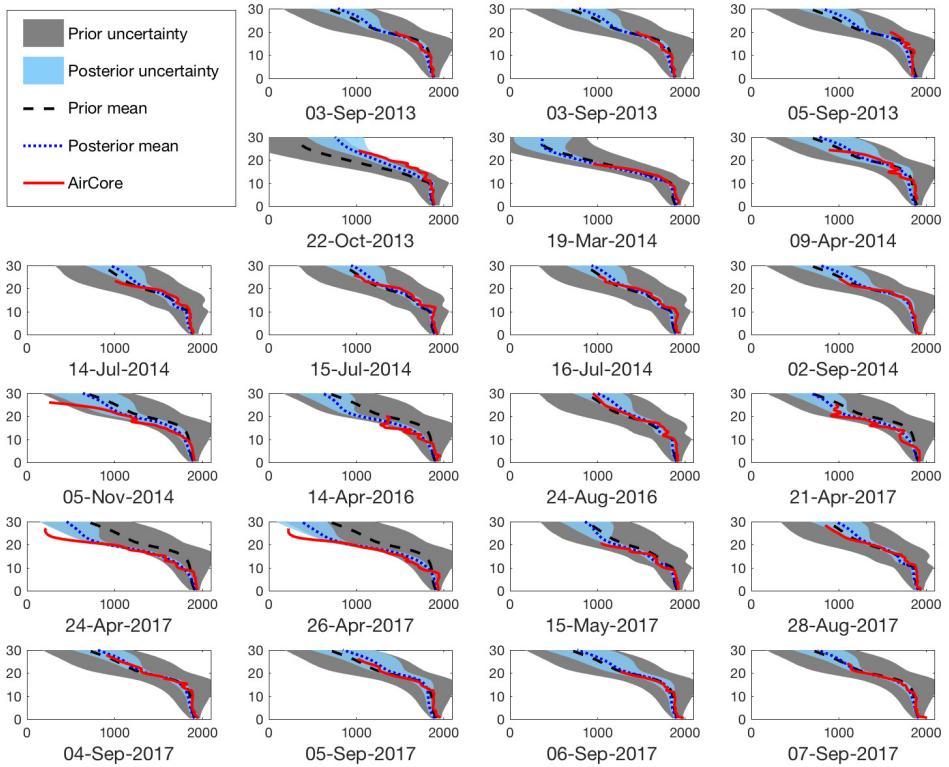


Figure 4.4: Comparison between retrieved posterior distributions obtained via MCMC, and AirCore profiles (red). Black dashed line is the prior profile with 95% prior variance in grey shading. Blue dotted line is the posterior mean with blue shading representing the 95% posterior confidence limits.

trend analyses due to lacking data, our results show a downwards trend for over 3 years which increases their credibility.

For three representative atmospheric altitudes at 11.5 km, 21.5 km and 26.5 km in Figure 4.5, the methane profiles we retrieved and ACE profiles agree very well within the 95% posterior confidence limits on the background level term. This shows the strength of the DLM approach: we have managed to detach seasonal variability

and autocorrelations from the data, obtaining a description of the mean background level of the atmospheric concentrations and related uncertainties.

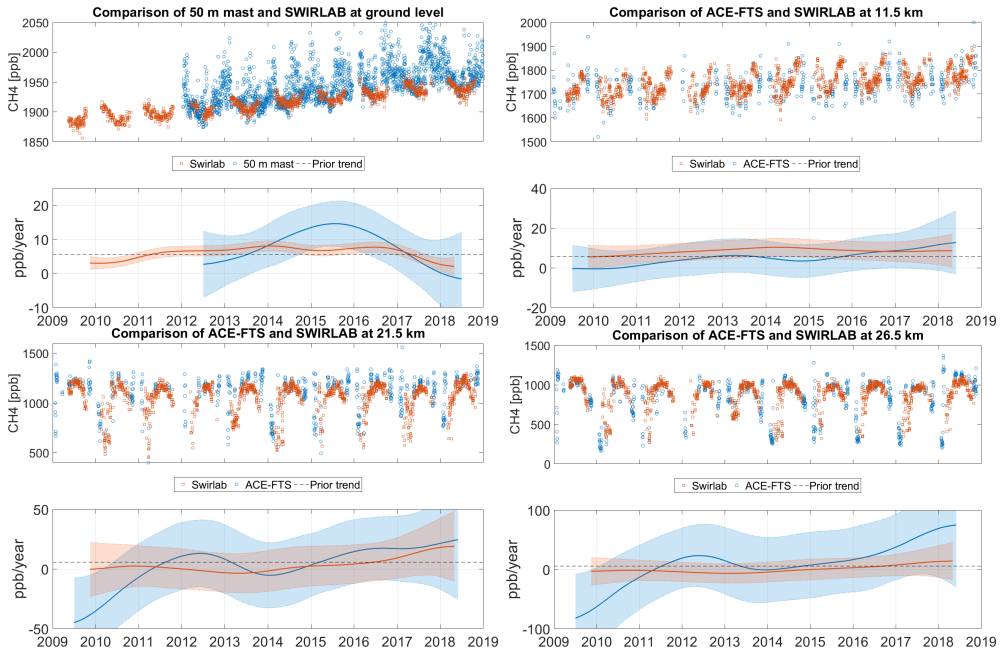


Figure 4.5: top left) Lowest layer of our retrieval compared against 50 m in situ measurements. Top right: Our retrieval compared to ACE-FTS instrument at 11.5 km. Bottom left: Our retrieval compared to ACE-FTS instrument at 21.5 km. Bottom right: our retrieval compared to ACE-FTS instrument at 26.5 km. In every panel the upper axis shows the daily averages from compared instruments and the lower axis shows the 1-year average trend and the 95 percent confidence limits evaluated from the DLM analysis. Our retrieval is in orange and the reference instrument in blue.

4.2 SATELLITE RETRIEVAL OF CO₂

Carbon dioxide retrievals carried out by NASA's *Orbiting Carbon Observatory 2* (OCO-2) satellite are required to have extremely reliable error estimates. Our work aims to aid in this task by providing a thorough quantification of uncertainty caused by the non-linear forward model used in the OCO-2 retrieval. These effects are assumed to be negligible in the operational retrieval algorithm, resulting in linear error estimates introduced in section 2.2.1. For rigorous UQ, we use a computationally feasible *surrogate forward model* together with Likelihood informed subspace dimension reduction to implement a fast and robust Markov chain Monte Carlo retrieval algorithm to sample directly from the obtained posterior distribution.

The strength of our analysis method is demonstrated by assessing the non-Gaussian shape of the retrieval's posterior distribution, and further exploring the sensitivity of the retrieval to operational OCO-2 prior covariance matrices' aerosol parameters. We base our implementation on existing code by Brynjarsdottir et al. (2018), where adaptive MCMC and Optimal Estimation for simulated OCO-2 retrievals were comprehensively compared. The results highlighted issues of slow convergence of MCMC and lack of consistent agreement of MCMC means and MAP estimates in retrievals. These issues are further addressed by our work.

In this section, we first introduce the OCO-2 satellite and the computational forward model used in this work, followed by results of applying adaptive MCMC and LIS dimension reduction introduced in earlier chapters to efficiently explore the posterior distribution obtained as the solution of the retrieval problem.

4.2.1 ORBITING CARBON OBSERVATORY 2

In atmospheric research (according to Miller et al. (2007)), an absolute accuracy of 1-2 ppm is required for reliable atmospheric carbon flux inversions and climatological predictions. NASA's *Orbiting Carbon Observatory 2* (OCO-2, Figure 4.6) (Crisp et al. (2004), Eldering et al. (2017)) satellite mission aims to satisfy these requirements and has been operationally measuring XCO₂ since July 2014. From orbit, OCO-2 measures radiances of reflected solar radiation, that is, absorption of sunlight in Earth's atmosphere on different wavelengths. These measurements are inverted to get an atmospheric state x that describes conditions in the light path, among which are CO₂ concentrations, aerosol distributions and surface properties. The resulting inverse problem is solved using a retrieval algorithm, which consists of a so called *Full Physics model* (O'Dell et al. (2012), Crisp et al. (2012), O'Dell et al. (2018)) that extensively describes the relevant physical phenomena in the atmosphere, and

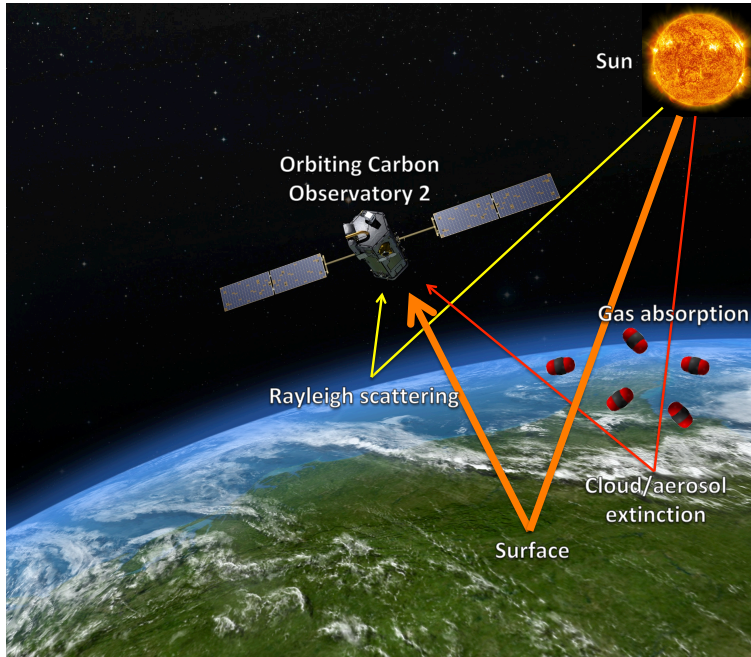


Figure 4.6: The Orbiting Carbon Observatory 2 satellite's measurement geometry. The satellite measures attenuation of reflected solar light that has passed through the atmosphere. The main beam (orange), reflected by Earth's surface, is the most significant light path hitting the detector. In addition, physics model describing the atmospheric radiative transfer has to take into account e.g. atmospheric scattering (yellow) and extinction by clouds and aerosols (red). Original image is credited to NASA/JPL and obtained from the OCO-2 public web page.

Optimal Estimation (OE) framework, in which an optimization algorithm minimizes the cost function related to the statistical inverse problem of the retrieval.

The OCO-2 instrument is composed of three spectrometers that measure light reflected from Earth's surface in the near-infrared part of the spectrum at three separate wavelength bands. These bands are centered around 0.765, 1.61 and 2.06 μm and are called the O_2 A-band, the weak CO_2 band and the strong CO_2 band, respectively. Each observation furthermore consists of 1016 radiances from each band, and together they form the observed sounding. The satellite flies in a polar,

sun-synchronous orbit that covers the whole Earth with a 233 orbit 16-day repeat cycle, during which it collects measurements with a surface footprint of less than 2.25 km down-track and 1.3 km cross-track every 0.333 seconds.

4.2.2 IMPLEMENTATION

For uncertainty quantification, we use Bayesian approach to obtain a posterior distribution as a solution to the OCO-2 inverse problem. We implement the retrieval as follows:

State vector: The state vector $x \in \mathbb{R}^n$ used in our study is based on the physics model used by Hobbs et al. (2017). In our case, $n = 39$ and the state consists of four parts:

- CO₂ vertical profile (1-20)
- Surface pressure (21)
- Surface albedo (22-27)
- Aerosol parameters (28-39)

This state vector partially represents the operational OCO-2 retrieval algorithm's state vector, but a thorough treatment of the Full Physics framework is not feasible in the scope of this thesis.

The main quantity of interest in the OCO-2 retrieval is the column-averaged dry air mole fraction of CO₂, denoted X_{CO_2} , which is a weighted average CO₂ concentration from the first 20 parameters of the state vector. X_{CO_2} is calculated as

$$X_{CO_2} = h^T x_\alpha, \quad (4.10)$$

where $h \in \mathbb{R}^{20}$ is a pressure weighting function (see Crisp et al. (2019)) assigning appropriate weights to each altitude, and $x_\alpha \in \mathbb{R}^{20}$ contains the first 20 elements of the surrogate state vector.

Prior: We use a Gaussian prior distribution $\mathcal{N}(x_0, \Sigma_0)$, where the prior mean $x_0 \in \mathbb{R}^n$ and prior covariance matrix $\Sigma_0 \in \mathbb{R}^{n \times n}$ are obtained from the operational retrieval (Crisp et al., 2019), while leaving out the elements that are not part of our state vector. The 20-dimensional CO₂ part of the state has a covariance that promotes smoothness, while the remaining part of the prior covariance matrix is diagonal.

Likelihood: We utilize likelihood derived from the assumed Gaussian distribution for the observation noise $\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$, $\varepsilon \in \mathbb{R}^m$, $\Sigma_\varepsilon \in \mathbb{R}^{m \times m}$. This results in

a likelihood proportional to $\exp\left(-\frac{1}{2}(y - F(x))^T \Sigma_\epsilon^{-1}(y - F(x))\right)$, where $y \in \mathbb{R}^m$ is the measured absorption spectrum and F is the forward model. In the case of OCO-2 data we have $m = 3048$. The error covariance Σ_ϵ is taken to be diagonal with entries corresponding to the magnitudes of radiances in data y multiplied by an appropriate constant given by Hobbs et al. (2017).

Forward model: The operational state-of-the-art OCO-2 physics model describes the light absorption and scattering by trace gases and aerosols in the atmosphere. While this is a comprehensive way to compute atmospheric radiative transfer, the model is computationally expensive to evaluate and hence is not suitable for Monte Carlo experiments. We ease this problem by using a computationally feasible surrogate forward model by Hobbs et al. (2017), which accounts for most of the major sources of error in the retrieval.

The surrogate model uses similar physics computations as the Full Physics model, but only a part of the operational state vector is taken into account as follows:

- The model considers the amount of molecules in 20 fixed pressure levels of the atmosphere. The concentration of CO_2 as well as pressure and temperature are taken to be constant inside each level.
- A single parameter of the state vector is retrieved to help identify the total amount of molecules in an atmospheric column. Since O_2 has a near constant concentration in each atmospheric layer, the total absorption on the O_2 band can be used to estimate the total amount of air at the measurement location.
- Earth's surface reflects radiation differently at different wavelengths, and the relation of reflected vs. total incoming radiation is given by albedo. In the surrogate model this corresponds to 6 parameters: a Lambertian albedo and its spectral slope at each of the 3 wavelength bands of the observed absorption spectrum.
- Small particles in the atmosphere absorb and scatter light in complex ways, which is taken into account by the model as an aerosol model. The vertical profile of aerosol concentration is modeled with 3 parameters per aerosol type, first of which is the logarithm of Aerosol Optical Depth (AOD) at 755nm which describes the total intensity of aerosol effects. The second parameter describes the height of the maximum aerosol concentration, and the third describes the thickness of the aerosol layer, with a small value corresponding to a thin layer.

The aerosol profile calculated from these parameters is proportional to

$$AOD \exp\left(-\frac{(x - x_a)^2}{2\sigma^2}\right), \quad (4.11)$$

where x_a is the layer height and σ is the layer thickness. In this work the number of atmospheric aerosol types is 4, which amounts to a total of 12 aerosol parameters in the state vector.

Data: Our study uses simulated data, which is given in locational templates. These templates are areas on land inside which most geophysical properties affecting the retrieval can be assumed to be constant. Geolocated state vectors from operational OCO-2 retrieval are used to calculate the empirical CO₂ mean and covariance (4.1) for each template. Together with climatological data about aerosol concentrations, surface properties etc, a Gaussian mixture model is constructed and ensemble of surrogate state vectors are simulated from it. Following Hobbs et al. (2017), radiances with additional noise are then computed using the surrogate forward model, and Optimal estimation retrieval (2.4) is conducted to obtain retrieved state vectors \hat{x} and posterior covariances Σ_{post} . All of this information is then saved on so called scene files.

Dimension reduction: It was observed that the number r of informative directions for the radiances varies from location to location. For each sounding, r is obtained from the SVD of the Gauss-Newton Hessian (3.17) evaluated at \hat{x} . LIS basis is then constructed by sampling from the Laplace approximation $\mathcal{N}(\hat{x}, \Sigma_{post})$ for the sample mean computation (3.32). The LIS basis is used to obtain the projection matrices $\mathbf{P}_r \in \mathbb{R}^{r \times n}$, $(\mathbf{I} - \mathbf{P}_r) \in \mathbb{R}^{n-r \times r}$, $\Phi_r \in \mathbb{R}^{n \times r}$ and $\Phi_{\perp} \in \mathbb{R}^{n \times n-r}$ such that

$$x = \mathbf{P}_r x + (\mathbf{I} - \mathbf{P}_r)x = \Phi_r x_r + \Phi_{\perp} x_{\perp}. \quad (4.12)$$

for the LIS parameter $x_r \in \mathbb{R}^r$ and the CS parameter $x_{\perp} \in \mathbb{R}^{n-r}$.

MCMC: The obtained LIS projector \mathbf{P}_r (3.21) is used to get $\mathbf{P}_r(\hat{x} - x_0) \in \mathbb{R}^r$ as the starting value, and $\mathbf{P}_r \Sigma_{post} \mathbf{P}_r^T \in \mathbb{R}^{r \times r}$ as the starting covariance for the Adaptive Metropolis algorithm (2.5.2). MCMC is run for 250 000 iterations to ensure the convergence of full-dimensional chains for adequate comparison against LIS MCMC. LIS samples are then projected back to full space according to formula (3.28), and a sample from the complement prior is added according to (3.30).

Parameter tuning for OE: As was shown by Brynjarsdottir et al. (2018) and Nelson (2015), the optimal estimation retrieval is sensitive to the initial guess of the optimization algorithm, and furthermore the result might not agree with the

overall posterior distribution obtained with MCMC. We performed trial-and-error based parameter tuning for the OE retrieval in order to rule out convergence problems due to bad starting parameter values. In previous work, the starting value for damping parameter γ for the Levenberg-Marquardt algorithm (2.25) was set to 10 (see Crisp et al. (2019)). For tuning, we set the maximum amount of iterations to 1000 so that the optimization will terminate only once one of the other conditions is met. As a starting state $x_0 \in \mathbb{R}^{39}$ for the optimization, we use the prior mean, the mean obtained from MCMC simulation, and the MCMC mean perturbed by $\pm\sigma \in \mathbb{R}^{39}$, where σ is a standard deviation obtained by taking the square root of the elements of the diagonal of \hat{S} . The results with both original $\gamma = 10$ and tuned $\gamma = 30$, which was found to be a reasonably good value, as starting values are shown in Figure 4.7.

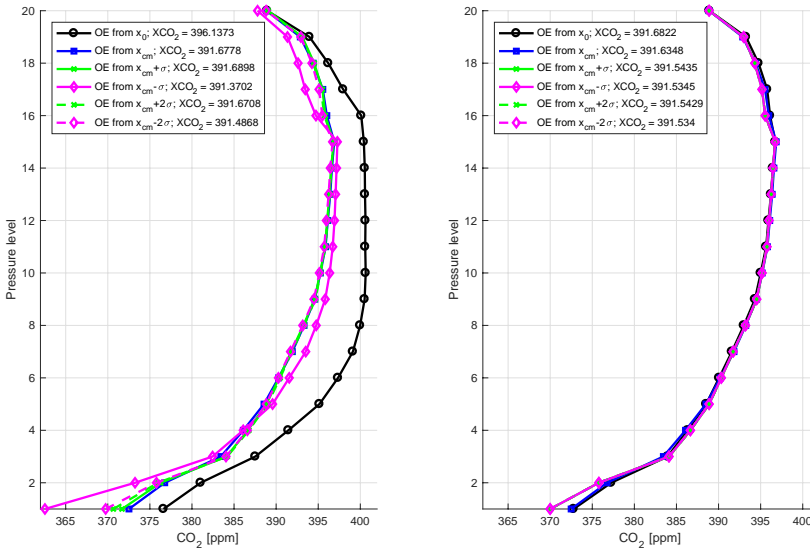


Figure 4.7: CO_2 profiles from OE retrievals, starting the optimization from various perturbed first guess profiles. **Left panel:** OE retrievals performed with parameter values of starting $\gamma = 10$, normalized step size tolerance = 40. **Right panel:** OE retrievals performed with parameter values of starting $\gamma = 30$, normalized step size tolerance = 0.0001.

4.2.3 OVERVIEW OF RESULTS

In **Paper III**, we implemented the LIS dimension reduction to OCO-2 retrieval introduced by Brynjarsdottir et al. (2018), using synthetic data, a computationally

efficient surrogate forward model and Adaptive MCMC. The novel LIS dimension reduction implementation was found to significantly increase the computational efficiency of MCMC, while yielding essentially a similar result as the full-dimensional retrieval. Using the MCMC samples we detected that the resulting posterior distribution is non-Gaussian, specifically in aerosol parameters of the state vector. This result gives important insights on the effectiveness of the linear error estimation (2.15) in the operational OCO-2 algorithm. Furthermore, our analysis showed that relaxing constraints on the prior covariance of aerosol parameters can significantly improve the results of the retrieval. This is to say that a misspecification of the prior for non-CO₂ elements of the state vector can cause large errors in the retrieved X_{CO_2} , which is in agreement with recent results in Nguyen et al. (2019).

In Figure 4.2.3, we show the MCMC chains of retrieved X_{CO_2} obtained from the CO₂ profile using formula 4.10. We clearly see that the implemented LIS dimension reduction results in more or less exactly the same posterior distribution as the full dimensional MCMC performed in Brynjarsdottir et al. (2018). Moreover, the obtained posterior in figure 4.2.3 is seen to be wider than the approximation obtained via OE. The lower panel of figure 4.2.3 also graphically demonstrates that convergence is obtained faster with LIS MCMC in comparison to the full dimensional case. This is useful for future work since, as stated in Brynjarsdottir et al. (2018), the OCO-2 Full Physics model is computationally too costly to perform a full dimensional MCMC analysis. As shown, LIS has the potential to aid in this task in the future.

We observed from the MCMC retrieval that while the Gaussian assumption holds well with the CO₂ part of the state vector, the distribution of aerosol parameters show clear non-Gaussian shape. Figure 4.9 shows the distribution of MCMC samples from the aerosol parameters of the state vector plotted against each other. We can see that the Gaussian posterior approximation from OE retrieval is too optimistic and doesn't provide wide enough confidence intervals for aerosol parameters. This error might propagate to the final X_{CO_2} estimate as an over confidence on the retrieval result. Moreover, the middle figure of the second row shows a so called "banana distribution" shape, which has been known to pose trouble for MCMC samplers and resulting in low acceptance rates (see e.g. Haario et al. (1999)). The third figure on the second row also reveals two dimensions that are strongly correlated, which results to the posterior being "flat" and difficult to sample from. This observation highlights one strength of the LIS approach, since correlated dimensions are automatically removed from the reduced state vector. We observed that numerous retrievals were overly constrained by the operational prior covariance, so that the true values of aerosol cloud width parameters (4.11) were not retrieved. This effect is seen in the middle panel of top row (parameters 30 vs. 33) in Figure 4.9. Guided by this

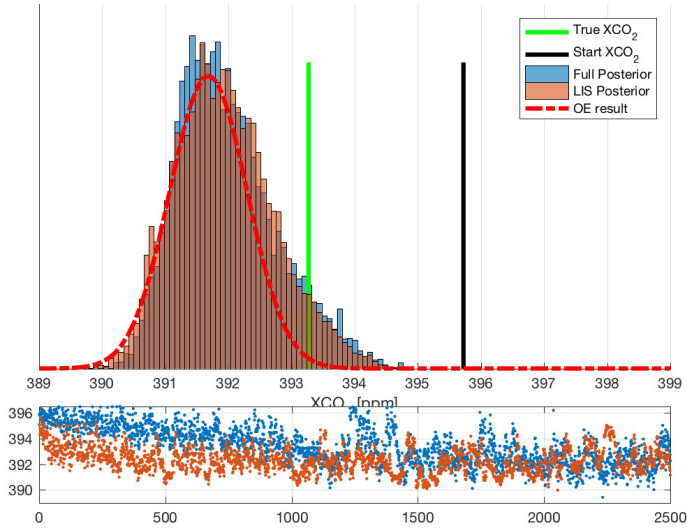


Figure 4.8: **Upper panel:** X_{CO_2} posterior histogram from full dimensional MCMC (blue) compared to LIS MCMC (red). Also shown the starting value for MCMC and simulated ground truth (green). **Lower panel:** every 100th sample of X_{CO_2} chain from full dimensional MCMC (blue) and LIS MCMC (red).

observation, we performed additional OE and MCMC retrievals on the same problematic cases while multiplying the prior covariance variance of the first aerosol cloud height parameter by 10. Figure 4.10 shows effects of this change to the histograms of aerosol parameters of the state vector obtained with MCMC. Using an inflated prior covariance is generally seen to make the results agree better with the true value of the aerosol parameters. Three MCMC test cases of retrieved X_{CO_2} are shown in Figure 4.11. The result is by no means universal, but suggests that in certain cases using a less strict prior can significantly improve the quality of both OE and MCMC based retrievals.

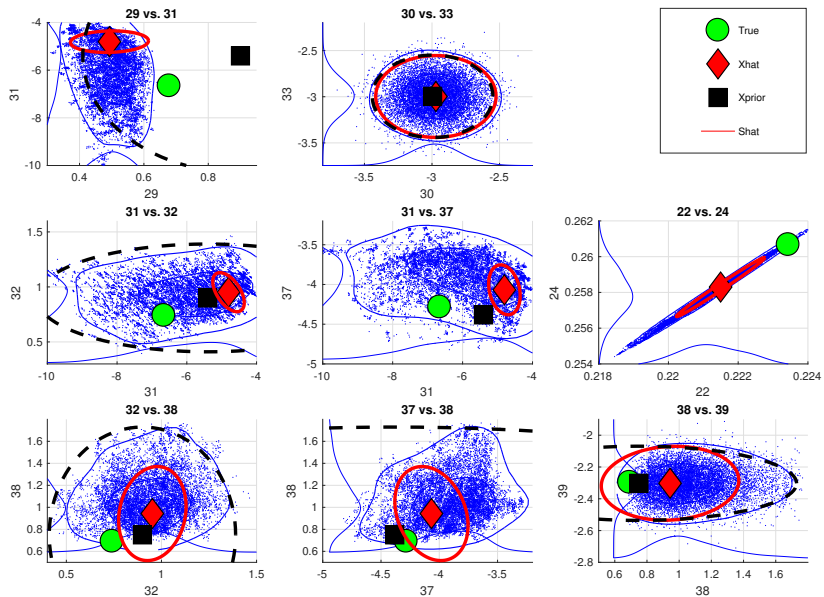


Figure 4.9: 2D posterior distributions of selected aerosol and surface parameters from LIS MCMC (blue with 95% posterior confidence interval on blue contour) compared with ground truth (green), prior (black with 95% confidence interval on black ellipse) and OE (red with 95% posterior confidence interval on red ellipse). Parameter numbers agree with the numbering in the surrogate model state vector.

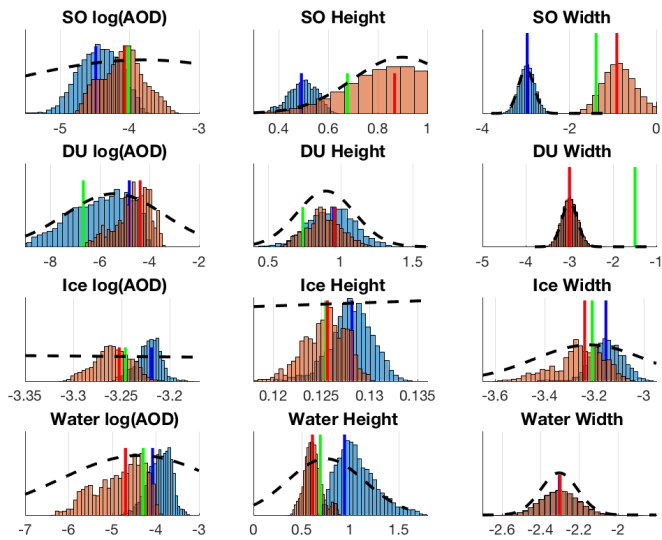


Figure 4.10: Aerosol parameter marginal histograms (corresponding to retrievals using operational prior (blue) and widened prior (orange)) with operational prior (dashed black), ground truth (green) and OE retrieval result (blue and red vertical lines). SO = sulphate, DU = dust, Ice = cloud ice and Water = cloud liquid water.

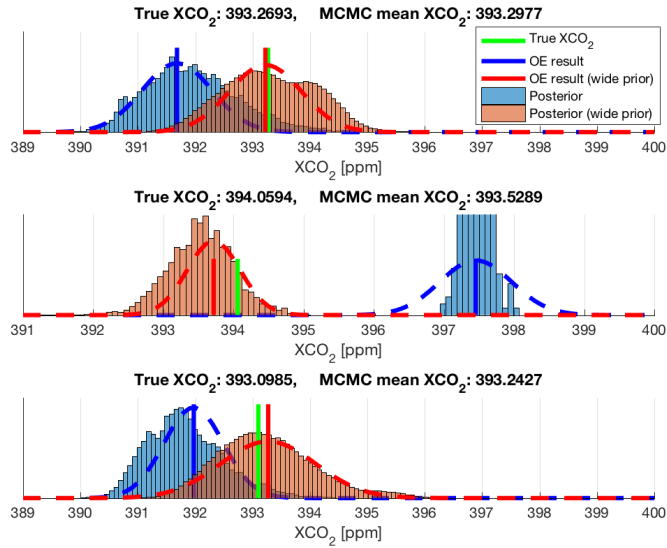


Figure 4.11: Three separate test cases showing X_{CO_2} posterior histograms (using operational prior (blue) and widened prior) from LIS MCMC compared to the OE retrieval (blue and red dashed lines) and simulated ground truth (green). Also shown the true value of X_{CO_2} with the corresponding MCMC mean using a widened prior. It should be noted that the example case 2 on the middle panel most likely did not converge in either the MCMC or OE when using the operational prior, which is not an unusual scenario even in the operational retrieval.

5 CONCLUSIONS

In the field of atmospheric remote sensing of greenhouse gases, rigorous uncertainty quantification is needed in order to ensure the accuracy and precision of concentration measurements. In this work we have developed and implemented efficient dimension reduction methods to facilitate the usage of Markov chain Monte Carlo algorithms to solve the related Bayesian inverse problem. Our methods were successfully used to gain insight into the retrieval problems of both atmospheric CH₄ profile from Sodankylä Fourier Transform Spectrometer, and CO₂ total column from NASA's Orbiting Carbon Observatory 2 satellite. In all our research cases, the Likelihood-informed subspace dimension reduction method was found to be robust and efficient in accelerating MCMC computation and making it feasible. In this work, LIS was used with real data for the first time in the FTS retrieval, which was further used to process the entire time series of measurements from 2009 to 2018. We were also able to detect the non-Gaussian shape of the posterior distribution of the OCO-2 inverse problem and suggest possible improvements to the retrieval algorithm via assessment of the prior covariance matrix.

The research of this thesis opens up some interesting further research questions as well. First, the LIS dimension reduction we used includes approximating the non-linear LIS with a Monte Carlo average. For a rigorous, general solution, a thorough analysis of *certified dimension reduction* was presented by Zahm et al. (2018), where the authors derived an expression for a general non-linear case. Certified dimension reduction could be implemented to our algorithms instead of LIS to improve the low-dimensional posterior approximations. Second, we used DLM to process the Sodankylä FTS CH₄ profile time series on an altitude-by-altitude basis. This implementation requires parameter and first guess tuning by hand, which is a time-consuming procedure. A possible improvement to our algorithm would be to perform the DLM analysis for the entire profile all at once. In order to do this, the state space model would need to be updated to include parameters for correlations between atmospheric layers, resulting in a larger hidden state vector than in our case.

Finally, our work with the OCO-2 algorithm relies on using a surrogate model which is an approximation of the operational Full Physics model and hence the results are not guaranteed to hold for the operational algorithm. A way to remedy this would be to use the *approximation error* approach (e.g. Kolehmainen et al. (2011), Lehtikoinen et al. (2007)) to extract the statistics of the error one makes when approximating a high fidelity model by a surrogate model. This statistical correction could then be included in the OCO-2 surrogate model retrieval to make it comparable with the Full Physics model.

SUMMARIES OF ORIGINAL PUBLICATIONS

1. Otto Lamminpää, Marko Laine, Simo Tukiainen, Johanna Tamminen: Likelihood informed dimension reduction for remote sensing of atmospheric constituent profiles

In **Paper I**, adaptive Metropolis algorithm with likelihood-informed subspace dimension reduction is implemented to a ground-based remote sensing problem. Rigorous uncertainty quantification is performed for the Bayesian inverse problem of inferring atmospheric CH₄ concentrations from solar radiation attenuation measurements by a Fourier transform spectrometer located in Sodankylä, Northern Finland. The results are compared with MCMC using prior based dimension reduction. Our novel method is found to offer a robust UQ framework with additional flexibility for prior covariance selection.

2. Tomi Karppinen, Otto Lamminpää, Simo Tukiainen, Rigel Kivi, Pauli Heikkinen, Juha Hatakka, Marko Laine, Huilin Chen, Hannakaisa Lindqvist and Johanna Tamminen: Vertical Distribution of Arctic Methane in 2009-2018 Using Ground-Based Remote Sensing.

The work of **Paper II** focuses on solving a long time series of inverse problems using a dimension reduction based algorithm, and then proceeds to analysing the obtained time series with a dynamic linear model. The entire time series of Sodankylä FTS measurements spanning years 2009-2018 is processed using prior-based dimension reduction algorithm and the obtained CH₄ concentration profiles are then corrected for solar zenith angle dependence. The time series analysis with DLM provides non-linear background trends with reliable uncertainty estimates, which are compared against independent satellite and ground-based concentration measurements of CH₄. Our method agrees well with existing instruments, and the time series analysis points out possible altitude dependence in the background trend of the time series.

3. Otto Lamminpää, Jonathan Hobbs, Jenny Brynjarsdottir, Marko Laine, Amy Braverman, Hannakaisa Lindqvist and Johanna Tamminen: Accelerated MCMC for Satellite-Based Measurements of Atmospheric CO₂.

In **Paper III**, adaptive Metropolis algorithm with likelihood-informed subspace dimension reduction is implemented to a satellite-based remote sensing problem. The uncertainty in NASA's Orbiting Carbon Observatory 2 satellite's CO₂ operational retrieval algorithm is assessed by using a computationally feasible surrogate model and simulated data. Our results show that the posterior distribution of the related Bayesian inverse problem is not Gaussian, which is caused by the parametrization of the aerosol elements of the OCO-2 state vector. We further use our analysis to show how relaxation in some prior constraints can significantly improve the accuracy and reliability of the retrieval.

Appendices

A PRELIMINARY MATHEMATICAL CONCEPTS

In this section, we give a short revision of key mathematical concepts required to understand the theory and applications in this work.

A.1 MATRICES

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is *positive definite* if $x^T \mathbf{A} x > 0$ for all $x \in \mathbb{R}^n$, and *symmetric* if $\mathbf{A}^T = \mathbf{A}$. For a symmetric, positive definite square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, there exists a unique *Cholesky factorization* (Golub and Loan, 1996)

$$\mathbf{A} = \mathcal{L}\mathcal{L}^T, \quad (\text{A.1})$$

where $\mathcal{L} \in \mathbb{R}^{n \times n}$ is lower-triangular and has positive diagonal entries.

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, there exists a *singular value decomposition (SVD)* (see e.g. Golub and Loan (1996), Mueller and Siltanen (2012)) given by

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T, \quad (\text{A.2})$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$, $\mathbf{U}^T \mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbb{I}_m$, $\mathbf{V}^T \mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbb{I}_n$ and $\mathbf{\Lambda} \in \mathbb{R}^{m \times n}$ is a diagonal matrix with *singular values* of matrix \mathbf{A} , $\lambda_1, \lambda_2, \dots, \in \mathbb{R}$, as diagonal entries. The matrix \mathbf{A} has n singular values if $m \geq n$, and m if $n > m$. The singular values are labelled in increasing order: $\lambda_1 \geq \lambda_2 \geq \dots$. Furthermore, the columns of u_i and v_i of matrices \mathbf{U} and \mathbf{V} are called *singular vectors* of \mathbf{A} . The right singular vectors v_i also form an orthonormal basis for \mathbb{R}^n (see e.g. Andrilli and Hecker (2016)).

Given an SVD $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$ as previously, the *eigenvalue decomposition* for symmetric matrices matrices $\mathbf{M} = \mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m \times m}$ and $\mathbf{N} = \mathbf{A}^T \mathbf{A} \in \mathbb{R}^{n \times n}$ is given by

$$\mathbf{M} = (\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T)(\mathbf{V}\mathbf{\Lambda}^T \mathbf{U}^T) = \mathbf{U}\mathbf{\Lambda}_m^2 \mathbf{U}^T. \quad (\text{A.3})$$

$$\mathbf{N} = (\mathbf{V}\mathbf{\Lambda}^T \mathbf{U}^T)(\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T) = \mathbf{V}\mathbf{\Lambda}_n^2 \mathbf{V}^T, \quad (\text{A.4})$$

where $\Lambda_m^2 = \Lambda\Lambda^T \in \mathbb{R}^{m \times m}$ and $\Lambda_n^2 = \Lambda^T\Lambda \in \mathbb{R}^{n \times n}$. Consider the matrix \mathbf{N} as above. The columns $v_i \in \mathbb{R}^n$ of matrix \mathbf{V} are called *eigenvectors* of \mathbf{N} and the diagonal entries $\lambda_i^2 \in \mathbb{R}$ of matrix Λ_n^2 are called *eigenvalues* of matrix \mathbf{N} , satisfying $\mathbf{N}v_i = \lambda_i^2 v_i$.

A.2 RANDOM VARIABLES

Let Ω be an abstract sample space, \mathfrak{G} be a sigma algebra over Ω and let P be a probability measure. Then the triple $(\Omega, \mathfrak{G}, P)$ is called a probability space (for a more comprehensive discussion, see e.g. Kallenberg (1997)). Given a probability space $(\Omega, \mathfrak{G}, P)$, a *random variable* X with values in \mathbb{R}^n is defined as a measurable mapping

$$X : \Omega \mapsto \mathbb{R}^n, \quad (\text{A.5})$$

or in other words, $X^{-1}(B) \in \mathfrak{G}$ for each open set $B \in \mathbb{R}^n$. We use the usual convention and denote random variables with upper case letters. Furthermore, the *realizations* of these random variables are denoted with lower case letters: $X(\omega) = x \in \mathbb{R}^n$, $\omega \in \Omega$.

A random variable $X \in \mathbb{R}^n$ induces a probability measure μ called *probability distribution of X* in \mathbb{R}^n such that

$$\mu(B) = P(X^{-1}(B)), \quad B \in \mathfrak{B}, \quad (\text{A.6})$$

where \mathfrak{B} is the Borel sigma-algebra of \mathbb{R}^n . In this work, we assume that all random variables introduced are absolutely continuous with respect to the Lebesgue measure on an appropriate space (see e.g. Rudin (1987) for details). The probability distribution of random variable X that is absolutely continuous with respect to the Lebesgue measure can be described with a *probability density function* $\pi(x)$, (also density function for short), which satisfies

$$\mu(B) = \int_B \pi(x) dx. \quad (\text{A.7})$$

Furthermore, for random variables $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$, the *joint density function* $\pi(x, y)$ for the product random variable $X \times Y : \Omega \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ satisfies

$$\mu(B_1, B_2) = \int_{B_1 \times B_2} \pi(x, y) dx dy. \quad (\text{A.8})$$

with open sets $B_1 \in \mathbb{R}^n$, $B_2 \in \mathbb{R}^m$.

Given a joint probability density function $\pi(x, y)$ for the variables X and Y , the *marginal density function* $\pi(x)$ for X is given by

$$\pi(x) = \int_{\mathbb{R}^m} \pi(x, y) dy. \quad (\text{A.9})$$

The *expected value* of $X \in \mathbb{R}^n$ over $\pi(x)$ is given by

$$\mathbb{E}[X] = \int_{\mathbb{R}^n} x \pi(x) dx, \quad (\text{A.10})$$

provided that the integral exists. Furthermore the *covariance* matrix of random variable $X \in \mathbb{R}^n$ is

$$\text{Cov}(X) = \mathbb{E}[XX^T] - \mathbb{E}[X](\mathbb{E}[X])^T \in \mathbb{R}^{n \times n} \quad (\text{A.11})$$

if the integral exists. A *n-variate Gaussian* random variable $X \in \mathbb{R}^n$ with mean $x_0 \in \mathbb{R}^n$ and a symmetric, positive-definite covariance matrix $\Sigma_0 \in \mathbb{R}^{n \times n}$ has the probability density function

$$\pi(x) = ((2\pi)^n |\Sigma_0|)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - x_0)^T \Sigma_0^{-1}(x - x_0)\right). \quad (\text{A.12})$$

We denote this distribution by $\mathcal{N}(x_0, \Sigma_0)$.

An assumption we usually make about the data is that observations x_i are *independent and identically distributed* (i.i.d.) random variables. Independence means that $P(x_1, x_2) = P(x_1)P(x_2)$, which also implies conditional independence: $P(x_1|x_2) = P(x_1)$.

Two random variables X_1 and X_2 are said to be independent $P(x_1, x_2) = P(x_1)P(x_2)$, which also implies conditional independence: $P(x_1|x_2) = P(x_1)$. If they are pairwise independent and all follow the same distribution.

Let X_1, \dots, X_N be N i.i.d. random variables with expected value \bar{x} and variance σ^2 and let S_n be the *sample mean* of the values, given by

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (\text{A.13})$$

The *central limit theorem* states that in the limit $n \rightarrow \infty$, the random variable $\sqrt{n}(S_n - \bar{x})$ converges in distribution to a zero-mean Gaussian distribution with variance σ^2 , i.e. $\mathcal{N}(0, \sigma^2)$. The *strong law of large numbers* states that in the limit $n \rightarrow \infty$, the sample mean converges to the expected value $S_n \rightarrow \bar{x}$.

Assume that $\mathbb{E}[X] < \infty$. For a random sample $x_i \stackrel{\text{i.i.d.}}{\sim} \pi(x), i = 1, \dots, N$, we can estimate $\mathbb{E}[X]$ using the sample mean

$$I_N = \frac{1}{N} \sum_{i=1}^N x_i. \quad (\text{A.14})$$

By the Strong law of large numbers, we obtain $\lim_{N \rightarrow \infty} I_N = \mathbb{E}[X]$.

Given a probability space $(\Omega, \mathfrak{G}, P)$, a *stochastic process* is a collection of \mathbb{R}^n -valued random variables

$$\{X(\omega, t) : \omega \in \Omega, t \in T\}, \quad (\text{A.15})$$

where T is some index set, usually taken to be time. While generalizations exist, in the scope this work we only need to consider the case $X \in \mathbb{R}^n$. Given the beginning and ending indices $t_1, t_2 \in T$, we will use the notation $\{X_t\}_{t=t_1}^{t_2}$ for stochastic processes.

Lastly, we need to use the *Hellinger distance* (e.g. Pollard (2002)) to quantify the difference between probability distributions. The Hellinger distance between two probability distributions with density functions $\pi_1(x)$ and $\pi_2(x)$ is defined as

$$\mathcal{H}^2(\pi_1, \pi_2) = \frac{1}{2} \int_{\mathbb{R}^n} \left(\sqrt{\pi_1(x)} - \sqrt{\pi_2(x)} \right)^2 dx \quad (\text{A.16})$$

It holds that $0 < \mathcal{H}(\pi_1, \pi_2) < 1$. The maximum distance 1 is achieved when $\pi_1(x)$ assigns probability zero to every set to which $\pi_2(x)$ assigns a positive probability, and vice versa.

REFERENCES

- S. Andrilli and D. Hecker, editors. *Elementary Linear Algebra (Fifth Edition)*. Academic Press, Boston, fifth edition edition, 2016. ISBN 978-0-12-800853-9. doi: <https://doi.org/10.1016/B978-0-12-800853-9.09977-1>.
- J. Bardsley. *Computational Uncertainty Quantification for Inverse Problems*. Society for Industrial and Applied Mathematics, 2018. ISBN 9781611975376.
- P. F. Bernath, C. T. McElroy, M. C. Abrams, C. D. Boone, M. Butler, C. Camy-Peyret, M. Carleer, C. Clerbaux, P.-F. Coheur, R. Colin, P. DeCola, M. DeMazière, J. R. Drummond, D. Dufour, W. F. J. Evans, H. Fast, D. Fussen, K. Gilbert, D. E. Jennings, E. J. Llewellyn, R. P. Lowe, E. Mahieu, J. C. McConnell, M. McHugh, S. D. McLeod, R. Michaud, C. Midwinter, R. Nassar, F. Nichitiu, C. Nowlan, C. P. Rinsland, Y. J. Rochon, N. Rowlands, K. Semeniuk, P. Simon, R. Skelton, J. J. Sloan, M.-A. Soucy, K. Strong, P. Tremblay, D. Turnbull, K. A. Walker, I. Walkty, D. A. Wardle, V. Wehrle, R. Zander, and J. Zou. Atmospheric Chemistry Experiment (ACE): Mission overview. *Geophysical Research Letters*, 32(15), 2005. ISSN 1944–8007. doi: 10.1029/2005GL022386.
- A. Beskos, A. Jasra, K. Law, Y. Marzouk, and Y Zhou. Multilevel sequential monte carlo with dimension-independent likelihood-informed proposals. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2), 2018. ISSN 2166-2525. doi: 10.1137/17m1120993.
- Å. Björck. *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics, 1996. doi: 10.1137/1.9781611971484.
- J. Brynjarsdottir, J. Hobbs, A. Braverman, and L. Mandrake. Optimal estimation versus MCMC for CO₂ retrievals. *Journal of Agricultural, Biological and Environmental Statistics*, 23(2):297–316, 2018.

- C. Chatfield. *The Analysis of Time Series, an Introduction*. Chapman & Hall, 4th edition edition, 1989.
- H. Chen, R. Kivi, P. Heikkinen, B. Kers, M. de Vreis, J. Hatakka, T. Laurila, C. Sweeney, and P. Tans. High-latitude balloon observations of CO₂/CH₄/CO using AirCore: evaluation of Sodankylä TCCON retrievals. *In preparation*, 2016.
- P. Constantine, E. Dow, and Q. Wang. Active subspace methods in theory and practice: Applications to kriging surfaces. *SIAM Journal on Scientific Computing*, 36(4):A1500–A1524, 2014. doi: 10.1137/130916138.
- P. Constantine, C. Kent, and T. Bui-Thanh. Accelerating markov chain monte carlo with active subspaces. *SIAM Journal on Scientific Computing*, 38(5):A2779–A2805, 2016. doi: 10.1137/15M1042127.
- National Research Council. *Ecological Impacts of Climate Change*. The National Academies Press, Washington, DC, 2008. ISBN 978-0-309-12710-3. doi: 10.17226/12491. URL <https://www.nap.edu/catalog/12491/ecological-impacts-of-climate-change>.
- N. Cressie and C. Wikle. *Statistics for Spatio-Temporal Data*. Wiley, 2011.
- D. Crisp, R. M. Atlas, F.-M. Breon, L. R. Brown, J. P. Burrows, P. Ciais, B. J. Connor, S. C. Doney, I. Y. Fung, D. J. Jacob, C. E. Miller, D. O'Brien, S. Pawson, J. T. Randerson, P. Rayner, R. J. Salawitch, S. P. Sander, B. Sen, G. L. Stephens, P. P. Tans, G. C. Toon, P. O. Wennberg, S. C. Wofsy, Y. L. Yung, Z. Kuang, B. Chudasama, G. Sprague, B. Weiss, R. Pollock, D. Kenyon, and S. Schroll. The Orbiting Carbon Observatory (OCO) mission. *Advances in Space Research*, 34: 700–709, January 2004. doi: 10.1016/j.asr.2003.08.062.
- D. Crisp, B. M. Fisher, C. O'Dell, C. Frankenberg, R. Basilio, H Boesch, L. R. Brown, R. Castano, B. Connor, N. M. Deutscher, A. Eldering, D. Griffith, M. Gunson, A. Kuze, L. Mandrake, J. McDuffie, J. Messerschmidt, C. E. Miller, I. Morino, V. Na-traj, J. Notholt, D. M. O'Brien, F. Oyafuso, I. Polonsky, J. Robinson, R. Salawitch, V. Sherlock, M. Smyth, H. Suto, T. E. Taylor, D. R. Thompson, P. O. Wennberg, D. Wunch, and Y. L. Yung. The acos co2 retrieval algorithm - part ii: Global xco2 data characterization. *Atmospheric Measurement Techniques*, 5(4):687–707, 2012.
- D. Crisp, H. Boesch, L. Brown, R. Castano, M. Christi, B. Connor, C. Frankenberg, J. McDuffie, C. E. Miller, V. Natraj, C. O'Dell, D. O'Brien, I. Polonsky, F. Oyafuso, D. Thompson, G. Toon, and R Spurr. Oco (orbiting carbon

- observatory)-2 level 2 full physics retrieval algorithm theoretical basis. Technical report, NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 2019. URL https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO_L2_ATBD.pdf.
- T. Cui, J. Martin, Y. M. Marzouk, A. Solonen, and A. Spantini. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems*, 30(11): 114015 pp. 28, 2014. ISSN 0021-9991. doi: 10.1088/0266-5611/30/11/114015.
- T. Cui, K. Law, and Y. Marzouk. Dimension-independent likelihood-informed mcmc. *Journal of Computational Physics*, 304:109 – 137, 2016a. ISSN 0021-9991. doi: <https://doi.org/10.1016/j.jcp.2015.10.008>.
- T. Cui, Y. Marzouk, and K. Willcox. Scalable posterior approximations for large-scale bayesian inverse problems via likelihood-informed parameter and state reduction. *Journal of Computational Physics*, 315:363 – 387, 2016b. doi: <https://doi.org/10.1016/j.jcp.2016.03.055>.
- M. Dameris and D. Loyola. Chemistry-climate connections – interaction of physical, dynamical, and chemical processes in earth atmosphere. In Juan Blanco and Houshang Kheradmand, editors, *Climate Change*, chapter 1. IntechOpen, Rijeka, 2011. doi: 10.5772/24210. URL <https://doi.org/10.5772/24210>.
- A. Eldering, C. W. O’Dell, P. O. Wennberg, D. Crisp, M. R. Gunson, C. Viatte, C. Avis, A. Braverman, R. Castano, A. Chang, L. Chapsky, C. Cheng, B. Connor, L. Dang, G. Doran, B. Fisher, C. Frankenberg, D. Fu, R. Granat, J. Hobbs, R. A. M. Lee, L. Mandrake, J. McDuffie, C. E. Miller, V. Myers, V. Natraj, D. O’Brien, G. B. Osterman, F. Oyafuso, V. H. Payne, H. R. Pollock, I. Polonsky, C. M. Roehl, R. Rosenberg, F. Schwandner, M. Smyth, V. Tang, T. E. Taylor, C. To, D. Wunch, and J. Yoshimizu. The orbiting carbon observatory-2: first 18 months of science data products. *Atmospheric Measurement Techniques*, 10(2):549–563, 2017. doi: 10.5194/amt-10-549-2017.
- D. Gamerman. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall, 1997. ISBN 9780412818202.
- A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, 1995. ISBN 0412039915.
- A. Gelman, G. O. Roberts, and W. R. Gilks. *Bayesian Statistics*, 5:599–608, 1996.

- P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Elsevier, 1981. ISBN 978-0122839528.
- G. H. Golub and C. F. Van Loan, editors. *Matrix Computations*. The Johns Hopkins University Press, Baltimore and London, 3 edition, 1996.
- I. E. Gordon, L. S. Rothman, C. Hill, R. V. Kochanov, Y. Tan, P. F. Bernath, M. Birk, V. Boudon, A. Campargue, K. V. Chance, B. J. Drouin, J. M. Flaud, R. R. Gamache, J. T. Hodges, D. Jacquemart, V. I. Perevalov, A. Perrin, K. P. Shine, M. A.H. Smith, J. Tennyson, G. C. Toon, H. Tran, V. G. Tyuterev, A. Barbe, A. G. Császár, V. M. Devi, T. Furtenbacher, J. J. Harrison, J. M. Hartmann, A. Jolly, T. J. Johnson, T. Karman, I. Kleiner, A. A. Kyuberis, J. Loos, O. M. Lyulin, S. T. Massie, S. N. Mikhailenko, N. Moazzen-Ahmadi, H. S.P. Müller, O. V. Naumenko, A. V. Nikitin, O. L. Polyansky, M. Rey, M. Rotger, S. W. Sharpe, K. Sung, E. Starikova, S. A. Tashkun, J. Vander Auwera, G. Wagner, J. Wilzewski, P. Wcisło, S. Yu, and E. J. Zak. The HITRAN2016 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 203:3–69, 2017. ISSN 00224073. doi: 10.1016/j.jqsrt.2017.06.038.
- H. Haario, E. Saksman, and J. Tamminen. Adaptive proposal distribution for random walk metropolis algorithm. *Computational Statistics*, 14(3):375–395, Sep 1999. ISSN 0943-4062. doi: 10.1007/s001800050022.
- H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001. doi: 10.2307/3318737.
- H. Haario, M. Laine, E. Saksman, J. Tamminen, and M. Lehtinen. Markov chain monte carlo methods for high dimensional inversion in remote sensing. *Journal of the Royal statistical Society, Series B*, 66:591–604, 2004.
- J. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- A. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1991.
- W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97 – 109, 1970. doi: doi:10.1093/biomet/57.1.97.
- J. Hobbs, A. Braverman, N. Cressie, R. Granat, and M. Gunson. Simulation-based uncertainty quantification for estimating atmospheric CO₂ from satellite data. *SIAM/ASA Journal on Uncertainty Quantification*, 5:956–985, 01 2017. doi: 10.1137/16M1060765.

- IPCC. *Climate Change 2013, The Scientific Basis*. Cambridge University Press, Cambridge, UK, 2013.
- IUPAC. *Compendium of Chemical Terminology 2nd ed. (the "Gold Book")*. Blackwell Scientific Publications, Oxford, 1997. ISBN 0-9678550-9-8. doi: <https://doi.org/10.1351/goldbook>.
- J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer-Verlag, New York, 2005. doi: 10.1007/b138659.
- O. Kallenberg. *Foundations of Modern Probability*. Springer, 1997. ISBN 0-387-94957-7.
- A. Karion, C. Sweeney, P. Tans, and T. Newberger. AirCore: An Innovative Atmospheric Sampling System. *Journal of Atmospheric and Oceanic Technology*, 27 (11):1839–1853, 2010. doi: 10.1175/2010JTECHA1448.1.
- R. Kivi and P. Heikkinen. Fourier transform spectrometer measurements of column CO₂ at Sodankylä, Finland. *Geoscientific Instrumentation, Methods and Data Systems*, 5(2):271–279, 2016. doi: 10.5194/gi-5-271-2016.
- E. Kivimäki, H. Lindqvist, J. Hakkarainen, M. Laine, R. Sussmann, A. Tsuruta, R. Detmers, N. Deutscher, E. Dlugokencky, F. Hase, O. Hasekamp, R. Kivi, I. Morino, J. Notholt, D. Pollard, C. Roehl, M. Schneider, M. Sha, V. Velazco, T. Warneke, D. Wunch, Y. Yoshida, and J. Tamminen. Evaluation and analysis of the seasonal cycle and variability of the trend from gosat methane retrievals. *Remote Sens*, 2019. doi: 10.3390/rs11070882.
- V. Kolehmainen, T. Tarvainen, S. Arridge, and J. Kaipio. Marginalization of uninteresting distributed parameters in inverse problems - application to diffuse optical tomography. *International Journal for Uncertainty Quantification*, 1, 01 2011. doi: 10.1615/Int.J.UncertaintyQuantification.v1.i1.10.
- M. Laine. Introduction to dynamic linear models for time series analysis. In J.-P. Montillet and M. Bos, editors, *Geodetic Time Series Analysis in Earth Sciences*, Springer Geophysics, pages 139–156. Springer, 2019. ISBN 978-3-030-21717-4. doi: 10.1007/978-3-030-21718-1_4.
- M. Laine and J. Tamminen. Aerosol model selection and uncertainty modelling by adaptive MCMC technique. *Atmospheric Chemistry and Physics*, 8(24):7697–7707, 2008. doi: 10.5194/acp-8-7697-2008.

- M. Laine, N. Latva-Pukkila, and E. Kyrölä. Analyzing time varying trends in stratospheric ozone time series using state space approach. *Atmospheric Chemistry and Physics*, 13:20503–20530, 2014. doi: 10.5194/acpd-13-20503-2013.
- A. Lehtikoinen, S. Finsterle, A. Voutilainen, L. Heikkinen, M. Vauhkonen, and J. Kaipio. Approximation errors and truncation of computational domains with application to geophysical tomography. *Inverse Problems and Imaging*, 2, 05 2007. doi: 10.3934/ipi.2007.1.371.
- K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164 – 168, 1944. ISSN 0021-9991.
- J. S. Liu and R. Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044, 1998.
- M. Loève. *Probability Theory II*. Springer, 4th edition edition, 1978. ISBN 978-0-387-90262-3.
- D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, 11(2):431 – 441, 1963. ISSN 2168-3484. doi: doi:10.1137/0111030.
- Y. M. Marzouk and H. N. Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862–1902, 2009. ISSN 0021-9991. doi: <http://dx.doi.org/10.1016/j.jcp.2008.11.024>.
- Jerry M. Melillo, Terese (T.C.) Richmond, and Gary W. Yohe. Climate change impacts in the united states: The third national climate assessment, 2014.
- N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: doi:10.1063/1.1699114.
- H. Migon, D. Gamerman, H. Lopes, and M. Ferreira. *Dynamic models, in: Handbook of Statistics, vol. 25, Bayesian Thinking: Modeling and Computation*. Elsevier, 2005. doi: 10.1016/S0169-7161(05)25019-8.
- C. E. Miller, D. Crisp, P. L. DeCola, S. C. Olsen, J. T. Randerson, A. M. Michalak, A. Alkhaled, P. Rayner, D. J. Jacob, P. Suntharalingam, D. B. A. Jones, A. S.

- Denning, M. E. Nicholls, S. C. Doney, S. Pawson, H. Boesch, B. J. Connor, I. Y. Fung, D. O'Brien, R. J. Salawitch, S. P. Sander, B. Sen, P. Tans, G. C. Toon, P. O. Wennberg, S. C. Wofsy, Y. L. Yung, and R. M. Law. Precision requirements for space-based data. *Journal of Geophysical Research: Atmospheres*, 112(D10314), 2007. doi: 10.1029/2006JD007659.
- J. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*. Society for Industrial and Applied Mathematics, Philadelphia, 2012.
- R. Nelson. The Impact of Aerosols on Space-Based Retrievals of Carbon Dioxide. Master's thesis, Colorado State University, Fort Collins, Colorado, 2015.
- H. Nguyen, N. Cressie, and J. Hobbs. Sensitivity of optimal estimation satellite retrievals to misspecification of the prior mean and covariance, with application to oco-2 retrievals. *Remote Sens.*, 2019. doi: 10.3390/rs11232770.
- C. W. O'Dell, B. Connor, H. Bösch, D. O'Brien, C. Frankenberg, R. Castano, M. Christi, D. Eldering, B. Fisher, M. Gunson, J. McDuffie, C. E. Miller, V. Natraj, F. Oyafuso, I. Polonsky, M. Smyth, T. Taylor, G. C. Toon, P. O. Wennberg, and D. Wunch. The ACOS CO₂ retrieval algorithm - Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Techniques*, 5(1): 99–121, 2012. doi: 10.5194/amt-5-99-2012.
- C. W. O'Dell, A. Eldering, P. O. Wennberg, D. Crisp, M. R. Gunson, B. Fisher, C. Frankenberg, M. Kiel, H. Lindqvist, L. Mandrake, A. Merrelli, V. Natraj, R. R. Nelson, G. B. Osterman, V. H. Payne, T. E. Taylor, D. Wunch, B. J. Drouin, F. Oyafuso, A. Chang, J. McDuffie, M. Smyth, D. F. Baker, S. Basu, F. Chevallier, S. M. R. Crowell, L. Feng, P. I. Palmer, M. Dubey, O. E. García, D. W. T. Griffith, F. Hase, L. T. Iraci, R. Kivi, I. Morino, J. Notholt, H. Ohyama, C. Petri, C. M. Roehl, M. K. Sha, K. Strong, R. Sussmann, Y. Te, O. Uchino, and V. A. Velazco. Improved retrievals of carbon dioxide from orbiting carbon observatory-2 with the version 8 acos algorithm. *Atmospheric Measurement Techniques*, 11(12):6539–6576, 2018. doi: 10.5194/amt-11-6539-2018.
- D. E. Pollard. *A user's guide to measure theoretic probability*. Cambridge University Press, 2002. ISBN 0-521-00289-3.
- B. D. Ripley. *Stochastic Simulation*. John Wiley & Sons, Ltd., 1987. ISBN 9780471818847.

- G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *Ann. Appl. Probab.*, 7:110–120, 1997. doi: 10.1214/aoap/1034625254.
- C. D. Rodgers. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific, Singapore, 2000.
- W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 3 edition, 1987. ISBN 0-07-054234-1.
- M. Saunio, A. R. Stavert, B. Poulter, P. Bousquet, J. G. Canadell, R. B. Jackson, P. A. Raymond, E. J. Dlugokencky, S. Houweling, P. K. Patra, P. Ciais, V. K. Arora, D. Bastviken, P. Bergamaschi, D. R. Blake, G. Brailsford, L. Bruhwiler, K. M. Carlson, M. Carrol, S. Castaldi, N. Chandra, C. Crevoisier, P. M. Crill, K. Covey, C. L. Curry, G. Etiope, C. Frankenberg, N. Gedney, M. I. Hegglin, L. Höglund-Isaksson, G. Hugelius, M. Ishizawa, A. Ito, G. Janssens-Maenhout, K. M. Jensen, F. Joos, T. Kleinen, P. B. Krummel, R. L. Langenfelds, G. G. Laruelle, L. Liu, T. Machida, S. Maksyutov, K. C. McDonald, J. McNorton, P. A. Miller, J. R. Melton, I. Morino, J. Müller, F. Murgia-Flores, V. Naik, Y. Niwa, S. Noce, S. O'Doherty, R. J. Parker, C. Peng, S. Peng, G. P. Peters, C. Prigent, R. Prinn, M. Ramonet, P. Regnier, W. J. Riley, J. A. Rosentreter, A. Segers, I. J. Simpson, H. Shi, S. J. Smith, L. P. Steele, B. F. Thornton, H. Tian, Y. Tohjima, F. N. Tubiello, A. Tsuruta, N. Viovy, A. Voulgarakis, T. S. Weber, M. van Weele, G. R. van der Werf, R. F. Weiss, D. Worthy, D. Wunch, Y. Yin, Y. Yoshida, W. Zhang, Z. Zhang, Y. Zhao, B. Zheng, Q. Zhu, Q. Zhu, and Q. Zhuang. The global methane budget 2000–2017. *Earth System Science Data Discussions*, 2019: 1–136, 2019. doi: 10.5194/essd-2019-128.
- S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng, editors. *Handbook of Markov chain Monte Carlo*. Chapman & Hall/CRC, 2011.
- S. L. Scott. Bayesian methods for hidden markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association*, 97:337–351, 2002.
- A. Solonen, T. Cui, J. Hakkarainen, and Y. Marzouk. On dimension reduction in Gaussian filters. *Inverse Problems*, 32(4):045003, 2016. doi: 10.1088/0266-5611/32/4/045003.

- A. Spantini, A. Solonen, T. Cui, J. Martin, L. Tenorio, and Y. Marzouk. Optimal low-rank approximations of Bayesian linear inverse problems. *SIAM Journal on Scientific Computing*, 37(6):A2451–A2487, 2015. doi: 10.1137/140977308.
- A. M. Stuart. Inverse problems: A bayesian perspective. *Acta Numerica*, 19:451–559, 2010. doi: 10.1017/S0962492910000061.
- S. Särkkä. *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks, Cambridge University Press, 2013.
- J. Tamminen. *Adaptive Markov chain Monte Carlo algorithms with geophysical applications*. PhD thesis, University of Helsinki Department of Mathematics and Statistics, 2004.
- J. Tamminen and E. Kyrölä. Bayesian solution for nonlinear and non-Gaussian inverse problem by Markov chain Monte Carlo method. *Journal of Geophysical Research*, 106(D13):14377–14390, 2001.
- A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. 2005. ISBN 9780898715729.
- S. Tukiainen, J. Railo, M. Laine, J. Hakkarainen, R. Kivi, P. Heikkinen, H. Chen, and J. Tamminen. Retrieval of atmospheric CH₄ profiles from Fourier transform infrared data using dimension reduction and MCMC. *Journal of Geophysical Research: Atmospheres*, 121:10312–10327, 2016. doi: 10.1002/2015JD024657.
- D. Wunch, G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, and P. O. Wennberg. The Total Carbon Column Observing Network. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 369(1943):2087–2112, 2011. ISSN 1364-503X. doi: 10.1098/rsta.2010.0240.
- D. Wunch, G. C. Toon, V. Sherlock, N. M. Deutscher, X. Liu, D. G. Feist, and P. O. Wennberg. The total carbon column observing network's GGG2014 data version. Technical report, Oak Ridge, Tennessee, U.S.A.: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, 2015.
- O. Zahm, T. Cui, K. Law, A. Spantini, and Y. Marzouk. Certified dimension reduction in nonlinear Bayesian inverse problems. working paper, July 2018. URL <https://hal.archives-ouvertes.fr/hal-01834039>.

PAPER I

Otto Lamminpää, Marko Laine, Simo Tukiainen, Johanna Tamminen

Likelihood informed dimension reduction for remote sensing of atmospheric constituent profiles

2017 MATRIX Annals, 2:65-78, 2019. doi: 10.1007/978-3-030-04161-8.

Copyright © Authors 2019. CC Attribution 3.0 License. Reprinted with permission.

Likelihood informed dimension reduction for remote sensing of atmospheric constituent profiles

Otto Lamminpää, Marko Laine, Simo Tukiainen, Johanna Tamminen

Abstract We use likelihood informed dimension reduction (LIS) [2] for inverting vertical profile information of atmospheric methane from ground based Fourier transform infrared (FTIR) measurements at Sodankylä, Northern Finland. The measurements belong to the world wide TCCON network for greenhouse gas measurements and, in addition to providing accurate greenhouse gas measurements, they are important for validating satellite observations.

LIS allows construction of an efficient Markov chain Monte Carlo sampling algorithm that explores only a reduced dimensional space but still produces a good approximation of the original full dimensional Bayesian posterior distribution. This in effect makes the statistical estimation problem independent of the discretization of the inverse problem. In addition, we compare LIS to a dimension reduction method based on prior covariance matrix truncation used earlier [16].

1 Introduction

Atmospheric composition measurements have an increasingly crucial role in monitoring the greenhouse gas concentrations in order to understand and predict changes in climate. The warming effect of greenhouse gases, such as carbon dioxide (CO_2) and methane (CH_4), is based on the absorption of electromagnetic radiation originating from the sun by these trace gases. This mechanism has a strong theoretical base and has been confirmed by recent observations [4].

Remote sensing measurements of atmospheric composition, and greenhouse gases in particular, are carried out by ground-based Fourier transform infrared (FTIR) spectrometers, and more recently by a growing number of satellites (for example SCIAMACHY, ACE-FTS, GOSAT, OCO-2). The advantage of satellite

Otto Lamminpää · Marko Laine · Simo Tukiainen · Johanna Tamminen
Finnish Meteorological Institute, P.O. BOX 503, FI-00101 Helsinki, Finland. e-mail: otto.lamminpaa@fmi.fi

measurements is that they provide global coverage. They are used for anthropogenic emission monitoring, detecting trends in atmospheric composition and studying the effects of biosphere, to name but a few examples. Accurate ground-based measurements are crucial to satellite measurement validation, and the global Total Carbon Column Observing Network (TCCON [17]) of FTIR spectrometers, consisting of around 20 measurement sites around the world, is widely used as a reference [3]. The FTIR instrument looks directly at sun, returning an absorption spectrum as measured data.

Determining atmospheric gas density profiles, or *retrieval*, from the absorption spectra is an ill-defined *inverse problem* as the measurement contains only a limited amount of information about the state of the atmosphere. Based on prior knowledge and using the Bayesian approach to regularize the problem, the profile retrieval is possible, provided that our prior accurately describes the possible states that may occur in the atmosphere. When retrieving a vertical atmospheric profile, the dimension of the estimation problem depends on the discretization. For accurate retrievals a high number of layers are needed, leading to a computationally costly algorithms. However, fast methods are required for the operational algorithm. For this purpose, different ways of reducing the dimension of the problem have been developed. The official operational TCCON GGG algorithm [17] solves the inverse problem by scaling the prior profile based on the measured data. This method is robust and computationally efficient, but only retrieves one piece of information and thus can give largely inaccurate results about the density profiles.

An improved dimension reduction method for the FTIR retrieval based on reducing the rank of the prior covariance matrix was used by Tukiainen et al. [16] using computational methods developed by Solonen et al. [13]. This method confines the solution to a subspace spanned by the non-negligible eigenvectors of the prior covariance matrix. This approach allows a retrieval using more basis functions than the operational method and thus gives more accurate solutions. However, the prior has to be hand tuned to have a number of non-zero singular values that correspond to the number of degrees of freedom for the signal in the measurement. Moreover, whatever information lies in the complement of this subspace remains unused.

In this work, we introduce an analysis method for determining the number of components the measurement can provide information from [12], as well as the *likelihood informed subspace* dimension reduction method for non-linear statistical inverse problems [2, 14]. We show that these two formulations are in fact equal. We then proceed to implement a dimension reduction scheme for the FTIR inverse problem using adaptive MCMC sampling [6, 5] to fully characterize the non-linear posterior distribution, and show that this method gives an optimal result with respect to Hellinger distance to the non-approximated full dimensional posterior distribution. In contrast with the previously implemented prior reduction method, the likelihood informed subspace method is also shown to give the user freedom to use a prior derived directly from an ensemble of previously conducted atmospheric composition measurements.

2 Methodology

We consider the atmospheric composition measurement carried out at the FMI Arctic Research Centre, Sodankylä, Finland [9]. The on-site Fourier transform infrared spectrometer (FTIR) measures solar light arriving to the device directly from the sun, or more precisely, the absorption of solar light at different wavelengths within the atmosphere. From the absorption spectra of different trace gases ($\text{CO}_2, \text{CH}_4, \dots$) we can compute the corresponding vertical density profiles, i.e. the fraction of the trace gas in question as a function of height.

Let us consider the absorption spectrum with m separate wavelengths. The solar light passing through the atmosphere and hitting the detector can be modeled using the *Beer-Lambert law*, which gives, for wavelengths $\lambda_j, j \in [1, \dots, m]$, the intensity of detected light as

$$I(\lambda_j) = I_0(\lambda_j) \exp\left(-\sum_{k=1}^K \int_0^\infty \mathcal{C}_k(\lambda_j, z) \rho_k(z) dz\right) (a\lambda_j^2 + b\lambda_j + c) + d, \quad (1)$$

where I_0 is the intensity of solar light when it enters the atmosphere, the atmosphere has K absorbing trace gases, $\mathcal{C}_k(\lambda_j, z)$ is the absorption coefficient of gas k , which depends on height z and on the wavelength λ_j , and $\rho_k(z)$ is the density of gas k at height z . The second degree polynomial and the constant d in (1) are used to describe instrument related features and the continuity properties of the spectrum. In reality, solar light is scattered on the way by atmospheric particles. This phenomenon is relatively weak in the wavelength band we are considering in this work, so it will be ignored for simplicity.

The absorption in continuous atmosphere is modeled by discretizing the integral in equation (1) into a sum over atmospheric layers and assuming a constant absorption for each separate layer. This way, a discrete computational *forward model* can be constructed, giving an absorption spectrum as data produced by applying the forward model to a state vector x describing the discretized atmospheric density profile for a certain trace gas. In this work, we limit ourselves to consider the retrieval of atmospheric methane (CH_4).

2.1 Bayesian formulation of the inverse problem

Consider an inverse problem of estimating unknown parameter vector $x \in \mathbb{R}^n$ from observation $y \in \mathbb{R}^m$,

$$y = F(x) + \varepsilon, \quad (2)$$

where our physical model is describe by the *forward model* $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and the random variable $\varepsilon \in \mathbb{R}^m$ represents the observation error arising from instrument noise and forward model approximations. In the Bayesian approach to inverse problems [7] our uncertainty about x is described by statistical distributions. The solution

to the problem is obtained as posterior distribution of x conditioned on a realization of the data y and depending on our prior knowledge. By the Bayes' formula, we have

$$\pi(x|y) \propto \pi(y|x)\pi_{pr}(x), \quad (3)$$

where $\pi(x|y)$ is the posterior distribution, $\pi(y|x)$ the likelihood and $\pi_{pr}(x)$ the prior distribution. The proportionality \propto comes from a constant that does not depend on the unknown x . In this work, we assume the prior to be Gaussian, $\mathcal{N}(x_0, \Sigma_{pr})$, e.g.

$$\pi_{pr}(x) \propto \exp\left(-\frac{1}{2}(x-x_0)^T \Sigma_{pr}^{-1}(x-x_0)\right). \quad (4)$$

Also, the additive noise is assumed to be zero-mean Gaussian with known covariance matrix, $\varepsilon \sim \mathcal{N}(0, \Sigma_{obs})$, so the likelihood will have form

$$\pi(y|x) \propto \exp\left(-\frac{1}{2}(y-F(x))^T \Sigma_{obs}^{-1}(y-F(x))\right). \quad (5)$$

When the forward model is non-linear, the posterior distribution can be explored by Markov chain Monte Carlo (MCMC) sampling. When the dimension of the unknown is high, for example by discretization of the inverse problem, MCMC is known to be inefficient. In this paper, we utilize dimension reduction to be able to make MCMC more efficient in high dimensional and high CPU problems.

2.2 Prior reduction

The operational GGG algorithm for the FTIR retrieval problem [17] is effectively one dimensional as it only scales the prior mean profile. However, there are about three degrees of freedom in the FTIR signal for the vertical profile information. To construct basis functions that could utilize this information a method that uses prior reduction was developed in [16]. It is based on the singular value decomposition on the prior covariance matrix,

$$\Sigma_{pr} = U\Lambda U^T = \sum_{i=1}^m \lambda_i u_i u_i^T, \quad (6)$$

which allows further decomposition as

$$\Sigma_{pr} = PP^T, \text{ with } P = \left(\sqrt{\lambda_1}u_1 + \dots + \sqrt{\lambda_m}u_m\right). \quad (7)$$

If the prior can be chosen so that most of the singular values are negligible, then the rank of the prior covariance matrix can be reduced by considering only the first r singular values and vectors:

$$\tilde{\Sigma}_{pr} = P_r P_r^T, \text{ with } P_r = \left(\sqrt{\lambda_1} u_1 + \dots + \sqrt{\lambda_r} u_r \right). \quad (8)$$

The unknown x has an approximate representation by r basis vectors from the columns of P_r and using a reduced dimensional parameter $\alpha \in \mathbb{R}^r$ as

$$x \approx x_0 + P_r \alpha. \quad (9)$$

By the construction, the random vector α has a simple Gaussian prior, $\alpha \sim \mathcal{N}(0, \mathbb{I})$, which allow us to write the approximate posterior as

$$\pi(x|y) \approx \tilde{\pi}(\alpha|y) \propto \exp \left(-\frac{1}{2} \left((y - F(x_0 + P_r \alpha))^T \Sigma_{obs}^{-1} (y - F(x_0 + P_r \alpha)) + \alpha^T \alpha \right) \right). \quad (10)$$

Now, instead running MCMC in the full space defined by x , we can sample the low dimensional parameter α and retain the approximation of the full posterior by equation (9).

2.3 Likelihood-informed subspace

The prior reduction approach depends on the ability to construct a realistic prior that can be described by only a few principle components. For the FTIR retrieval problem this is possible to some extent [16]. However, there are several possible caveats. We have to manually manipulate the prior covariance matrix to have a lower rank, which can lead to information loss as the solution will be limited to a subspace defined by the reduced prior only.

In atmospheric remote sensing the information content of the measurement is an important concept to be considered when designing the instruments and constructing the retrieval methodology, we refer to book by Rodgers [12].

Consider a linearized version of the inverse problem in equation (2),

$$y = J(x - x_0) + \varepsilon, \quad (11)$$

with Gaussian prior and noise. The forward model is assumed to be differentiable, and J denotes the Jacobian matrix of the forward model with elements $J_{ij} = \frac{\partial}{\partial x_j} F_i$. Using Cholesky factorizations for the known prior and error covariances,

$$\Sigma_{pr} = \mathcal{L}_{pr} \mathcal{L}_{pr}^T, \quad \Sigma_{obs} = \mathcal{L}_{obs} \mathcal{L}_{obs}^T, \quad (12)$$

we can perform pre-whitening of the problem by setting

$$\tilde{y} = \mathcal{L}_{obs}^{-1} y, \quad \tilde{J} = \mathcal{L}_{obs}^{-1} J \mathcal{L}_{pr}, \quad \tilde{x} = \mathcal{L}_{pr}^{-1} (x - x_0) \text{ and } \tilde{\varepsilon} = \mathcal{L}_{obs}^{-1} \varepsilon. \quad (13)$$

Now the problem can be written as

$$\tilde{y} = \tilde{J}\tilde{x} + \tilde{\varepsilon}, \quad (14)$$

with $\tilde{\varepsilon} \sim \mathcal{N}(0, \mathbb{I})$ and a priori $\tilde{x} \sim \mathcal{N}(0, \mathbb{I})$.

As the unknown x and the error ε are assumed to be independent, the same holds for the scaled versions. We can compare the prior variability of the observation depending on x and that coming from the noise ε by

$$\tilde{\Sigma}_y = \mathbb{E}[\tilde{y}\tilde{y}^T] = \mathbb{E}[(\tilde{J}\tilde{x} + \tilde{\varepsilon})(\tilde{J}\tilde{x} + \tilde{\varepsilon})^T] = \tilde{J}\tilde{J}^T + \mathbb{I}. \quad (15)$$

The variability in y that depends only on the parameter x depends itself on $\tilde{J}\tilde{J}^T$ and it can be compared to the unit matrix \mathbb{I} that has the contribution from the scaled noise. The directions in $\tilde{J}\tilde{J}^T$ which are larger than unity are those dominated by the signal. Formally this can be seen by diagonalizing the scaled problem by the singular value decomposition,

$$\tilde{J} = W\Lambda V^T, \quad (16)$$

and setting

$$y' = W^T\tilde{y} = W^T\tilde{J}\tilde{x} + W^T\tilde{\varepsilon} = \Lambda V^T\tilde{x} + \tilde{\varepsilon}' = \Lambda x' + \tilde{\varepsilon}'. \quad (17)$$

The transformations ε' and x' conserve the unit covariance matrix. In other words, y' is distributed with covariance $\Lambda^2 + \mathbb{I}$. This is a diagonal matrix, and the elements of vector y' that are not masked by the measurement error are those corresponding to the singular values $\lambda_i \geq 1$ of the pre-whitened Jacobian \tilde{J} . Furthermore, degrees of freedom for signal and noise are invariant under linear transformations [12], so the same result is also valid for the original y .

Another way to compare the information content of the measurement relative to the prior was used in [2]. This is to use the Rayleigh quotient

$$\mathcal{R}(\mathcal{L}_{pr}a) = \frac{a^T \mathcal{L}_{pr}^T H \mathcal{L}_{pr} a}{a^T a}, \quad (18)$$

where $a \in \mathbb{R}^n$ and $H = J^T \Sigma_{obs}^{-1} J$ is the Gauss-Newton approximation of Hessian matrix of the data misfit function

$$\eta(x) = \frac{1}{2} ((y - F(x))^T \Sigma_{obs}^{-1} (y - F(x))). \quad (19)$$

Directions for which $\mathcal{R}(\mathcal{L}_{pr}a) > 1$ are the ones in which the likelihood contains information relative to the prior. This follows from the fact that the i th eigenvector v_i of the prior-preconditioned Gauss-Newton Hessian

$$\tilde{H} := \mathcal{L}_{pr}^T H \mathcal{L}_{pr} \quad (20)$$

maximizes the Rayleigh quotient over a subspace $\mathbb{R}^n \setminus \text{span}\{v_1, \dots, v_{i-1}\}$ and the r directions v_i for which $\mathcal{R}(\mathcal{L}_{pr}v) > 1$ correspond to the first r eigenvalues of \tilde{H} . We call these vectors the *informative directions of the measurement*.

To see the correspondence for the two approaches for the informative directions we notice that for $\tilde{H}(x)$ it holds that

$$\begin{aligned}\mathcal{L}_{pr}^T H(x) \mathcal{L}_{pr} &= \mathcal{L}_{pr}^T J(x)^T \Sigma_{obs}^{-1} J(x) \mathcal{L}_{pr}^T \\ &= (\mathcal{L}_{obs}^{-1} J(x) \mathcal{L}_{pr})^T (\mathcal{L}_{obs}^{-1} J(x) \mathcal{L}_{pr}) \\ &= \tilde{J}^T(x) \tilde{J}(x).\end{aligned}\quad (21)$$

The eigenvalues λ^2 of matrix $\tilde{H}(x)$ less than unity correspond to the singular values λ less than unity of the scaled Jacobian $\tilde{J}(x)$. The corresponding eigenvectors are the same as the right singular vectors v of \tilde{J} . The informative and non-informative directions for a simple 2-dimensional Gaussian case are illustrated in Figure 1.

Next, we use the informative directions of the measurement to reduce the dimension of the inverse problem. Consider approximations for the posterior of the form

$$\tilde{\pi}(x|y) \propto \pi(y|\Pi_r x) \pi_{pr}(x), \quad (22)$$

where Π_r is rank r projection matrix. In [2] and [14] it was shown that for any given r , there exists a unique optimal projection Π_r that minimizes the Hellinger distance between the approximative rank r posterior and the full posterior. Furthermore, using the connection to Rodgers' formalism, the optimal projection can be obtained explicitly with the following definition.

Definition 1 (LIS). Let $V_r \in \mathbb{R}^{n \times r}$ be a matrix containing the first r left singular vectors of the scaled Jacobian \tilde{J} . Define

$$\Phi_r := \mathcal{L}_{pr} V_r \text{ and } \Theta_r := \mathcal{L}_{pr}^{-T} V_r. \quad (23)$$

The rank r LIS projection for the posterior approximation (22) is given by

$$\Pi_r = \Phi_r \Theta_r^T. \quad (24)$$

The range \mathbb{X}_r of projection $\Pi_r : \mathbb{R}^n \rightarrow \mathbb{X}_r$ is a subspace of state space \mathbb{R}^n spanned by the column vectors of matrix Φ_r . We call the subspace \mathbb{X}_r the *likelihood-informed subspace (LIS)* for the linear inverse problem, and its complement $\mathbb{R}^n \setminus \mathbb{X}_r$ the *complement subspace (CS)*.

Definition 2. The matrix of singular vectors $V = [V_r, V_\perp]$ forms a complete orthonormal system in \mathbb{R}^n and we can define

$$\Phi_\perp := \mathcal{L}_{pr} V_\perp \text{ and } \Theta_\perp := \mathcal{L}_{pr}^{-T} V_\perp \quad (25)$$

and the projection $\mathbb{I} - \Pi_r$ can be written as

$$\mathbb{I} - \Pi_r = \Phi_\perp \Theta_\perp^T. \quad (26)$$

Define the LIS-parameter $x_r \in \mathbb{R}^r$ and the CS-parameter $x_\perp \in \mathbb{R}^{n-r}$ as

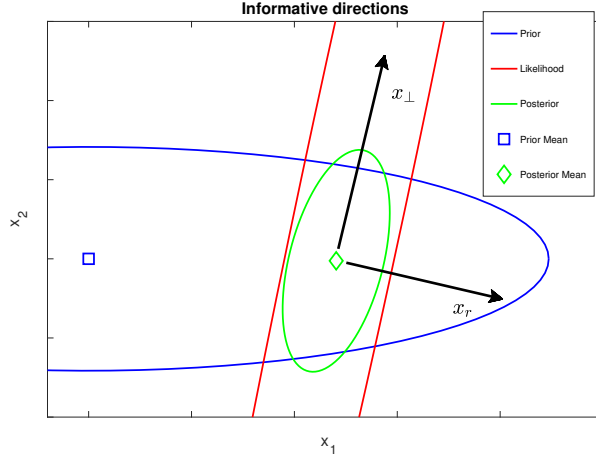


Fig. 1 Illustration of an informative direction x_r and a non-informative direction x_\perp using a 2-dimensional Gaussian case. Here, the likelihood has only one informative component, so the remaining direction for the posterior is obtained from the prior.

$$x_r := \Theta_r^T x, \quad x_\perp := \Theta_\perp^T x. \quad (27)$$

The parameter x can now be naturally decomposed as

$$\begin{aligned} x &= \Pi_r x + (\mathbb{I} - \Pi_r)x \\ &= \Phi_r x_r + \Phi_\perp x_\perp. \end{aligned} \quad (28)$$

Using this decomposition and properties of multivariate Gaussian distributions, we can write the prior as

$$\pi_{pr}(x) = \pi_r(x_r) \pi_\perp(x_\perp) \quad (29)$$

and approximate the likelihood by using the r informative directions,

$$\pi(y|x) = \pi(y|\Phi_r x_r) \pi(y|\Phi_\perp x_\perp) \approx \pi(y|\Phi_r x_r), \quad (30)$$

which leads us to the approximate posterior

$$\tilde{\pi}(x|y) = \pi(y|\Phi_r x_r) \pi_r(x_r) \pi_\perp(x_\perp). \quad (31)$$

When the forward model is not linear, the Jacobian and Hessian matrices depend on the parameter x and the criterion (18) only holds point wise. To extend this local condition into a global one, we consider the expectation of the local Rayleigh quotient $\mathcal{R}(\mathcal{L}_{pr} v; x)$ over the posterior,

$$\mathbb{E}[\mathcal{R}(\mathcal{L}_{pr}; v; x)] = \frac{v^T \hat{J}^T \hat{J} v}{v^T v}, \quad \hat{J} = \int_{\mathbb{R}^n} \tilde{J}(x) \pi(x|y) dx. \quad (32)$$

The expectation is with respect to the posterior distribution, which is not available before the analysis. In practice, an estimate is obtained by Monte Carlo,

$$\hat{J}_n = \frac{1}{n} \sum_{k=1}^n \tilde{J}(x^{(k)}), \quad (33)$$

where $x^{(k)}$ is a set of samples from some reference distribution which will be discussed later in this work. We can now use the singular value decomposition $\hat{J}_n = W A V^T$ to find a basis for the global LIS analogously to the linear case.

The advantage of LIS dimension reduction is that it is sufficient to use MCMC to sample the low-dimensional x_r from the reduced posterior $\pi(y|\Phi_r x_r) \pi_r(x_r)$, and form the full space approximation using the known analytic properties of the Gaussian complement prior $\pi_{\perp}(x_{\perp})$.

3 Results

To solve the inverse problem related to the FTIR measurement [16], we use adaptive MCMC [5, 10] and SWIRLAB [15] toolboxes for Matlab. The results from newly implemented LIS-algorithm as well as from the previous prior reduction method are compared against a full dimensional MCMC simulation using the Hellinger distance of approximations to the full posterior. We use a prior derived from an ensemble of atmospheric composition measurements by the ACE satellite [1]. The vertical prior distribution, prior covariance and prior singular values are illustrated in Figure 2.

In Figure 3, we show the results of our retrievals using full-space MCMC, compared with LIS dimension reduction and prior reduction using 4 basis vectors in each method. The retrievals are further compared against accurate in-situ measurements made using AirCore balloon soundings [8] which are available for the selected cases, also included in Figure 3. In this example, the Monte-Carlo estimator (33) for \hat{J}_n in equation (33) was computed using 1000 samples drawn from the Laplace approximation $\mathcal{N}(\hat{x}, \hat{\Sigma}_{post})$, where \hat{x} and $\hat{\Sigma}_{post}$ are the posterior MAP and covariance, respectively, obtained using optimal estimation [12].

In order to compare the performance of MCMC methods, we define the *sample speed* of a MCMC run as

Definition 3. The *effective sample size* N_{eff} of a MCMC chain is given by

$$N_{\text{eff}} = \frac{N_M}{1 + s \sum_{k=1}^{\infty} \rho_k(x)}, \quad (34)$$

where N_M is the length of the MCMC chain and $\rho_k(x)$ is lag- k autocorrelation for parameter x [11]. Define the *sample speed* of an MCMC chain as

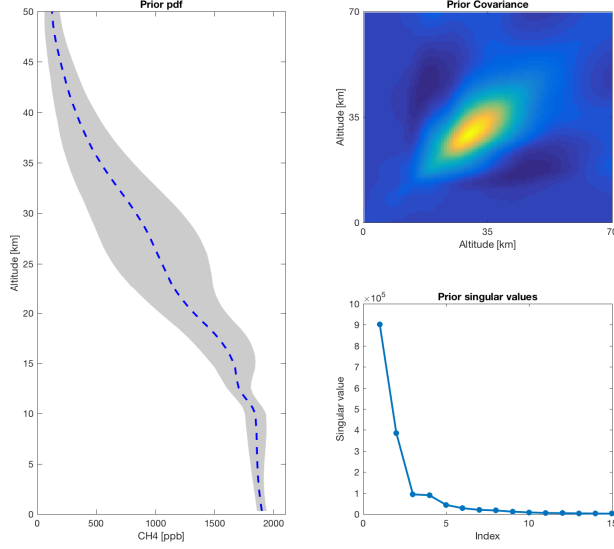


Fig. 2 The prior derived from an ensemble of ACE satellite measurements. Left: Full prior profile, mean with dashed line and 95% probability limits in grey. Top right: covariance matrix derived from the measurements. Bottom right: first 20 singular values of the prior covariance matrix.

$$\mathbb{V} = \frac{N_{\text{eff}}}{t_M}, \quad (35)$$

where t_M is the total computation time of the MCMC chain.

For the MCMC runs shown in Figure 3, we get as corresponding sample speeds as samples per second:

$$\mathbb{V}(\text{full}) = 1.56 \text{ s}^{-1}, \quad \mathbb{V}(\text{LIS}) = 19.01 \text{ s}^{-1}, \quad \mathbb{V}(\text{PriRed}) = 19.66 \text{ s}^{-1}. \quad (36)$$

In order to compare the approximate posteriors obtained from prior reduction and LIS-dimension reduction against the full posterior, we use the discrete Hellinger distance,

$$\mathcal{H}(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{p_i} - \sqrt{q_i})^2}, \quad (37)$$

where $P = (p_1, \dots, p_k)$ and $Q = (q_1, \dots, q_i)$ are discrete representations of the full and approximate posterior distributions obtained from histograms of corresponding

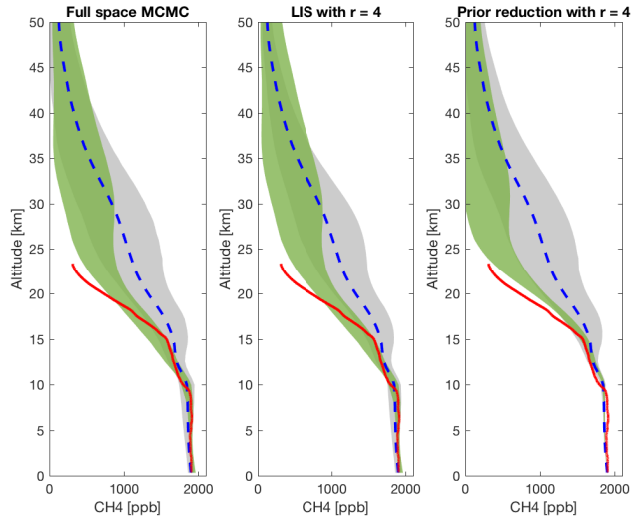


Fig. 3 Atmospheric CH_4 density profile retrieval results. Retrieved posterior in green, prior in gray, and in-situ AirCore measurement in red. The color shading indicates areas where 95% of the profiles are. Right: MCMC with in full space. Middle: MCMC with LIS. Right: MCMC with prior reduction.

MCMC runs. The Hellinger distances of both approximations to the full posterior can be seen in Figure 4 together with the corresponding sample speeds, both as a function of the number of singular vectors used. In Figure 4 have also visualized the first four singular vectors used in prior reduction and LIS method for the example retrieval in Figure 3.

4 Conclusions

Although both of the discussed dimension reduction methods provide roughly the same computational gains in the performance of the MCMC sampler, we see from Figure 4 that while using an empirical prior, the prior reduction method requires a lot more singular vectors to achieve the same Hellinger distance from the full posterior as the LIS method, which gets really close already with 4 singular vectors. We conclude that the LIS method gives an efficient MCMC sampling algorithm to solve the inverse problem arising from the FTIR retrieval, with an additional improvement of allowing the direct usage of an empirical prior.

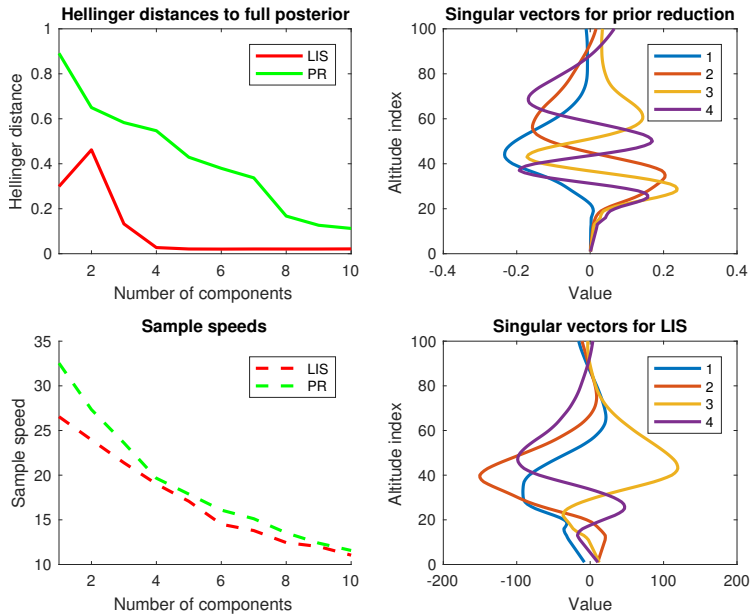


Fig. 4 Left: Hellinger distances to full posterior and sample speeds of corresponding MCMC runs as functions of singular vectors used in the approximation. Top right: first 4 singular vectors from prior reduction. Bottom right: first four singular vectors of \tilde{J} forming the LIS basis.

Acknowledgements We thank Dr. Rigel Kivi from FMI Arctic Research Centre, Sodankylä, Finland for the AirCore and TCCON data. We thank Dr. Tiangang Cui from Monash University and the mathematical research institute MATRIX in Australia for organizing a workshop where a part of this research was performed. This work has been supported by Academy of Finland (projects INQUIRE, IIDA-MARI and CoE in Inverse Modelling and Imaging) and by EU's Horizon 2020 research and innovation programme (project GAIA-CLIM).

References

1. Bernath, P.F., McElroy, C.T., Abrams, M.C., Boone, C.D., Butler, M., Camy-Peyret, C., Carleer, M., Clerbaux, C., Coheur, P.F., Colin, R., DeCola, P., DeMazière, M., Drummond, J.R., Dufour, D., Evans, W.F.J., Fast, H., Fussen, D., Gilbert, K., Jennings, D.E., Llewellyn, E.J., Lowe, R.P., Mahieu, E., McConnell, J.C., McHugh, M., McLeod, S.D., Michaud, R., Midwinter, C., Nassar, R., Nichitiu, F., Nowlan, C., Rinsland, C.P., Rochon, Y.J., Rowlands, N., Semeniuk, K., Simon, P., Skelton, R., Sloan, J.J., Soucy, M.A., Strong, K., Tremblay, P., Turnbull, D., Walker, K.A., Walkty, I., Wardle, D.A., Wehrle, V., Zander, R., Zou, J.: Atmospheric Chemistry Experiment (ACE): Mission overview. *Geophysical Research Letters*

- 32(15) (2005). DOI 10.1029/2005GL022386
2. Cui, T., Martin, J., Marzouk, Y.M., Solonen, A., Spantini, A.: Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems* **30**(11), 114,015 pp. 28 (2014). DOI 10.1088/0266-5611/30/11/114015
 3. Dils, B., Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Parker, R., Guerlet, S., Aben, I., Blumenstock, T., Burrows, J.P., Butz, A., Deutscher, N.M., Frankenberg, C., Hase, F., Hasekamp, O.P., Heymann, J., De Mazière, M., Notholt, J., Sussmann, R., Warneke, T., Griffith, D., Sherlock, V., Wunch, D.: The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparative validation of GHG-CCI CHY/ENVISAT and TANSO-FTS/GOSAT CO₂ and CH₄ retrieval algorithm products with measurements from the TCCON. *Atmospheric Measurement Techniques* **7**(6), 1723–1744 (2014). DOI 10.5194/amt-7-1723-2014
 4. Feldman, D.R., Collins, W.D., Gero, P.J., Torn, M.S., Mlawer, E.J., Shippert, T.R.: Observational determination of surface radiative forcing by CO₂ from 2000 to 2010. *Nature* **519**, 339–343 (2015). DOI 10.1038/nature14240
 5. Haario, H., Laine, M., Mira, A., Saksman, E.: DRAM: Efficient adaptive MCMC. *Statistics and Computing* **16**(4), 339–354 (2006). DOI 10.1007/s11222-006-9438-0
 6. Haario, H., Saksman, E., Tamminen, J.: An adaptive Metropolis algorithm. *Bernoulli* **7**(2), 223–242 (2001). DOI 10.2307/3318737
 7. Kaipio, J., Somersalo, E.: *Statistical and Computational Inverse Problems*. Springer-Verlag, New York (2005). DOI 10.1007/b138659
 8. Karion, A., Sweeney, C., Tans, P., Newberger, T.: AirCore: An Innovative Atmospheric Sampling System. *Journal of Atmospheric and Oceanic Technology* **27**(11), 1839–1853 (2010). DOI 10.1175/2010JTECHA1448.1
 9. Kivi, R., Heikkinen, P.: Fourier transform spectrometer measurements of column CO₂ at Sodankylä, Finland. *Geoscientific Instrumentation, Methods and Data Systems* **5**(2), 271–279 (2016). DOI 10.5194/gi-5-271-2016
 10. Laine, M.: MCMC toolbox for Matlab (2013). URL <http://helios.fmi.fi/~lainema/mcmc/>
 11. Ripley, B.D.: *Stochastic Simulation*. Wiley (1987)
 12. Rodgers, C.D.: *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific, Singapore (2000)
 13. Solonen, A., Cui, T., Hakkarainen, J., Marzouk, Y.: On dimension reduction in Gaussian filters. *Inverse Problems* **32**(4), 045,003 (2016). DOI 10.1088/0266-5611/32/4/045003
 14. Spantini, A., Solonen, A., Cui, T., Martin, J., Tenorio, L., Marzouk, Y.: Optimal low-rank approximations of Bayesian linear inverse problems. *SIAM Journal on Scientific Computing* **37**(6), A2451–A2487 (2015). DOI 10.1137/140977308
 15. Tukiainen, S.: Swirlab toolbox for Matlab (2017). URL <https://github.com/tukiains/swirlab>
 16. Tukiainen, S., Railo, J., Laine, M., Hakkarainen, J., Kivi, R., Heikkinen, P., Chen, H., Tamminen, J.: Retrieval of atmospheric CH₄ profiles from Fourier transform infrared data using dimension reduction and MCMC. *Journal of Geophysical Research: Atmospheres* **121**, 10,312–10,327 (2016). DOI 10.1002/2015JD024657
 17. Wunch, D., Toon, G.C., Sherlock, V., Deutscher, N.M., Liu, X., Feist, D.G., Wennberg, P.O.: The total carbon column observing network’s GGG2014 data version. Tech. rep., Oak Ridge, Tennessee, U.S.A.: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory (2015). DOI 10.14291/tcon.ggg2014.documentation.R0/1221662

PAPER II

Tomi Karppinen, Otto Lamminpää, Simo Tukiainen, Rigel Kivi, Pauli Heikkinen, Juha Hatakka, Marko Laine, Huilin Chen, Hannakaisa Lindqvist and Johanna Tamminen

Vertical Distribution of Arctic Methane in 2009-2018 Using Ground-Based Remote Sensing

Remote Sens. 2020, 12, 917. doi: 10.3390/rs12060917.

Copyright © Authors 2020. CC Attribution 3.0 License. Reprinted with permission.

Article

Vertical Distribution of Arctic Methane in 2009–2018 Using Ground-Based Remote Sensing

Tomi Karppinen ^{1,*}, Otto Lamminpää ², Simo Tukiainen ², Rigel Kivi ¹, Pauli Heikkinen ¹, Juha Hatakka ², Marko Laine ², Huilin Chen ³, Hannakaisa Lindqvist ² and Johanna Tamminen ²

¹ Finnish Meteorological Institute, 99600 Sodankylä, Finland; rigel.kivi@fmi.fi (R.K.); Pauli.Heikkinen@fmi.fi (P.H.)

² Finnish Meteorological Institute, 00560 Helsinki, Finland; otto.lamminpaa@fmi.fi (O.L.); simo.tukiainen@fmi.fi (S.T.); juha.hatakka@fmi.fi (J.H.); marko.laine@fmi.fi (M.L.); hannakaisa.lindqvist@fmi.fi (H.L.); Johanna.Tamminen@fmi.fi (J.T.)

³ Centre for Isotope Research, Energy and Sustainability Research Institute Groningen, University of Groningen, 9747 AG Groningen, The Netherlands; huilin.chen@rug.nl (H.C.)

* Correspondence: tomi.karppinen@fmi.fi

Received: 31 January 2020; Accepted: 10 March 2020; Published: 12 March 2020



Abstract: We analyzed the vertical distribution of atmospheric methane (CH₄) retrieved from measurements by ground-based Fourier Transform Spectrometer (FTS) instrument in Sodankylä, Northern Finland. The retrieved dataset covers 2009–2018. We used a dimension reduction retrieval method to extract the profile information, since each measurement contains around three pieces of information about the profile shape between 0 and 40 km. We compared the retrieved profiles against Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) satellite measurements and AirCore balloon-borne profile measurements. Additional comparison at the lowest tropospheric layer was done against in-situ measurements from a 50-m-high mast. In general, the ground-based FTS and ACE-FTS profiles agreed within 10% below 20 km and within 30% in the stratosphere between 20 and 40 km. Our method was able to accurately capture reduced methane concentrations inside the polar vortex in the Arctic stratosphere. The method produced similar trend characteristics as the reference instruments even when a static prior profile was used. Finally, we analyzed the time series of the CH₄ profile datasets and estimated the trend using the dynamic linear model (DLM).

Keywords: arctic methane; vertical profiles; growth rate; ground-based remote sensing; Fourier Transform Spectrometer; TCCON; dynamic linear model

1. Introduction

After carbon dioxide (CO₂), methane (CH₄) is the second most important anthropogenic greenhouse gas in the atmosphere. While methane is not as abundant, it traps about 28 times more heat per unit mass than carbon dioxide [1]. In comparison to carbon dioxide, methane is relatively short lived with a lifetime of around nine years [2]. Despite this, the concentration of atmospheric methane has more than doubled since pre-industrial times, with global average surface concentration being 1863 ppb in May 2019 [3,4].

Atmospheric methane is produced on the surface by both human activities and natural processes. Major sources of methane include agriculture, waste management, production and use of fossil fuels and different processes in the biosphere involving microorganisms and plants, especially in the wetlands [5]. Increased methane concentrations introduce a potential positive climate feedback in the Arctic in the form of emission changes from the wetlands and potential release of methane trapped in the permafrost caused by a warmer climate [6,7].

The main loss mechanism of methane, accounting for about 90% of the sink, is oxidation by the hydroxyl radical (OH), mainly in the troposphere [5]. In the troposphere, other notable sinks are oxidation in soils and loss by photochemistry in the marine boundary layer [5]. In the stratosphere, methane is lost in the reactions with OH, but also with atomic oxygen ($O(^1D)$) and chlorine (Cl). The latter two account for 3% of the total methane sink, with $O(^1D)$ being more important closer to the equator and Cl at high latitudes [8].

While the main source of methane is on the surface, the sink is spread more evenly in the troposphere. This creates a negative vertical gradient in the tropospheric profile, especially close to a source where the emitted methane is not completely mixed vertically [9,10]. The stratospheric air in the Arctic originates from equatorial troposphere from where it slowly ascends to stratosphere and is transported towards the poles [11,12]. During the ascent and transport, methane is subject to chemical loss, which causes a negative vertical gradient in the stratosphere. During winter, the temperature of the polar stratosphere decreases because of reduced solar radiation. This creates a pressure gradient between high latitudes with the polar night and mid-latitudes. This pressure gradient together with the Earth's rotation produces a circumpolar belt of westerly winds in the stratosphere called polar vortex [13]. Horizontal transport through the edge of the polar vortex is restricted and the air inside the vortex is descending. This results in an even larger gradient inside the vortex as the descending air has resided longer in the stratosphere, leading to a stronger depletion of methane [14].

The global growth rate of methane has varied in the recent decades. In the beginning of the 2000s, methane concentrations leveled off to what was suggested to be a possible steady state [15]. In 2007, however, methane concentration started to increase again [16]. The reasons for the stagnation and the renewed growth have been widely speculated but, thus far, a consensus has not been reached [5]. Saunio et al. [5] concluded that the most probable reasons for the latest rising trend are increased microbial and fossil emissions, with the tropics being the most probable region with the highest increase in emissions and the Arctic having a negligible effect. Changes in the OH sink can provide a partial explanation: Rigby et al. [17] showed that the OH sink was likely stronger during the stagnation period, which could explain the stable concentrations with a smaller decrease in emissions.

The concentrations of methane in the Arctic have been observed to follow the global behavior, increasing about 6 ppb/year (2008–2013) [7]. Since the recent increase in methane concentrations after 2007, the growth rate in the Arctic has been found to have a tendency to lag behind the rest of the world [18], suggesting that the increase is not of local or regional Arctic origin. During winter, methane levels in the Arctic are impacted by transport from mid-latitude source regions. During summer, local sources consist mostly of wetlands, confirmed by the isotopic data from Arctic measurement sites [7]. While Arctic methane emissions are expected to increase due to rising average temperatures, to date, no definitive changes in the Arctic methane emissions have been detected by the existing observational network [7].

Atmospheric methane can be measured with a variety of different techniques. Close to the ground level, a gas analyzer can be set up to sample surrounding air and analyze the methane concentration. Alternatively, an air sample can be collected to a flask and analyzed in a laboratory [19,20]. Using remote sensing, the total column of methane (XCH_4) can be retrieved measuring the absorption of direct solar radiation at near-infrared or mid-infrared wavelengths. This method is used by the Total Carbon Column Observing Network (TCCON [21]) and the Network for the Detection of Atmospheric Composition Change (NDACC [22]). The infrared radiation back-scattered by the atmosphere and the surface is used by satellite instruments, such as TROPOMI onboard Sentinel 5 Precursor satellite (S5P [23]) and TANSO-FTS onboard GOSAT [24], to retrieve XCH_4 from space.

While surface concentrations and total columns are measured continuously at numerous stations, the atmospheric profile of methane is usually measured in campaigns or rather infrequently by flying an instrument on an aircraft or using an AirCore [25]. An AirCore is a long, narrow stainless steel tube, which is lifted with an atmospheric sounding balloon up to about 30 km altitude. The tube then ingests the surrounding air while descending through the atmosphere. AirCore measurements are

crucial for vertical distribution research; most aircraft are unable to reach altitudes above roughly 13 km [26] and therefore cannot measure the stratospheric profile that is highly variable for methane, especially at high latitudes inside the polar vortex [14].

Another way to measure atmospheric profiles is using a satellite with limb viewing geometry, which allows the satellite to scan through different layers of the atmosphere as the sun (or other star) is ascending from behind the horizon or descending behind it. This method is used by the FTS instrument onboard the Atmospheric Chemistry Experiment (ACE) satellite to retrieve atmospheric profiles of CH₄ in the upper atmospheric layers. In this work, the ACE-FTS profile retrievals are used to evaluate our ground-based profile retrievals. A large ensemble of co-located ACE-FTS profiles is also used in the construction of the extra-tropospheric part of prior profiles.

In this work, we retrieved a time series of methane profiles from high-resolution short-wave infrared spectra measured by a ground-based Fourier Transform Spectrometer (FTS). Retrieval was done by modeling the temperature and pressure broadened absorption lines in short-wave infrared (SWIR) wavelength range using the Swirlab code that is publicly available [27]. Other remote sensing profile retrieval codes for ground-based FTS spectra were described and demonstrated by Hase et al. [28] and Zhou et al. [29]. The inverse problem related to the profile retrieval is ill-posed, thus, to retrieve physically meaningful profiles and to prevent artificial oscillations in the retrieved methane profile, the retrieval needs to be constrained. The other codes, PROFFIT ([28]) and SFIT4TCCON ([29]), use Tikhonov regularization [30] to obtain an appropriate prior for the retrieval. Swirlab on the other hand utilizes prior based dimension reduction [31], where the retrieved profile is constructed using only the largest singular vectors calculated from the prior covariance matrix. The prior itself is first built to reflect the mean and variability of the ACE-FTS measurements. An advantage of this method is that, in addition to regularization, the dimension of the problem is reduced, leading to a more computationally efficient retrieval algorithm [32].

The most significant advantage of the FTS measurements over the other profile measurements is the time resolution. The aircraft and balloon-borne measurements are superior in accuracy and vertical resolution. However, they are time consuming and resource intensive, and therefore not made on a daily or even weekly basis. Solar occultation measurements by ACE-FTS also have a higher vertical resolution than the ground-based remote sensing methods but the profiles cover only altitudes from roughly 10 km upwards. Due to the satellite's orbit and measurement geometry, the revisit time over Sodankylä varies between 1 and 70 days. Temporary data gaps in the ground-based FTS measurements are caused by clouds blocking the direct view of the sun. Excluding the polar night, data gaps longer than two days are very rare.

While the profile information is useful in itself, the profile retrieval method could also improve the total column measurements. The averaging kernels show better sensitivity to the variability in the stratosphere than in the case of a profile scaling method [32]. In addition, the solar zenith angle (SZA) dependence of the averaging kernels is lower, thus resulting in reduced SZA-induced biases when the true profile differs significantly from the prior profile used for the retrieval [32].

To evaluate the performance of Swirlab retrievals, the profiles were here compared to concurrent ACE-FTS profiles above 10 km altitude, AirCore balloon measurements below 30 km and to 50-m mast measurements for the lowest atmospheric level of the profile. In addition to one-to-one comparisons, a dynamic linear model (DLM) trend analysis [33] was carried out to each dataset, and the growth rates of the recent years were compared. The growth rates were also evaluated in the light of recently published Arctic methane growth rates [7,10,18].

2. Data Description

The study concentrated on data measured at or above Finnish Meteorological Institute's Arctic Space Centre in Sodankylä. The site is located at 67.37°N, 26.63°E (see Figure 1) and hosts a high-quality measurement infrastructure covering several research areas such as upper-air chemistry and physics,

atmospheric column measurements, cryospheric and hydrological processes, biosphere—atmosphere interaction and satellite calibration and validation.

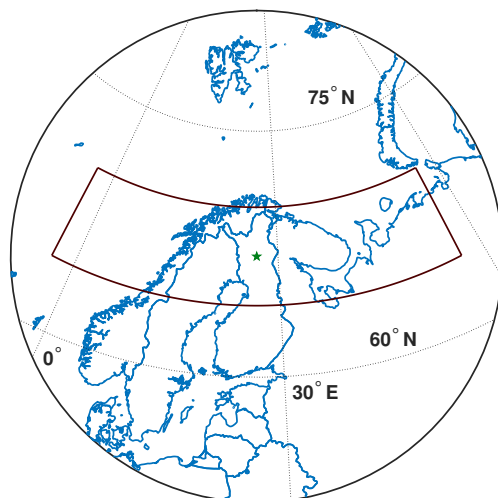


Figure 1. Sodankylä (67.367°N , 26.629°E) marked with a star and the co-location criteria (3° in latitude, 30° in longitude) for ACE-FTS satellite measurements marked with dark rectangle.

2.1. Ground-Based FTS Dataset

The Fourier Transform Spectrometer at Sodankylä became operational in 2009 [34]. The instrument participates in the Total Carbon Column Observing Network (TCCON). The TCCON facility at Sodankylä is equipped with a Bruker IFS 125HR spectrometer and a large solar tracker A547N (Bruker Optics, Ettlingen, Germany). The instrumentation is designed to record high-resolution solar spectra in the near-infrared spectral region. From the spectra, column-averaged abundances of CO_2 , CH_4 , N_2O , HF, CO, H_2O , HDO and other gases are derived.

The FTS instrument at Sodankylä has two room temperature detectors: an indium gallium arsenide (InGaAs, wavenumber range is $4000\text{--}11,000\text{ cm}^{-1}$) and a silicon diode (Si, range is $9000\text{--}15,000\text{ cm}^{-1}$). In addition, a liquid nitrogen cooled indium antimonide detector (InSb, detector covers $1800\text{--}6000\text{ cm}^{-1}$) was installed in July 2013. The additional detector expands the wavelength region covered by the instrument and provides a possibility to retrieve more atmospheric species. For stability and to keep water vapor low, the measurements are performed in vacuum. For the TCCON measurements, the spectral resolution is 0.02 cm^{-1} , with an optical path difference of 45 cm and the collection time for a single scan is 78 s. The instrument has been operated in an automated mode: after two InGaAs/Si scans, one InSb scan is taken. The instrument characterization has been performed using a hydrogen chloride gas cell.

For the TCCON retrieval, GGG2014 algorithm is used [35]. The algorithm retrieves total column dry air mole fraction of methane using three wavelength windows centered at 5938, 6002 and 6076 cm^{-1} (1684, 1666 and 1647 nm). The spectroscopy is based on a modified HITRAN2012 line list [36]. The National Centers for Environmental Protection and National Center for Atmospheric Research (NCEP/NCAR) reanalysis data are used to generate prior profiles of pressure, temperature, geopotential height and water vapor. Methane prior profiles are generated by empirical functions that include a secular increase, interhemispheric gradient, seasonal cycle and stratospheric decay based on the age of air. These functions are optimized to fit existing information from balloon-borne platforms, satellites and aircraft campaigns. The prior CH_4 profiles are iteratively scaled to generate forward-modeled spectra that best fit the measured spectra. To calculate the total column dry air mole fractions (XCH_4), the total dry air columns are calculated from retrieved O_2 columns, with a fair

assumption of constant O₂ dry air mole fraction (20.95%). An empirical correction is applied to correct the solar zenith angle (SZA) dependence, determining any variations that are symmetric about local noon as SZA-dependent artefacts [21]. The TCCON data are publicly available at the TCCON data archive, hosted by Caltech-DATA [37]. The algorithm to retrieve profile information from the same spectra is described in Section 3.1.

2.2. AirCore Measurements

The AirCore was constructed as a 100-m-long coated stainless steel tube, with a volume of about 1400 mL. A data logger was built to record the temperature of the AirCore and the ambient pressure. An automatic valve was used to close the inlet valve of the AirCore system shortly after the landing of the payload. The AirCore was lifted by a meteorological balloon to the altitude of typically 25–30 km, and was filled during descent from the stratosphere down to the surface. The gas analysis was performed typically within 2–3 h after the landing of the AirCore using a cavity ring-down spectrometer (CRDS). Calibration of the CRDS was performed before and after analyzing the sample to ensure the traceability of the measurements to the World Meteorological Organization scale of CH₄ (X2004A).

2.3. ACE Data

Atmospheric Chemistry Experiment (ACE), also known as SCISAT, is a Canadian-led satellite mission for remote sensing of the Earth's atmosphere. It was launched in August 2003. The primary instrument onboard is the infrared Fourier transform spectrometer (ACE-FTS). It measures infrared spectrum in the wavenumber range of 750–4400 cm⁻¹ with a resolution of 0.02 cm⁻¹. The spectra are measured during sunsets and sunrises in the limb viewing geometry with different slant paths and tangent heights from upper troposphere up to 150 km [38]. From the measured spectra, several tens of microwindows are analyzed to retrieve CH₄ profiles with the vertical resolution of about 4 km. In the comparison studies, the accuracy has been found to be within 10% in the upper troposphere–lower stratosphere region [39].

To gather a comprehensive dataset, the co-location criteria for ACE-FTS measurements was set quite loose. Measurements within 30 degrees in longitude and 3 degrees in latitude from Sodankylä were included in the analysis (see map in Figure 1). The comparison between ACE-FTS and Swirlab was done on the spatially closest ACE-FTS profile of the day and the daily average of SZA-corrected Swirlab profiles. We used the Level 2 ACE-FTS data version 3.5 and 3.6 interpolated to a 1 km grid with center altitudes from 0.5 to 149.5 km. The analysis was restricted to altitudes from 10 to 40 km as the ACE-FTS data below 10 km are sparse and the sensitivity of the Swirlab retrieval is very low above 40 km.

2.4. Continuous Mast Measurements

Sodankylä 50-m-high mast has gas line inlets at four different heights: 2, 10, 23 and 50 m a.g.l. Sample air from highest level inlet is measured for 50 min/h, and air from two other (2 and 23 m) levels for 5 min/h each. The 10 m level is not measured regularly. Gas lines are 8-mm OD stainless steel, and they are flushed with flow rate of ca. 2.5 L/min, which corresponds to less than 1-min residence time in the lines.

Sample air is measured with a CRDS instrument, which measures CO₂, CH₄, CO and water vapor concentrations at the same time. This instrument is automatically calibrated with WMO/GAW/CCL (World Meteorological Organization/Global Atmosphere Watch/Central Calibration Laboratory, i.e., NOAA ESRL GMD) traceable calibration gases ca. 2 times a month. In addition, two cylinders, “working standard” and “target”, are measured every 11 and 34 h, respectively. The methane scale in use is WMO X2004A.

Humid air samples are measured directly, i.e., no dryer is used. This is compensated by determining and using individual water vapor correction factors for the instrument [40,41]. The raw data are first calculated as 1-min averages, and further to uncalibrated dry concentrations

based on water vapor correction factors. Calibration is applied to these values, and hourly average concentrations are calculated from 1-min calibrated dry concentrations. For the time series analysis, a daily average value is calculated from the hourly values.

3. Methods

3.1. Profile Retrieval Method

Traditionally, ground-based direct solar radiation measurements at SWIR wavelengths have been used to retrieve methane total column. This is done by scaling the prior profile to produce a modeled spectrum that has the closest match to the measured spectrum (e.g., the method used in the TCCON retrievals [21]). However, the shapes of absorption lines include information about the vertical structure of the methane profile. The lines are broadened by temperature and pressure, resulting in wider absorption features at low altitudes and narrower at high altitudes. The algorithm described in this section finds the best fit between the modeled and measured spectra by allowing the methane profile shape to vary. This allows better fit between the measured and modeled spectra as also the width, in addition to the depth, of the absorption features in the modeled spectra is modified. The retrieved profiles, the vertically resolved concentrations and their changes can be studied. There are also advantages in using this method for total column retrievals: the averaging kernels show better sensitivity to the variability in the stratosphere than in the case of profile scaling method [32]. In addition, the SZA dependence of the averaging kernels is lower, thus resulting smaller SZA induced biases when there is a large difference between the shapes of the true profile and the prior profile used for the retrieval [32].

To retrieve vertical profiles of methane, we used the dimension reduction method described in detail in [32]. In this section, we give a brief description of the method. The goal has been to develop a method that would be only weakly dependent on the prior assumptions but would still provide information about the vertical structure and realistic profiles. The entire algorithm, hereafter referred to as Swirlab, is programmed to be run in Matlab and is publicly available [27].

The measured FTS data, described in Section 2.1, allows us to retrieve around 2–3 independent pieces of information about the vertical structure of CH₄, as demonstrated in [32]. This means that we can actually extract more information from the measurement than the operational TCCON retrieval, which only has one degree of freedom. Thus, we construct the inverse problem so that there are 2–3 estimated parameters only. This kind of parameterization simplifies and stabilizes the inverse problem, and eases the computation in comparison to a layer-by-layer retrieval. The parameters need to be chosen so that they manifest the information content of the measurement, and we need a way to project the low-parameter profile back to the full space.

The profile of the atmospheric constituent of interest, in our cases CH₄, is defined by first discretizing the atmosphere from 0 to 70 km into n layers and identifying the relative contribution of methane to the total atmospheric density (in parts per billion, ppb) per layer as $\vec{x} \in \mathbb{R}^n$. The inverse problem of inferring methane concentrations from measured data $\vec{y} \in \mathbb{R}^m$, also known as a retrieval, is derived from the relation

$$\vec{y} = f(\vec{x}) + \vec{\epsilon}, \quad (1)$$

where f is a non-linear atmospheric radiative transfer model and $\vec{\epsilon} \in \mathbb{R}^m$ is a noise term describing the measurement and modeling error. The measured spectra used in this work are the same used for the TCCON retrieval but from a small microwindow with a wavenumber range from 6003 to 6005.5 cm⁻¹ including five methane absorption lines. An example spectrum is given in Figure 2.

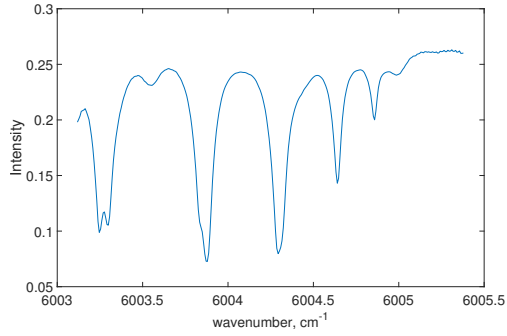


Figure 2. An example spectrum measured using ground-based Fourier Transform Spectrometer at noon on 20 June 2016. The figure covers the wavenumber range used in the Swirlab profile retrieval including strong and weak CH₄ absorbing features.

We use the Bayesian analysis for solving the retrieval problem and assume Gaussian distributions for the prior and noise, $\vec{x}_{pr} \sim \mathcal{N}(\vec{x}_0, \mathbf{C})$, $\vec{\epsilon} \sim \mathcal{N}(0, \mathbf{C}_y)$, where $\vec{x}_0 \in \mathbb{R}^n$ is a priori mean profile, $\mathbf{C} \in \mathbb{R}^{n \times n}$ positive definite prior covariance matrix and $\mathbf{C}_y \in \mathbb{R}^{m \times m}$ is the measurement error covariance matrix. The solution is obtained as posterior density of form

$$\bar{p}(\vec{x}|\vec{y}) \propto \exp\left(-\frac{1}{2}\left(\|\vec{y} - f(\vec{x})\|_{\mathbf{C}_y}^2 + \|\vec{x} - \vec{x}_0\|_{\mathbf{C}}^2\right)\right). \quad (2)$$

We obtain the retrieved methane profile $\hat{\vec{x}}$ from the retrieval method \mathbf{R} as

$$\hat{\vec{x}} = \mathbf{R}(\vec{y}, \mathbf{C}_y, \vec{\theta}), \quad (3)$$

where \mathbf{R} includes our prior assumptions, forward model and Bayesian formulation as well as additional parameters $\vec{\theta}$ such as pressure, solar parameters, etc. These parameters can be retrieved as a part of the state vector, but to keep the notation compact we keep them fixed.

The model atmosphere consists of a non-scattering atmosphere with 100 spherical layers from ground level up to 70 km. Absorption lines are calculated for each layer separately as Voigt profiles taking into account the temperature and pressure of the layer. Atmospheric model data from National Centers for Environmental Prediction (NCEP) are used as source for these atmospheric parameters. As an update to description in Tukiainen et al. [32], the absorption coefficients are calculated using HITRAN2016 line database [42].

The dimension reduction approach that we adapted is based on those of Marzouk and Najm [43] and Solonen et al. [31]. The fundamental idea is to constrain the problem to a subspace that contains most of the variability allowed by the prior. This is achieved by using a smoothing prior covariance that has only a small amount of singular values significantly greater than zero, and thus most of this variability can be represented by a linear combination of the corresponding leading singular vectors. Thus, we start with the singular value decomposition of the prior covariance matrix

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad (4)$$

where the diagonal matrix $\mathbf{\Lambda}$ contains the singular values of the prior covariance matrix and \mathbf{U} contains the corresponding singular vectors $\vec{u}_1, \dots, \vec{u}_n$ as columns. The dimension of the problem is then reduced by using the k ($1 \leq k < n$) largest singular values and the corresponding singular vectors and representing the unknown as a linear combination

$$\vec{x} = \vec{x}_0 + \mathbf{P}_k \vec{z}_k, \quad \mathbf{P}_k = \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} = \left[\sqrt{\lambda_1} \vec{u}_1, \sqrt{\lambda_2} \vec{u}_2, \dots, \sqrt{\lambda_k} \vec{u}_k \right]. \quad (5)$$

Here, the new unknown \vec{z}_k is a k -dimensional vector whose prior distribution is a k -dimensional Gaussian $\mathcal{N}(0, \mathbf{I}_k)$ and $\mathbf{P}_k \in \mathbb{R}^{n \times k}$, $k < n$, is a projection matrix from \mathbb{R}^k to \mathbb{R}^n . Inserting this new parameterization to the problem leads to a posterior distribution for the subspace parameter \vec{z}_k of the form

$$\vec{p}(\vec{z}|\vec{y}) \propto \exp\left(-\frac{1}{2}\left(\|\vec{y} - f(\vec{x}_0 + \mathbf{P}_k \vec{z}_k)\|_{\hat{\mathbf{C}}_y}^2 + \|\vec{z}\|_{\mathbf{I}_k}^2\right)\right). \quad (6)$$

The prior mean, \vec{x}_0 , is estimated by combining the stratospheric part from the ACE-FTS satellite data and the tropospheric part from the TCCON prior data. The prior is calculated separately for three domains: summer (from May to August), polar vortex and the rest (denoted as winter). The location inside the polar vortex is defined using ERA-Interim atmospheric model data. The station is assumed to be inside the polar vortex when potential vorticity at 475 K potential temperature level exceeds $3 \times 10^{-5} \text{ K m}^2 \text{ s}^{-1} \text{ kg}$. The seasonal prior means are depicted in Figure 3. A trend of 5.57 ppb per year, determined from the TCCON prior profiles 2009–2016, is added to each layer. In addition, retrieval with trendless prior is made and the effect is discussed in Section 4.2.

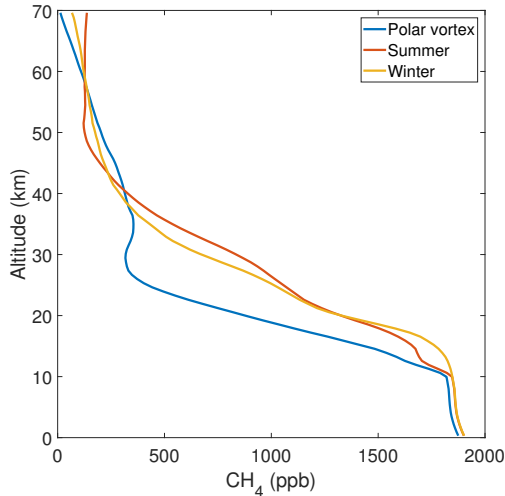


Figure 3. The shapes of the three seasonally selected prior mean profiles. Summer profile is used from May to August and the polar vortex profile when potential vorticity exceeds a threshold value. For the rest of the cases, winter prior profile is used.

For the prior covariance construction, we follow the procedure from Tukiainen et al. [32] to produce a variability that matches well to the empirical covariance derived from an ensemble of ACE-FTS satellite based measurements. In short, this is done by first computing the diagonal elements of \mathbf{C} as

$$\sigma(h) = \sigma_1 \exp\left(-\frac{(h - h_1)^2}{s_1^2}\right) + \sigma_2 \exp\left(-\frac{(h - h_2)^2}{s_2^2}\right), \quad (7)$$

where h are the atmospheric layers and the parameters $\sigma_1, \sigma_2, h_1, h_2, s_1$ and s_2 are chosen appropriately ($\sigma_1 = 300, \sigma_2 = 30, h_1 = 25, h_2 = 5, s_1 = 10$ and $s_2 = 5$). The off-diagonal terms are further given by

$$c_{ij} = \sigma(i)\sigma(j) \exp\left(-\frac{1}{2} \left(\frac{\text{dist}(i, j)}{l}\right)^2\right), \quad (8)$$

where i and j are two atmospheric layers, $\text{dist}(i, j)$ is the distance between these layers in kilometers and l is an appropriately chosen correlation length ($l = 12 \text{ km}$). The full covariance matrix and the first four singular vectors are presented in Figure 4.

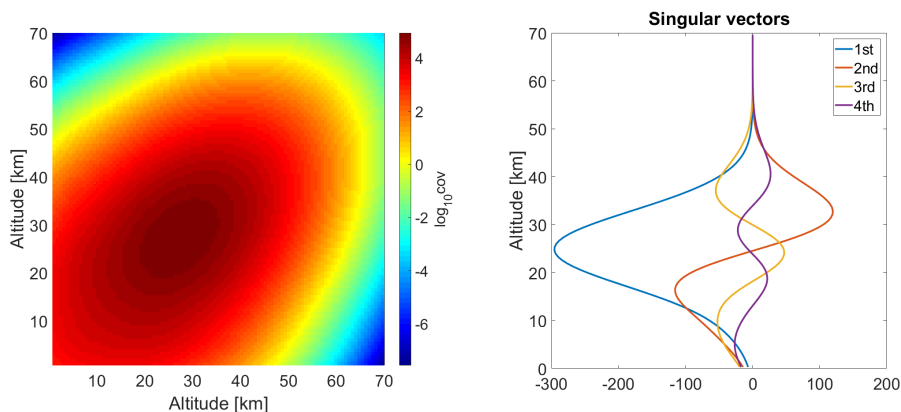


Figure 4. (left) Visualization of the prior covariance structure used for the retrieval. (right) The first four singular vectors calculated from the prior covariance matrix.

3.1.1. Solar Zenith Angle Dependence

Due to errors in radiative transfer modeling and spectroscopy, and because the measurements sensitivity to atmospheric composition at different altitudes changes when solar zenith angle (SZA) changes, there is a systematic difference between volume mixing ratios (VMR_{CH_4}) measured at noon and at early morning or late evening (see Figure 5). This difference is compensated for by going through data for days when the SZA has had both small and large values. The measurement with the smallest SZA for the day is regarded as the reference value (REF_{CH_4}) and the ratio (R) between all the daily measurements and the reference value is calculated.

$$R = \frac{\text{VMR}_{\text{CH}_4}}{\text{REF}_{\text{CH}_4}} \quad (9)$$

The R is first calculated for days with the minimum SZA smaller than 44.5 degrees. Before calculating the ratios for the rest of the days, an exponential function

$$f_{\text{corr}}(\text{SZA}) = a \exp(b\text{SZA}) + c \exp(d\text{SZA}) \quad (10)$$

is fitted to ratios R. For the days with the minimum SZA larger than 44.5 degrees, the reference value is corrected with this function and the following ratios are defined as

$$R = \frac{\text{VMR}_{\text{CH}_4}}{\text{REF}_{\text{CH}_4}} f_{\text{corr}}(\text{SZA}_{\text{ref}}), \quad (11)$$

where SZA_{ref} is the smallest SZA of the evaluated day. The days are evaluated in order of minimum daily SZA. After each evaluated day, the f_{corr} is updated and the evaluation ends when 10% of the time series (about 13,000 measurements) have been evaluated. Then, the final f_{corr} is regarded as the zenith angle correction function for the time series. The correction function is calculated for each retrieved layer separately. The final corrected volume mixing ratio ($\text{cVMR}_{\text{CH}_4}$) at layer i is then calculated as

$$\text{cVMR}_{\text{CH}_4,i} = \text{VMR}_{\text{CH}_4,i} f_{\text{corr},i}(\text{SZA}). \quad (12)$$

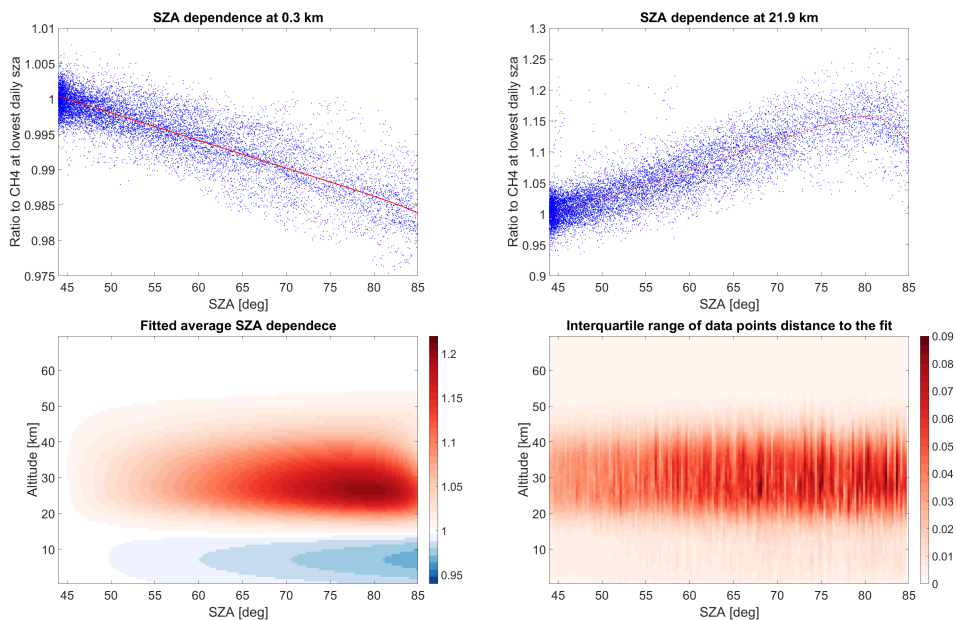


Figure 5. (top left) Dependence of the methane mixing ratio on solar zenith angle for the layer with the center altitude of 0.3 km. (top right) Dependence of the methane mixing ratio on solar zenith angle for the layer with the center altitude of 21.9 km. (bottom left) The SZA correction fits color coded for each layer. (bottom right) The interquartile range of the data points' distance to the fitted curve depicted as a color plot for each layer.

As an example, the SZA dependence of the volume mixing ratios at two different layers are depicted in Figure 5. The Figure also depicts the fits for all the layers and the spread of the difference between data points and the fit. The variability of the data points is described with interquartile range as there are some outliers skewing the distribution so that standard deviation would not describe the data well. Assuming that on average the methane concentration is stable and should not vary as a function of the solar zenith angle, it seems that at the layers close to the ground the mixing ratios are underestimated when the SZA increases, and in the stratosphere the values are overestimated when the sun is close to the horizon.

3.2. Averaging Kernel Correction to ACE Data

To compare the atmospheric profiles obtained from ACE-FTS measurements with Swirlab retrievals, the results must first be projected into the same space. This is because the ill-defined inverse problem is regulated by prior information and the retrieval is hence a biased estimator of true CH_4 . Moreover, the instruments naturally have different sensitivities at different altitudes, which needs to be accounted for as well. A widely used way to make two separate measurements by different instruments comparable is smoothing using the averaging kernel

$$\mathbf{A} = \frac{\partial \hat{x}}{\partial x}, \quad (13)$$

which measures the sensitivity of the retrieved value $\hat{\vec{x}}$ to the actual true atmospheric state \vec{x} . Averaging kernels for \vec{z} parameters can be estimated from the known Swirlab measurements using Jacobians, i.e., the derivatives of forward model f of the form $\mathbf{K}_{ij} = \frac{\partial_i}{\partial x_j} f$, as

$$\mathbf{A}_z = \frac{\partial \hat{\vec{z}}}{\partial \vec{x}} = \left(\mathbf{K}_k^T \mathbf{C}_y^{-1} \mathbf{K}_k + \mathbf{I}_k \right)^{-1} \left(\mathbf{K}_k^T \mathbf{C}_y^{-1} \mathbf{K}_v \right) \quad (14)$$

where $\mathbf{A}_z \in \mathbb{R}^{k \times n}$, $\mathbf{C}_y \in \mathbb{R}^{m \times m}$ is the error covariance matrix, $\mathbf{I}_k \in \mathbb{R}^{k \times k}$ is the prior covariance of \vec{z}_k ,

$$\begin{aligned} \mathbf{K}_k &= \mathbf{K} \text{diag}(\vec{l}(\mathbf{P}_k \vec{z}_k + \vec{x}_0)) \mathbf{P}_k \\ \mathbf{K}_v &= \mathbf{K} \text{diag}(\vec{l}) \end{aligned} \quad (15)$$

and $\vec{l} \in \mathbb{R}^n$ contains the slant lengths of different atmospheric layers (for details, please refer to [32]). The full-state averaging kernel can then be computed as

$$\mathbf{A} = \text{diag}(\mathbf{P}_k \vec{z}_k + \vec{x}_0) \mathbf{P}_k \mathbf{A}_z \quad (16)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$. The averaging kernels are SZA dependent, and thus can partly contribute to diurnal and also seasonal biases. A Swirlab averaging kernel matrix at SZA = 45° is visualized in Figure 6. We see that the changes in the lowest altitudes have a large effect to the higher levels also. As the covariance structure allows the profile to vary the most at altitudes between 20 and 30 km, many of the changes at levels that are more restricted to follow the prior propagate to these altitudes.

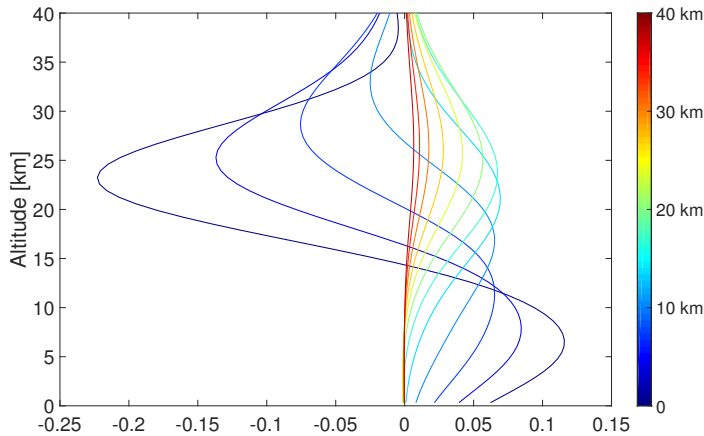


Figure 6. Every fifth layer of a full Swirlab kernel matrix at SZA = 45° depicted in ppb units. This describes how a change at a given altitude, color coded on the right, propagates to other altitudes in the retrieval.

Using the Swirlab averaging kernel, we follow [44] and get the smoothed ACE-FTS profile as

$$\vec{x}_{\text{smooth}} = \vec{x}_0 + \mathbf{A}(\vec{x}_{\text{ACE}} - \vec{x}_0). \quad (17)$$

3.3. Time Series Analysis

In this study, our main interest is the altitude-dependent trend in the Sodankylä FTS time series. In general terms, trend can be defined as a change of distributional processes (e.g., mean) of the process that generates the observations. Typically, the interesting part is the more slowly varying background level after better-known variabilities, such as seasonality, have been accounted for. For time series trend analysis, we follow the approach by Laine et al. [45] and apply dynamical linear regression to the

FTS data by using Dynamic Linear Model (DLM) toolbox for Matlab by Marko Laine [46]. In contrast to the more usual linear regression, dynamical regression allows the regression coefficient to evolve in time. This evolution is controlled by model variance parameters that are estimated from the data. DLM is extremely useful in atmospheric time series analysis since a lot of factors, such as external forcing, make the background mean trend nonlinear. On top of this, atmospheric processes are often a combination of short term rapid changes, such as seasonal cycles, and more long term background processes.

We present a short description of the DLM formulation used in our analysis. DLM with Gaussian errors is best described as a general linear state space model with an observation equation and an evolution equation as

$$\begin{aligned}\bar{y}_t &= \mathbf{H}_t \bar{u}_t + \bar{v}_t, & \bar{v}_t &\sim \mathcal{N}(0, \mathbf{R}_t), \\ \bar{u}_t &= \mathbf{M}_t \bar{u}_{t-1} + \bar{w}_t, & \bar{w}_t &\sim \mathcal{N}(0, \mathbf{Q}_t),\end{aligned}\quad (18)$$

where \bar{y}_t contains the observations and \bar{u}_t the hidden and unobserved model state at time step t , the observations are gained via the observation operator \mathbf{H}_t , the state is evolving along the model evolution operator \mathbf{M}_t . In this formulation, the observation error \bar{v}_t and the model error \bar{w}_t are assumed to be zero mean Gaussian with covariance matrices given by \mathbf{R}_t and \mathbf{Q}_t , respectively.

The hidden state vector \bar{u} has components for each process that is fitted in the time series analysis. In our case, we get

$$\bar{u}_t = [\mu_t \ \alpha_t \ \beta_{t,1} \ \beta_{t,2} \ \eta_t], \quad (19)$$

where μ_t is the parameter for the local mean, α_t is for local trend, $\beta_{t,1}$ and $\beta_{t,2}$ represent the seasonal components and η_t is the autoregressive component of the state vector.

In DLM, the background trend is modeled as a random walk process defined by local mean μ_t and local trend α_t . For these parameters, we can write the observation operator, model evolution operator and model error as

$$\begin{aligned}\mathbf{M}_{\text{trend}} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \\ \mathbf{H}_{\text{trend}} &= \begin{bmatrix} 1 & 0 \end{bmatrix}, \\ \mathbf{Q}_{\text{trend}} &= \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{\text{trend}}^2 \end{bmatrix},\end{aligned}\quad (20)$$

where σ_{trend}^2 is the error variance of the local trend. Similarly, the annual seasonal cycles are modeled using harmonic functions. For seasonal components $\beta_{t,1}$ and $\beta_{t,2}$, we thus have

$$\begin{aligned}\mathbf{M}_{\text{seas}} &= \begin{bmatrix} \cos(2\pi/s) & \sin(2\pi/s) \\ -\sin(2\pi/s) & \cos(2\pi/s) \end{bmatrix}, \\ \mathbf{H}_{\text{seas}} &= \begin{bmatrix} 1 & 0 \end{bmatrix}, \\ \mathbf{Q}_{\text{seas}} &= \begin{bmatrix} \sigma_{\text{seas}}^2 & 0 \\ 0 & \sigma_{\text{seas}}^2 \end{bmatrix},\end{aligned}\quad (21)$$

where $s = 365.242$ is the number of days in a year (accounting also for leap years) and σ_{seas}^2 is the error variance of the seasonal components. In this work, we also allow autocorrelations in the residuals by

using a first-order autoregressive model. The corresponding observation operator, model evolution operator and model error are given by

$$\begin{aligned}\mathbf{M}_{\text{AR}} &= [\rho], \\ \mathbf{H}_{\text{AR}} &= [1], \\ \mathbf{Q}_{\text{AR}} &= [\sigma_{\text{AR}}^2],\end{aligned}\tag{22}$$

Combining the independent parts for trend, seasonality and autoregression, we get the full observation operator, model evolution operator and model error for the hidden state vector \vec{u} as

$$\begin{aligned}\mathbf{M} &= \begin{bmatrix} \mathbf{M}_{\text{trend}} & 0 & 0 \\ 0 & \mathbf{M}_{\text{seas}} & 0 \\ 0 & 0 & \mathbf{M}_{\text{AR}} \end{bmatrix}, \\ \mathbf{H} &= [\mathbf{H}_{\text{trend}} \quad \mathbf{H}_{\text{seas}} \quad \mathbf{H}_{\text{AR}}], \\ \mathbf{Q} &= \begin{bmatrix} \mathbf{Q}_{\text{trend}} & 0 & 0 \\ 0 & \mathbf{Q}_{\text{seas}} & 0 \\ 0 & 0 & \mathbf{Q}_{\text{AR}} \end{bmatrix}.\end{aligned}\tag{23}$$

With the previous formulation, the DLM algorithm proceeds with Kalman Filter and Simulation Smoother to generate an ensemble of states that fit the observed data. This step is repeated with Markov chain Monte Carlo (MCMC), sampling the parameters σ_{trend} , σ_{seas} , σ_{AR} and ρ . This procedure explores various choices of tuning parameters for the problem and gives the values that fit the observations the best, minimizing arbitrary fixing that is often done manually in other applications. For a detailed description of the Simulation Smoother and the MCMC, we refer the reader to the documentation in [45,46].

4. Results

4.1. The Time Series

The Swirlab profile retrieval algorithm was used to retrieve methane profiles from all the SWIR spectra measured when non-obstructed sunlight was available in Sodankylä between May 2009 and November 2018. For further analysis and illustrative purposes, a time series of daily averages was calculated. Figure 7 shows a surface plot of the time series. In the figure, gaps shorter than seven days are filled and only few data gaps remain during the measurement season. The measurement season covers the time when the daily maximum solar elevation angle exceeds eight degrees, in Sodankylä roughly from mid-February to the end of October. On average there were 148 days with measurements per year between 2010 and 2018. The algorithm produces nicely the seasonal variability with summer minimum in the lower troposphere and really low values in the stratosphere throughout the spring when Sodankylä lies beneath the polar vortex. The seasonal behavior and trends at various altitudes are more closely studied in Section 4.6.

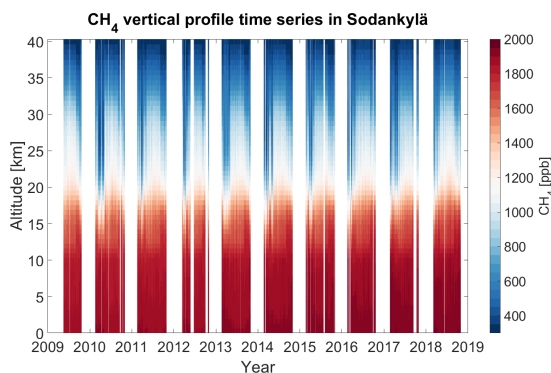


Figure 7. Swirlab time series. Each daily average profile is drawn as one column with the volume mixing ratio of methane depicted with a color. Data gaps shorter than seven days have been filled using the previous existing daily average profile.

4.2. Sensitivity to Prior Profile

To demonstrate the profile information content in the ground-based FTS spectrum, we ran the Swirlab retrieval with different prior profiles. Even when using a single prior profile (Figure 8), the algorithm is able to detect the dip in stratospheric methane concentration inside the polar vortex. The use of a seasonal prior lets the retrievals inside the vortex reach a bit lower level. The introduction of the seasonal prior for summer also makes Swirlab retrieval agree better with the ACE-FTS mixing ratios (see (Figure 8)) when the retrieval with a single prior leaves the summer values somewhat low compared to ACE-FTS.

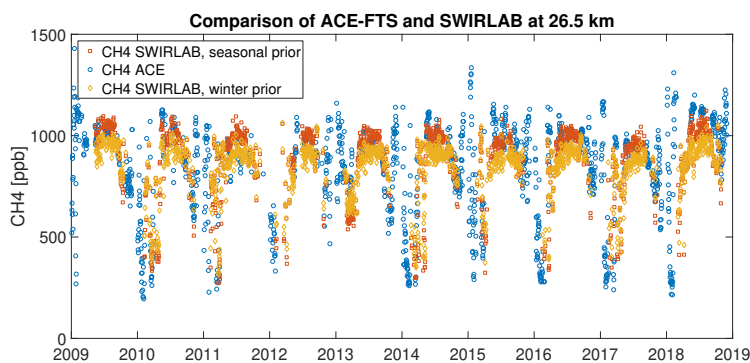


Figure 8. The effect of seasonal prior profile in the stratosphere (26.5 km), where the seasonal variability is the largest. The red squares represent Swirlab retrievals using seasonal priors. Yellow diamonds show the Swirlab time series when a single prior profile shape is used throughout the time series, however including a trend of 5.57 ppb/year. Blue circles show the ACE-FTS data used as a reference.

Figure 9 shows the difference in total column mixing ratios calculated from Swirlab profiles with and without a trend in the prior. Swirlab is able to retrieve a positive trend in the methane total column even when the trend is excluded from the prior. However, the trend is roughly 0.3 ppb/year smaller than the one retrieved when a prior trend of 5.57 ppb/year is introduced. The prior trend is the average trend of total methane column calculated from TCCON prior profiles. The total column methane retrieved with prior trend agrees well with the TCCON data (see Section 4.5).

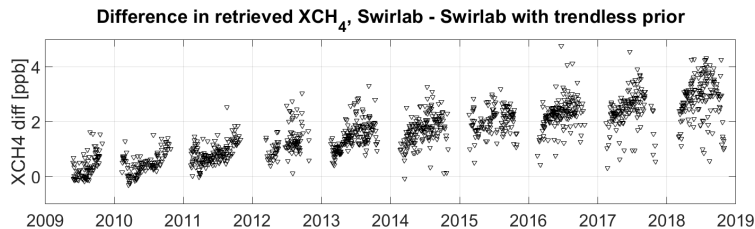


Figure 9. Difference between the total columns calculated from the Swirlab profiles when the retrieval is made including or neglecting the trend in the prior profile.

4.3. Profile Comparison to AirCore Measurements

The concurrent Swirlab and AirCore profiles are shown in Figure 10. The AirCore profiles were obtained during 2013–2017. During multiyear flights, all seasons have been represented. In 2017, the flights were performed within ESA FRM4GHG Project ([47,48]). There is, in general, a good agreement between AirCore and Swirlab posterior profile. For some retrievals, another prior profile could have resulted in a better agreement. For example, on 24 and 26 April 2017, the polar vortex prior would follow the AirCore profiles more closely but the potential vorticity in the ECMWF data suggests Sodankylä was not inside the polar vortex so the winter prior profile is used. Nevertheless, the methane deficit in the stratosphere is quite well detected by Swirlab even when using the winter prior profile.

For quantitative analysis, we calculated the root mean square error (RMSE) of Swirlab retrievals against the AirCore profiles. The RMSE is defined as

$$\text{RMSE}(x) = \left(\sum_{i=1}^N (z_i - x_i)^2 / N \right)^{\frac{1}{2}} \quad (24)$$

where x is the Swirlab profile, z is the AirCore profile and N is the number of layers where we have co-occurring data points. We performed the same RMSE analysis also for another set of profiles, retrieved using the prior scaling method from [32]. The prior scaling retrieval is similar to the one used in TCCON and thus serves as a good reference point. The results are presented in Table 1. We can conclude that, in the sense of RMSE, Swirlab tends to give results that are on average closer to the AirCore profile. In addition, looking at the measurements where the profile scaling yields a smaller RMSE than Swirlab, the differences between the methods are still relatively small. In contrast, in the majority of cases where Swirlab provided a better retrieval, it was significantly better. We would also like to note that, as can be seen in Figure 10, the retrievals where profile scaling performs better are generally times when the measured AirCore profile is very close to the prior profile.

Table 1. Root mean square errors of the posterior mean obtained by Swirlab profile retrieval and prior scaling method similar to the one used by TCCON, relative to the AirCore profile.

Date	AirCore-Swirlab	AirCore-Prior Scaling
03 Sep 2013	98.7	135.6
05 Sep 2013	188.4	259.5
22 Oct 2013	106.6	289.4
19 Mar 2014	61.1	46.3
09 Apr 2014	76.3	87.5
14 Jul 2014	81.4	61.9
15 Jul 2014	64.6	63.8
16 Jul 2014	60.9	53.6
02 Sep 2014	48.5	32.6
05 Nov 2014	158.5	222.9
14 Apr 2016	185.0	272.4
24 Aug 2016	65.5	43.7
21 Apr 2017	74.4	146.3
24 Apr 2017	167.3	384.5
26 Apr 2017	150.7	372.8
15 May 2017	54.1	97.8
28 Aug 2017	94.0	49.3
04 Sep 2017	40.3	44.1
05 Sep 2017	73.2	44.6
06 Sep 2017	41.0	28.4
07 Sep 2017	23.3	31.3
Mean	87.5	126.3

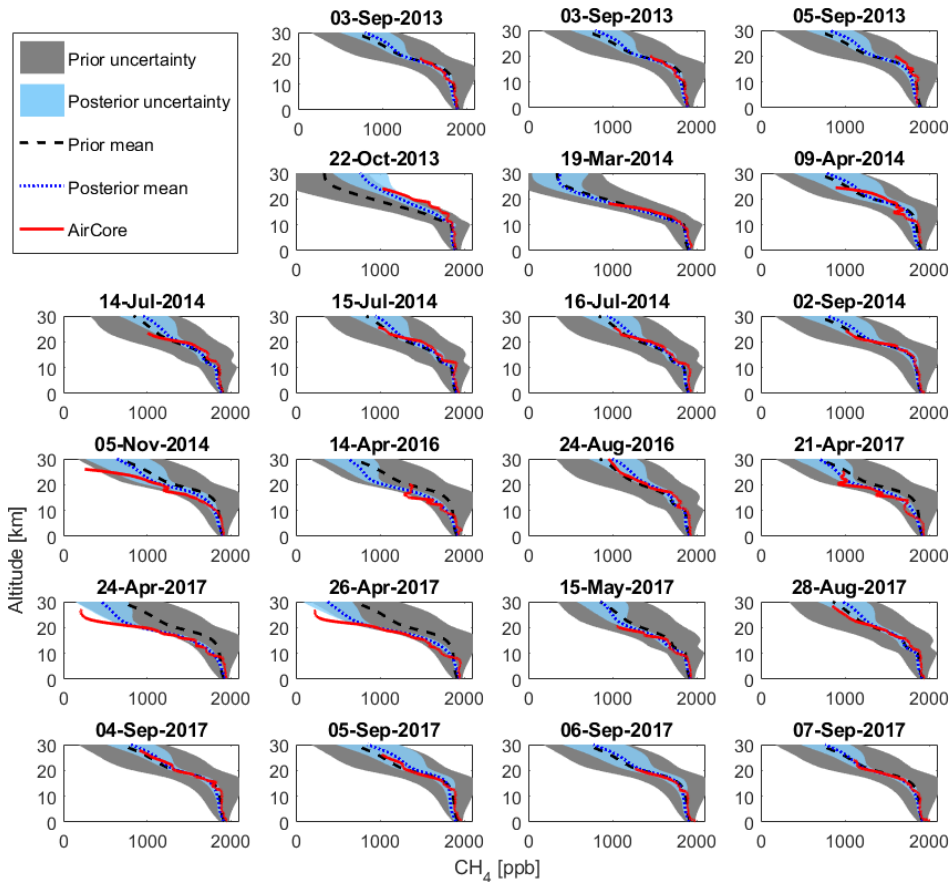


Figure 10. Comparison between Swirlab and AirCore methane profiles. The altitudes on the vertical axes are in kilometers (km) and the concentrations on horizontal axes in parts per billion (ppb). Red lines are the AirCore profiles. Black dashed line is the prior profile with prior variance in grey shading. Blue dotted line is the posterior mean with blue shading representing the 95% posterior.

4.4. Profile Comparison with ACE-FTS Instrument

Swirlab retrievals were compared to both original and averaging kernel corrected ACE-FTS profiles (Figure 11). The comparison was made between the closest ACE-FTS retrieval of the day within the co-location region (see Figure 1) and the concurrent Swirlab retrieval over Sodankylä. For the averaging kernel correction, the missing values in ACE-FTS profiles have been filled with the Swirlab prior profile. The gap-filling method might partly explain the small negative bias below 25 km and small positive bias above 25 km when comparing averaging kernel corrected ACE-FTS to Swirlab.

A large set of ACE-FTS profiles was used to create the prior means for Swirlab and also the prior covariance was built to resemble the empirical covariance of ACE-FTS profiles. Therefore, it was expected that on average ACE-FTS and Swirlab would agree well. However, the information from ACE-FTS for the priors is condensed to three seasonal prior means. The variability around these priors, even if constrained to a realistic and smooth solution by the covariance, is interpreted from the spectra. In this regard, using ACE-FTS as a reference is justified.

The raw, non-smoothed, ACE-FTS profiles and Swirlab profiles agree well below 25 km with Swirlab retrievals staying within 25% from the satellite data. Between 25 and 35 km the profiles agree well on average but the variability becomes larger. The distribution of differences against

non-smoothed ACE-FTS profiles is slightly asymmetric with the median difference being close to zero. The median suggests that there is an equal amount of cases where Swirlab concentrations are lower than cases where they are higher than ACE-FTS values. The asymmetry suggests that, when ACE values are lower, the discrepancies between the two retrievals are larger. This could be caused by Swirlab being unable to sense large methane deficits occurring on a limited altitude range that can occur close to the polar vortex edge when air from outside and inside the vortex are mixed [14]. Close to 40 km altitude, the Swirlab retrievals get closer to prior profiles and on average compare well with ACE-FTS but the variability ACE-FTS can detect is much larger than what is possible with Swirlab.

The Swirlab profiles compare well also to the averaging kernel corrected ACE-FTS profiles, the distribution of the differences resembling the comparison to raw ACE-FTS profiles but with slightly smaller variability. However, the asymmetry above 25 km is even larger than in the comparison with raw ACE-FTS. This again could also, at least partly, be caused by the gap-filling with the Swirlab prior information that might not represent the true profile. At levels close to 40 km, the variability becomes smaller as the sensitivity of Swirlab retrieval is reduced and the averaging kernel smoothed ACE-FTS profiles are forced closer to the prior profiles as are the Swirlab retrievals.

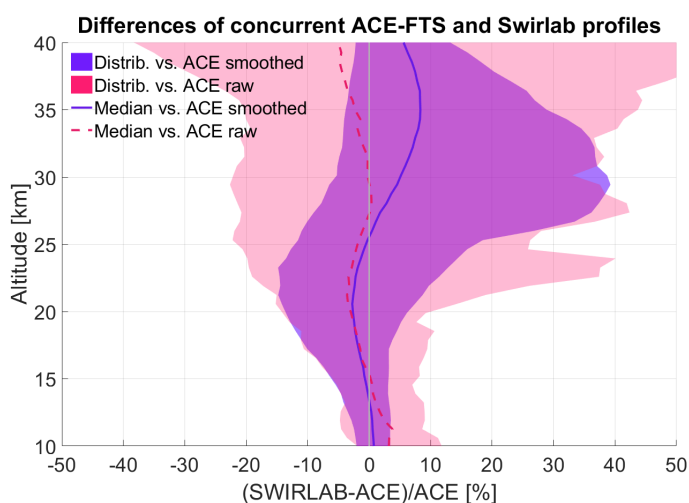


Figure 11. Median differences between daily average profiles from ACE-FTS and Swirlab. The curtain plots show 95th percentile of comparison between ACE-FTS profiles and Swirlab profiles. ACE smoothed has been averaging kernel corrected as described in Section 3.2

4.5. Total Column Comparison to the TCCON Algorithm

We used the pressure profile and water vapor profile from NCEP/NCAR data to calculate the total column dry air mole fraction (X_{CH_4}) from the Swirlab profiles. The daily averages of TCCON and Swirlab total columns are compared in Figure 12. The agreement is good with a mean bias of -2.8 ppb and standard deviation of 6.1 ppb. The trend of TCCON is followed well by Swirlab data, and based on the lack of seasonality in the residual time series, the SZA correction implemented to the Swirlab data is sufficient. A small step in the difference can be seen since the beginning of 2015 but its origin is unresolved.

The main difference of the two algorithms is the degree of freedom in finding the methane profile that would produce the measured spectra. TCCON scales the prior shape and Swirlab uses four singular vectors to modify also the shape of the profile. This improves the accuracy in cases where the prior profile significantly differs from the true profile. In addition, the residuals in the fit between the measured and modeled spectra become smaller as the absorption line shapes are allowed to vary [32].

There are also other differences. The wavelength range used by TCCON is wider as Swirlab uses one short microwindow (see Figure 2), whereas TCCON has three wavelength windows for its methane retrieval. The algorithms use slightly different absorption line data as Swirlab uses HITRAN2016 [42] and GGG2014 uses a modified HITRAN2012 [36]. Differences can also be partly attributed to the different way of defining the dry air column. As the Swirlab algorithm retrieves neither the O₂ column nor the water vapor column, the dry air column used to derive the dry air mole fraction (DMF) is calculated from the NCEP/NCAR model data. In TCCON retrieval, the air column is calculated using the retrieved O₂ column assuming that the DMF of oxygen is constant. The use of O₂ column for air column estimation improves the precision as it reduces the effects of instrumental or measurement errors that are common to both gases CH₄ and O₂ [49].

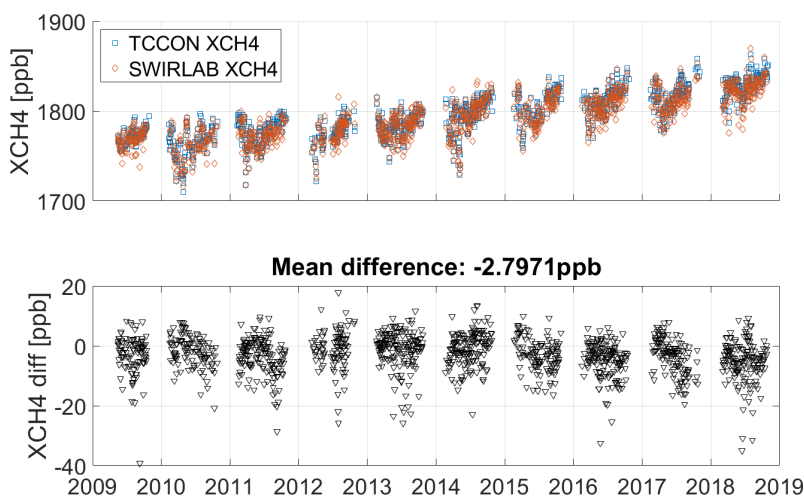


Figure 12. Comparison of total methane column calculated from Swirlab profiles against the total column from official TCCON GGG2014 product. **(top panel)** The daily averages of both retrievals. **(bottom panel)** The difference between Swirlab and TCCON.

4.6. CH₄ Trends

The CH₄ profiles retrieved with Swirlab allow us to quantify the growth rate of atmospheric CH₄ in every level during the 10-year time series. For this analysis, we drew samples from the posterior distribution of the level parameter μ_t in our DLM analysis. For each sample, we computed the yearly trend at a timestep t as a change

$$\text{trend}_t = \mu_{t+s/2} - \mu_{t-s/2}, \quad (25)$$

where s is the number of days considered. To assess the trend and the uncertainty, we drew 1000 samples and calculated the average trend and the standard deviation of the trend for every time step.

The Swirlab CH₄ time series at four different levels (at the ground level, 11.5 km, 21.5 km and 26.5 km) and respective growth rates are shown in Figure 13. In the troposphere (upper panels in Figure 13; orange color), the results show a positive growth rate throughout the time series with moderate interannual variability, in particular at the ground level. The maximum growth rate of 8.1 ± 1.5 ppb/year occurs around winter 2013–2014 at the ground level, and also simultaneously higher in the troposphere. This is also clearly shown by Figure 14 (left panels), where the growth rate in all tropospheric levels increases until 2014, after which it begins to decrease but remains positive. It is notable that the growth rate has decreased significantly since 2017 at the ground level; however, this cannot be seen with confidence higher in the troposphere. The uncertainty estimates of the growth rate increase with altitude, as well as towards the edges of the time series. In the stratosphere (lower panels in Figure 13; orange color), the seasonal variability of CH₄ is stronger, as well as the interannual

variability; thus, weak trends are more challenging to quantify reliably. The variability of the growth rate is mostly smaller-scale than the estimated uncertainty range. The average stratospheric trend is estimated to be weak, and above roughly 20 km altitude, systematically weakly negative; however, the uncertainty estimates are large. Contrary to the ground level and the troposphere, the stratospheric CH₄ growth rate based on Swirlab retrieval levels has been increasing since 2017 and even earlier in the lowest part of the stratosphere. In addition, the Swirlab data show a weak positive trend in the levels close to 40 km but at these high altitudes this is likely due to Swirlab retrieval closely following the prior with a predefined growth rate of 5.57 ppb/year.

We evaluated the Swirlab CH₄ trend signatures described above through comparisons with appropriate reference data at different atmospheric levels. Figure 13 (upper left panel) shows a comparison of Swirlab to the 50-metre mast measurements. In general, the time series from the Swirlab retrieval agrees well with the in-situ measurements, albeit the local variability in the concentration is stronger for the mast measurements. The large variability of the in-situ data results in large uncertainty estimates for the growth rate, exceeding those of Swirlab. Both measurements agree that the trend is lower, almost zero, at the end of the time series, although the uncertainty also increases. For the overlapping time period 2012–2018, the estimated growth rate is 6.3 ± 0.4 ppb/year (the uncertainty denoting the 95% confidence interval) for Swirlab and 6.8 ± 2.6 ppb/year for the mast measurements.

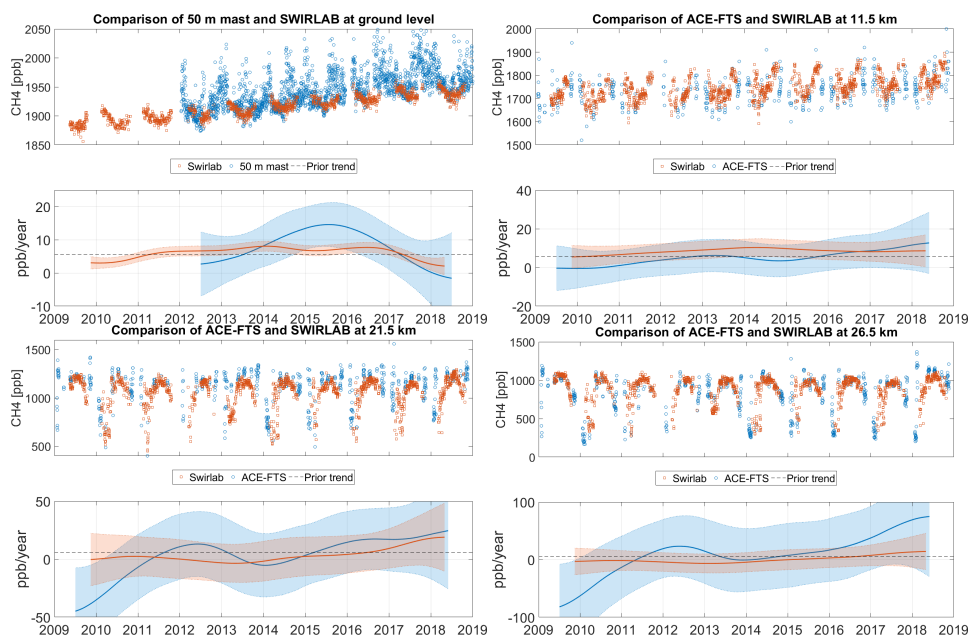


Figure 13. (top left) The lowest layer of Swirlab retrieval compared against 50-m mast measurements. (top right) Swirlab compared to ACE-FTS instrument at 11.5 km. (bottom left) Swirlab compared to ACE-FTS instrument at 21.5 km. (bottom right) Swirlab compared to ACE-FTS instrument at 26.5 km. For every altitude, the upper axis shows the daily averages from compared instruments and the lower axis shows the one-year average trend and the 95% confidence limits evaluated from the DLM analysis. Swirlab data are in orange and the reference instrument in blue. The dashed black line shows the Swirlab prior trend.

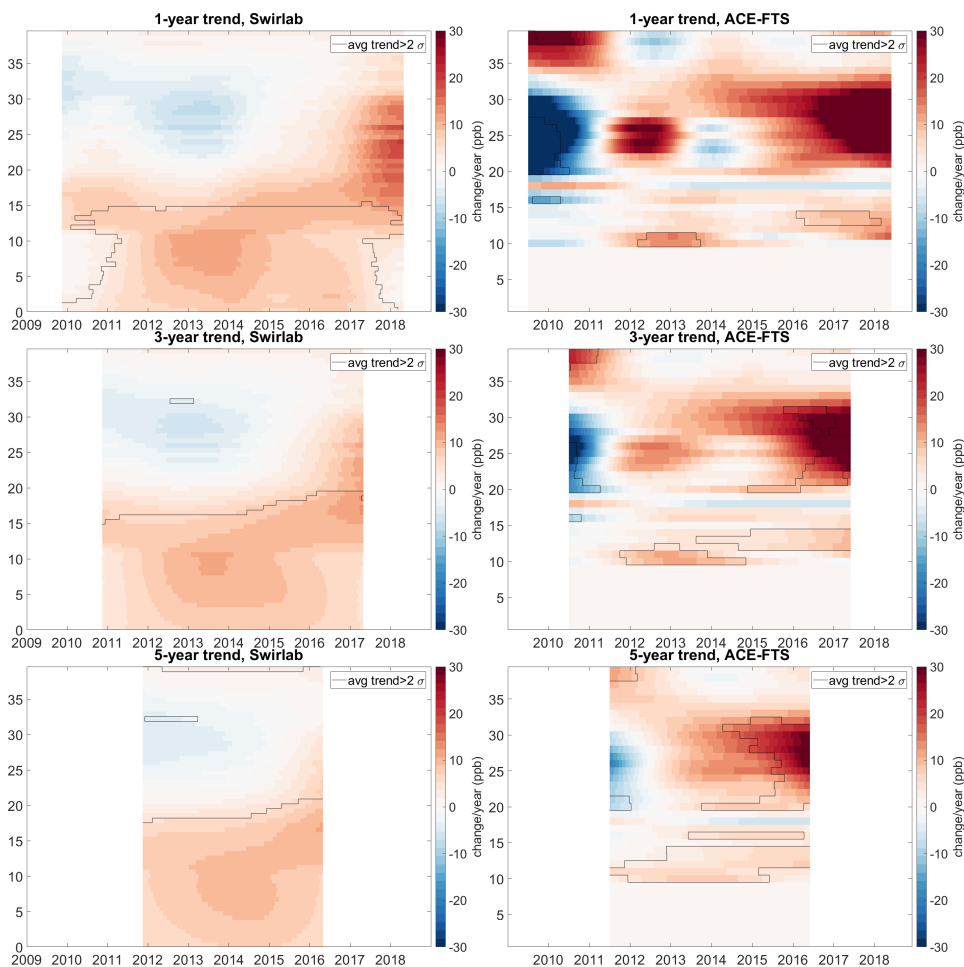


Figure 14. Surface plots of DLM trends at altitudes from 0 to 40 km based on two different methods: Swirlab retrieval using ground-based remote sensing and ACE-FTS retrieval using satellite-based solar occultation measurements. Swirlab trends are on the left and ACE-FTS trends on the right. The top row shows trends calculated for change over a one-year period, the middle row for a three-year range and the bottom row for a five-year range. The black line confines the areas where the trend differs from zero with a 95% confidence level.

At higher altitudes, the reference instrument is the ACE-FTS. As shown in Figure 13, the ACE-FTS and Swirlab time series agree well but the overall trends differ, although they are mostly within the uncertainty envelopes. As Swirlab is more constrained than ACE-FTS, the uncertainty estimates are larger for the latter. At 11.5 km, the eight-year trend from the beginning of 2010 until the end of 2017 (neglecting the edges of the DLM fit) shows a change of 4.9 ± 1.8 ppb/year in ACE-FTS measurements and 8.4 ± 1.2 ppb/year in Swirlab measurements. In the stratosphere, the ACE-FTS-derived growth rates are more subject to interannual variability. For the eight-year period of 2010–2017, overall, the trends at 21.5 km are 4.6 ± 4.8 ppb/year for ACE-FTS and 3.1 ± 3.4 ppb/year for Swirlab. At 26.5 km, the trends for the same period are 11.1 ± 11.0 ppb/year for ACE-FTS and -0.2 ± 4.4 ppb/year for Swirlab. The increasing growth rate towards the end of the ACE-FTS time series seems to be driven by the large increase from 2017 to 2018. This change is also noticeable in the Swirlab data albeit not as large as in the ACE-FTS data.

In Figure 14, the Swirlab growth rate is compared to that of the ACE-FTS at all levels and for different time scales. The ACE-FTS data are presented from 10 km upwards as the data are very sparse below that level. The analysis is also confined to the altitudes below 40 km as Swirlab is not very sensitive at altitudes above that and thus mostly follows the prior. Below 15 km, the instruments agree that there is a small but significant positive trend. However, between 15 and 20 km, ACE-FTS does not detect consistent growth while Swirlab does. From 20 to 35 km, Swirlab data show very stable methane concentrations with a weak yet insignificant negative trend until 2015 when the trend turns to weak positive, while ACE-FTS shows more variable trends in the beginning. In the latest years, Swirlab and ACE-FTS agree on a large but not statistically significant positive trend; however, in the ACE-FTS DLM fit, the trend is found significant in the three- and five-year time spans.

When comparing DLM trends for two datasets, it should be noted that the differences in concurrent measurements are not the only cause for differences in trends. The sampling of the data also has strong influence to the DLM fit. Random gaps in the data can create artifacts in the trends, especially if the measurement frequency is already sparse and the data gap hides a yearly maximum or minimum that is usually observed. Missing data in fall 2011 and early spring 2012 are partly the cause for the oscillations in the ACE-FTS stratospheric trends around winter 2011–2012. The larger amount of data in the Swirlab time series partially stabilizes the DLM fit but the missing values during the polar night make the fitting of the seasonal cycle challenging, as was also pointed out by Kivimäki et al. [50].

5. Discussion and Summary

We produced a ten-year time series of atmospheric methane profiles using a dimension reduction retrieval method to ground-based short wave infrared spectra measured with a Fourier Transform Spectrometer. After statistically correcting the solar zenith angle dependence, the retrieved profiles agree well with reference measurements: ACE-FTS above 10 km up to 40 km, AirCore from the surface to 25 km and 50-m mast air sampling measurements at the lowest altitude levels. The retrieval method succeeds in producing CH₄ profiles with realistic trends and seasonal variability even when using a single prior throughout the time series but the comparison to reference instruments performs better when a prior trend assumption and seasonal prior profiles are introduced.

The DLM approach in the paper results in realistic trend estimates. Missing data, especially in the ACE-FTS time series with less frequent measurements over Sodankylä, can induce short-term variability. The data gap due to the polar night mostly influences the seasonal cycle fit rather than the trend.

The tropospheric trends in the Swirlab time series agree with the trend in local 50-m mast measurements. In a wider perspective, the weaker tropospheric trends until 2011 and higher trends after that agree with other Arctic stations reported by Nisbet et al. [18]. Interestingly, the weaker trend in the end of the time series from 2017 to 2018 does not follow the global average CH₄ increase as reported by the Global Monitoring Division of NOAA's Earth System Research Laboratory [51]; however, this weaker trend is shown by both mast measurements and Swirlab. The slowdown of the growth rate could be interpreted as: (1) the decrease of surface fluxes in the influencing areas of the mast; (2) decreased biomass signals from e.g., Siberia; and (3) a transport pattern change. However, confirming the reason demands a thorough model study outside the scope of this work and thus can only be speculated here.

In the stratosphere, the CH₄ growth rates from ACE-FTS and Swirlab retrievals are generally in a good agreement. However, the confidence limits of the trend are larger than in the troposphere because of a larger uncertainty of the retrieved values, and a larger seasonal and interannual variability in the methane concentrations. Therefore, to be reliably detected from other variability, the growth rate would need to be consistent over a long time period. The significance of the estimated growth rate was evaluated by its deviation from zero with 95% confidence. As shown in Figure 14, both instruments detect mostly insignificant changes in the stratospheric CH₄ over Sodankylä. However,

in the ACE-FTS time series, the positive stratospheric trends towards the end of the time series are found to be significant, contrary to the trends estimated for Swirlab. This could originate from higher vertical resolution of ACE-FTS but could also be caused by the differences in the DLM fitting due to less frequent data points in the ACE-FTS time series.

Albeit appearing towards the edge of the time series, the rising stratospheric trend from 2017 to 2018 draws attention. There is a slight suspicion that this could be an edge effect in the DLM fit, especially as there is a similar negative trend in the beginning of the time series. However, the late rise can be identified in both ACE-FTS and Swirlab data at the 26.5 km altitude as a high, sudden increase in CH₄ from 2017 to 2018, suggesting that the DLM-inferred positive trends are indeed described by the data. The stratospheric air at high latitudes mostly originates from the tropical troposphere. Therefore, the stratospheric methane concentrations in the Arctic are not expected to follow the changes in the local troposphere: the stratospheric air may have been transported several years after being in contact with methane sources at the surface. Hence, the recent sudden increase of CH₄ in the Arctic stratosphere could point to enhanced circulation from the tropics or to a decrease in stratospheric methane sink. However, continuing observations are needed to confirm and monitor this increase, and further studies involving atmospheric inverse modeling are necessary for pinpointing the reason for the observed growth rate.

Overall, the method discussed in this paper is a valuable addition to other methane profile measurements. While it does not reach the accuracy and vertical resolution of aircraft and balloon-borne measurements, the temporal coverage is much better. The information provided by the method will be a valuable addition to the existing continuous monitoring of the total column and the surface concentrations of methane. These continuous measurements can further help to develop models and reliably track the changes in the sources and sinks of methane. In this work, we considered a local scale but the method is publicly available and applicable to any ground-based, high-resolution spectrometer measuring in the short-wave infrared wavelength range.

Author Contributions: Conceptualization, T.K., S.T, J.T., O.L and H.L.; methodology and software, T.K., S.T., O.L. and M.L.; Formal analysis and Investigation, T.K. and O.L.; data curation, R.K., P.H., J.H. and H.C.; writing—original draft preparation, T.K., O.L., S.T., R.K. and J.H.; writing—review and editing, all authors; and visualization, T.K. and O.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by Finnish Meteorological institute, the Academy of Finland's Carbon Balance under Changing Processes of Arctic and Subarctic Cryosphere (CARB-ARC) project (285630), the Academy of Finland's Centre of Excellence in Inverse Modelling and Imaging (312125), and European Space Agency project Solving METHane fluxes at northern latitudes using atmospheric and soil EO data (MethEO). Hannakaisa Lindqvist acknowledges funding from the Academy of Finland via project number 285421. TCCON, FTS and AirCore work at Sodankylä have been supported by the ESA FRM4GHG, EU GAIA-CLIM and EU RINGO projects. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT, is a Canadian-led mission mainly supported by the Canadian Space Agency.

Acknowledgments: We thank the University of Waterloo for providing the access to ACE-FTS data.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Myhre, G.; Shindell, D.; Bréon, F.M.; Collins, W.; Fuglestedt, J.; Huang, J.; Koch, D.; Lamarque, J.F.; Lee, D.; Mendoza, B.; et al. 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*; Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Doschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK, 2013; pp. 659–740, doi:10.1017/CBO9781107415324.018.
2. Prather, M.J.; Holmes, C.D.; Hsu, J. Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry. *Geophys. Res. Lett.* **2012**, *39*, doi:10.1029/2012GL051440.

3. Etheridge, D.M.; Steele, L.P.; Francey, R.J.; Langenfelds, R.L. Atmospheric methane between 1000 A.D. and present: Evidence of anthropogenic emissions and climatic variability. *J. Geophys. Res. Atmos.* **1998**, *103*, 15979–15993, doi:10.1029/98JD00923.
4. Dlugokencky, E.J.; Crotwell, A.M.; Mund, J.W.; Crotwell, M.J.; Thoning, K.W. *Atmospheric Methane Dry Air Mole Fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network*; ESRL/GMD CCGG Group: Boulder, CO, USA, 2019, doi:10.15138/VNCZ-M766.
5. Saunio, M.; Stavert, A.R.; Poulter, B.; Bousquet, P.; Canadell, J.G.; Jackson, R.B.; Raymond, P.A.; Dlugokencky, E.J.; Houweling, S.; Patra, P.K.; et al. The Global Methane Budget 2000–2017. *Earth Syst. Sci. Data Discuss.* **2019**, *2019*, 1–136, doi:10.5194/essd-2019-128.
6. Dean, J.F.; Middelburg, J.J.; Röckmann, T.; Aerts, R.; Blauw, L.G.; Egger, M.; Jetten, M.S.; de Jong, A.E.; Meisel, O.H.; Rasigraf, O.; et al. Methane Feedbacks to the Global Climate System in a Warmer World. *Rev. Geophys.* **2018**, *56*, 207–250, doi:10.1002/2017RG000559.
7. Worthy, D.; Taylor, C.; Dlugokencky, E.; Chan, E.; Nisbet, E.G.; Laurila, T. *Long-term monitoring of atmospheric methane*; Arctic Monitoring and Assessment Programme (AMAP): Oslo, Norway, 2015; Chapter 6, pp. 61–75.
8. Thanwerdas, J.; Saunio, M.; Berchet, A.; Pison, I.; Hauglustaine, D.; Ramonet, M.; Crevoisier, C.; Baier, B.; Sweeney, C.; Bousquet, P. C : 12 C isotopic ratio at global scale. *Atmos. Chem. Phys. Discuss.* **2019**, 1–28, doi:10.5194/acp-2019-925.
9. Bruhwiler, L.M.; Basu, S.; Bergamaschi, P.; Bousquet, P.; Dlugokencky, E.; Houweling, S.; Ishizawa, M.; Kim, H.S.; Locatelli, R.; Maksyutov, S.; et al. U.S. CH₄ emissions from oil and gas production: Have recent large increases been detected? *J. Geophys. Res.* **2017**, *122*, 4070–4083, doi:10.1002/2016JD026157.
10. Sasakawa, M.; Machida, T.; Ishijima, K.; Arshinov, M.; Patra, P.K.; Ito, A.; Aoki, S.; Petrov, V. Temporal Characteristics of CH₄ Vertical Profiles Observed in the West Siberian Lowland Over Surgut From 1993 to 2015 and Novosibirsk From 1997 to 2015. *J. Geophys. Res. Atmos.* **2017**, *122*, 11261–11273, doi:10.1002/2017JD026836.
11. Brewer, A.W. Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Q. J. R. Meteorol. Soc.* **1949**, *75*, 351–363, doi:10.1002/qj.49707532603.
12. Holton, J.R.; Haynes, P.H.; McIntyre, M.E.; Douglass, A.R.; Rood, R.B.; Pfister, L. Stratosphere-troposphere exchange. *Rev. Geophys.* **1995**, *33*, 403–439, doi:10.1029/95RG02097.
13. Schoeberl, M.R.; Hartmann, D.L. The dynamics of the stratospheric polar vortex and its relation to springtime ozone depletions. *Science* **1991**, *251*, 46–52, doi:10.1126/science.251.4989.46.
14. Röckmann, T.; Brass, M.; Borchers, R.; Engel, A. The isotopic composition of methane in the stratosphere: High-altitude balloon sample measurements. *Atmos. Chem. Phys.* **2011**, *11*, 13287–13304, doi:10.5194/acp-11-13287-2011.
15. Dlugokencky, E.J.; Houweling, S.; Bruhwiler, L.; Masarie, K.A.; Lang, P.M.; Miller, J.B.; Tans, P.P. Atmospheric methane levels off: Temporary pause or a new steady-state? *Geophys. Res. Lett.* **2003**, *30*, 3–6, doi:10.1029/2003GL018126.
16. Nisbet, E.G.; Dlugokencky, E.J.; Manning, M.R.; Lowry, D.; Fisher, R.E.; France, J.L.; Michel, S.E.; Miller, J.B.; White, J.W.; Vaughn, B.; et al. Rising atmospheric methane: 2007–2014 growth and isotopic shift. *Global Biogeochemical Cycles* **2016**, doi:10.1002/2016GB005406.
17. Rigby, M.; Montzka, S.A.; Prinn, R.G.; White, J.W.C.; Young, D.; O’Doherty, S.; Lunt, M.F.; Ganesan, A.L.; Manning, A.J.; Simmonds, P.G.; et al. Role of atmospheric oxidation in recent methane growth. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 5373–5377, doi:10.1073/pnas.1616426114.
18. Nisbet, E.G.; Manning, M.R.; Dlugokencky, E.J.; Fisher, R.E.; Lowry, D.; Michel, S.E.; Myhre, C.L.; Platt, S.M.; Allen, G.; Bousquet, P.; et al. Very Strong Atmospheric Methane Growth in the 4 Years 2014–2017: Implications for the Paris Agreement. *Glob. Biogeochem. Cycles* **2019**, *33*, 318–342, doi:10.1029/2018GB006009.
19. Steele, L.P.; Fraser, P.J.; Rasmussen, R.A.; Khalil, M.A.K.; Conway, T.J.; Crawford, A.J.; Gammon, R.H.; Masarie, K.A.; Thoning, K.W. The global distribution of methane in the troposphere. *J. Atmos. Chem.* **1987**, *5*, 125–171, doi:10.1007/BF00048857.
20. Dlugokencky, E.J.; Steele, L.P.; Lang, P.M.; Masarie, K.A. The growth rate and distribution of atmospheric methane. *J. Geophys. Res.* **1994**, *99*.
21. Wunch, D.; Toon, G.C.; Blavier, J.F.L.; Washenfelder, R.A.; Notholt, J.; Connor, B.J.; Griffith, D.W.T.; Sherlock, V.; Wennberg, P.O. The Total Carbon Column Observing Network. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **2011**, *369*, 2087–2112, doi:10.1098/rsta.2010.0240.

22. De Mazière, M.; Thompson, A.M.; Kurylo, M.J.; Wild, J.D.; Bernhard, G.; Blumenstock, T.; Braathen, G.O.; Hannigan, J.W.; Lambert, J.C.; Leblanc, T.; et al. The Network for the Detection of Atmospheric Composition Change (NDACC): History, status and perspectives. *Atmos. Chem. Phys.* **2018**, *18*, 4935–4964, doi:10.5194/acp-18-4935-2018.
23. Hu, H.; Hasekamp, O.; Butz, A.; Galli, A.; Landgraf, J.; Aan de Brugh, J.; Borsdorff, T.; Scheepmaker, R.; Aben, I. The operational methane retrieval algorithm for TROPOMI. *Atmos. Meas. Tech.* **2016**, *9*, 5423–5440, doi:10.5194/amt-9-5423-2016.
24. Parker, R.; Boesch, H.; Cogan, A.; Fraser, A.; Feng, L.; Palmer, P.I.; Messerschmidt, J.; Deutscher, N.; Griffith, D.W.T.; Notholt, J.; et al. Methane observations from the Greenhouse Gases Observing SATellite: Comparison to ground-based TCCON data and model calculations. *Geophys. Res. Lett.* **2011**, *38*, doi:10.1029/2011GL047871.
25. Karion, A.; Sweeney, C.; Tans, P.; Newberger, T. AirCore: An Innovative Atmospheric Sampling System. *J. Atmos. Ocean. Technol.* **2010**, *27*, 1839–1853, doi:10.1175/2010JTECHA1448.1.
26. Verma, S.; Marshall, J.; Parrington, M.; Agustí-Panareda, A.; Massart, S.; Chipperfield, M.P.; Wilson, C.; Gerbig, C. Extending methane profiles from aircraft into the stratosphere for satellite total column validation using the ECMWF C-IFS and TOMCAT/SLIMCAT 3-D model. *Atmos. Chem. Phys.* **2017**, *17*, 6663–6678, doi:10.5194/acp-17-6663-2017.
27. Tukiainen, S. Swirlab. Available online: <https://github.com/tukiains/swirlab> (accessed on 2 August 2019).
28. Hase, F.; Hannigan, J.; Coffey, M.; Goldman, A.; Höpfner, M.; Jones, N.; Rinsland, C.; Wood, S. Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements. *J. Quant. Spectrosc. Radiat. Transf.* **2004**, *87*, 25–52, doi:10.1016/j.jqsrt.2003.12.008.
29. Zhou, M.; Langerock, B.; Sha, M.K.; Kumps, N.; Hermans, C.; Petri, C.; Warneke, T.; Chen, H.; Metzger, J.M.; Kivi, R.; et al. Retrieval of atmospheric CH₄ vertical information from ground-based FTS near-infrared spectra. *Atmos. Meas. Tech.* **2019**, *12*, 6125–6141, doi:10.5194/amt-12-6125-2019.
30. Mueller, J.; Siltanen, S. *Linear and Nonlinear Inverse Problems with Practical Applications*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2012.
31. Solonen, A.; Cui, T.; Hakkarainen, J.; Marzouk, Y. On dimension reduction in Gaussian filters. *Inverse Problems* **2016**, *32*, 045003.
32. Tukiainen, S.; Railo, J.; Laine, M.; Hakkarainen, J.; Kivi, R.; Heikkinen, P.; Chen, H.; Tamminen, J. Retrieval of atmospheric CH₄ profiles from Fourier transform infrared data using dimension reduction and MCMC. *J. Geophys. Res. Atmos.* **2016**, *121*, 10312–10327. 2015JD024657, doi:10.1002/2015JD024657.
33. Laine, M. Introduction to Dynamic Linear Models for Time Series Analysis. In *Geodetic Time Series Analysis in Earth Sciences*; Montillet, J.P.; Bos, M., Eds.; Springer: Berlin, Germany, 2019; pp. 139–156, doi:10.1007/978-3-030-21718-1_4.
34. Kivi, R.; Heikkinen, P. Fourier transform spectrometer measurements of column CO₂ at Sodankylä, Finland. *Geosci. Instrum. Methods Data Syst.* **2016**, *5*, 271–279, doi:10.5194/gi-5-271-2016.
35. Wunch, D.; Toon, G.C.; Sherlock, V.; Deutscher, N.M.; Liu, C.; Feist, D.G.; Wennberg, P.O. *The Total Carbon Column Observing Network's GGG2014 Data Version*; The Total Carbon Column Observing Network (TCCON): 2015, doi:10.14291/tcon.ggg2014.documentation.r0/1221662. Available online: <https://data.caltech.edu/records/249> (accessed on 8 August 2019)
36. Rothman, L.; Gordon, I.; Babikov, Y.; Barbe, A.; Benner, D.C.; Bernath, P.; Birk, M.; Bizzocchi, L.; Boudon, V.; Brown, L.; et al. The HITRAN2012 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **2013**, *130*, 4–50, doi:10.1016/j.jqsrt.2013.07.002.
37. Kivi, R.; Heikkinen, P.; Kyrö, E. TCCON data from Sodankylä (FI), Release GGG2014.R0. 2017. doi:10.14291/tcon.ggg2014.sodankyla01.r0/1149280. Available online: <https://data.caltech.edu/records/289> (accessed on 28 June 2019)
38. Bernath, P.F.; McElroy, C.T.; Abrams, M.C.; Boone, C.D.; Butler, M.; Camy-Peyret, C.; Carleer, M.; Clerbaux, C.; Coheur, P.F.; Colin, R.; et al. Atmospheric chemistry experiment (ACE): Mission overview. *Geophys. Res. Lett.* **2005**, *32*, 1–5, doi:10.1029/2005GL022386.
39. De Mazière, M.; Vigouroux, C.; Bernath, P.F.; Baron, P.; Blumenstock, T.; Boone, C.; Brogniez, C.; Catoire, V.; Coffey, M.; Duchatelet, P.; et al. Validation of ACE-FTS v2.2 methane profiles from the upper troposphere to the lower mesosphere. *Atmos. Chem. Phys.* **2008**, *8*, 2421–2435, doi:10.5194/acp-8-2421-2008.

40. Chen, H.; Winderlich, J.; Gerbig, C.; Hoefler, A.; Rella, C.W.; Crosson, E.R.; Van Pelt, A.D.; Steinbach, J.; Kolle, O.; Beck, V.; et al. High-accuracy continuous airborne measurements of greenhouse gases (CO₂ and CH₄) using the cavity ring-down spectroscopy (CRDS) technique. *Atmos. Meas. Tech.* **2010**, *3*, 375–386, doi:10.5194/amt-3-375-2010.
41. Rella, C.W.; Chen, H.; Andrews, A.E.; Filges, A.; Gerbig, C.; Hatakka, J.; Karion, A.; Miles, N.L.; Richardson, S.J.; Steinbacher, M.; et al. High accuracy measurements of dry mole fractions of carbon dioxide and methane in humid air. *Atmos. Meas. Tech.* **2013**, *6*, 837–860, doi:10.5194/amt-6-837-2013.
42. Gordon, I.E.; Rothman, L.S.; Hill, C.; Kochanov, R.V.; Tan, Y.; Bernath, P.F.; Birk, M.; Boudon, V.; Campargue, A.; Chance, K.V.; et al. The HITRAN2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *203*, 3–69, doi:10.1016/j.jqsrt.2017.06.038.
43. Marzouk, Y.M.; Najm, H.N. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.* **2009**, *228*, 1862–1902, doi:10.1016/j.jcp.2008.11.024.
44. Rodgers, C.D.; Connor, B.J. Intercomparison of remote sounding instruments. *J. Geophys. Res.* **2003**, *108*, doi:10.1029/2002JD002299.
45. Laine, M.; Latva-Pukkila, N.; Kyrölä, E. Analysing time-varying trends in stratospheric ozone time series using the state space approach. *Atmos. Chem. Phys.* **2014**, *14*, 9707–9725, doi:10.5194/acp-14-9707-2014.
46. Laine, M. DLM toolbox for Matlab. Available online: <https://mjlane.github.io/dlm/> (accessed on 8 July 2019).
47. Sha, M.K.; De Mazière, M.; Notholt, J.; Blumenstock, T.; Chen, H.; Griffith, D.; Hase, F.; Heikkinen, P.; Hoffmann, A.; Huebner, M.; et al. FRM4GHG level2 dataset from the Sodankylä campaign. 2019. doi:10.18758/71021048. Available online: <http://frm4ghg.aeronomie.be/index.php/outreach/datasetlevel2doi> (accessed on 18 November 2019).
48. Sha, M.K.; De Mazière, M.; Notholt, J.; Blumenstock, T.; Chen, H.; Dehn, A.; Griffith, D.W.T.; Hase, F.; Heikkinen, P.; Hermans, C.; et al. Intercomparison of low and high resolution infrared spectrometers for ground-based solar remote sensing measurements of total column concentrations of CO₂, CH₄ and CO. *Atmos. Meas. Tech. Discuss.* **2019**, *2019*, 1–67, doi:10.5194/amt-2019-371.
49. Wunch, D.; Toon, G.C.; Wennberg, P.O.; Wofsy, S.C.; Stephens, B.B.; Fischer, M.L.; Uchino, O.; Abshire, J.B.; Bernath, P.; Biraud, S.C.; et al. Calibration of the total carbon column observing network using aircraft profile data. *Atmos. Meas. Tech.* **2010**, *3*, 1351–1362, doi:10.5194/amt-3-1351-2010.
50. Kivimäki, E.; Lindqvist, H.; Hakkarainen, J.; Laine, M.; Sussmann, R.; Tsuruta, A.; Detmers, R.; Deutscher, N.M.; Dlugokencky, E.J.; Hase, F.; et al. Evaluation and Analysis of the Seasonal Cycle and Variability of the Trend from GOSAT Methane Retrievals. *Remote Sens.* **2019**, *11*, 882, doi:10.3390/rs11070882.
51. Dlugokencky, E. NOAA/ESRL Trends in Atmospheric Methane. Available online: https://www.esrl.noaa.gov/gmd/ccgg/trends_ch4/ (accessed on 20 January 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

PAPER III

III

Otto Lamminpää, Jonathan Hobbs, Jenny Brynjarsdottir, Marko Laine, Amy Braverman, Hannakaisa Lindqvist and Johanna Tamminen

Accelerated MCMC for Satellite-Based Measurements of Atmospheric CO₂

Remote Sens. 2019, 11(17):2061, 2019. doi: 10.3390/rs11172061.

Copyright © Authors 2019. Reprinted with permission.



Article

Accelerated MCMC for Satellite-Based Measurements of Atmospheric CO₂

Otto Lamminpää ^{1,*}, Jonathan Hobbs ², Jenný Brynjarsdóttir ³, Marko Laine ¹, Amy Braverman ², Hannakaisa Lindqvist ¹ and Johanna Tamminen ¹

¹ Finnish Meteorological Institute, 00560 Helsinki, Finland

² Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

³ Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University, Cleveland, OH 44106-7058, USA

* Correspondence: otto.lamminpaa@fmi.fi

Received: 26 June 2019; Accepted: 26 August 2019; Published: 2 September 2019



Abstract: Markov Chain Monte Carlo (MCMC) is a powerful and promising tool for assessing the uncertainties in the Orbiting Carbon Observatory 2 (OCO-2) satellite's carbon dioxide measurements. Previous research in comparing MCMC and Optimal Estimation (OE) for the OCO-2 retrieval has highlighted the issues of slow convergence of MCMC, and furthermore OE and MCMC not necessarily agreeing with the simulated ground truth. In this work, we exploit the inherent low information content of the OCO-2 measurement and use the Likelihood-Informed Subspace (LIS) dimension reduction to significantly speed up the convergence of MCMC. We demonstrate the strength of this analysis method by assessing the non-Gaussian shape of the retrieval's posterior distribution, and the effect of operational OCO-2 prior covariance's aerosol parameters on the retrieval. We further show that in our test cases we can use this analysis to improve the retrieval to retrieve the simulated true state significantly more accurately and to characterize the non-Gaussian form of the posterior distribution of the retrieval problem.

Keywords: OCO-2; Markov Chain Monte Carlo; carbon dioxide; aerosols; retrieval; uncertainty quantification

1. Introduction

One of the greatest challenges to the future of planet Earth is climate change arising from global warming, which in turn is accelerated by human-originated emissions of greenhouse gases. The major warming effects are due to atmospheric carbon dioxide (CO₂), which is emitted in increasing amounts since industrialization. A majority of these emissions comes from e.g., usage of fossil fuels in transport, in manufacturing and in agricultural industries. In order to model climate change reliably and to predict future scenarios, accurate carbon flux estimates are crucial. In addition to source estimation, this also means quantifying the natural sinks of carbon, e.g., uptake by oceans and forests. Fluxes are estimated by a procedure called inverse modeling, which in addition to computational models needs observations of atmospheric greenhouse gas concentrations. While notable ground based measurement networks exist for this purpose (e.g., GLOBALVIEW-CO₂: Cooperative Atmospheric Data Integration Project—Carbon Dioxide for in-situ, flask, and aircraft measurements [1] used in flux inversions, and TCCON [2,3] for total atmospheric column measurements), their coverage is not extensive enough to account for all of the global phenomena relating to the carbon cycle and in particular CO₂ emissions.

One way to improve the coverage and resolution of the measurements is to globally collect air column-averaged measurements of CO₂, denoted X_{CO₂}, using satellites. Measurements from orbit offer a

great coverage around the globe, but special care is needed with quantifying the accuracy and precision of these measurements. Importantly, systematic errors that vary in space and/or time can lead to errors in flux estimates, which are inferred from CO₂ gradients. Simulation experiments with inversion systems (e.g., [4,5]) indicate that regional or seasonal biases of a few tenths of a part per million (ppm) can induce substantial errors in flux estimates. Although local CO₂ variations in the lower troposphere close to sinks and sources can be large, these variations are confined to low altitudes and thus have a relatively low contribution to the total variability of X_{CO₂}. NASA's Orbiting Carbon Observatory 2 (OCO-2) satellite mission ([6–9]) aims to satisfy these requirements and has been operationally measuring X_{CO₂} since July 2014.

From orbit, OCO-2 measures radiances of reflected solar radiation, that is, absorption of sunlight in Earth's atmosphere on different wavelengths. These measurements are inverted to get an atmospheric state x that describes conditions on the light path, such as CO₂ concentrations and aerosol and surface properties. The resulting inverse problem is solved using a retrieval algorithm, which consists of a state-of-the-art full physics model that extensively describes the related physical phenomena in the atmosphere, and the so called Optimal Estimation (OE) framework [10], in which the algorithm minimizes the cost function related to the posterior distribution of the retrieval. This is obtained via statistical formulation of inverse problems, where in short error bars are thought of as probability distributions, and the famous Bayes' Formula is used to regularize the problem by implementing prior knowledge of the underlying statistics and physics (operational implementation: ACOS Algorithm [7,9]). OE is an efficient tool for computing maximum A posteriori (MAP) estimates and related uncertainties, but these quantities rely on a Gaussian approximation of the posterior distribution. It is, however, clear that the forward model in this case is non-linear, which is known to result in a non-Gaussian posterior.

Non-Gaussian posterior distributions can be explored using Markov chain Monte Carlo (MCMC) methods, in which a set of samples is created such that they are distributed along the desired posterior distribution. Quantities of interest, e.g., mean, variance, distribution shapes etc. can then further be inferred from the samples. MCMC was first introduced in atmospheric retrievals by [11], the advantages of MCMC method for validating operational satellite retrievals were demonstrated using real Envisat/GOMOS observations in [12]. Other work includes [13–15]. An application of MCMC to the OCO-2 retrieval problem was recently performed by [16] as a simulation study following the approach of [17] with a simplified and computationally efficient surrogate forward model that captures all the essential effects of the Full Physics Model's CO₂, surface and atmospheric aerosol parameters. It was noticed that results from MCMC and OE can differ from each other, and furthermore in some cases not even agree with the simulated ground truth.

It was further discussed in [9,18] that perturbations in the first guesses of aerosol part of the state vector lead to unpredictable changes in retrieved X_{CO₂}. The retrievals in the study were performed using simulated OCO-2 measurements ([19]) with a known ground truth using the full physics model and OE. As a linear problem would have a unique minimum for the cost function, this suggests that the aerosol parametrization in the model causes problems in the retrieval that might significantly affect the accuracy of retrieved X_{CO₂} values. This is due to scattering by aerosols being a substantial source of non-linearity in the forward model (as was found for example by [17,20]), which might translate to a non-Gaussian posterior distribution.

In addition to the potential effect of aerosol parametrization of the forward model, it was found in [16] that the MCMC on OCO-2 surrogate model has a very low acceptance ratio (number of accepted samples over number of computed samples), and that chains can often converge extremely slowly. These are symptoms of non-linearity, high dimensionality and strong correlations between the parameters in MCMC. Recent work on dimension reduction for MCMC in Bayesian inverse problems (e.g., [21,22]) that are solved using MCMC sampling has successfully addressed this problem. This approach is called the Likelihood

Informed Subspace (LIS) dimension reduction and it is based on identifying a subspace of the state space that contains all the information that is possible to obtain from the measurement. It was shown in [15] that this methodology actually aligns with the informative directions formalism by [10] that is widely used as a standard in the atmospheric remote sensing community.

In this work we focus on the analysis of single retrievals and their posterior distributions. We implement the LIS dimension reduction for the OCO-2 surrogate model retrieval to speed up the convergence of MCMC algorithm and further to characterize the non-Gaussian parts of the full posterior distribution of the state vector, particularly in the parameters relating to the aerosols in the physics model. In this study, we use the same simulated radiances and OE retrievals as in [16]. The rest of the article is organized as follows: in Section 2, we present an overview of the OCO-2 measurement and surrogate forward model, as well as introducing the MCMC and dimension reduction framework used in this study. Section 3 summarizes our results comparing the dimension-reduced MCMC with a full dimensional one, results related to the form of the posterior distribution, and results of an experiment with a relaxed retrieval prior covariance. Finally, Section 4 contains discussion of the results, conclusions and further research topics.

2. Methodology

In this section, we give an overview of NASA's Orbiting Carbon Observatory 2 satellite and the simplified surrogate forward model used to investigate uncertainties of the actual operational retrieval algorithm. We then describe briefly the Markov Chain Monte Carlo method and the Adaptive Metropolis-Hastings implementation used in this study. Lastly, we introduce the mathematical basis for Bayesian formulation of inverse problems as well as the principles of using the low information content of the OCO-2 measurement to construct a Likelihood-Informed Subspace dimension reduction scheme for accelerating our MCMC simulations.

2.1. Remote Sensing of CO_2 by OCO-2

We will give a brief overview of the instrument since this is crucial in understanding the nature of the inverse problem at hand. The OCO mission is described in detail by [6,8,23].

The OCO-2 instrument is composed of three spectrometers that measure light reflected from Earth's surface in the near-infrared part of the spectrum at three separate wavelength bands. These bands are centered around 0.765, 1.61 and 2.06 μm and are called the O_2 A-band, the weak CO_2 band and the strong CO_2 band, respectively. Each observation furthermore consists of 1016 radiances from each band, and together they form the observed sounding, denoted as $y \in \mathbb{R}^{3048}$ in this work. The satellite flies in a polar, sun-synchronous orbit that covers the whole Earth with a 233 orbit 16-day repeat cycle, during which it collects measurements with a surface footprint of less than 2.25 km down-track and 1.3 km cross-track every 0.333 s.

2.2. Surrogate forward Model

We consider the inverse problem of estimating an unknown atmospheric state vector $x \in \mathbb{R}^n$ from observed radiances $y \in \mathbb{R}^m$,

$$y = F(x) + \varepsilon, \quad (1)$$

where the *forward model* $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ describes the physics relating the atmospheric state to measured radiances and the random variable $\varepsilon \in \mathbb{R}^m$ is the observation error caused by forward model approximations and instrument noise.

The state-of-the-art Full Physics Model describes the light absorption and scattering by trace gases and aerosols in the atmosphere. While this is a comprehensive way to compute atmospheric radiative transfer,

the model is computationally heavy to evaluate and hence is not suitable for Monte Carlo experiments. We ease this problem by following the approach in [16] and use the surrogate forward model by [17] to account for the major sources of error in the retrieval. A more comprehensive look can be found in Appendix A.2, but in short, the state vector used in the surrogate model can be split into four parts:

- **CO₂ vertical profile:** The model considers the amount of molecules in 20 fixed pressure levels of the atmosphere. The concentration of CO₂ as well as pressure and temperature are taken to be constant inside each level.
- **Surface pressure:** A single parameter of the state vector is retrieved to help identify the total amount of molecules in an atmospheric column. Since O₂ has a near constant concentration in each atmospheric layer, the total absorption on the O₂ band can be used to estimate the total amount of air at the measurement location.
- **Surface albedo:** Earth's surface reflects radiation differently at different wavelengths, and the relation of reflected vs. total incoming radiation is given by albedo. In the surrogate model this corresponds to 6 parameters: a Lambertian albedo and its spectral slope at each of the 3 wavelength bands of the observed absorption spectrum.
- **Aerosol parameters:** Small particles in the atmosphere absorb and scatter light in complex ways, which is taken into account by the model as an aerosol model. The vertical profile of aerosol concentration is modeled with 3 parameters per aerosol type, first of which is the logarithm of Aerosol Optical Depth (AOD) at 755 nm which describes the total intensity of aerosol effects. The second parameter describes the height of the maximum aerosol concentration, and the third describes the thickness of the aerosol layer, with a small value corresponding to a thin layer. The aerosol profile calculated from these parameters is proportional to

$$AOD \exp\left(-\frac{(x-x_a)^2}{2\sigma^2}\right), \quad (2)$$

where x_a is the layer height and σ is the layer thickness. In this work the number of atmospheric aerosol types is 4, which amounts to a total of 12 aerosol parameters in the state vector.

The main quantity of interest in the OCO-2 inverse problem is the column-averaged dry air mole fraction of CO₂, denoted X_{CO_2} , which is a weighted average CO₂ concentration from the first 20 parameters of the state vector. Motivated by this, we partition the state vector as

$$x = \begin{bmatrix} x_\alpha \\ x_\beta \end{bmatrix}, \quad (3)$$

where x_α contains the CO₂ profile and x_β has the rest. We can now calculate X_{CO_2} as

$$X_{CO_2} = h^T x_\alpha, \quad (4)$$

where $h \in \mathbb{R}^{20}$ is a pressure weighting function [23] assigning appropriate weights to the CO₂ profile.

As noted in [17], the surrogate model does not include several elements of the operational retrieval's state vector, including a temperature profile offset, a water vapor profile scale factor, band-specific dispersion parameters for wavelength offsets, band-specific empirical orthogonal function scale factors, and solar-induced fluorescence (SIF) in the O₂A-band. As such, the purpose of the surrogate model is to act as a computationally feasible means for quantifying the propagation of uncertainty in input arguments into the algorithm outputs. Based on these insights, further more quantitative experiments can then be carried out for the Full Physics model.

2.3. Bayesian Formulation of the Inverse Problem

We follow the Bayesian approach [24] to solving inverse problems, which gives us a solution and the related uncertainties as a statistical distribution of states x conditioned on observed data y and regularized by a choice of prior distribution. This is obtained through Bayes' formula as

$$\mathbb{P}(x|y) \propto \mathbb{P}(y|x)\mathbb{P}_{pr}(x), \quad (5)$$

where $\mathbb{P}(x|y)$ is the posterior distribution, $\mathbb{P}(y|x)$ the likelihood and $\mathbb{P}_{pr}(x)$ the prior distribution. The proportionality \propto comes from a constant that does not depend on the unknown x . In this work, we assume the prior to be Gaussian, which is denoted as $\mathcal{N}(x_0, S_{pr})$, i.e.

$$\mathbb{P}_{pr}(x) \propto \exp\left(-\frac{1}{2}(x-x_0)^T S_{pr}^{-1}(x-x_0)\right). \quad (6)$$

For comparability, we use the same prior mean x_0 and covariance S_{pr} as the operational OCO-2 retrieval. Also, the additive noise is assumed to be zero-mean Gaussian with known covariance matrix, $\varepsilon \sim \mathcal{N}(0, S_{obs})$. This way, the likelihood function will have the form

$$\mathbb{P}(y|x) \propto \exp\left(-\frac{1}{2}(y-F(x))^T S_{obs}^{-1}(y-F(x))\right). \quad (7)$$

In this work, we use the same error statistics as [17], where the variance of y for the diagonal S_{obs} is given by

$$\text{Var}(y_{i,j}) = c_j F_{i,j}(z), \quad (8)$$

where $z \in \mathbb{R}^n$ is a simulated state vector, $j = 1, 2, 3$ is the index of wavelength band, $i = 1, \dots, m_j$ is the index of wavelengths in given band j and c_j denotes a band specific constant to yield broadly comparable signal-to-noise ratios to the operational OCO-2 retrieval. This parameterization is a simplification relative to the operational retrieval, which includes channel-specific parameters based on instrument calibration [25].

To solve this inverse problem, also referred to as *retrieval*, a common strategy in atmospheric remote sensing is to use an optimization algorithm such as Levenberg-Marquadt (see e.g., [10]) solver to find the posterior mode. This corresponds to minimizing the cost function

$$-2 \log \mathbb{P}(x|y) = (y-F(x))^T S_{obs}^{-1}(y-F(x)) + (x-x_0)^T S_{pr}^{-1}(x-x_0). \quad (9)$$

This procedure is often called *Optimal Estimation* (OE). We denote the minimizer obtained with OE by \hat{x} . On top of this, OE yields a linear Gaussian approximation of the posterior covariance, denoted $\hat{\hat{S}}$, such that

$$\hat{\hat{S}} = \left(K(\hat{x})^T S_{obs}^{-1} K(\hat{x}) + S_{pr}^{-1}\right)^{-1}, \quad (10)$$

where $K_{ij} = \frac{\partial}{\partial x_j} F_i$ is the Jacobian of the forward model.

2.4. Markov Chain Monte Carlo

While OE can provide a fast and computationally inexpensive solution to the retrieval problem, it is still a linear Gaussian approximation of the true posterior. In the case of OCO-2 retrieval, the forward model is non-linear and hence we need a way to characterize the actual underlying distribution. When the forward model is non-linear, the posterior distribution can be explored by Markov chain Monte Carlo (MCMC) sampling. In this work, we follow [16] and utilize adaptive Metropolis-Hastings MCMC

from [26]. The algorithm progresses by moving from state x_t to a proposed state x_{t+1} drawn from a proposal distribution $\mathcal{N}(x_t, C_t)$ with a density denoted $q(x, \cdot)$. The new point is then accepted or rejected with probability

$$\min \left(1, \frac{\mathbb{P}(x_{t+1}|y)q(x_t, x_{t+1})}{\mathbb{P}(x_t|y)q(x_t, x_{t+1})} \right) \tag{11}$$

In the adaptive Metropolis algorithm, we are additionally updating the proposal covariance matrix by empirically calculating the covariance from the already obtained chain:

$$C_t = \begin{cases} C_0, & t < t_0, \\ s_d \text{cov}(x_0, \dots, x_{t-1}) + \epsilon \mathbb{I}, & t \geq t_0 \end{cases} \tag{12}$$

where t_0 is the training length and s_d is a scaling parameter from [26] that optimizes the sampling efficiency. Also, as was noted in that article, we extend the adaptation length to updating the covariance every 100th time step after the initial t_0 .

2.5. Likelihood-Informed Subspace

In atmospheric remote sensing the information content of the measurement is an important concept to be considered e.g., when designing the instruments and constructing the retrieval methods. Most of the time, the actual amount of informative directions in the measurement is small compared to the dimension of the state vector. This is discussed in detail by Rodgers [10]. Consistent with this idea, the OCO-2 Algorithm Theoretical Basis Document (ATBD) [23] states that the CO₂ part of the state vector only has 2 degrees of freedom for the signal. Motivated by these observations, we use the likelihood-informed subspace (LIS) Dimension reduction ([21,22]) to truncate the dimension of the inverse problem in order to speed up the MCMC computation. It was shown in [15] that initial derivation of LIS is equal to Rodgers’ discussion on informative directions of the measurement, so we present the basics of the formulation using the Rodgers’ formalism.

We identify a subspace of the state space that has all the measurable information of the retrieval. This can be found by first linearizing the inverse problem and then using linear transformations to rotate and scale the space so that we can see how much of the natural variability of x , coming from the prior, can actually be distinguished from the noise ϵ . This is accomplished by first scaling the problem so that we end up with unit covariances for prior and error, and then diagonalizing the resulting whitened problem. We can then see how many independent degrees of freedom we can obtain from our measurement for the state x by finding the directions of variance greater than unity, since the rest are indistinguishable from the measurement noise.

Consider a linearized version of the inverse problem in Equation (1),

$$y = K(x - x_0) + \epsilon, \tag{13}$$

where K denotes the Jacobian matrix of the forward model with elements $K_{ij} = \frac{\partial}{\partial x_j} F_i$. Next, we use the Cholesky factorizations of prior and error covariances,

$$S_{pr} = \mathcal{L}_{pr} \mathcal{L}_{pr}^T, \quad S_{obs} = \mathcal{L}_{obs} \mathcal{L}_{obs}^T, \tag{14}$$

and perform pre-whitening of the problem by setting

$$\tilde{y} = \mathcal{L}_{obs}^{-1} y, \quad \tilde{K} = \mathcal{L}_{obs}^{-1} K \mathcal{L}_{pr}, \quad \tilde{x} = \mathcal{L}_{pr}^{-1} (x - x_0) \text{ and } \tilde{\epsilon} = \mathcal{L}_{obs}^{-1} \epsilon. \tag{15}$$

Pre-whitening transforms the error and prior distributions to have zero mean and unit covariance, that is, $\tilde{\varepsilon} \sim \mathcal{N}(0, \mathbb{I})$ and $\tilde{x} \sim \mathcal{N}(0, \mathbb{I})$. Now the problem can be written as

$$\tilde{y} = \tilde{K}\tilde{x} + \tilde{\varepsilon}. \tag{16}$$

We now consider the prior predictive distribution for whitened measurements \tilde{y} , obtained by evaluating the linearized forward model at states \tilde{x} drawn from the whitened prior distribution with added white noise. It follows that \tilde{y} will be distributed with covariance

$$\tilde{S}_y = \mathbb{E}[\tilde{y}\tilde{y}^T] = \mathbb{E}[(\tilde{K}\tilde{x} + \tilde{\varepsilon})(\tilde{K}\tilde{x} + \tilde{\varepsilon})^T] = \tilde{K}\tilde{K}^T + \mathbb{I}. \tag{17}$$

To get rid of possible non-diagonal elements in the resulting covariance matrix, we further diagonalize the scaled problem by using the singular value decomposition of the whitened Jacobian,

$$\tilde{K} = W\Lambda U^T. \tag{18}$$

We denote this scaling by a superscript $'$. This gives us the diagonalized problem

$$\begin{aligned} y' &= W^T\tilde{y} = W^T\tilde{K}\tilde{x} + W^T\tilde{\varepsilon} \\ &= \Lambda U^T\tilde{x} + W^T\tilde{\varepsilon} \\ &= \Lambda\tilde{x}' + \tilde{\varepsilon}'. \end{aligned} \tag{19}$$

The transformations ε' and x' conserve the unit covariance matrices for the scaled error and prior distributions, since W^TW and U^TU both result in unit matrix. It follows that the scaled data y' is distributed with covariance $\Lambda^2 + \mathbb{I}$. This is a diagonal matrix with unit contribution from the noise in each diagonal element. From this covariance we can deduce that only the variability coming from states \tilde{x} corresponding to singular values λ_i of scaled Jacobian \tilde{K} that are greater than unity can be distinguished from the measured data; the rest is masked by measurement noise. In other words, the number of singular values greater than unity correspond to the number of informative directions or degrees of freedom for signal in the measurement. Furthermore, degrees of freedom for signal and noise are invariant under linear transformations [10], and as the unknown x and the error ε are assumed to be independent, the same result is also valid for the original y .

On the other hand, the linear approximation of the posterior distribution given by Equation (16) yields the following expression for the posterior covariance of \tilde{x} :

$$\tilde{S}_{post} = (\tilde{K}^T\tilde{K} + \mathbb{I})^{-1}. \tag{20}$$

Using the same reasoning as before, we see that only the directions in the likelihood corresponding to singular values of scaled Jacobian that are greater than unity can be obtained from the measurement, and the rest are given by the prior. This leads to a connection with the work in [21,22], given also by the fact that $\tilde{K}\tilde{K}^T$ has the same eigenvalues as $\tilde{K}^T\tilde{K}$, all of which are the squared singular values of \tilde{K} . The matrix $\tilde{K}^T\tilde{K}$ in turn is the same thing as the prior-preconditioned Gauss-Newton Hessian

$$\tilde{H} := \mathcal{L}_{pr}^T H \mathcal{L}_{pr} = \mathcal{L}_{pr}^T K^T S_{obs}^{-1} K \mathcal{L}_{pr} = \tilde{K}^T \tilde{K} \tag{21}$$

from [21]. Here, H denotes the Gauss-Newton approximation of the Hessian of $-2 \log \mathbb{P}(x|y)$.

Informative directions of the measurement can now be identified with the first r eigenvectors of \tilde{H} corresponding to the first r eigenvalues greater than unity. These are the first r columns of matrix V given by

$$\tilde{H} = V\Lambda^2V^T. \quad (22)$$

We now use the informative directions to reduce the dimension of the inverse problem. Consider the low-rank approximations for the posterior of the form

$$\tilde{\mathbb{P}}(x|y) \propto \mathbb{P}(y|P_r x)\mathbb{P}_{pr}(x), \quad (23)$$

where P_r is rank r projection matrix. In [21,22] it was shown that the optimal P_r that minimizes the Hellinger distance between rank r approximations and the full posterior is given by the eigendecomposition of \tilde{H} as follows:

Definition 1. [LIS]: Let $V_r \in \mathbb{R}^{n \times r}$ be a matrix containing the first r eigenvectors of the prior preconditioned Gauss-Newton Hessian \tilde{H} . Define

$$\Phi_r := \mathcal{L}_{pr}V_r \text{ and } \Theta_r := \mathcal{L}_{pr}^{-T}V_r. \quad (24)$$

The rank r LIS projection for the low-rank posterior approximation in Equation (23) is given by

$$P_r = \Phi_r\Theta_r^T. \quad (25)$$

The range \mathbb{X}_r of projection $P_r : \mathbb{R}^n \rightarrow \mathbb{X}_r$ is a subspace of state space \mathbb{R}^n spanned by the column vectors of matrix Φ_r . We call the subspace \mathbb{X}_r the *likelihood-informed subspace (LIS)* for the linear inverse problem, and its complement $\mathbb{R}^n \setminus \mathbb{X}_r$ the *complement subspace (CS)*.

Definition 2. The matrix of singular vectors $V = [V_r V_\perp]$, where $V_\perp \in \mathbb{R}^{n \times n-r}$ contains the rest of the eigenvectors not included in V_r , forms a complete orthonormal system in \mathbb{R}^n and we can define

$$\Phi_\perp := \mathcal{L}_{pr}V_\perp \text{ and } \Theta_\perp := \mathcal{L}_{pr}^{-T}V_\perp \quad (26)$$

and the projection $\mathbb{I} - P_r$ can be written as

$$\mathbb{I} - P_r = \Phi_\perp\Theta_\perp^T. \quad (27)$$

Define the LIS-parameter $x_r \in \mathbb{R}^r$ and the CS-parameter $x_\perp \in \mathbb{R}^{n-r}$ as

$$x_r := \Theta_r^T x, \quad x_\perp := \Theta_\perp^T x. \quad (28)$$

The parameter x can now be naturally decomposed as

$$\begin{aligned} x &= P_r x + (\mathbb{I} - P_r)x \\ &= \Phi_r x_r + \Phi_\perp x_\perp. \end{aligned} \quad (29)$$

Using the previous definitions, we can write the approximate posterior as

$$\tilde{\mathbb{P}}(x|y) = \mathbb{P}(y|\Phi_r x_r)\mathbb{P}_{pr}(x_r)\mathbb{P}_{pr}(x_\perp). \quad (30)$$

2.6. MCMC Sampler

When dealing with a non-linear forward model, the Hessian matrix depends on the point x and thus is not constant on the state space. It was suggested in [21] that a way to get around this issue is to take a Monte Carlo average of \hat{H} over some reference distribution. The effect of different distributions was further explored in [27]. Since we have obtained the posterior mode and the related covariance approximation from the OE retrieval, we use these to get an average

$$\hat{H}_N = \mathcal{L}_{pr}^T \left(\frac{1}{N} \sum_{k=1}^N H(x^{(k)}) \right) \mathcal{L}_{pr}, \tag{31}$$

where the N samples $x^{(k)}$ are drawn from $\mathcal{N}(\hat{x}, \hat{S})$. We then use the eigenvalue decomposition $\hat{H}_N = \hat{V} \hat{\Lambda}^2 \hat{V}^T$ to find a basis for the global LIS analogously to the linear case. We then proceed to sample x_r from the reduced posterior $\mathbb{P}(y|\Phi_r, x_r) \mathbb{P}_{pr}(x_r)$, and add a sample drawn from the Gaussian complement prior $\mathbb{P}_{pr}(x_{\perp})$ to get an approximation of the full posterior.

In adaptive MCMC, a good starting value for C_0 as a training covariance is $s_d \hat{S}$. However, since we already have computed the expected value of the Gauss-Newton Hessian in the LIS computation, we can use this Monte Carlo estimator to better capture the variability over the non-linear posterior and set

$$C_0 = s_d \left(\hat{H} + S_{pr}^{-1} \right)^{-1}, \tag{32}$$

where \hat{H} is the expectation of the Gauss-Newton Hessian over the reference distribution $\mathcal{N}(\hat{x}, \hat{S})$. Next, we project C_0 into LIS and use that as a starting covariance for the MCMC. The projection is given by

$$\Theta_r C_0 \Theta_r^T := \tilde{C}_r \in \mathbb{R}^{r \times r}. \tag{33}$$

After dimension reduction, the MCMC sampling then proceeds with the standard Adaptive Metropolis algorithm where we draw at time step t a new candidate point from proposal $\mathcal{N}(x_{t-1}, C_t)$. Acceptance probability is computed by first projecting the LIS parameter x_r into full space as

$$\Phi_r x_r + x_0 \in \mathbb{R}^n. \tag{34}$$

The likelihood function is then evaluated using the full-dimensional vector as an input to the surrogate model. Finally, after completing the MCMC chain, we draw a sample with the same length as the chain from the complement prior

$$x_{\perp} \sim \mathcal{N}(0, \mathbb{I}), x_{\perp} \in \mathbb{R}^{n-r}. \tag{35}$$

The sample is then projected back to full space as

$$\Phi_{\perp} x_{\perp} \in \mathbb{R}^n \tag{36}$$

and added to the LIS-chain. It follows from (23) that this way we get a sample from the optimal rank r posterior approximation, as can be seen from the CO₂ profiles in Figure 3.

3. Results

We present our results in 3 subsections. The example soundings we use are chosen from the data used in [16], which is based on the work in [17]. In short, this study used a clustering method to group the first 18 months of global OCO-2 measurements into 100 geolocation templates based on physical, climatological

and temporal properties. Within a template, synthetic data was created by randomly sampling a Gaussian mixture model, the parameters of which were empirically gained from OCO-2 CO₂ profiles and MERRA climatology. We specifically chose soundings from a template in South America (centered at Southern Bolivia) in which the initial results of OE and MCMC in [16] did not agree. The dominating aerosol species for this template are sulphate and dust (see the dominating aerosol type map in [23]). First, we focus on comparing the LIS implementation and its results with previous MCMC results from [16], as well as showing the relative contributions of LIS and CS to the sampling, as well as improvements achieved in computational speed of the MCMC when using LIS. In the second section we inspect the shapes of posterior marginal distributions of model aerosol parameters and compare the MCMC results with OE. Lastly, we perform MCMC retrievals for three example cases using both the operational OCO-2 prior covariance matrix and a modification of it, in which the prior has been widened to allow a greater variability on aerosol layer width parameters.

3.1. LIS Implementation

For constructing the LIS basis for the retrieval problem, we use the posterior approximation obtained from OE and sample 200 points from $\mathcal{N}(\hat{x}, \hat{S})$ which are then used to compute the Monte Carlo average (see Section 2.6). As seen in Figure 1, the total number of eigenvalues greater than unity is found to be 18, which is in accordance with [23] stating that the degrees of freedom for signal in the full physics model is around 20. Moreover, the number of eigenvalues greater than unity can be seen to converge to 18 even when using 100 samples, so using 200 samples will reasonably ensure this convergence. In order to ensure that all the variability is captured by LIS we set $r = 19$ as the effective dimension of the problem, as the 19th eigenvalue is still relatively close to 1. Our tests showed that using less than 19 eigenvectors resulted in a significantly different posterior distribution than the full dimensional case, whereas using 19 or more resulted in the same posterior as in the full dimensional case.

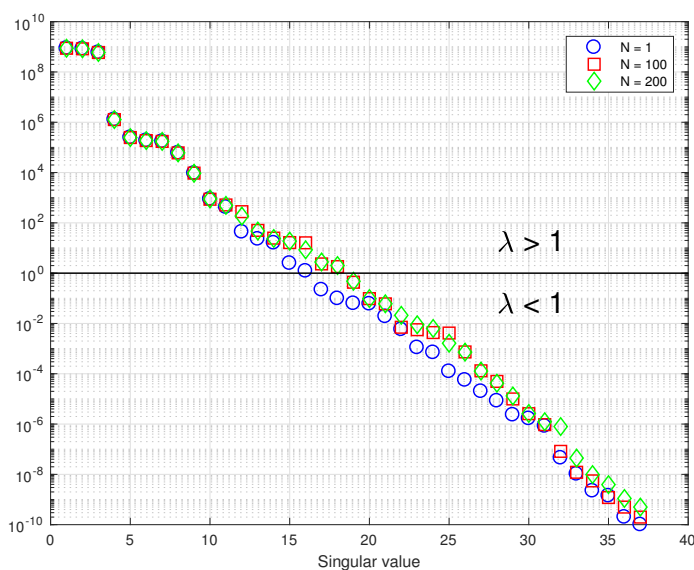


Figure 1. Eigenvalues of the Hessian matrix with different numbers of samples used in the calculation of the Monte Carlo average 2.6. The eigenvalues are separated to less than and greater than unity by a black line.

As an example case, we start the comparative MCMC chain from a value that significantly disagrees with the simulated true value of X_{CO_2} . We perform the full dimensional MCMC with a chain length of 250,000 samples with a training time of 1000 samples as was done in previous work. For comparability, we perform the dimension reduced MCMC runs using the same sample sizes. The resulting X_{CO_2} chains for both runs are shown in Figure 2. We also show the histograms of posterior X_{CO_2} for both runs as well as the simulated ground truth value. When comparing the last 100 000 samples of the chains, we get acceptance rates of 1.7% for the full-dimensional MCMC and 5.1% for LIS MCMC, and effective sample sizes (number of samples that can be considered uncorrelated; see e.g., [28], pp. 5–7) of 54 and 134, respectively. The scaling parameter for adaptive MCMC was set to $s_d = 0.3 \times 2.4^2/d$, where d is the dimension of the estimated parameter. This choice led to a better acceptance rate for the chain which is better for data visualization, but did not otherwise affect the efficiency of the sampler.

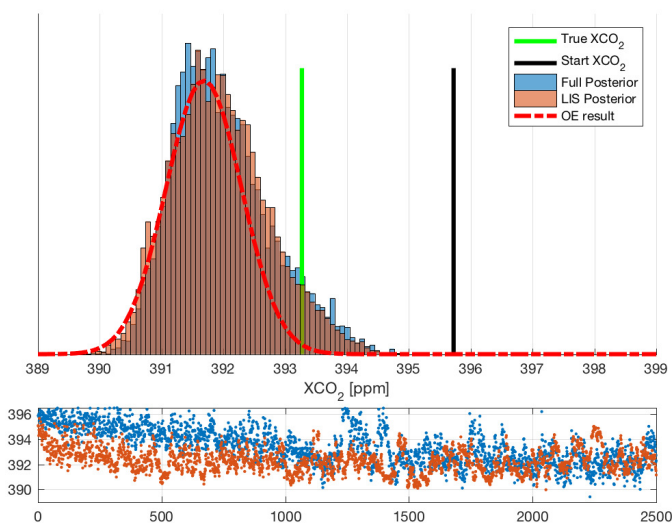


Figure 2. Upper panel: X_{CO_2} posterior histogram from full dimensional MCMC (blue) compared to LIS MCMC (red). Also shown the starting value for MCMC and simulated ground truth (green). Lower panel: every 100th sample of X_{CO_2} chain from full dimensional MCMC (blue) and LIS MCMC (red).

Although the OCO-2 algorithm retrieves a full CO_2 profile on 20 levels, it only validates its full-column retrievals, as there are typically only two degrees of freedom for CO_2 information in the retrieved profile. Nevertheless, we show the Optimal Estimation's retrieved 20 pressure levels in the left panel of Figure 3 plotted against the MCMC results and the ground truth. The contributions from the LIS parameter x_r and CS parameter x_{\perp} projected back to the full space are also shown in the right panel of Figure 3. Together, these results show that the LIS implementation works well and succeeds in capturing all the information present in the measured data with an increased computational efficiency. Furthermore, dimension reduced MCMC seems to give a decent estimation of the CO_2 profile as well.

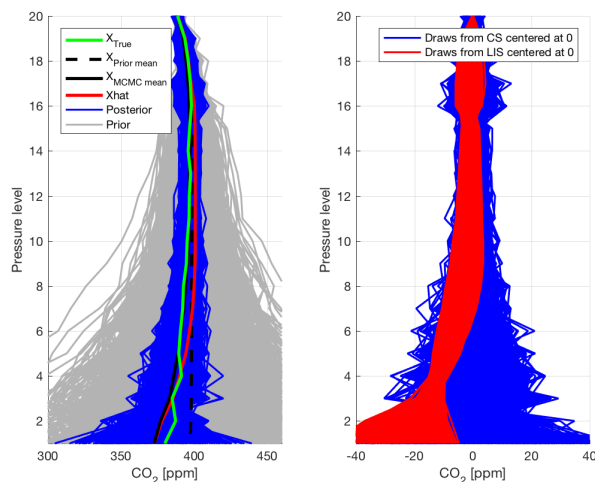


Figure 3. Left panel: CO₂ profiles from the MCMC retrieval compared with prior mean, ground truth and OE retrieval result. Right panel: CO₂ parts of samples from LIS and CS projected back to full space.

3.2. MCMC Results

The previous subsection offers an opportunity to conduct further analysis with dimension reduction accelerated MCMC. With the MCMC approach we can compute the full posterior distribution, and a natural question is whether the OE optimization algorithm actually finds the mode of the posterior distribution and how well this corresponds to the expected value obtained from MCMC sampling.

In the studied case the initial termination criteria for OE suggested that the cost function did not have a unique minimum, but our tests concluded that by utilizing the information of the MCMC run, the OE algorithm could be tuned so that differing starting values did actually converge to same minimum given enough computation time. Details of this parameter tuning can be found in Appendix A.1. The parameter tuning we performed achieved an approximate agreement with MCMC mean and \hat{x} estimate obtained from OE, but this resulted in significantly more iterations required for the optimization. However, this effect is not further addressed in this work as we focus on comparing the results of a well-performing optimization algorithm and MCMC.

Next, we perform MCMC simulation for the same example sounding as before (from the South American template from [16]) using the improved optimization algorithm parameter values from Appendix A.1 to initialize the MCMC with a more reliable OE. Figure 4 shows the marginal posterior distributions for all state vector aerosol parameters. A clear non-Gaussian shape is detected in 1D marginal posterior distributions of dust and cloud liquid water AOD parameters, which also correlates to the OE not being able to pick up the full posterior width. The true value of the aerosol parameters is relatively poorly captured by both MCMC and OE mean estimates. Especially the parameters for sulphate and water coefficient 2 (aerosol altitude) have the true value outside the edge of the posterior histogram, which means that these parameters were not identified by MCMC.

We further illustrate the non-Gaussian behavior by plotting the 2D distributions of these state vector elements in Figure 5, and the uncertainties coming to X_{CO_2} from aerosol parameters as 2D distributions in Figure 6. Both of these figures show that the posterior estimate as 95% confidence intervals obtained from OE is consistently not able to contain the full uncertainty in these parameters as shown by the MCMC

simulation. Compared to OE, the probability region given by MCMC still contains the true value in most cases, which means that the uncertainty estimate from OE is again inadequate. We can further see that parameters 29 and 38 (sulphate and water altitudes) have the MCMC posterior tail away from the true value, which further shows that these directions are not identified and cause error to X_{CO_2} .

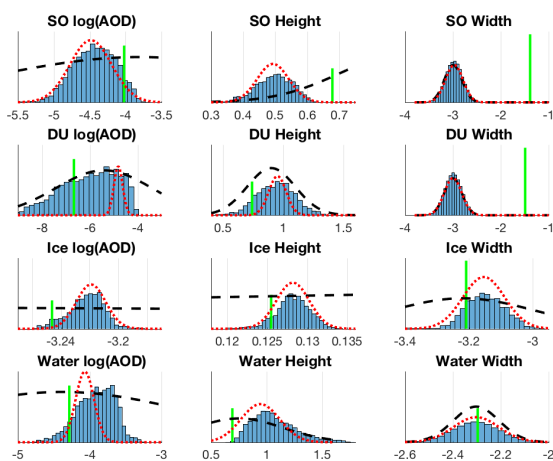


Figure 4. Aerosol parameter marginal histograms with operational prior (black dashed line), ground truth (green) and OE retrieval result (red dotted line). SO = sulphate, DU = dust, Ice = cloud ice and Water = cloud liquid water.

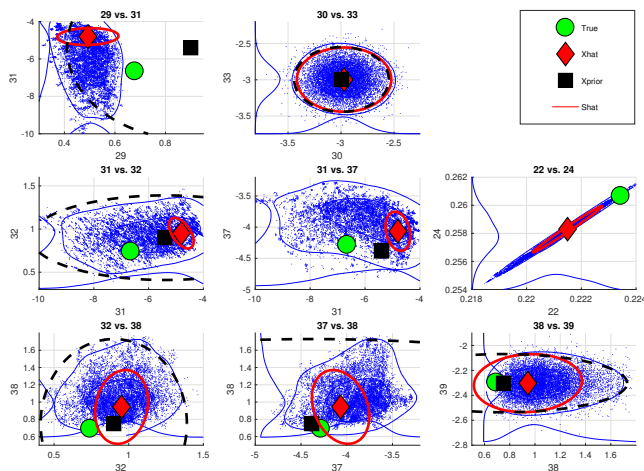


Figure 5. 2D posterior distributions of selected aerosol and surface parameters from LIS MCMC (blue with 95% posterior confidence interval on blue contour) compared with ground truth (green), prior (black with 95% confidence interval on black ellipse) and OE (red with 95% posterior confidence interval on red ellipse). Parameter numbers agree with the numbering in the surrogate model description (see Table A3).

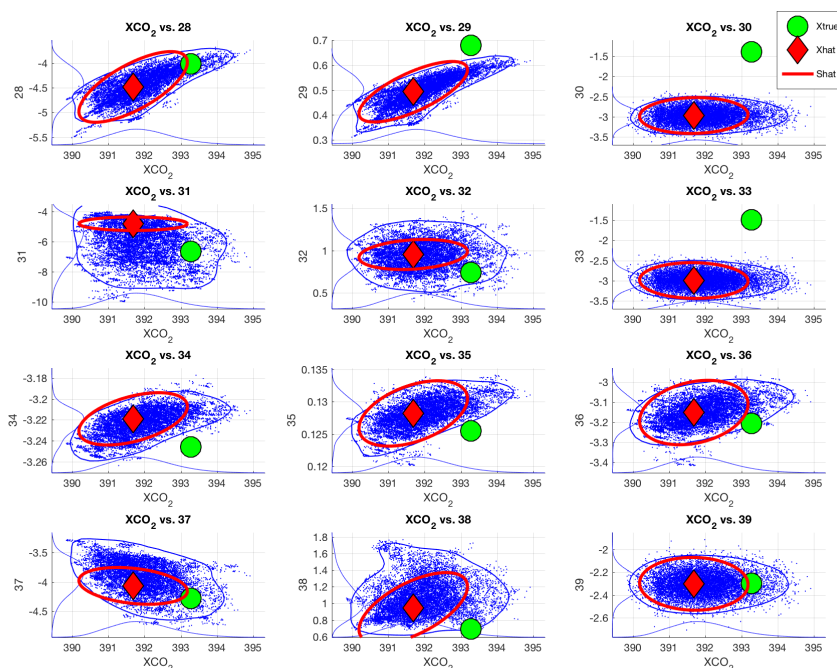


Figure 6. 2D posterior distributions of X_{CO_2} against aerosol parameters from LIS MCMC (blue with 95% posterior confidence interval on blue contour) compared with ground truth (green) and OE (red with 95% posterior confidence interval on red ellipse). Parameter numbers agree with the numbering in the surrogate model description and correspond to those in Figure 4.

3.3. Wider Prior

In addition to gaining information about the shape of the posterior distribution, we can see from Figure 5 that the simulated state vector's parameters 30 and 33 (corresponding to sulphate and dust layer thickness parameters) are outside the prior distribution induced by the operational prior. Since the state vectors were simulated using an empirical covariance derived from a credible ensemble of simulated and meteorological data ([17], Section 4.1), this phenomenon might possibly occur in the operational OCO-2 retrieval as well, because both retrievals use the same prior. It is worth noting that in the operational retrieval, the prior constraint on the aerosol width is so tight that this parameter is effectively not retrieved. This was based on early work [29] which showed that the spectra had little sensitivity to it.

Focusing on the 3 parameters describing the sulphate aerosol cloud, the posterior correlation matrix of \hat{S} in Figure 7 shows that parameters 28 and 29 (sulphate log AOD and aerosol altitude) have a strong correlation with the CO_2 part of the state vector. Additionally, we can deduce that if the aerosol profile width (parameter 30) is outside the prior range and hence cannot be retrieved, the retrieval might try to compensate for this with the remaining sulphate parameters. This in turn could cause the observed MCMC mean in Figure 2 to not align with the simulated ground truth. To test this hypothesis, we first select 2 additional example soundings from the same template as before such that they as well have state

vector parameter 30 valued outside the operational prior distribution. We then perform a second round of OE and MCMC but this time we increase the entry in the prior covariance's diagonal element 30 so that the prior is in effect non-informative for this parameter, allowing variability that captures the true simulated value. Furthermore, since the prior is diagonal for the aerosol parameters, any cross-correlation adjustment is not necessary.

We illustrate the observed effect of widening the prior covariance to the retrieved X_{CO_2} in Figure 8. The X_{CO_2} computed from both chains (using operational (blue) and widened (red) prior) are plotted in comparison with the corresponding OE results and the simulated ground truth. We can clearly see that the wider prior covariance had, in these cases, the effect of significantly improving the accuracy of both MCMC and OE retrievals of X_{CO_2} . In Figures 9–11, we have summarized the MCMC histograms, for each test case, of parameters relating to log AOD, height, and width of the modelled aerosol layer for all 4 aerosol types present in the retrieval. We also show the ground truth, prior distribution and OE results for comparison. The parameter SO Coefficient 3 on the upper right panel is the one with widened prior, and in comparison to previous histograms in Figure 4, we see that this parameter is now retrieved relatively well. As we can see, this change also had the effect of improving the overlap of resulting posterior histograms and the true value.

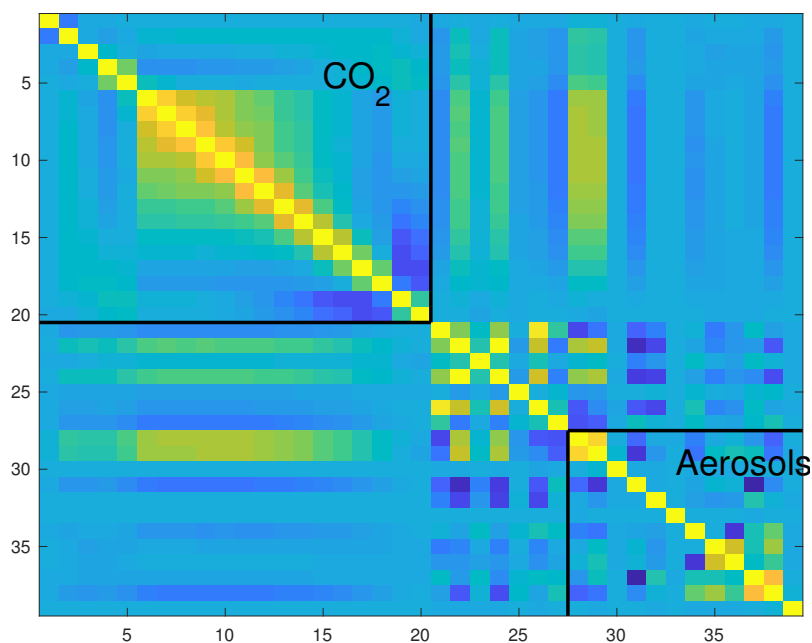


Figure 7. Posterior correlation matrix of \hat{S} . The elements of the matrix corresponding to CO_2 and aerosol parameters of the state vector are separated with black lines. It should be noted that the first aerosol parameters have a high correlation with the CO_2 part of the state vector.

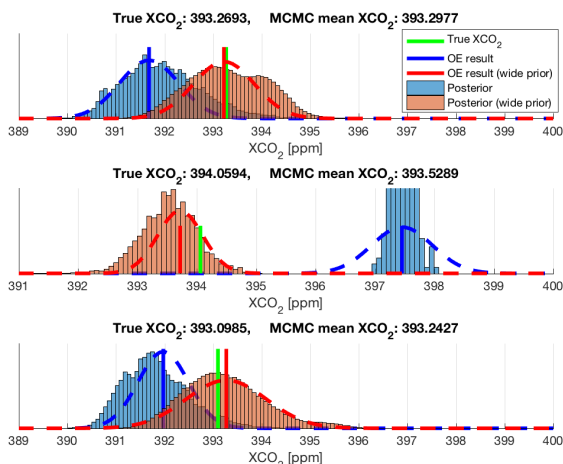


Figure 8. Three separate test cases showing X_{CO_2} posterior histograms (using operational prior (blue) and widened prior) from LIS MCMC compared to the OE retrieval (blue and red dashed lines) and simulated ground truth (green). Also shown the true value of X_{CO_2} with the corresponding MCMC mean using a widened prior. It should be noted that the example case 2 on the middle panel most likely did not converge in either the MCMC or OE when using the operational prior, which is not an unusual scenario even in the operational retrieval.

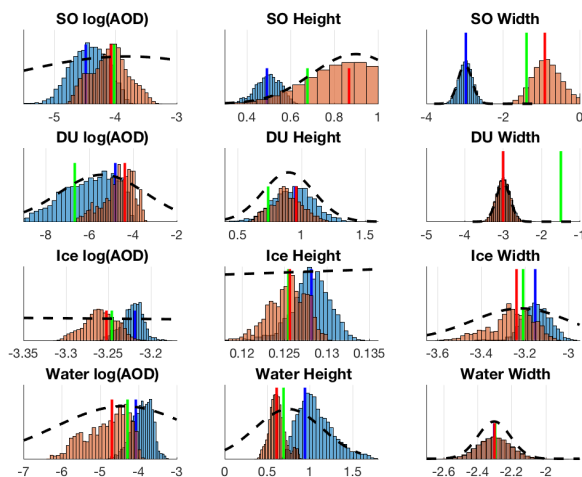


Figure 9. Test case 1: aerosol parameter marginal histograms (corresponding to retrievals using operational prior (blue) and widened prior (orange)) with operational prior (dashed black), ground truth (green) and OE retrieval result (blue and red vertical lines). SO = sulphate, DU = dust, Ice = cloud ice and Water = cloud liquid water.

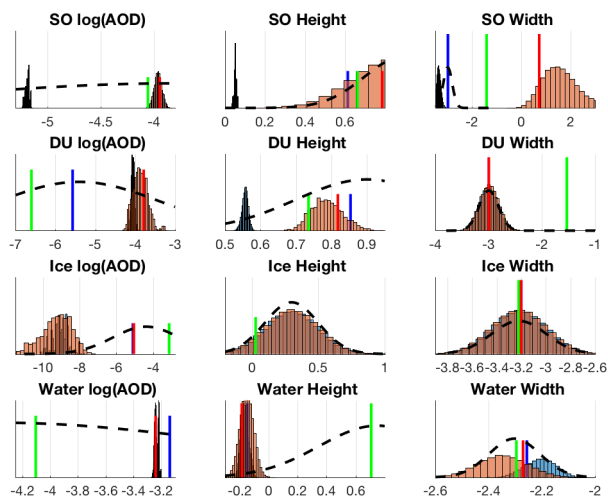


Figure 10. Test case 2: aerosol parameter marginal histograms (corresponding to retrievals using operational prior (blue) and widened prior (orange)) with operational prior (dashed black), ground truth (green) and OE retrieval result (blue and red vertical lines). SO = sulphate, DU = dust, Ice = cloud ice and Water = cloud liquid water.

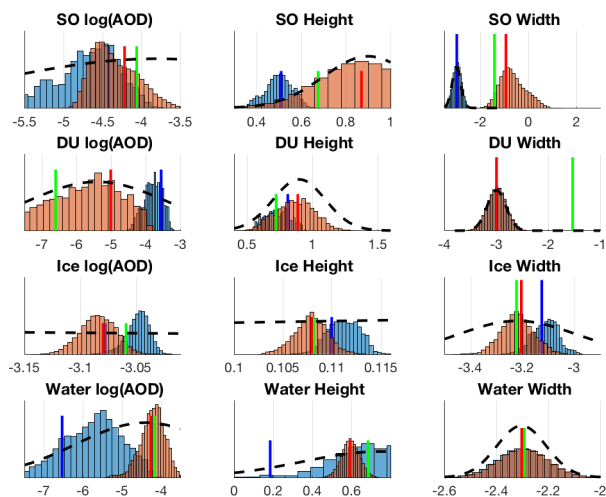


Figure 11. Test case 3: aerosol parameter marginal histograms (corresponding to retrievals using operational prior (blue) and widened prior (orange)) with operational prior (dashed black), ground truth (green) and OE retrieval result (blue and red vertical lines). SO = sulphate, DU = dust, Ice = cloud ice and Water = cloud liquid water.

4. Conclusions

Previous work on Monte Carlo experiments with the OCO-2 retrieval problem laid out the need to accelerate the convergence of these demanding computations. We have shown that in this case, where the problem is actually over-parametrized as the parameter space has a relatively low amount of informative directions, LIS dimension reduction can accelerate convergence of the MCMC simulation. In addition to the known fact that the CO₂ profile has only 2 degrees of freedom, the low information content of the aerosol part is also supported by recent research in e.g., [20]. The low information content is apparent when comparing the full-dimensional X_{CO₂} MCMC chain (Figure 2) against the corresponding low dimensional LIS chain. Both were initiated at a value just outside the resulting posterior distribution, and upon visual inspection of the chains we can conclude that the LIS chain converges to the same result at least three times faster. Faster convergence is also supported by increased acceptance rate and effective sample size, as is to be expected thanks to the reduced dimension. This gain in computational efficiency makes it more feasible to conduct further studies with different geolocation templates.

The MCMC simulations allowed us to see the shape of the posterior distribution's aerosol and surface parameters that have repeatedly caused problems in previous work relating to the OCO-2 retrieval. The marginal distributions shown in Figure 4 show clear non-Gaussian shapes for some aerosol parameters, as well as the OE posterior being too narrow to contain the actual error limits. The most pathological cases are assembled in Figure 5 as 2D plots. In addition to showing again the inadequate coverage of \hat{S} , two observations can be made: first of all, parameters 31 and 37, both of which correspond to log(AOD), form a so-called "banana" distribution, which is a curved part of the distribution. This is known to cause problems for the adaptive MCMC (see e.g., [30]) which in part explains the low acceptance rates we obtained. Although we were mainly focused on aerosol parameters, we have included a plot of parameters 22 vs. 24 in this figure as well. These parameters correspond to albedo in the strong and weak CO₂ bands. As we can see from Figure 5, these parameters are strongly correlated, which is also a known feature that causes a "flat" posterior distribution which is harder to be characterized by MCMC algorithms.

The X_{CO₂} histogram shown in Figure 2 suggests that, at least in this case, the well-tuned OE retrieval agrees quite well with MCMC in finding the posterior mean, but gives a too narrow uncertainty in comparison. This is in contrast to the findings of [16], but could be explained by OE not having converged to a global minimum. We have plotted the X_{CO₂} distribution against the aerosol parameters in Figure 6, which further illustrates that OE underestimates the uncertainties propagated to X_{CO₂} especially from the optical depth parameters of dust and cloud water. Although we have tested this with several other retrievals (see e.g., Figure 8), they were all from the same template with same dominating aerosol species and geophysical conditions. These results suggest there would be added value to conduct a comprehensive further thorough examination of different geolocation templates in the future.

Also, from panel 30 vs. 33 and from panels of SO and DU Coefficient 3 in Figure 4, we can see that the simulated ground truth is outside of the range of the operational prior distribution in the third parameter of both SO (sulphate) and DU (dust). This indicates that the retrieval tries to compensate the out-of-range-truth with other parameters and thus we might end up seeing possibly erroneous OE and MCMC results. As is apparent from Figure 7, these aerosol parameters (log AOD and profile height) are strongly correlated with the 20 CO₂ parameters and as such they can induce significant errors to the profile if retrieved incorrectly. Although this observation can in part explain the inconsistencies in the previous simulation experiments, it can also give a clue to the properties of the operational retrieval covariance matrix, since the state vectors used in the studies are based on real-world empirical physical parameters (see [17]). This also highlights the importance of prior validation and the use of MCMC to inspect the actual posterior distribution, since the OE approximation can give a misleading picture of the actual posterior uncertainties: a non-smooth likelihood function that might have a non-Gaussian

shape or multiple local minima, caused by e.g., the use of lookup tables in the model or the non-linear forward model.

Lastly, we perform the same MCMC retrieval for three separate test cases all having parameters 30 and 33 outside the range of the prior distribution, but with the difference that we inflate the prior covariance of above-mentioned SO Coefficient 3 parameter so that the resulting prior is practically non-informative in this direction and the state vector can vary freely. This naturally changes how the posterior distribution looks like, since the posterior is a product of the prior and likelihood distributions (see Equation (5)). The resulting X_{CO_2} histograms are plotted in Figure 8 and we can clearly see that the ground truth is this time found roughly in the middle of the posterior distribution. Following [9,29], the parameters describing the prior width of distributions of aerosols 1 and 2 in the operational retrieval algorithm are intentionally left small. This way they are practically not retrieved, as the effect of the total aerosol profile is assumed to be compensated by the log(AOD) and aerosol height parameters. However, as this means that the resulting posterior does not contain the original ground truth in these selected test cases, we end up with a faulty posterior mean.

Allowing more variability with a wider prior covariance can result in finding the correct CO_2 concentration, but it still does not mean that the retrieved aerosol parameters correspond credibly to the ground truth. As can be seen from Figures 9–11 as some marginal histograms do actually converge to include the ground truth, some on the other hand seem worse after changing the prior. Our experiments also showed that this improved behavior is not generally observed in all retrievals, and factors such as OE convergence, simulated profile shape, other out-of-bounds parameters and even possible aerosol model problems (see e.g., [9] for added stratospheric aerosol) may lead to the OE and MCMC not agreeing with each other and/or the ground truth, as was observed in [16], even when the prior is relaxed.

5. Discussion

Results of this work open up several interesting topics for further research. Repeating the earlier experiments of [16] with LIS MCMC would probably aid in getting a better understanding of differences between different templates, as aerosols used in the state vector and their models change according to geolocation. In this work we have only seen the effect of dust and sulphate aerosol types, whereas there exist three more types in the operational OCO-2 retrieval algorithm: organic carbon, black carbon and sea salt. Investigating the combined effect of aerosols is also of interest, as different aerosol types seem to together affect the retrieval in a correlated and non-trivial way according to our results. To further assess the effect of aerosol parameters on X_{CO_2} errors, a comprehensive study on several hundreds of different conditions should be conducted.

This work has demonstrated insights that can be gained with a strategic implementation of the LIS MCMC retrieval. The surrogate model illustrated here differs from the operational full physics retrieval in several ways, including the complexity of the state vector and radiance noise properties. A more efficient sampling algorithm also makes it more feasible to perform MCMC on the operational OCO-2 full physics model, and identify which features seen in this work are also present in the operational retrieval; although one might expect the results to be broadly applicable to the operational retrieval, there could be important differences not captured by the simplified state vector and forward model.

Author Contributions: Conceptualization, O.L., J.H., J.B., M.L., A.B. and J.T.; Data curation, O.L.; Formal analysis, O.L.; Funding acquisition, A.B. and J.T.; Investigation, O.L. and J.T.; Methodology, O.L., J.H., M.L. and J.T.; Project administration, A.B. and H.L.; Resources, J.H. and J.B.; Software, O.L., J.H. and J.B.; Supervision, M.L. and J.T.; Visualization, O.L.; Writing—original draft, O.L.; Writing—review & editing, O.L., J.H., J.B., M.L., J.T. and H.L.

Funding: This work has been financially supported by Finnish Academy Centre of Excellence in Inverse Modelling and Imaging (312125) and Academy of Finland project number 285421. Additional research was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. Support was provided by the Orbiting Carbon Observatory-2 (OCO-2) mission.

Acknowledgments: The work presented here contributes to FMI activities on the OCO-2 Science Team. The authors thank Annmarie Eldering, Michael Gunson, and James McDuffie for valuable suggestions and technical assistance. We would also like to thank the two anonymous reviewers for their extremely useful comments and feedback.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AOD	Aerosol optical depth
ATBD	Algorithm Theoretical Basis Document
CO ₂	Carbon Dioxide
CS	Complement Subspace
LIS	Likelihood-Informed Subspace
MCMC	Markov chain Monte Carlo
MDPI	Multidisciplinary Digital Publishing Institute
OCO-2	Orbiting Carbon Observatory 2
OE	Optimal Estimation
X _{CO₂}	Column averaged dry air mole fraction of CO ₂

Appendix A

Appendix A.1. OE Convergence

As was shown in [16,18], the optimal estimation retrieval seems to be sensitive to the first guess of the optimization algorithm, and furthermore might not agree with the overall posterior distribution obtained with MCMC. In order to rule out the possibilities of issues with the convergence of the Levenberg-Marquadt algorithm, we perform parameter tuning for the retrieval and compare the tuned version of the retrieval in our example case with the corresponding one used in [16,17].

In the previous work, the starting value for γ parameter in the retrieval was set to 10 ([23]). Also, the tolerance for normalized step size was set to 4. To test the effect of these parameters, we set the maximum amount of iterations to 1000 so that the optimization will terminate when one of the other conditions is met. As a starting value for the optimization, we use the prior mean, the mean obtained from MCMC simulation, and the mean perturbed by $\pm\sigma \in \mathbb{R}^{39}$, where σ is a standard deviation obtained by taking the square root of the elements of the diagonal of \hat{S} . The results are shown in the left panel of Figure A1 and Table A1. Next, we repeat the experiment by setting the starting value of γ to 30 and reducing the tolerance of normalized step size to 0.0001. The results are shown in the right panel of Figure A1 for the CO₂ part and in and Table A2 for the aerosol part of the state vector.

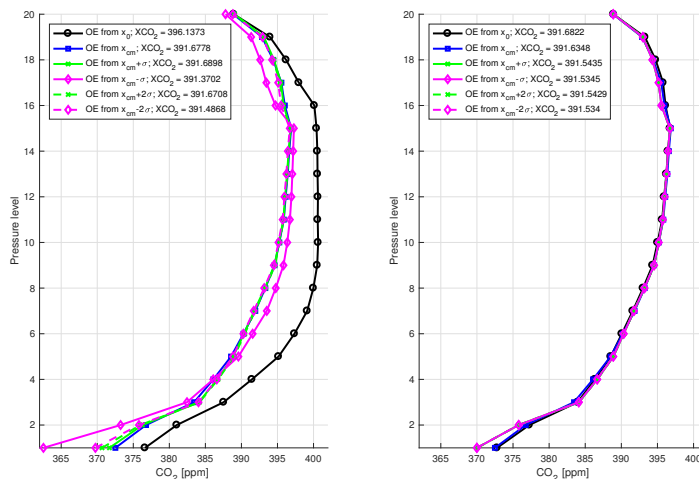


Figure A1. CO₂ profiles from OE retrievals with starting the optimization from different first guesses. **Left panel:** starting $\gamma = 10$, normalized step size tolerance = 40. **Right panel:** starting $\gamma = 30$, normalized step size tolerance = 0.0001.

Table A1. Aerosol parameter values and resulting X_{CO_2} values from OE retrievals shown on the left panel of Figure A1. SO = sulphate, DU = dust, Ic = cloud ice and Wa = cloud liquid water.

	OE 1	OE 2	OE 3	OE 4	OE 5	OE 6
SO 1	-3.42	-4.46	-4.47	-4.57	-4.47	-4.53
SO 2	0.68	0.49	0.49	0.48	0.49	0.48
SO 3	-2.99	-2.98	-2.95	-3.13	-2.95	-2.98
DU 1	-4.67	e4.72	-4.60	-5.28	-4.53	-5.32
DU 2	0.87	0.94	0.95	0.81	0.96	0.90
DU 3	-2.99	-3.00	-2.97	-3.14	-2.98	-3.00
WA 1	-3.33	-3.21	-3.21	-3.21	-3.21	-3.22
WA 2	0.13	0.12	0.12	0.12	0.12	0.12
WA 3	-3.13	-3.15	-3.15	-3.14	-3.15	-3.15
IC 1	-5.37	-4.08	-4.13	-3.93	-4.17	-3.86
IC 2	0.06	0.95	0.94	0.94	0.93	0.96
IC 3	-2.29	-2.30	-2.29	-2.37	-2.29	-2.30
X_{CO_2}	396.13	391.67	391.68	391.37	391.67	391.48

Table A2. Aerosol parameter values and resulting X_{CO_2} values from OE retrievals shown on the right panel of Figure A1. SO = sulphate, DU = dust, Ic = cloud ice and Wa = cloud liquid water.

	OE 1	OE 2	OE 3	OE 4	OE 5	OE 6
SO 1	-4.48	-4.48	-4.52	-4.52	-4.52	-4.52
SO 2	0.49	0.49	0.48	0.48	0.48	0.48
SO 3	2.97	-2.98	-2.97	-2.97	-2.97	-2.97
DU 1	-4.80	-4.72	-4.96	-5.08	-4.97	-5.12
DU 2	0.95	0.94	0.94	0.93	0.94	0.93
DU 3	-2.99	-3.00	-2.99	-2.99	-2.99	-2.99
WA 1	-3.21	-3.21	-3.22	-3.22	-3.22	-3.22
WA 2	0.12	0.12	0.12	0.12	0.12	0.12
WA 3	-3.15	-3.15	-3.15	-3.15	-3.15	-3.15
IC 1	-4.06	-4.06	-3.95	-3.92	-3.95	-3.91
IC 2	0.94	0.95	0.95	0.95	0.95	0.95
IC 3	-2.30	-2.30	-2.30	-2.30	-2.30	-2.30
X_{CO_2}	391.68	391.63	391.54	391.53	391.54	391.53

Appendix A.2. Surrogate Model State Vector

Table A3. Surrogate forward model state vector description: element names and the prior values for the sounding performed in this study.

No.	Name	Prior Value
CO2 Volume Mixing Ratio [ppm]		
1	Vertical Level 1 (Top of Atmosphere)	388.9123
2	Vertical Level 2	393.7508
3	Vertical Level 3	395.8462
4	Vertical Level 4	397.4091
5	Vertical Level 5 (Tropopause)	398.5947
6	Vertical Level 6	398.5761
7	Vertical Level 7	398.5522
8	Vertical Level 8	398.5237
9	Vertical Level 9	398.4913
10	Vertical Level 10	398.4556
11	Vertical Level 11	398.4172
12	Vertical Level 12	398.3766
13	Vertical Level 13	398.3341
14	Vertical Level 14	398.2900
15	Vertical Level 15	398.2445
16	Vertical Level 16	398.1977
17	Vertical Level 17	398.1495
18	Vertical Level 18	398.1000
19	Vertical Level 19	397.7311
20	Vertical Level 20 (Surface)	397.2732
21	Surface Pressure [hPa]	1000
Lambertian Albedo		
22	Strong CO ₂ Band Mean Albedo	0.2296
23	Strong CO ₂ Band Albedo Spectral Slope	0
24	Weak CO ₂ Band Mean Albedo	0.2577
25	Weak CO ₂ Band Albedo Spectral Slope	0
26	O ₂ A Band Mean Albedo	0.2132
27	O ₂ A Band Albedo Spectral Slope	0
Aerosols		
28	Sulphate Log Aerosol Optical Depth	-3.8054
29	Sulphate Profile Height	0.9000
30	Sulphate Log Profile Thickness	-2.9957
31	Dust Log Aerosol Optical Depth	-5.4027
32	Dust Profile Height	0.9000
33	Dust Log Profile Thickness	-2.9957
34	Cloud Ice Log Aerosol Optical Depth	-4.3820
35	Cloud Ice Profile Height	0.3000
36	Cloud Ice Log Profile Thickness	-3.2189
37	Cloud Water Log Aerosol Optical Depth	-4.3820
38	Cloud Water Profile Height	0.7500
39	Cloud Water Log Profile Thickness	-2.3026

References

1. Peylin, P.; Law, R.; Gurney, K.; Chevallier, F.; Jacobson, A.; Maki, T.; Niwa, Y.; Patra, P.; Peters, W.; Rayner, P.; Roedenbeck, C.; Zhang, X. Global atmospheric carbon budget: Results from an ensemble of atmospheric CO₂ inversions. *Biogeosciences* **2013**, *10*, 6699–6720. [[CrossRef](#)]
2. Wunch, D.; Toon, G.C.; Blavier, J.F.L.; Washenfelder, R.A.; Notholt, J.; Connor, B.J.; Griffith, D.W.T.; Sherlock, V.; Wennberg, P.O. The Total Carbon Column Observing Network. *Philos. Trans. R. Soc. Lond. Math. Phys. Eng. Sci.* **2011**, *369*, 2087–2112, doi:10.1098/rsta.2010.0240. [[CrossRef](#)] [[PubMed](#)]
3. Wunch, D.; Toon, G.C.; Sherlock, V.; Deutscher, N.M.; Liu, X.; Feist, D.G.; Wennberg, P.O. *The Total Carbon Column Observing Network's GGG2014 Data Version*; Technical Report; Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory: Oak Ridge, TN, USA, 2015, doi:10.14291/tcon.ggg2014.documentation.R0/1221662. [[CrossRef](#)]
4. Miller, C.E.; Crisp, D.; DeCola, P.L.; Olsen, S.C.; Randerson, J.T.; Michalak, A.M.; Alkhaled, A.; Rayner, P.; Jacob, D.J.; Suntharalingam, P.; et al. Precision requirements for space-based data. *J. Geophys. Res. Atmos.* **2007**, *112*, doi:10.1029/2006JD007659. [[CrossRef](#)]
5. Chevallier, F.; Bréon, F.M.; Rayner, P.J. Contribution of the Orbiting Carbon Observatory to the estimation of CO₂ sources and sinks: Theoretical study in a variational data assimilation framework. *J. Geophys. Res. Atmos.* **2007**, *112*, doi:10.1029/2006JD007375. [[CrossRef](#)]
6. Crisp, D.; Atlas, R.M.; Breon, F.M.; Brown, L.R.; Burrows, J.P.; Ciaia, P.; Connor, B.J.; Doney, S.C.; Fung, I.Y.; Jacob, D.J.; et al. The Orbiting Carbon Observatory (OCO) mission. *Adv. Space Res.* **2004**, *34*, 700–709, doi:10.1016/j.asr.2003.08.062. [[CrossRef](#)]
7. O'Dell, C.W.; Connor, B.; Bösch, H.; O'Brien, D.; Frankenberg, C.; Castano, R.; Christi, M.; Eldering, D.; Fisher, B.; Gunson, M.; et al. The ACOS CO₂ retrieval algorithm—Part 1: Description and validation against synthetic observations. *Atmos. Meas. Tech.* **2012**, *5*, 99–121, doi:10.5194/amt-5-99-2012. [[CrossRef](#)]
8. Eldering, A.; O'Dell, C.W.; Wennberg, P.O.; Crisp, D.; Gunson, M.R.; Viatte, C.; Avis, C.; Braverman, A.; Castano, R.; Chang, A.; et al. The Orbiting Carbon Observatory-2: First 18 months of science data products. *Atmos. Meas. Tech.* **2017**, *10*, 549–563, doi:10.5194/amt-10-549-2017. [[CrossRef](#)]
9. O'Dell, C.W.; Eldering, A.; Wennberg, P.O.; Crisp, D.; Gunson, M.R.; Fisher, B.; Frankenberg, C.; Kiel, M.; Lindqvist, H.; Mandrake, L.; et al. Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm. *Atmos. Meas. Tech.* **2018**, *11*, 6539–6576, doi:10.5194/amt-11-6539-2018. [[CrossRef](#)]
10. Rodgers, C.D. *Inverse Methods for Atmospheric Sounding: Theory and Practice*; World Scientific: Singapore, 2000.
11. Tamminen, J.; Kyrölä, E. Bayesian solution for nonlinear and non-Gaussian inverse problem by Markov chain Monte Carlo method. *J. Geophys. Res.* **2001**, *106*, 14377–14390. [[CrossRef](#)]
12. Haario, H.; Laine, M.; Lehtinen, M.; Saksman, E.; Tamminen, J. MCMC methods for high dimensional inversion in remote sensing. *J. R. Stat. Soc. Ser. B* **2004**, *66*, 591–607. [[CrossRef](#)]
13. Wang, Y.; Jiang, X.; Yu, B.; Jiang, M. A Hierarchical Bayesian Approach for Aerosol Retrieval Using MISR Data. *J. Am. Stat. Assoc.* **2013**, *108*, 483–493. [[CrossRef](#)]
14. Tukiainen, S.; Railo, J.; Laine, M.; Hakkarainen, J.; Kivi, R.; Heikkinen, P.; Chen, H.; Tamminen, J. Retrieval of atmospheric CH₄ profiles from Fourier transform infrared data using dimension reduction and MCMC. *J. Geophys. Res. Atmos.* **2016**, *121*, 10312–10327, doi:10.1002/2015JD024657. [[CrossRef](#)]
15. Lamminpää, O.; Laine, M.; Tukiainen, S.; Tamminen, J. Likelihood Informed Dimension Reduction for Remote Sensing of Atmospheric Constituent Profiles. *2017 Matrix Ann.* **2019**, *2*, 65–78. doi:10.1007/978-3-030-04161-8. [[CrossRef](#)]
16. Brynjarsdottir, J.; Hobbs, J.; Braverman, A.; Mandrake, L. Optimal Estimation versus MCMC for CO₂ Retrievals. *J. Agric. Biol. Environ. Stat.* **2018**, *23*, 297–316. [[CrossRef](#)]
17. Hobbs, J.; Braverman, A.; Cressie, N.; Granat, R.; Gunson, M. Simulation-Based Uncertainty Quantification for Estimating Atmospheric CO₂ from Satellite Data. *SIAM/ASA J. Uncertain. Quantif.* **2017**, *5*, 956–985, doi:10.1137/16M1060765. [[CrossRef](#)]

18. Nelson, R. The Impact of Aerosols on Space-Based Retrievals of Carbon Dioxide. Master's Thesis, Colorado State University, Fort Collins, CO, USA, 2015.
19. D. M. O'Brien, Igor Polonsky, C.O.A.C. *Orbiting Carbon Observatory (OCO) Algorithm Theoretical Basis Document The OCO Simulator*; Technical Report; Cooperative Institute for Research in the Atmosphere, Colorado State University: Fort Collins, CO, USA, 2009.
20. Nelson, R.R.; O'Dell, C.W. The impact of improved aerosol priors on near-infrared measurements of carbon dioxide. *Atmos. Meas. Tech.* **2019**, *12*, 1495–1512. [[CrossRef](#)]
21. Cui, T.; Martin, J.; Marzouk, Y.M.; Solonen, A.; Spantini, A. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Probl.* **2014**, *30*, 114015, doi:10.1088/0266-5611/30/11/114015. [[CrossRef](#)]
22. Spantini, A.; Solonen, A.; Cui, T.; Martin, J.; Tenorio, L.; Marzouk, Y. Optimal Low-rank Approximations of Bayesian Linear Inverse Problems. *Siam J. Sci. Comput.* **2015**, *37*, A2451–A2487, doi:10.1137/140977308. [[CrossRef](#)]
23. Crisp, D.; Boesch, H.; Brown, L.; Castano, R.; Christi, M.; Connor, B.; Frankenberg, C.; McDuffie, J.; Miller, C.E.; Natraj, V.; et al. *OCO (Orbiting Carbon Observatory)-2 Level 2 Full Physics Retrieval Algorithm Theoretical Basis*; Technical Report; NASA Jet Propulsion Laboratory, California Institute of Technology: Pasadena, CA, USA, 2019.
24. Kaipio, J.; Somersalo, E. *Statistical and Computational Inverse Problems*; Springer: New York, NY, USA, 2005, doi:10.1007/b138659.
25. Eldering, A.; Osterman, G.; Pollock, R.; Lee, R.; Rosenberg, R.; Oyafuso, F.; Crisp, D.; Chapsky, L.; Granat, R. *Orbiting Carbon Observatory (OCO-2) Level 1B Algorithm Theoretical Basis*; Technical Report; NASA Jet Propulsion Laboratory, California Institute of Technology: Pasadena, CA, USA, 2017.
26. Haario, H.; Saksman, E.; Tamminen, J. An adaptive Metropolis algorithm. *Bernoulli* **2001**, *7*, 223–242. doi:10.2307/3318737. [[CrossRef](#)]
27. Zahm, O.; Cui, T.; Law, K.; Spantini, A.; Marzouk, Y. Certified Dimension Reduction in Nonlinear Bayesian Inverse Problems. Working Paper. *arXiv*, **2018**, arXiv:1807.03712.
28. Sokal, A.D. *Lecture Notes: Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*; Springer US, Boston, MA, USA, 1997.
29. Butz, A.; Hasekamp, O.P.; Frankenberg, C.; Aben, I. Retrievals of atmospheric CO₂ from simulated space-borne measurements of backscattered near-infrared sunlight: accounting for aerosol effects. *Appl. Opt.* **2009**, *48*, 3322–3336, doi:10.1364/AO.48.003322. [[CrossRef](#)]
30. Haario, H.; Saksman, E.; Tamminen, J. Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Stat.* **1999**, *14*, 375–395, doi:10.1007/s001800050022. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).